ACTA UNIVERSITATIS LODZIENSIS
FOLIA OECONOMICA 34, 1984

*Iwona Konarzewska**, *Władysław Milo***

ON RIDGE ESTIMATORS OF THE PARAMETER VECTOR IN THE GENERAL LINEAR MODEL

I. INTRODUCTION

Our analysis is focussed on a subclass of biased estimators, i.e., the ridge estimators of the parameter vector β of the model

$$\mathcal{M}_{\mathcal{N}} = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi, \nu(X) = k_o, \nu(\mathcal{D}(Y)) = n_o,$$

$$\mathcal{P}_Y = \mathcal{N}_Y(X\beta, \sigma^2 I)),$$

where:

$\mathcal{R}^{n\times k}$ - a set of real $(n\times k)$ matrices,

$\mathcal{S} = (\mathcal{U}, \mathcal{F}, \mathcal{P})$ - a complete probability space with $\mathcal{P}$ as the measure for which $\mathcal{P}(\mathcal{U}) = 1$, $\mathcal{U}$ it the set of elementary events, $\mathcal{F}$ is the Borel σ-field of subsets of $\mathcal{U}$, Y, Ξ : $(\mathcal{U}, \mathcal{F}) \to (\mathcal{R}^n, \mathcal{F}_{\mathcal{R}^n})$, $\mathcal{R}^n$ is the n-dimensional space of reals,

$\mathcal{F}_{\mathcal{R}^n}$ - the Borel σ-field of subsets of $\mathcal{R}^n$,

$y, \xi \in \mathcal{R}^n$ - sample values of Y, Ξ,

$\mathcal{E}, \mathcal{D}$ - expectations and dispersion operators,

* Senior Assistant, Institute of Econometrics and Statistics, University of Łódź.

** Lecturer, Institute of Econometrics and Statistics, University of Łódź.

$\mathcal{E}(Y) = X\beta, \mathcal{D}(Y) = \sigma^2 I = \mathcal{D}(\Xi), \beta \in \mathcal{R}^k, X \in \mathcal{R}^{n\times k}, \sigma^2, k_o, k, n_o, n \in \mathcal{R},$
$\nu(X) = \text{rank}(X) = k_o < k,\ \nu(\mathcal{D}(Y)) = \text{rank}(\mathcal{D}(Y)) = n_o < n,$

"$\mathcal{P}_Y = \mathcal{N}_Y(X\beta, \sigma^2 I)$" read as "probability distribution of Y is multivariate n-dimensional normal distribution with $\mathcal{E}(Y) = X\beta$ and $\mathcal{D}(Y) = \sigma^2 I$".

Minimizing the quadratic form $\Phi_o = \| Y - X\beta\|^2$ with respect to β we obtain the least squares estimator B_o of the parameter vector β in the model $\mathcal{M}_{\mathcal{N}}$, i.e.,

$$B_o = (X'X)^{-1}X'Y.$$

For the model $\mathcal{M}_{\mathcal{N}}$, in the case of bad-conditioning of the matrix $X'X$, one can propose the following estimation criteria functions: $\Phi_{01} = \| Y - X\beta\|^2 + c\beta'\beta$ and $\Phi_{02} = \| Y - X\beta\|^2 + \beta' C\beta$. Minimizing the form Φ_{01} with respect to β we have the ridge estimator

$$B_{01} = (X'X + cI)^{-1}X'Y,\ c \in R_+,\ R_+ = \{\delta \in \mathcal{R}:\ \delta > 0\}$$

(X' is the transpose of X), and doing the same with the form Φ_{02} we obtain

$$B_{02} = (X'X + C)^{-1}X'Y,\ C = \text{diag}\,(c_1, \ldots, c_k),\ c_i \in R_+,\ C \in R^{k\times k}.$$

For the reasons of easy handling the derivation of theoretical results concerning the properties of B_{01} and B_{02} it is necessary to reparametrize the model $\mathcal{M}_{\mathcal{N}}$ in the following way. Let $X_* = XQ$, $\alpha = Q'\beta$, $X\beta = XQQ'\beta = X_*\alpha$, $Q'Q = QQ' = I_{(k)}$, $Q'X'XQ = \Lambda$, where $\Lambda = \text{diag}\,(\lambda_1, \ldots, \lambda_k)$, $\lambda_i = \lambda_i(X'X)$ are the eigen value matrix and the i-th eigen value of the matrix $X'X$. Then the model $\mathcal{M}_{\mathcal{N}}$ can be written as

$$\mathcal{M}^*_{\mathcal{N}} = (\mathcal{R}^{n\times k}, \mathcal{B}, Y = X_*\alpha + \Xi, \nu(X) = k_o, \nu(\mathcal{D}(Y)) =$$
$$= n_o, \mathcal{P}_Y = \mathcal{N}_Y(X_*\alpha, \sigma^2 I)).$$

Minimizing the form $\Phi^*_o = \| Y - X_*\alpha\|^2$ with respect to α we obtain the least-squares estimator (l.s.e.) of the parameter vector α for the model $\mathcal{M}^*_{\mathcal{N}}$, i.e.,

$$A_o = \Lambda^{-1} X_*' Y.$$

As was shown by Theobald [13] the relation $MSE(B_o) > MSE(B_{01})$ holds if $0 < c < 2\sigma^2 k / (\sigma^2 \sum_{i=1}^{k} \lambda_i^{-1} - \sum_{i=1}^{k} \alpha_i^2)$, where $MSE(B_{(i)}) = \mathcal{E}(B_{(i)} - \beta)'(B_{(i)} - \beta) = \mathcal{E}(\| B_{(i)} - \beta \|^2)$, and the relation $MSE(B_o) > MSE\ B_{02}$ holds iff $\forall c_i : c_i > \lambda_i (\alpha_i^2 - 2\sigma^2)\sigma^{-2} > 0$, $i = 1, \ldots, k$.

These relations concerning MSE are to be true if c, $\{c_i, i = 1, \ldots, k\}$ are given. If they are not one can propose:

1.1. Estimators for c [2, 8, 4]:

$$\hat{c}_1 = S_E^2 / A_{o(j^*)}^2, \quad A_{o(j^*)}^2 = \max_{j=1, \ldots, k} \{A_{oj}^2\}, \quad S_E^2 =$$

$$= E'E/(n-k), \quad E = Y - X_* A_o,$$

$$\hat{c}_2 = k\, S_E^2 \Big/ \left(\sum_{i=1}^{k} \lambda_i A_{oi}^2 \right),$$

$$\hat{c}_3 = k\, S_E^2 \Big/ \left(\sum_{i=1}^{k} A_{oi}^2 \right).$$

Using the definitions of $\hat{c}_1$, $\hat{c}_2$, $\hat{c}_3$ we can define the following ridge estimators

$$B_{011} = (X'X + \hat{c}_1 I)^{-1} X'Y,$$

$$B_{012} = (X'X + \hat{c}_2 I)^{-1} X'Y,$$

$$B_{013} = (X'X + \hat{c}_3 I)^{-1} X'Y.$$

1.2. Estimators for C [1]:

$$\hat{C}_1 = \mathrm{diag}\,(\hat{c}_{11}, \ldots, \hat{c}_{k1}), \quad \hat{c}_{i1} = S_E^2 / A_{o(i)}^2, \quad i = 1, \ldots, k,$$

$$\hat{C}_2 = \operatorname{diag}(\hat{c}_{12}, \ldots, \hat{c}_{k2}), \quad \hat{c}_{i2} = \lambda_i S_E^2/(B_o'X'XB - \delta S_E^2),$$

$$0 < \delta < \frac{2(k-2)(n-k)}{n-k+2},$$

$$\hat{C}_3 = \operatorname{diag}(\hat{c}_{13}, \ldots, \hat{c}_{k3}), \quad \hat{c}_{i3} = 2/\left(\left(1 - \frac{2}{n-k}\right)\frac{A_{o(i)}^2}{S_E^2} - 1\right).$$

Using these definitions we can define

$$B_{O21} = (X'X + \hat{C}_1)^{-1}X'Y,$$

$$B_{O22} = (X'X + \hat{C}_2)^{-1}X'Y,$$

$$B_{O23} = (X'X + \hat{C}_3)^{-1}X'Y.$$

The purpose of this paper is to present some results on chosen properties of ridge estimators.

In § 2 we will present some analytical results for B_O, B_{O1}, B_{O2}, A_O, A_{O1}, A_{O2}.

In § 3 we will present some of the results concerning precision and predictive power for B_{O11}, B_{O12}, B_{O13}, B_{O21}, B_{O22}, B_{O23} obtained by the use of Monte-Carlo experiments. The plan of experiments was prepared jointly. Konarzewska carried out the experiments and their analysis, and wrote § 3. Milo wrote § 1, § 2 and § 4.

2. SOME RESULTS ON THE PROPERTIES OF B_O, B_{O1}, B_{O2}

In what follows we will use

Theorem 1 (see [11]). Let $\varphi_Y = \mathcal{N}_Y(\mu, V)$. Then

i) $\mathcal{E}(Y'AY) = \operatorname{tr}(AV) + \mu'A\mu$,

ii) the r-th cumulant of $Y'AY$ is

$$\mathcal{K}_r(Y'AY) = 2^{r-1}(r-1)!\left[\operatorname{tr}(AV)^r + r\mu'A(VA)^{r-1}\right],$$

(hence: $\text{var}(Y'AY) = 2\text{tr}(AV)^2 + 4\mu'AVA\mu)$,

iii) $\text{cov}(LY, Y'AY) = 2LVA\mu$. ◆

Theorem 2 (see [9]). Let B be the linear estimator of β, $\beta \in \mathcal{R}^k$ with the dispersion matrix $\mathcal{D}(B) = \mathcal{E}(B - \mathcal{E}(B))(B - \mathcal{E}(B))'$ and the total $\text{MSE}(B) = \mathcal{E}(B - \beta)'(B - \beta)$. Then if $\mathcal{E}(B) \neq \beta$ holds we have

$$\text{MSE}(B) = \text{tr}(D),\ D = \mathcal{E}(B - \beta)(B - \beta)' = \mathcal{D}(B) + \text{bias}(B)\,\text{bias}(B'),$$

$$\text{bias}(B) = \mathcal{E}(B) - \beta,\ \text{bias}(B') = [\text{bias}(B)]'. \quad ◆$$

Theorem 3 (see [10]). Let $\mathcal{P}_U = \mathcal{N}_U(\mu, \Omega)$ and $Y = d + CU, \mu, d \in \mathcal{R}, \Omega \in \mathcal{R}^{n\times n}$. Then

$$\mathcal{P}_Y = \mathcal{N}_Y(C\mu + d, C\Omega C'). \quad ◆$$

Theorem 4 (see [5]). Let $\mathcal{P}_Y = \mathcal{N}_Y(\mu, \Omega)$ and $Q = Y'AY + 2a'Y + \bar{\alpha}$, $A \in \mathcal{R}^{n\times n}$, $a \in \mathcal{R}^n$, $\bar{\alpha} \in \mathcal{R}$. Then $\mathcal{P}_Q = \chi^2(s, \lambda)$ if

$$s = \text{tr}(AV) = \nu(VAV),\ VAVAV = VAV,$$

$$V(a + A\mu) = VAV(a + A\mu),\quad \lambda = \bar{\alpha} + 2a'\mu + \mu'A\mu. \quad ◆$$

Using the definitions of $\mathcal{E}$, $\mathcal{D}$, $\mathcal{P}$, MSE, var, B_o, $\hat{Y}_o$, E_o, B_{o1}, $\hat{Y}_{o1}$, E_{o1}, B_{o2}, $\hat{Y}_{o2}$, $\mathcal{M}_{\mathcal{N}}$ and the theorems 1÷4 we obtain

<u>a) the case B_o, $\hat{Y}_o$, E_o:</u>

1a) $B_o = \mathcal{X}^{-1}X'Y,\quad \mathcal{X} = X'X;$

2a) $\mathcal{E}(B_o) = \beta,\ \mathcal{D}(B_o) = \sigma^2\mathcal{X}^{-1};$

3a) $\mathcal{P}_{B_o} = \mathcal{N}_{B_o}(\beta, \sigma^2\mathcal{X}^{-1}),\ \text{MSE}(B_o) = \sigma^2\text{tr}(\mathcal{X}^{-1});$

4a) $B_o'B_o = Y'X\mathcal{X}^{-2}X'Y,\ \mathcal{E}(B_o'B_o) = \sigma^2\text{tr}(\mathcal{X}^{-1}) + \beta'\beta;$

5a) $\text{var}(B_o'B_o) = 2\sigma^4\text{tr}(\mathcal{X}^{-2}) + 4\sigma^2\beta'\mathcal{X}^{-2}\beta;$

6a) $\mathcal{P}_{B_o'B_o/\sigma^2} \neq \chi^2_{B_o'B_o/\sigma^2}(.,.)$ since $\sigma^6(X\mathcal{X}^{-2}X')^2 \neq \sigma^4X\mathcal{X}^{-2}X;$

7a) $\hat{Y}_o = XB_o = (I - M)Y, \quad M = I - X\mathcal{X}^{-1}X';$

8a) $\mathcal{E}(\hat{Y}_o) = XB, \quad \mathcal{D}(\hat{Y}_o) = \sigma^2(I - M);$

9a) $\mathcal{P}_{\hat{Y}_o} = \mathcal{ON}_{\hat{Y}_o}(XB, \sigma^2 (I - M)), \quad MSE(\hat{Y}_o) = \sigma^2 k;$

10a) $\hat{Y}_o'\hat{Y}_o = Y'(I - M)Y, \quad \mathcal{E}(\hat{Y}_o'\hat{Y}_o) = \sigma^2 k + B'\mathcal{X} B;$

11a) $\text{var}\,(\hat{Y}_o'\hat{Y}_o) = 2\sigma^4 k + 4\sigma^2 B'\mathcal{X} B;$

12a) $\mathcal{P}_{\hat{Y}_o'\hat{Y}_o/\sigma^2} = \chi^2_{\hat{Y}_o'\hat{Y}_o/\sigma^2}\left(k, \frac{1}{2\sigma^2} B'\mathcal{X} B\right);$

13a) $E_o = MY = M\Xi, \quad \mathcal{E}(E_o) = 0;$

14a) $\mathcal{D}(E_o) = \sigma^2 M, \quad \mathcal{P}_{E_o} = \mathcal{ON}_{E_o}(0, \sigma^2 M);$

15a) $MSE(E_o) = \sigma^2(n-k), \quad E_o'E_o = Y'MY = \Xi'M\Xi;$

16a) $\mathcal{E}(E_o'E_o) = \sigma^2(n-k), \quad \text{var}\,(E_o'E_o) = 2\sigma^4(n-k);$

17a) $\mathcal{P}_{E_o'E_o/\sigma^2} = \chi^2_{E_o'E_o/\sigma^2}((n-k), 0);$

b) the case of $B_{o1}, \hat{Y}_{o1}, E_{o1}$:

1b) $B_{o1} = (\mathcal{X} + cI)^{-1}X'Y;$

2b) $\mathcal{E}(B_{o1}) = (\mathcal{X} + cI)^{-1}\mathcal{X}'B, \quad \mathcal{D}(B_{o1}) = \sigma^2 (\mathcal{X} + cI)^{-1}\mathcal{X}(\mathcal{X} +$

$+ cI)^{-1};$

3b) $\mathcal{P}_{B_{o1}} = \mathcal{N}_{B_{o1}}(\mathcal{E}(B_{o1}), \mathcal{D}(B_{o1})), \quad MSE(B_{o1}) = \sigma^2 \text{tr}\,(\mathcal{X}(\mathcal{X} +$

$+ cI)^{-2}) + c^2 B'(\mathcal{X} + cI)^{-2}B;$

4b) $B_{o1}B_{o1} = Y'X(\mathcal{X} + cI)^{-2}X'Y, \quad \mathcal{E}(B_{o1}'B_{o1}) = \sigma^2 \text{tr}\,(\mathcal{X}(\mathcal{X}+cI)^{-2}) +$

$+ B'\mathcal{X}(\mathcal{X} + cI)^{-2}\mathcal{X} B;$

5b) $\text{var}\,(B_{o1}'B_{o1}) = 2\sigma^4 \text{tr}\,(\mathcal{X}^2(\mathcal{X} + cI)^{-4} + 4\sigma^2 B'\mathcal{X}(\mathcal{X} +$

$+ cI)^{-2}\mathcal{X}(\mathcal{X} + cI)^{-2}\mathcal{X} B;$

6b) $\mathcal{P}_{B'_{01}B_{01}/\sigma^2} \neq \chi^2_{B'_{01}B_{01}/\sigma^2}(.,.)$ because

$\sigma^6 X(\mathfrak{X} + cI)^{-2}\mathfrak{X}(\mathfrak{X} + cI)^{-2}X' \neq \sigma^4 X(\mathfrak{X} + cI)^{-2}X'$;

7b) $\hat{Y}_{01} = XB_{01} = X(\mathfrak{X} + cI)^{-1}X'Y$;

8b) $\mathcal{E}(\hat{Y}_{01}) = X(\mathfrak{X} + cI)^{-1}\mathfrak{X}\beta$, $\mathfrak{D}(\hat{Y}_{01}) = \sigma^2 X(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} +$

$+ cI)^{-1}X'$;

9b) $\mathcal{P}_{\hat{Y}_{01}} = \mathcal{N}_{\hat{Y}_{01}}(\mathcal{E}(\hat{Y}_{01}), \mathfrak{D}(\hat{Y}_{01}))$, $\mathrm{MSE}(\hat{Y}_{01}) = \sigma^2 \mathrm{tr}(\mathfrak{X}^2(\mathfrak{X} +$

$+ cI)^{-2}) + c^2\beta'(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}\beta$;

10b) $\hat{Y}'_{01}\hat{Y}_{01} = Y'X(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}X'Y$,

$\mathcal{E}(\hat{Y}'_{01}Y_{01}) = \sigma^2 \mathrm{tr}(\mathfrak{X}^2(\mathfrak{X} + cI)^{-2}) + \beta'\mathfrak{X}(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} +$

$+ cI)^{-1}\mathfrak{X}\beta$;

11b) $\mathrm{var}(\hat{Y}'_{01}\hat{Y}_{01}) = 2\sigma^4 \mathrm{tr}(\mathfrak{X}^4(\mathfrak{X} + cI)^{-4}) + 4\sigma^2\beta'(\mathfrak{X}(\mathfrak{X} +$

$+ cI)^{-1})^4\mathfrak{X}\beta$;

12b) $\mathcal{P}_{\hat{Y}'_{01}\hat{Y}_{01}/\sigma^2} \neq \chi^2_{\hat{Y}'_{01}\hat{Y}_{01}/\sigma^2}(.,.)$ since

$\sigma^6 X(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}X' \neq$

$\neq \sigma^4 X(\mathfrak{X} + cI)^{-1}\mathfrak{X}(\mathfrak{X} + cI)^{-1}X'$;

13b) $E_{01} = M_{01}Y$, $M'_{01} = M_{01}$, $M^2_{01} \neq M_{01}$, $M_{01} = I - X(\mathfrak{X} +$

$+ cI)^{-1}X'$,

$\mathcal{E}(E_{01}) = M_{01}X\beta$;

14b) $\mathfrak{D}(E_{01}) = \sigma^2 M_{01}M'_{01}$, $\mathcal{P}_{E_{01}} = \mathcal{N}_{E_{01}}(M_{01}X\beta, \sigma^2 M_{01}M'_{01})$;

15b) $\mathrm{MSE}(E_{01}) = \sigma^2 \mathrm{tr}(M_{01}M'_{01}) + \beta'X'M'_{01}M_{01}X\beta$, $E'_{01}E_{01} = Y'M'_{01}M_{01}Y$;

16b) $\mathcal{E}(E'_{01}E_{01}) = \sigma^2 \mathrm{tr}(M_{01}M'_{01}) + \beta'X'M'_{01}M_{01}X\beta$,

$var(E'_{o1}E_{o1}) = 2\sigma^2 tr(M'_{o1}M_{o1})^2 + 4\sigma^2 \beta' X'(M'_{o1}M_{o1})^2 X\beta;$

17b) $\mathcal{P}_{E'_{o1}E_{o1}/\sigma^2} \neq \chi^2_{E'_{o1}E_{o1}/\sigma^2}(.,.)$ since $\sigma^6(M'_{o1}M_{o1})^2 \neq \sigma^4 M'_{o1}M_{o1};$

c) <u>the case of B_{o2}, $\hat{Y}_{o2}$, E_{o2}:</u>

1c) $B_{o2} = (\mathcal{X} + C)^{-1}X'Y;$

2c) $\mathcal{E}(B_{o2}) = (\mathcal{X} + C)^{-1}\mathcal{X}\beta, \mathcal{D}(B_{o2}) = \sigma^2 (\mathcal{X} + C)^{-1}\mathcal{X}(\mathcal{X} + C)^{-1};$

3c) $\mathcal{P}_{B_{o2}} = \mathcal{N}_{B_{o2}}(\mathcal{E}(B_{o2}), \mathcal{D}(B_{o2})),$ $MSE(B_{o2}) = \sigma^2 tr((\mathcal{X} + C)^{-2}\mathcal{X}) +$

$+ (\beta'(I + C^{-1}\mathcal{X})^{-1}(I + \mathcal{X}C^{-1})^{-1}\beta);$

4c) $B'_{o2}B_{o2} = Y'X(\mathcal{X} + C)^{-2}X'Y,$ $\mathcal{E}(B'_{o2}B_{o2}) = \sigma^2 tr(\mathcal{X}(\mathcal{X} + C)^{-2}) +$

$+ \beta'\mathcal{X}(\mathcal{X} + C)^{-2}\mathcal{X}\beta;$

5c) $var(B'_{o2}B_{o2}) = 2\sigma^4 tr(\mathcal{X}(\mathcal{X} + C)^{-2})^2 + 4\sigma^2\beta'\mathcal{X}(\mathcal{X} + C)^{-2}\mathcal{X}(\mathcal{X} +$

$+ C)^{-2}\mathcal{X}\beta;$

6c) $\mathcal{P}_{B'_{o2}B_{o2}/\sigma^2} \neq \chi^2_{B'_{o2}B_{o2}/\sigma^2}(.,.)$ since

$\sigma^6 X(\mathcal{X} + C)^{-2}\mathcal{X}(\mathcal{X} + C)^{-2}X' \neq \sigma^4 X(\mathcal{X} + C)^{-2}X';$

7c) $\hat{Y}_{o2} = YB_{o2} = X(\mathcal{X} + C)^{-1}X'Y;$

8c) $\mathcal{E}(\hat{Y}_{o2}) = X(\mathcal{X} + C)^{-1}\mathcal{X}\beta,$ $\mathcal{D}(\hat{Y}_{o2}) = \sigma^2 X(\mathcal{X} + C)^{-1}\mathcal{X}(\mathcal{X} +$

$+ C)^{-1}X';$

9c) $\mathcal{P}_{\hat{Y}_{o2}} = \mathcal{N}_{\hat{Y}_{o2}}(\mathcal{E}(\hat{Y}_{o2}), \mathcal{D}(\hat{Y}_{o2}),$

$MSE(\hat{Y}_{o2}) = \sigma^2 tr(\mathcal{X}^2(\mathcal{X} + C)^{-2}) + \beta'\mathcal{X}(\mathcal{X} + C)^{-1}\mathcal{X}(\mathcal{X} +$

$+ C)^{-1}\mathcal{X}\beta;$

10c) $\hat{Y}'_{o2}\hat{Y}_{o2} = Y'X(\mathcal{X} + C)^{-1}\mathcal{X}(\mathcal{X} + C)^{-1}X'Y,$

$$\mathcal{E}(\hat{Y}_{o2}'\hat{Y}_{o2}) = \sigma^2 tr\ (\mathfrak{X}^2(\mathfrak{X}+C)^{-2}) + \beta'\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X} + $$
$$+ C)^{-1}\mathfrak{X}\beta;$$

11c) $$\mathrm{var}\ (\hat{Y}_{o2}'\hat{Y}_{o2}) = 2\sigma^4 tr\ (\mathfrak{X}^4(\mathfrak{X}+C)^{-4}) + 4\sigma^2\beta'\mathfrak{X}(\mathfrak{X}+C)^{-1}$$
$$\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}\beta;$$

12c) $$\mathcal{P}_{\hat{Y}_{o2}'\hat{Y}_{o2}/\sigma^2} \neq \chi^2_{\hat{Y}_{o2}'\hat{Y}_{o2}/\sigma^2}(.,.) \quad \text{since}$$
$$\sigma^6 X(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}X' \neq$$
$$\neq \sigma^4 X(\mathfrak{X}+C)^{-1}\mathfrak{X}(\mathfrak{X}+C)^{-1}X';$$

13c) $$E_{o2} = M_{o2}Y,\ M_{o2}' = M_{o2},\ M_{o2}^2 \neq M_{o2},\ M_{o2} = I - X(\mathfrak{X}+C)^{-1}X',$$
$$\mathcal{E}(E_{o2}) = M_{o2}X\beta;$$

14c) $$\mathcal{D}(E_{o2}) = \sigma^2 M_{o2}M_{o2}',\quad \mathcal{P}_{E_{o2}} = \mathcal{N}_{E_{o2}}(M_{o2}X\beta, \sigma^2 M_{o2}M_{o2}');$$

15c) $$\mathrm{MSE}(E_{o2}) = \sigma^2 tr(M_{o2}M_{o2}') + \beta'X'M_{o2}'M_{o2}X\beta,\quad E_{o2}'E_{o2} = Y'M_{o2}'M_{o2}Y;$$

16c) $$\mathcal{E}(E_{o2}'E_{o2}) = \sigma^2 tr(M_{o2}M_{o2}') + \beta'X'M_{o2}'M_{o2}X\beta,$$
$$\mathrm{var}(E_{o2}'E_{o2}) = 2\sigma^4 tr(M_{o2}'M_{o2})^2 + 4\sigma^2\beta'X'(M_{o2}'M_{o2})^2X\beta;$$

17c) $$\mathcal{P}_{E_{o2}'E_{o2}/\sigma^2} \neq \chi^2_{E_{o2}'E_{o2}/\sigma^2}(.,.) \quad \text{since}\quad \sigma^6(M_{o2}'M_{o2})^2 \neq$$
$$\neq \sigma^4 M_{o2}'M_{o2}.$$

From the comparisons of (1a)-(6a), (1b)-(6b), (1c)-(6c) it follows

S t a t e m e n t 1. Let the assumptions of model $\mathcal{M}_{\mathcal{N}}$ and Theorems 1÷4 hold. Then (according to (1a)-(6a), (1b)-(6b), (1c)-(6c)) the replacement of B_o with the B_{o1} (or B_{o2}) signifies the change of analytical form (with respect to the expressions connected with B_o) of expressions defining the expectation value, dispersion matrix, normal density functions, the total MSE, the "length" of estimator, the expectation value of the "length", variance of the "length", density function of the "length".

S t a t e m e n t 2. Under the assumptions of model $\mathcal{M}_{\mathcal{N}}$ and Theorems 1÷4 (according to the relations (7a)-(12a), (7b)-(12b), (7c)-(12c)) the replacement of predictor $\hat{Y}_o$ with the biased ridge predictor $\hat{Y}_{o1}$ (or $\hat{Y}_{o2}$) signifies the change in the analytical form (with respect to the relevant expressions connected with $\hat{Y}_o$) of the expressions defining predictor's expectation value, dispersion matrix, normal density function, the total MSE, the "length", the expectation value of the "length", variance of the "length", density function of the "length". ◆

S t a t e m e n t 3. Under the assumptions of model $\mathcal{M}_{\mathcal{N}}$ and Theorems 1÷4 (according to (13a)-(17a), (13b)-(17b), (13c)-(17c)) the replacement of residual vector E_o with the ridge residual vector E_{o1} (or E_{o2}) signifies the change in the analytical form of the expressions defining the residual vector's expectation value, dispersion matrix, normal density function, the total MSE, the "length", expectation value of the "length", variance of the "length", density function of the "length". ◆

S t a t e m e n t 4. Under the assumptions of model $\mathcal{M}_{\mathcal{N}}$ and Theorems 1÷4: a) estimators B_o, B_{o1}, B_{o2} have multivariate normal distributions, b) the predictor $\hat{Y}_o$ and the residual vector E_o have multivariate singular normal distributions and the predictors $\hat{Y}_{o1}$, $\hat{Y}_{o2}$ and the residual vectors E_{o1}, E_{o2} have multivariate normal distributions, c) the only random quantities which have χ^2 distributions are $\hat{Y}_o'\hat{Y}_o$ and $E_o'E_o$; other quantities as $B_o'B_o$, $B_{o1}'B_{o1}$, $B_{o2}'B_{o2}$, $\hat{Y}_{o1}'\hat{Y}_{o1}$, $\hat{Y}_{o2}'\hat{Y}_{o2}$, $E_{o1}'E_{o1}$, $E_{o2}'E_{o2}$, do not have χ^2 distribution. ◆

The above statements 1-4 state the fact of change in the analytical form of the formulae defining some of the chosen real or random functions that describe the performance of estimators, predictors and residuals. For the model $\mathcal{M}_{\mathcal{N}}$ we were not able to find out the precise relation (equality or inferiority or superiority) between chosen real and random functions of B_o, B_{o1}, B_{o2}; Y_o, Y_{o1}, Y_{o2}; E_o, E_{o1}, E_{o2}. It is easy to do it for the model $\mathcal{M}^*_{\mathcal{N}}$ (and in fact it was done by M i l o in the previous works [6, 7]). Here we present only some final statements.

T h e o r e m 5. Let the assumptions of $\mathcal{M}^*_{\mathcal{N}}$ and Theorems 1÷4 be fulfilled. Then:

a) $A_{o1} = (\Lambda + cI)^{-1}X'_*Y$ is biased estimator;

b) if $cI > \Lambda$, then $\mathcal{D}(A_o) \geqslant \mathcal{D}(A_{o1})$;

c) $\mathrm{MSE}(A_{o1}) < \mathrm{MSE}(A_o)$ if $0 < c < 2\sigma^2 \sum_{i=1}^{k} \lambda_i \Big/ \left[\left(\sum_{i=1}^{k} \lambda_i \alpha_i^2\right)^2 - k\sigma^2\right]$;

d) $\mathcal{E}(A'_{o1}A_{o1}) < \mathcal{E}(A'_oA_o)$ if $c > -2\left(\sigma^2 \sum_{i=1}^{k} \lambda_i + \sum_{i=1}^{k} \alpha_i^2 \lambda_i^2\right) \Big/ \left(k\sigma^2 + \sum_{i=1}^{k} \alpha_i^2 \lambda_i\right)$;

e) $\mathrm{var}\,(A'_{o1}A_{o1}) < \mathrm{var}\,(A'_oA_o)$ if $c^3(2k + \sigma^2 k) +$

$$+ c^2\left(8 \sum_{i=1}^{k} \alpha_i^2 \lambda_i^2 + 4\sigma^2 \sum_{i=1}^{k} \lambda_i\right) + c\left(12 \sum_{i=1}^{k} \alpha_i^2 \lambda_i^3 + 6\sigma^2 \sum_{i=1}^{k} \lambda_i^2\right) + \left(8 \sum_{i=1}^{k} \alpha_i^2 \lambda_i^4 + 4\sigma^2 \sum_{i=1}^{k} \lambda_i^3\right) > 0. \qquad \blacklozenge$$

T h e o r e m 6. Under the assumptions of Theorem 5 we have:

a) $\tilde{Y}_{o1} = X_*A_{o1}$ is biased predictor;

b) $\mathcal{D}(\tilde{Y}_{o1}) \leqslant \mathcal{D}(\tilde{Y}_o)$ if $c^2 I \geqslant \Lambda^2$;

c) $\mathrm{MSE}(\tilde{Y}_{o1}) \leqslant \mathrm{MSE}(\tilde{Y}_o)$ if $c < \dfrac{2\sigma^2 \sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{k} \alpha_i^2 \lambda_i - \sigma^2 k}$;

d) $\mathcal{E}(\tilde{Y}_{o1}'\tilde{Y}_{o1}) < \mathcal{E}(\tilde{Y}_o'\tilde{Y}_o)$ if condition (d) from Theorem 5 holds;

e) $\mathrm{var}\,(\tilde{Y}_{o1}'\tilde{Y}_{o1}) < \mathrm{var}\,(\tilde{Y}_o'\tilde{Y}_o)$ if condition (e) from Theorem 5 holds;

f) $\tilde{E}_{o1}$ is biased estimator of $\mathcal{E}(\Xi)$, $\tilde{E}_{o1} = Y_{o1} - \tilde{Y}_{o1}$. ◆

It is easy to find out that:

18a) $\mathrm{MSE}(\tilde{E}_{o1}) < \mathrm{MSE}(\tilde{E}_o)$ if $c\left(\sum_{i=1}^{k} \alpha_i^2 \lambda_i + \sigma^2 k\right) < 0$;

18b) $\mathcal{E}(\tilde{E}_{o1}', \tilde{E}_{o1}) < \mathcal{E}(\tilde{E}_o'\tilde{E}_o)$ if $c\left(\sum_{i=1}^{k} \alpha_i^2 \lambda_i + \sigma^2 k\right) < 0$;

18c) $\mathfrak{D}(\tilde{E}_{o1}) < \mathfrak{D}(\tilde{E}_o)$ if $(I + c\Lambda)^2 \leqslant 0$.

Because $c > 0$, $k > 0$, $\sigma^2 > 0$, $\forall i : \lambda_i > 0$ therefore the relations $\mathcal{E}(A_{o1}'A_{o1}) < \mathcal{E}(A_o'A_o)$, $\mathrm{var}\,(A_{o1}'A_{o1}) < \mathrm{var}\, A_o'A_o$, $\mathcal{E}(\tilde{Y}_{o1}'\tilde{Y}_{o1}) < \mathcal{E}(\tilde{Y}_o'\tilde{Y}_o)$, $\mathrm{var}\,(\tilde{Y}_{o1}'\tilde{Y}_{o1}) < \mathrm{var}\,(\tilde{Y}_o'\tilde{Y}_o)$ hold independently of conditions attached to them. This is due to the analytical form of these conditions. By the same arguments, that is $c, k, \sigma^2 > 0$, $\forall i : \lambda_i > 0$ and the form of conditions given in order to save the truth of relations (g), (h), (i) in the Theorem 6, the relations (g), (h), (i) do not hold. Hence:

T h e o r e m 7. Under the assumptions of Theorem 5 the following relations hold:

a) $MSE(\tilde{E}_{o1}) > MSE(\tilde{E}_o)$;

b) $\mathfrak{E}(\tilde{E}'_{o1}\tilde{E}_{o1}) > \mathfrak{E}(\tilde{E}'_o\tilde{E}_o)$;

c) $\mathfrak{D}(\tilde{E}_{o1}) > \mathfrak{D}(\tilde{E}_o)$. ◆

3. PROPERTIES OF RIDGE ESTIMATORS OBTAINED BY THE USE OF MONTE-CARLO EXPERIMENTS

In experiments we have tested the behaviour of the estimators B_o, B_{o11}, B_{o12}, B_{o13}, B_{o21}, B_{o22}, B_{o23} in B_{o22} the parameter δ was equal $\delta = (k-2)(n-k)/(n-k+2)$. We have made an attempt to compare these estimators considering the precision of estimates of the vector β of the model $\mathcal{M}_o$. The base of experiments were empirical values of X and β obtained from the least squares estimation. We took under consideration 4 matrices X for two explanatory variables. The descriptive correlation coefficients were 0.9989, 0.9977, 0.9859, 0.9580. The values of σ^2 were taken as equal to those from the least squares estimation. In the second variant the values were increased by about 20% (we assumed that bad-conditioning caused underestimation of σ^2 by the use of least squares residuals). In each experiment we have generated 100 realizations of the random vector Ξ, calculated the values of estimates and chosen measures. Then we have calculated the mean values of these measures after all 100 realizations. We have considered the following measures of estimators precision and prognostical properties: standard error of estimate, mean bias of estimate, mean absolute bias of estimate, square root of the MSE, mean absolute percentage error, mean square prediction error, mean absolute error, mean square percentage error, standard prediction error, mean absolute relative error. We now present some of the results of experiments in Tab. 1 and 2. Square root of MSE (RMSE), sum of prediction errors (SPE), mean relative absolute error (MARE) were the measures which differentiated best the estimators which we have analyzed.

T a b l e 1

The results - MSE and SPE criterion (ranking)

Value of correlation coefficient / Estimator	0.9989		0.9977		0.9859		0.9580		Sums of ranks
	1	2	1	2	1	2	1	2	
B_o	3	3	3	7	1	1	1	1	20
B_{o11}	7	7	4	1*	6	6	6	6	43
B_{o12}	1	1	1	4	3	3	2	1	16
B_{o13}	6	5	7	5	5	5	5	5	43
B_{o21}	4	4	5	2	2	2	3	3	25
B_{o22}	2	2	2	6	1	1	1	2	17
B_{o23}	5	6	6	3	4	4	4	4	36

* It was the only example when the estimator B_{o11} had not bad MSE properties; at the same time it showed great bias.

T a b l e 2

The results - MARE criterion (ranking)

Value of correlation coefficient / Estimator	0.9989		0.9977		0.9859		0.9580		Sums of ranks
	1	2	1	2	1	2	1	2	
B_o	1	1	1	1	1	1	1	1	8
B_{o11}	6	5	5	5	5	6	4	5	41
B_{o12}	2	1	1	1	4	5	1	2	17
B_{o13}	5	4	4	4	3	4	3	4	31
B_{o21}	3	2	2	2	2	2	2	2	17
B_{o22}	2	1	1	1	1	1	1	1	9
B_{o23}	4	3	3	3	3	3	3	3	25

Considering MSE and SPE the best were the estimators B_{o12} and B_{o22} and then B_o, B_{o21}. Differences between B_{o12} and B_{o22} in absolute values of RMSE and SPE were not great. Considering MARE (and other prediction errors) the best was B_o estimator and then B_{o22}. Greater values of prediction errors were for B_{o21}, B_{o12}. The estimators B_{o11}, B_{o13}, B_{o23} behaved very poorly - in every experiment they were worse than ordinary least squares estimator.

We have considered models with matrices of observations X taken from practice of econometric modelling. In many experimental studies of ridge estimators the matrices X are the matrices of standardized observations. The problem of standardization is at least controversial (see [12]). We have conducted experiments without standardization - it is not indispensable in defining and computing ridge estimators. However, by applying standardized observation matrices (for explanatory variables) we can generalize results of experiments for a wider class of matrices X and parameters ß. In designing future experiments we will include cases of standardization, and interpretability of estimates of standardized model parameters and consider models with greater (than two) number of explanatory variables. It will need more precise definition of bad-conditioned matrices of observations.

4. FINAL REMARKS

The results presented give us some idea about the studied quality of considered estimators. Studies are far from the end. In an analytical realm such estimators as A_{o2}, B_{o11}, B_{o12}, B_{o13}, B_{o21}, B_{o22}, B_{o23} need further studies. In doing further Monte-Carlo experiments the most difficult problem we have faced is to find out possible ways of parameter space reduction but with keeping in mind one of the ultimate goals of experiments: to cover as dense as possible the parameter space.

It is an open question how to study and compare different results of studies, the properties of estimators in the case of models with and without standardization in Y and X, standardization and orthogonalization in Y and X.

REFERENCES

[1] B a r a n c h i k A. J. (1973): *Inadmissibility of Maximum Likelihood Estimators in Some Multiple Regression Problems with Three or More Independent Variables*, Ann. of Statist., 1, p. 312-321.

[2] H o e r l A., K e n n a r d R. (1970a): *Ridge Regression: Bias Estimation*, "Technometrics", 12, p. 55-67.

[3] H o e r l A., K e n n a r d R. (1970b): *Ridge Regression: Applications*, "Technometrics", 12, p. 69-82.

[4] H o e r l A., K e n n a r d R., B a l d w i n K. (1975): *Ridge Regression: Some Simulation*, Comm. in Statist., 4, p. 105-123.

[5] K h a t r i C., S r i v a s t a v a M. (1969): *Introduction to Multivariate Analysis*, Amsterdam, North-Holland.

[6] K o n a r z e w s k a I., M i l o W. (1979): *Własności estymatorów obciążonych. Cz. I*, work within the contract R. III.9.5.7.

[7] K o n a r z e w s k a I., M i l o W. (1980): *Własności estymatorów obciążonych. Cz. II*, work within the contract R. III.9.5.7.

[8] L a w l e s s J., W a n g P. (1976): *A Simulation Study on Ridge Regression Estimators*, Comm. in Statist., 5, p. 307-323.

[9] M i l o W. (1978): *Estymacja parametrów ogólnych modeli liniowych. Cz. III*, work within the contract R. III.9.5.7.

[10] R a o C. (1965): *Linear Statistical Inference and Its Applications*, New York, Wiley and Sons.

[11] S e a r l S. (1971): *Linear Models*, New York, Wiley and Sons.

[12] S m i t h G., C a m p b e l l F. (1980): *Critique of Some Ridge Regression Methods*, Journal of Am. Statist. Assoc., 75, p. 74-81.

[13] T h e o b a l d C. (1974): *Generalizations of Mean Square Errors Applied to Ridge Regression*, J. Roy. Statist. Soc., Ser. B, 36, p. 103-106.

Iwona Konarzewska, Władysław Milo

O ESTYMATORACH GRZBIETOWYCH WEKTORA PARAMETRÓW
OGÓLNEGO MODELU LINIOWEGO

Artykuł zawiera nowe wyniki analityczne dotyczące konsekwencji zastąpienia metody najmniejszych kwadratów B_o, predyktora $\hat{Y}_o = XB_o$, wektora reszt $E = Y - \hat{Y}_o$ przez ich grzbietowe analogony zarówno przy założeniu, gdy wykorzystujemy macierz korekty cI, $C = \text{diag}(c_1, \ldots, c_k)$ z tytułu złego uwarunkowania macierzy $X'X$, jak i przy założeniu, że korzystamy z oszacowań macierzy cI, C.

W przypadku losowych macierzy cI, C przy użyciu metod Monte-Carlo zbadano zachowanie się wybranych estymatorów i dokonano ich zrangowania względem miar predykcyjno-dokładnościowych.