A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 34, 1984

*Władysław Milo**

ON BIASED REGULARIZING ESTIMATORS. PART I

1. INTRODUCTION

We will introduce some subclass of biased estimators for the parameter vector β in the following linear models:

$$\mathcal{NM}_o = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o = k,\ n_o = n,\ \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \sigma^2 I)),$$

$$\mathcal{NM}_1^{<,} = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o < k,\ n_o = n,\ \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \sigma^2 I)),$$

$$\mathcal{NM}_2 = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o = k,\ n_o = n,\ \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \Omega)),$$

$$\mathcal{NM}_3^{<,} = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o < k,\ n_o = n,\ \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \Omega)),$$

$$S\mathcal{NM}_4^{<,} = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o = k,\ n_o < n,\ \mathcal{P}_Y = S\mathcal{N}_Y(X\beta, \Omega)),$$

$$S\mathcal{NM}_5^{<<} = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi,\ k_o < k,\ n_o < n,\ \mathcal{P}_Y = S\mathcal{N}_Y(X\beta, \Omega)),$$

where:

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

$\mathcal{S}$ - a complete probability space, $\mathcal{S} = (\mathcal{U}, \mathcal{F}, \mathcal{P})$,

$\mathcal{U}$ - a set of elementary events, $\mathcal{F}$ - the Borel σ-field of subsets of $\mathcal{U}$, $\mathcal{P}$ - a complete measure with $\mathcal{P}(\mathcal{U}) = 1$,

$Y, \Xi : (\mathcal{U}, \mathcal{F}) \longrightarrow (\mathcal{R}^n, \mathcal{F}_{\mathcal{R}^n})$,

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

$X \in \mathcal{R}^{n \times k}$, $\beta \in \mathcal{R}^k$, k_o, k, n_o, n, $\sigma^2 \in \mathcal{R}$, rank $(X) = k_o$, rank $(\mathcal{D}(Y)) = n_o$, $\mathcal{E}(Y) = X\beta$, $\mathcal{D}(Y) = \sigma^2 I$ (or Ω), $\mathcal{E}$ and $\mathcal{D}$ are expectation and dispersion operators,

"$\mathcal{P}_Y = S\mathcal{N}_Y(X\beta, \Omega)$" read as "probability distribution of Y is a singular multidimensional normal distribution with mean vector $\mathcal{E}(Y) = X\beta$ and dispersion matrix $\mathcal{D}(Y) = \Omega$" (note that singularity of $\mathcal{P}_Y$ comes from $n_o < n$).

Let B_l denote the l-th biased estimator B of the vector β. The word "biased" will be understood in the sense of

D e f i n i t i o n 1. The estimator B_l is biased estimator of β if $\mathcal{E}(B_l) - \beta \neq 0$, $0 \in \mathcal{R}^k$, l is an identifier of the analytical form of estimator.

Using Definition 1 we can define the total bias as

$$\mathrm{TBIAS}(B_l) = \| \mathcal{E}(B_l) - \beta \|^2,$$

where $\| \cdot \|^2$ denotes the square of euclidean norm of a given vector or matrix. ◆

One can distinguish two groups of causes for the biasedness of a given estimator. The <u>first group</u> is resulting from the model assumptions:

a1) under $k < n$ we can assume $k_o < k$ or $n_o < k_o$,

a2) a shift in the mean vector of Y, i.e. an atypicality (outliers) shift $\sum_{i \in N_m} \mathring{\mu}_i j_i$, $N_m = N \setminus N_{n-m}$, $N = \{1, \ldots, n\}$, N_{m-n} is the set of indices for typical results of observations; N_m is the set of indices for atypical results of observations generated by

$$\mathcal{NM}_{o1} = \Big(\mathcal{R}^{n \times k}, \mathcal{S}, Y = X\beta + \sum_{i \in N_m} \mathring{\mu}_i j_i + \Xi, \; k_o = k, \; n_o = n,$$

$$\mathcal{P}_Y = \mathcal{N}_Y\Big(X\beta + \sum_{i \in N_m} \mathring{\mu}_i j_i, \sigma^2 I\Big)\Big),$$

where $j_i \in \mathcal{R}^n$ is the unit vector with 1 at the i-th coordinate of j_i.

The second group of biasedness causes is resulting from the estimator form. An obvious example is a ridge estimator of the form $B(c) = (X'X + cI)^{-1}X'Y$, $c \in \mathcal{R}$ which under $\mathcal{N}\mathcal{M}_o$ is biased since $\mathcal{E}(B(c)) \neq \beta$.

2. ON CONSEQUENCES OF BAD-CONDITIONING OF X'X

In § 1 it was shown that $B(c)$ is biased due to its analytical form. Before giving a justification for using such a kind of estimator we must first understand some numerical points of error analysis. We first recall that under $\mathcal{N}\mathcal{M}_o$ the l.s.e. of β is the solution of homogeneous system of equations $\frac{\partial\Phi}{\partial\beta} = 0$, $\Phi_o^{(0)} = \|Y - X\beta\|^2$. This solution can be written as:

(1a) $X'XB_o^{(o)} = X'Y$, $X'X\, b_o^{(o)} = X'y$, or under $\det(X'X) \neq 0$,

(1b) $$B_o^{(o)} = (X'X)^{-1}X'Y, \quad b_o^{(o)} = (X'X)^{-1}X'y.$$

Using singular value decomposition of matrix X, i.e. $X = U\Lambda V'$; $U \in \mathcal{R}^{n\times k}$, $V \in \mathcal{R}^{k\times k}$, $U'U = I_k = V'V = VV'$ we rewrite (1b) as:

(1c) $$B_o^{(o)} = V\Lambda^{-1}U'Y = \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{\cdot i}U'_{\cdot i}Y,$$

or its sample value analogue

(1d) $$b_o^{(o)} = V\Lambda^{-1}U'y = \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{\cdot i}U'_{\cdot i}y.$$

The l-th coordinate of vector $b_o^{(o)}$ is

(1e) $$b_{o,l} = \sum_{i=1}^{k} \frac{1}{\lambda_i} v_{li} \sum_{t=1}^{n} u_{ti}y_t, \quad l = 1, \dots, k.$$

Suppose now that $\lambda_1 > \lambda_2 > \ldots \lambda_k$, $0 < \lambda_k \leqslant 0.01$ and other $\lambda_i \geqslant 1$, $i = 1, \ldots, k-1$. In this case the only part of (1e) which makes a large increase in the value of $b_{o,l}$ is

$$\frac{1}{\lambda_k} v_{lk} \sum_{j=1}^{n} u_{jk} y_j$$

due to the factor λ_k^{-1} (for instance if $\lambda_{k_1} = 0.01$, then $\lambda_{k_1}^{-1} = 100$ or if $\lambda_{k_2} = 0.0001$, then $\lambda_{k_2}^{-1} = 10\,000$). This increase is especially big if

$$v_{lk} \sum_{j=1}^{n} u_{jk} y_j > 1$$

which is the usual case for non-standardized data.

Until now we have discussed the situation of lack of measurement errors. Let us assume now the existence of such errors in the process of measurement of Y. Let $y_* = y + \xi_*$, and $\xi_* = (\xi_{*1}, \ldots, \xi_{k,n})'$ be the vector of measurement errors with respect to the unobserved vector y. We assume that the size of total measurement error is $\|\xi_*\|^2 = \gamma \|y\|^2$. Under (1a) and $y_* = y + \xi_*$ we obtain $X'X(b_o + q_*) = X'y + X'\xi_*$ and hence

$$\text{(2a)} \qquad X'Xq_* = X'\xi_*,\ q_* \in \mathcal{R}^k,\ \xi_* \in \mathcal{R}^n \text{ or}$$

$$\text{(2b)} \qquad q_* = (X'X)^{-1}X'\xi_* = V\Lambda^{-1}U'\xi_* = \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{.i}U_{.i}\xi_*.$$

Let us consider the simplest case of a change in the measurement error, i.e., $\mathring{\xi}_* = \xi_* + \Delta\xi_*$, $\Delta\xi_* = (\xi_{*,1}, 0, \ldots, 0)'$. Then $\mathring{q}_* = V\Lambda^{-1}U'\mathring{\xi}_*$,

$$\mathring{q}_* = \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{.i}U'_{.i}\xi_* + \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{.i}U'_{.i}\Delta\xi_* = q_* + \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{.i}U'_{.i}\Delta\xi_*,$$

(3a) $$\Delta\mathring{q}_* = \sum_{i=1}^{k} \frac{1}{\lambda_i} V_{\cdot i} u_{1i} \xi_{*,1} = \sum_{i=1}^{k} \frac{\xi_{*,1}}{\lambda_i} u_{1i} V_{\cdot i},$$

or for the fixed l-th coordinate of $\Delta\mathring{q}_*$ we get

(3b) $$\Delta\mathring{q}_{*,l} = \sum_{i=1}^{k} \frac{\xi_{*,1}}{\lambda_i} u_{1i} v_{li}, \quad l = 1, \ldots, k.$$

The increase in the measurement error of Y_1 by $\xi_{*,1}$ due to the existence of bad conditioning (expressed by, for example, the fact that $0 < \lambda_k \leqslant 0.001$) is especially magnified by the factor λ_k^{-1}. Under other conditions unchanged if $\xi_{*,1} = 0.1$ and $\lambda_k = 0.001$ we have $\xi_{*,1}\lambda_k^{-1} = 100$ times amplification of the value $u_{1i}v_{li}$ in the value of $\Delta\mathring{q}_{*,l}$. This increment is especially big for $u_{1i}v_{li} > 1, \ \forall_i$.

We have made some efforts in explaining the size of influence of measurement errors and bad conditioning on the size of solution error for the system (1a). In practice one should have ways of measuring the level of bad conditioning in X'X in order to make assessment whether we really have this phenomenon or not. The most popular measures are

(4a) $$\nu_{X'X}^{(1)} = \frac{\lambda_{max}(X'X)}{\lambda_{min}(X'X)}, \quad \lambda(X'X) \text{ is the eigen value of } X'X,$$

(4b) $$\nu_{X'X}^{(2)} = \lambda_{max}\,(|A| \cdot |A^{-1}|), \quad |A| = [|a_{ij}|]_{i,j=1}^{k}, \ A = X'X,$$

$\lambda_{max}(|A||A^{-1}|)$ is the maximum eigen value of the product of matrices A and A^{-1} modulus. The measure $\nu_{X'X}^{(2)}$ is called the minimal size measure of bad-conditioning of A,

(4c) $$\nu_{X'X}^{(3)} = \max_i \sum_{i=1}^{k} |a_{ij}| \max_i \sum_{j=1}^{k} |\alpha_{ij}|, \quad A^{-1} = [\alpha_{ij}]_{i,j=1}^{k},$$

(4d) $$\nu^{(4)}_{X'X} = k \max_{i,j} |a_{ij}| \max_{i,j} |\alpha_{ij}|,$$

(4e) $$\nu^{(5)}_{X'X} = k^{-1} \Big(\sum_{i,j} a^2_{ij} \Big)^{1/2} \Big(\sum_{i,j} \alpha^2_{ij} \Big)^{1/2}, \quad i,j = 1, \dots, k.$$

For orthogonal and orthonormal matrices X'X the considered measures $\{\nu^{(i)}_{X'X}\}$, $i = 1,2,5$ all take value 1, i.e., $\nu^{(5)}_{X'X} = 1$, $i = 1,2,5$. For singular matrices X'X the measure $\nu^{(1)}_{X'X}$ takes value $\nu^{(1)}_{X'X} = \infty$ since $\lambda_{min}(X'X) = 0$ (other measures, for det $(X'X) = 0$, are not defined). The problem of determining the threshold value (level)of $\nu^{(i)}_{X'X}$ for which the matrix X'X is to be bad-conditioned is not generally solved. There is a view that this level should be linked somehow with the measurement errors of X and/or y.

One approach (proposed by L a n c z o s [3]) is to use a kind of iteration procedure the iterate function of which depends directly on the fixed level of $\nu^{*}_{X'X}$. The second approach is based on the use of regularization principle build-in to the estimation criterion function. The purpose of this regularization is to smoothe the range of estimation quality function because of bad-conditioning of matrix X'X or large measurement errors in y. The estimators derived under the regularization principle are called regularizing estimators. They are relatively robust against bad-conditioning (or in more general setting against ill-posing the problem of solution of operator equations) or/and measurement errors. If we include the autocorrelation matrix into an estimation criterion function, then the estimators derived from such a function would be relatively robust against autocorrelation. The described two approaches do not change the specification list of explanatory variables (columns of X) in order to reduce the influence of bad-conditioning (strong correlation) in X'X can have on the solution errors of 1a. We assume that in the situation when such a list is given on the grounds of outside statistical reasons (the assumptions of tested theory of some part of science) these approaches are fully justified. Other approaches which include dropping some strongly correlated variables will not be analysed.

In this paper we will present the formal characterization of ill-posed linear model estimation problems and the derivation of solving regularizing procedures (estimators).

3. SOME CONCEPTS OF ILL-POSED ESTIMATION PROBLEMS IN THE CASE OF GENERAL LINEAR MODELS

Let $A = X'X$, $Z = X'Y$, $Z : (\mathcal{U}, \mathcal{F}) \to (\mathcal{R}^k, \mathcal{F}_{\mathcal{R}^k})$, $z = X'y$, $z \in \mathcal{R}^k$ be a sample value of Z, $\mathcal{Z} = (\mathcal{R}^k, \rho_{\mathcal{R}^k})$. Then the system (1a), being an implicit form of the estimation problem

$$(5) \qquad \min_{\beta} \| Y - X\beta \|^2,$$

can be written (for the sample-value case) as

$$(6) \qquad Ab = z.$$

For (6) we can formulate the following definition:

D e f i n i t i o n 2. The problem of solving (6) is <u>well posed</u> on a pair $(\mathcal{B}, \mathcal{Z})$ of metric spaces if:

a) for each $z \in \mathcal{Z}$ there is a vector $b \in \mathcal{B}$ being a solution vector of (6),

b) vector b is the unique solution of (6),

c) the problem of solving (6) is stable on $(\mathcal{B}, \mathcal{Z})$, if

$$(7) \qquad \forall \varepsilon > 0\ \exists\ \delta(\varepsilon) > 0\ \forall z_2 \in \mathcal{Z} (\forall z_1 \in \mathcal{Z} : \rho_{\mathcal{Z}}(z_1, z_2) \leqslant \delta(\varepsilon)) \Longrightarrow$$
$$\Longrightarrow (\rho_{\mathcal{B}}(b_1, b_2) \leqslant \varepsilon),$$

$z_1, z_2 \in \mathcal{Z}$, $b_1, b_2 \in \mathcal{B}$, $z_1 \neq z_2$, $b_1 \neq b_2$, $\rho_{\mathcal{Z}}(\rho_{\mathcal{B}})$ is a metric in the space $\mathcal{Z}$ (or $\mathcal{B}$). ◆

D e f i n i t i o n 3. The problem of solving the system (6) is ill-posed if it is not well-posed, that is, if one (or more) of the conditions {a,b,c} is (are) not fulfilled. ◆

D e f i n i t i o n 4. The estimation problem (5) is ill-posed on $(\mathcal{B}, \mathcal{Z})$ if the problem of solving (6) is ill-posed on $(\mathcal{B}, \mathcal{Z})$. ◆

Remarks on Def. 2:

r1) checking the condition (a) in Def. 2 consists in checking the inconsistency of (6), i.e. whether $b \in \mathcal{R}(A)$, $\mathcal{R}(A) = \{z \in \mathcal{Z} : Ab = z,\ b \in \mathcal{B},\ \mathcal{A} : \mathcal{B} \to \mathcal{Z}\}$, where $\mathcal{A}$ is a linear mapping (operator) of $\mathcal{B}$ into $\mathcal{Z}$ or whether the rank $(A) =$ rank $(A : z)$ for A being the matrix of operator $\mathcal{A}$;

r2) checking the condition (b) of Def. 2 consists in checking whether $\det(A) \neq 0$;

r3) the truth of (c) in Def. 2 depends on assumptions about the metric spaces $\mathcal{B}$, $\mathcal{Z}$, forms of metrics $\rho_{\mathcal{B}}, \rho_{\mathcal{Z}}$, values of ε, $\delta(\varepsilon)$;

r4) the concept "stability of solution" used in (c) of Def. 2 is equivalent to the concept "uniform continuity of solution b on $\mathcal{Z}$", where $b \in \mathcal{A}^{\square} z$, $\mathcal{A}^{\square} : \mathcal{Z} \to \mathcal{B}$; "stability (uniform-continuity) of b with respect to z" will be also called "stability (uniform-continuity) of first kind";

r5) limited applicability of the concept "stability of the first kind" (for (6)) results from the fact that it "catches" the cases of small changes in $z \in \mathcal{Z}$ but it does not catch the small changes in A;

r6) from the definition of metric $\rho_{\mathcal{Z}}(z_1, z_2)$ and $\mathcal{A} b = z$ we have $\rho_{\mathcal{Z}}(z_1, z_2) = \rho_{\mathcal{Z}}(\mathcal{A} b_1, b_2)$; for $\rho_{\mathcal{Z}}(z_1, z_2) = \| z_1 - z_2 \|^2$, it is easy to show that $\| z_1 - z_2 \|^2 = \| \mathcal{A}(b_1 - b_2) \|^2 \leqslant \| \mathcal{A} \|^2 \cdot \| b_1 - b_2 \|^2$.

Therefore in metric $\| \cdot \|^2$ the dependence of distance between z_1 and z_2 on the distance between b_1 and b_2 and $\| \mathcal{A} \|^2$ is seen immediately. It motives a natural modification of the concept of stability of the first kind. We have

D e f i n i t i o n 5. By stability of b on $(\mathcal{Z}, \mathcal{A})$ we mean such a property of b that

$$(8)\quad \forall z_2 \in \mathcal{Z}\ \forall A_2 \in \mathcal{A}\ \forall \varepsilon > 0\ \exists \delta_1(\varepsilon) > 0\ \exists \delta(\varepsilon) > 0\ \forall A_1 \in \mathcal{A}\ \forall z_1 \in \mathcal{Z}:$$

$$((\rho_{\mathcal{A}}(A_1, A_2) \leqslant \delta_1(\varepsilon)) \wedge (\rho_{\mathcal{Z}}(z_1, z_2) \leqslant \delta(\varepsilon))) \Rightarrow (\rho_{\mathcal{B}}(b_1, b_2) \leqslant \varepsilon),$$

$$A_1 \neq A_2,\quad z_1 \neq z_2,\quad b_1 \neq b_2. \quad \blacklozenge$$

N o t e 1: $\mathcal{A} \subset \mathcal{B} \times \mathcal{Z}$; stability of b in the sense of Def. 5 will be called "stability (uniform-continuity) of the second kind" (with respect to (A,z)); stability (uniform-continuity) of the second kind, by A, is strictly connected with the concept of "bad-conditioning of system Ab = z" or "bad-conditioning of matrix A".

D e f i n i t i o n 2a. The problem of solution of the system (6) is well-posed on the triple $(\mathcal{B},\mathcal{Z},\mathcal{A})$ if the conditions (a), (b) of Def. 2 and the condition (8) of Def. 4 are fulfilled. ◆

D e f i n i t i o n 3a. The problem of solution of the system (6) is ill-posed on $(\mathcal{B},\mathcal{Z},\mathcal{A})$ if it is not well posed, i.e., one (or more) of the conditions of Def. 2a does (do) not hold. ◆

D e f i n i t i o n 4a. The estimation problem (5) is ill-posed on $(\mathcal{B},\mathcal{Z},\mathcal{A})$ if the problem of solving (6) is ill-posed on $(\mathcal{B},\mathcal{Z},\mathcal{A})$. ◆

N o t e 2: if one changes the assumption of uniform continuity ((c) in Def. 2 and (8) in Def. 5) on the assumption of continuity he will obtain the concept of "classical (or Adamard) well-posed problem of solving the system (6) and using this modified definition one can arise at Adamard's analogues of Def. 3, ..., Def. 4a.

The definitions introduced above concern estimation problems for all the models presented in § 1.

4. ANALYTICAL FORMS OF ESTIMATION QUALITY FUNCTIONALS

Among the methods of solving ill-posed estimation problems there is a class called "regularization methods". In the case of ill-posed problem (5) they consist in regularizing the Legendre-Gauss functional $\Phi_o^{(o)}$

$$\Phi_o^{(o)} = \| Y - X\beta \|^2 = \sum_{t=1}^{n} \left(Y_t - \sum_{j=1}^{k} X_{tj}\beta_j \right)^2. \tag{9}$$

The least-squares solution of ill-posed problem $\min_{\beta} \Phi_o^{(o)}(\beta)$ is $B_o^{\,o} = A^{-1}X'Y$. It is non-robust on the existence of $(\mu, \mathcal{N})$, $(\sigma, \mathcal{N})$, $(\mu, \sigma, \mathcal{N})$-outliers and bad-conditioning of matrix A, that is, many of the known estimator B_o performance's measures are getting slightly worse. To get rid of these troubles with robustness there were introduced some outliers and bad-conditioning (or in general ill-posedness) smoothers (smoothing functions) into (9). They smooth the range of $\Phi_o^{(o)}$. In order to unify their presentation we write down a distance between Y and $X\beta$ as

(10)
$$\Phi^{(o)} = \rho_y^{(o)}(Y, X\beta),$$

where y is a metric space with a metric ρ_y, i.e. $y = (\mathcal{R}^n, \rho_{\mathcal{R}^n})$, $y \in \mathcal{R}^n$.

In this general setting one cannot find practically meaningful solutions of $\min \Phi^{(o)}$. It is, however, easy to do it if one establishes the concrete functional form of metric ρ_y. There are, for example, the following options:

(10a)
$$\Phi_1^{(o)} = \sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j| = |(Y - X\beta)'| \cdot 1,$$

(10b)
$$\Phi_2^{(o)} = \left[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^2\right]^{1/2},$$

(10c)
$$\Phi_3^{(o)} = \left[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^3\right]^{1/3},$$

(10d)
$$\Phi_4^{(o)} = \Phi_1^{(o)} + \Phi_2^{(o)}, \quad \Phi_{4a}^{(o)} = \Phi_4^{(o)} + \Phi_3^{(o)},$$

(10e)
$$\Phi_5^{(o)} = \left[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^{\infty}\right]^{1/\infty} = \max_{1 \leq t \leq n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|,$$

(10f) $$\Phi_6^{(o)} = \frac{\Phi_1^{(o)}}{E_1'1 + \Phi_1^{(o)}}, \quad \Phi_7^{(o)} = \frac{\Phi_2^{(o)}}{E_2'1 + \Phi_2^{(o)}}, \quad \Phi_8^{(o)} = \frac{\Phi_3^{(o)}}{E_3'1 + \Phi_3^{(o)}},$$

$$\Phi_9^{(o)} = \frac{\Phi_4^{(o)}}{E_4'1 + \Phi_4^{(o)}}, \quad \Phi_{10}^{(o)} = \frac{\Phi_{4a}^{(o)}}{E_{4a}'1 + \Phi_{4a}^{(o)}}, \quad \Phi_{11}^{(o)} = \frac{\Phi_5^{(o)}}{E_5'1 + \Phi_5^{(o)}},$$

E_i, $i = 1, \ldots, 5$, are corresponding to $\Phi_i^{(o)}$, $i = \overline{1,5}$ residuals vectors.

(10g) $$\Phi_{12}^{(o)} = \begin{cases} \sum_{t=1}^{n} t & \text{if } Y_t \neq \sum_{j=1}^{k} X_{tj}\beta_j, \quad t = 1, \ldots, n^o, \quad 1 \leqslant n^o \leqslant n, \\ 0 & \text{if } Y_t = \sum_{j=1}^{k} X_{tj}\beta_j, \quad t = 1, \ldots, n^o, \quad 1 \leqslant n^o \leqslant n, \end{cases}$$

(10h) $$\Phi_{13}^{(o)} = \begin{cases} \rho_y(Y, X\beta) & \text{if } \rho_y(Y, X\beta) < 1, \\ 1 & \text{if } \rho_y(Y, X\beta) \geqslant 1, \end{cases}$$

(10i) $$\Phi_{14}^{(o)} = \begin{cases} \frac{1}{2} \| Y - X\beta \|^2 & \text{if } \| Y - X\beta \|^2 < \delta, \\ \delta \| Y - X\beta \|^2 - \frac{1}{2}\delta^2 & \text{if } \| Y - X\beta \|^2 > \delta, \quad \delta \in \mathcal{R}_1. \end{cases}$$

It is seen that the metrics $\Phi_1^{(o)}$, $\Phi_2^{(o)}$, $\Phi_3^{(o)}$, $\Phi_5^{(o)}$ are easily generalized in writing by

(11) $$\Phi_I^{(o)} = \left[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^p \right]^{1/p}, \quad p \in (0, \infty].$$

One can easily check that:

if $p = 1$, then $\Phi_I^{(o)} := \Phi_1^{(o)}$, if $p = 2$, then $\Phi_I^{(o)} := \Phi_2^{(o)}$,

if $p = 3$, then $\Phi_I^{(o)} := \Phi_3^{(o)}$,

if $p = 5$, then $\Phi_I^{(o)} := \Phi_5^{(o)}$.

Each of the above metrics gives rise to the formulation of one estimation problem

$$\min_{\beta} \Phi_l^{(o)}(\beta), \quad l = 1, \ldots, 14, \tag{12}$$

and derivation of corresponding estimator.

The properties of these estimators are not well recognized (up to now the relatively richest experimental and theoretical results concern $B_2^{(o)} = A^{-1}X'Y$). The results [6] about the performance of estimators $B_1^{(o)}$, $B_5^{(o)}$ show that they are relatively robust against outliers.

The estimation quality functionals (10)-(10h) do not lead to estimators which are robust against autocorrelation in Σ (as it is introduced in the models $\mathcal{NM}_2$, $\mathcal{NM}_3^{<'}$; $S\mathcal{NM}_4^{'<}$; $S\mathcal{NM}_5^{<'<}$). These functionals do not contain expressions which represent the autocorrelation in Σ. Before going to some detailed propositions it must be noticed that autocorrelation relationships should somehow "weight" coordinate-wise distances $\rho_y(Y_i, \sum_j X_{ij}\beta_j)$ and a total distance between Y and $X\beta$. Suppose that the weight function is $w^{(2)}(\Omega)$. Then

$$\Phi^{(2)} = w^{(2)}(\Omega)\, \rho_y^{(o)}(Y, X\beta) = \rho_y^{(2)}(Y, X\beta). \tag{13}$$

Some of the possible forms of Φ^2 are as follows

$$\Phi_1^{(2)} = \sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j| \sum_{l=1}^{n} \omega^{lt}, \tag{13a}$$

where ω_{lt} is the (l,t)-element of matrix Ω and ω^{lt} is the (l,t)-element of matrix Ω^{-1} or Ω^{+} if $n_o < n$, (13b) $$\Phi_2^{(2)} = \Big[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^2 \omega^{tt} + \sum_{t=1}^{n}\sum_{l=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|\,\omega^{tl}|Y_l - \sum_{j=1}^{k} X_{lj}\beta_j|\Big]^{1/2} = \|\Omega^{-1/2}(Y - X\beta)\|,$$

(13c) $$\Phi_3^{(2)} = \Big[\sum_{t=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^3 \omega^{tt} + \sum_{t=1}^{n}\sum_{l=1}^{n} |Y_t - \sum_{j=1}^{k} X_{tj}\beta_j|^2\,\omega^{tl}|Y_l - \sum_{j=1}^{k} X_{lj}\beta_j|\Big]^{1/3},$$

(13d) $$\Phi_4^{(2)} = \Phi_1^{(2)} + \Phi_2^{(2)},$$

(13e) $$\Phi_5^{(2)} = \max_{1\leq t\leq n} |Y_t - \sum_{j=1}^{n} X_{lj}\beta_j| \sum_{l=1}^{n} \omega^{lt}, \quad \Phi_6^{(2)} = \frac{\Phi_1^{(2)}}{E_1'\mathbf{1} + \Phi_1^{(2)}},$$

(13f) $$\Phi_7^{(2)} = \frac{\Phi_2^{(2)}}{E_2'\mathbf{1} + \Phi_2^{(2)}}, \quad \Phi_8^{(2)} = \frac{\Phi_3^{(2)}}{E_3'\mathbf{1} + \Phi_3^{(2)}}, \quad \Phi_9^{(2)} = \frac{\Phi_4^{(2)}}{E_4'\mathbf{1} + \Phi_4^{(2)}},$$

$$\Phi_{10}^{(2)} = \frac{\Phi_5^{(2)}}{E_5'\mathbf{1} + \Phi_5^{(2)}},$$

E_i, $i = \overline{2,5}$, are corresponding to the Φ_i, $i = \overline{2,5}$, residuals vectors.

$$(13g)\quad \Phi_{11}^{(2)} = \begin{cases} \sum_{t=1}^{n} t \sum_{l=1}^{n^o} \omega^{lt} & \text{if } Y_t \neq \sum_{j=1}^{k} X_{tj}\beta_j, \\ & t = 1, \ldots, n^o,\ 1 \leqslant n^o \leqslant n, \\ 0 & \text{if } Y_t = \sum_{j=1}^{k} X_{tj}\beta_j, \\ & t = 1, \ldots, n^o,\ 1 \leqslant n^o \leqslant n, \end{cases}$$

$$(13h)\quad \Phi_{12}^{(2)} = \begin{cases} \sum_{t=1}^{n} t\lambda_t & \text{if } Y_t \neq \sum_{j} X_{tj}\beta_j, \\ & t = 1, \ldots, n^o,\ 1 \leqslant n^o \leqslant n, \\ 0 & \text{otherwise, } \lambda_t = \lambda_t(\Omega) \text{ is the t-eigen value of } \Omega, \end{cases}$$

$$(13i)\quad \Phi_{13}^{(2)} = \begin{cases} \varrho_y^{(2)}(Y,X\beta) & \text{if } \varrho_y^{(2)}(Y,X\beta) < 1, \\ 1 & \text{if } \varrho_y^{(2)}(Y,X\beta) \geqslant 1, \end{cases}$$

where

$$w^{(2)}(\Omega) = \sum_{t=1}^{n} \lambda_t^{-1}.$$

The estimators derived from the weighted estimation quality functionals $\{\Phi_l^{(2)}\}_{l=1}^{13}$ are more insensitive to autocorrelation effects than the estimators derived from the functionals $\{\Phi_l^{(0)}\}_{l=1}^{12}$.

In both families of functionals there is no "stabilizer" which will smooth the range of Φ because of ill-posedness of the estimation problem caused by a strong pair-correlation between columns of X (bad-conditioning of data). Such a stabilizer (regulizer), regularizing functionals may have different forms. In general it should be some weight function of distance between β and a priori value of β (that is $b^{(i)}$) or some weighted function of distance between a priori value of Y (i.e. $Y^{(p)}$), and $X\beta$. In both cases, weights themselves should be some functions of stabilizing (regularizing) parameter(s) γ ($\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_k)$), γ, $\gamma_i \in \mathfrak{R}$. General forms of regularizers may be as follows

(14) $$w(\gamma)\, \rho_{\mathcal{B}}(\beta, b^{(i)}),$$

(15) $$w(\gamma)\, \rho_{y}(Y^{(p)}, X\beta),$$

(16) $$w(\Gamma)\, \rho_{\mathcal{B}}(\beta, b^{(i)}),$$

(17) $$w(\Gamma)\, \rho_{y}(Y^{(p)}, X\beta),$$

(18) $$w(\gamma)\, \rho_{\mathcal{B}}(\beta, 0),$$

(19) $$w(\Gamma)\, \rho_{\mathcal{B}}(\beta, 0).$$

Under (14) and $w(\gamma) = \gamma$ we may distinguish

(14a) $$\Phi^{(o)}_{I,i} = \gamma \sum_{j=1}^{k} |\beta_j - b_j^{(i)}| = \gamma |(\beta - b^{(i)})'| \cdot \mathbf{1},$$

(14b) $$\Phi^{(o)}_{II,i} = \gamma \left[\sum_{j=1}^{k} |\beta_j - b_j^{(i)}|^2 \right]^{1/2} = \gamma \| \beta - b^{(i)} \|,$$

(14c) $$\Phi^{(o)}_{III,i} = \gamma \left[\sum_{j=1}^{k} |\beta_j - b_j^{(i)}|^{3/2} \right]^{2/3},$$

(14d) $$\Phi^{(o)}_{IV,i} = \Phi^{(o)}_{I,i} + \Phi^{(o)}_{II,i},$$

(14e) $$\Phi^{(o)}_{V,i} = \Phi^{(o)}_{I,i} + \Phi^{(o)}_{II,i} + \Phi^{(o)}_{III,i},$$

(14f) $$\Phi^{(o)}_{VI,i} = \gamma \max_{1 \leqslant j \leqslant k} |\beta_j - b_j^{(i)}|, \quad \Phi^{(o)}_{VII,i} = \frac{\Phi^{(o)}_{I,i}}{E_I'\mathbb{1} + \Phi^{(o)}_{I,i}},$$

$$\Phi^{(o)}_{VIII,i} = \frac{\Phi^{(o)}_{II,i}}{E_{II}'\mathbb{1} + \Phi^{(o)}_{II,i}},$$

(14g) $$\Phi^{(o)}_{IX,i} = \frac{\Phi^{(o)}_{III,i}}{E_{III}'\mathbb{1} + \Phi^{(o)}_{III,i}}, \quad \Phi^{(o)}_{X,i} = \frac{\Phi^{(o)}_{IV,i}}{E_{IV}'\mathbb{1} + \Phi^{(o)}_{IV,i}},$$

$$\Phi^{(o)}_{XI,i} = \frac{\Phi^{(o)}_{V,i}}{E_V'\mathbb{1} + \Phi^{(o)}_{V,i}}, \quad \Phi^{(o)}_{XII,i} = \frac{\Phi^{(o)}_{VI,i}}{E_{VI}'\mathbb{1} + \Phi^{(o)}_{VI,i}},$$

E_j, $j = I, \ldots, VI$ are corresponding to the $\Phi_{j,i}$, $i = I, \ldots, VI$ residuals vectors.

(14h) $$\Phi^{(o)}_{XIII,i} = \begin{cases} \gamma \sum_{j=1}^{k^o} j \text{ if } \beta_j \neq b_j^{(i)}, \ j = 1, \ldots, j^o, \ 1 \leqslant j^o \leqslant k \\ 0 \quad \text{if } \beta_j = b_j^{(i)}, \ j = 1, \ldots, j, \ 1 \leqslant j^o \leqslant k. \end{cases}$$

Under (15) and $w(\gamma) = \gamma$ we have

(15a) $$\Phi^{(o)}_{I} = \gamma \sum_{t=1}^{n} Y_t^{(p)} - \sum_{j=1}^{k} X_{tj}\beta_j = \gamma(|Y^{(p)} - X\beta|)'\mathbb{1},$$

(15b) $$\Phi^{(o)}_{II} = \gamma\left[\sum_{t=1}^{n} |Y_t^{(p)} - \sum_{j=1}^{k} X_{tj}\beta_j|^2\right]^{1/2} = \gamma \| Y^{(p)} - X\beta \|,$$

(15c) $$\Phi^{(o)}_{III} = \gamma \sum_{t=1}^{n} |Y_t^{(p)} - \sum_{j=1}^{k} X_{tj}\beta_j|^{3/2}\Big]^{2/3},$$

(15d) $$\Phi_{IV}^{(o)} = \Phi_{I}^{(o)} + \Phi_{II}^{(o)},$$

(15e) $$\Phi_{V}^{(o)} = \gamma \max_{1\leq t\leq n} |Y_t^{(p)} - \sum_{j=1}^{k} X_{tj}\beta_j|, \quad \Phi_{VI}^{(o)} = \frac{\Phi_{I}^{(o)}}{E_1'\mathbb{1} + \Phi_{I}^{(o)}},$$

(15f) $$\Phi_{VII}^{(o)} = \frac{\Phi_{II}^{(o)}}{E_{II}'\mathbb{1} + \Phi_{II}^{(o)}}, \quad \Phi_{VIII}^{(o)} = \frac{\Phi_{III}^{(o)}}{E_{III}'\mathbb{1} + \Phi_{III}^{(o)}},$$

$$\Phi_{IX}^{(o)} = \frac{\Phi_{IV}^{(o)}}{E_{IV}'\mathbb{1} + \Phi_{IV}^{(o)}}, \quad \Phi_{X}^{(o)} = \frac{\Phi_{V}^{(o)}}{E_{V}'\mathbb{1} + \Phi_{V}^{(o)}}.$$

Under (16) and $\Gamma = \mathrm{diag}\,(\gamma_1, \ldots, \gamma_k)$ we have

(16a) $$\Phi_{XIV,i}^{(o)} = \sum_{j=1}^{k} \gamma_j|\beta_j - b_j^{(i)}| = (\Gamma|\beta - b^{(i)}|)'\mathbb{1},$$

(16b) $$\Phi_{XV,i}^{(o)} = \left[\sum_{j=1}^{k} \gamma_j|\beta_j - b_j^{(i)}|^2\right]^{1/2} = \|\Gamma^{1/2}|\beta - b^{(i)}|\|,$$

(16c) $$\Phi_{XVI,i}^{(o)} = \left[\sum_{j=1}^{k} \gamma_j|\beta_j - b_j^{(i)}|^{3/2}\right]^{2/3},$$

(16d) $$\Phi_{XVII,i}^{(o)} = \Phi_{XIV,i}^{(o)} + \Phi_{XV,i}^{(o)}, \quad \Phi_{XVIII,i}^{(o)} = \Phi_{XIV,i}^{(o)} + \Phi_{XV,i}^{(o)} + \Phi_{XVI,i}^{(o)},$$

(16e) $$\Phi_{XIX,i}^{(o)} = \max_{1\leq j\leq k} \gamma_j|\beta_j - b_j^{(i)}|, \quad \Phi_{XX,i}^{(o)} = \frac{\Phi_{XIV,i}^{(o)}}{E_{XIV}'\mathbb{1} + \Phi_{XIV,i}^{(o)}},$$

$$\Phi_{XXI,i}^{(o)} = \frac{\Phi_{XV,i}^{(o)}}{E_{XV}'\mathbb{1} + \Phi_{XV,i}^{(o)}},$$

(16f) $$\Phi^{(o)}_{XXII} = \frac{\Phi^{(o)}_{XVI,i}}{E'_{XVI}\mathbf{1} + \Phi^{(o)}_{XVI,i}}, \quad \Phi^{(o)}_{XXIII} = \frac{\Phi^{(o)}_{XVII,i}}{E'_{XVII}\mathbf{1} + \Phi^{(o)}_{XVII,i}},$$

$E^{(o)}_j$, j = XIV, ..., XVII, are corresponding residuals vectors.

Under the general regularizer (17) and w (Γ) = Γ we have

(17a) $$\Phi^{(o)}_{XI} = \sum_{t=1}^{n} |Y^{(p)}_t - \sum_{j=1}^{k} \gamma_j X_{tj}\beta_j|,$$

(17b) $$\Phi^{(o)}_{XII} = \left[\sum_{t=1}^{n} |Y^{(p)}_t - \sum_{j=1}^{k} \gamma_j X_{tj}\beta_j|^2\right]^{1/2},$$

(17c) $$\Phi^{(o)}_{XIII} = \left[\sum_{t=1}^{n} |Y^{(p)}_t - \sum_{j=1}^{k} \gamma_j X_{tj}\beta_j|^{3/2}\right]^{2/3},$$

(17d) $$\Phi^{(o)}_{XIV} = \Phi^{(o)}_{XI} + \Phi^{(o)}_{XII},$$

(17e) $$\Phi^{(o)}_{XV} = \max_{1<t<n} |Y^{(p)}_t - \sum_{j=1}^{k} \gamma_j X_{tj}\beta_j|, \quad \Phi^{(o)}_{XVI} = \frac{\Phi^{(o)}_{XI}}{E'_{XI}\mathbf{1} + \Phi^{(o)}_{XI}},$$

(17f) $$\Phi^{(o)}_{XVII} = \frac{\Phi^{(o)}_{XII}}{E'_{XII}\mathbf{1} + \Phi^{(o)}_{XII}}, \quad \Phi^{(o)}_{XVIII} = \frac{\Phi^{(o)}_{XIII}}{E'_{XIII}\mathbf{1} + \Phi^{(o)}_{XIII}},$$

$$\Phi^{(o)}_{XIX} = \frac{\Phi^{(o)}_{XIV}}{E'_{XIV}\mathbf{1} + \Phi^{(o)}_{XIV}}, \quad \Phi^{(o)}_{XX} = \frac{\Phi^{(o)}_{XV}}{E'_{XV}\mathbf{1} + \Phi^{(o)}_{XV}},$$

E_j, j = XI, ..., XV, are corresponding residuals vectors.

Under the general regularizer (18) and w (γ) = γ we have the following explicit forms

(18a) $$\Phi^{(o)}_{XXI} = \gamma \sum_{j=1}^{k} |\beta_j|,$$

(18b) $$\Phi^{(o)}_{XXII} = \left[\sum_{j=1}^{k} \gamma |\beta_j|^2 \right]^{1/2},$$

(18c) $$\Phi^{(o)}_{XXIII} = \left[\sum_{j=1}^{k} \gamma |\beta_j|^{3/2} \right]^{2/3},$$

(18d) $$\Phi^{(o)}_{XXIV} = \Phi^{(o)}_{XXI} + \Phi^{(o)}_{XXII}, \quad \Phi^{(o)}_{XXV} = \Phi^{(o)}_{XXIV} + \Phi^{(o)}_{XXIII},$$

(18e) $$\Phi^{(o)}_{XXVI} = \max_{1 \leq j \leq k} \gamma |\beta_j|, \quad \Phi^{(o)}_{XXVII} = \frac{\Phi^{(o)}_{XXI}}{E'_{XXI}\mathbb{1} + \Phi^{(o)}_{XXI}},$$

(18f) $$\Phi^{(o)}_{XXVIII} = \frac{\Phi^{(o)}_{XXII}}{E'_{XXII}\mathbb{1} + \Phi^{(o)}_{XXII}}, \quad \Phi^{(o)}_{XXIX} = \frac{\Phi^{(o)}_{XXIII}}{E'_{XXIII}\mathbb{1} + \Phi^{(o)}_{XXIII}},$$

$$\Phi^{(o)}_{XXX} = \frac{\Phi^{(o)}_{XXIV}}{E'_{XXIV}\mathbb{1} + \Phi^{(o)}_{XXIV}},$$

E_j, j = XXI, ..., XXIV, are corresponding to $\Phi^{(o)}_j$ residuals vectors.

For (19) and $w(\Gamma) = \Gamma$ we have

(19a) $$\Phi^{(o)}_{XXXI} = \sum_{j=1}^{k} \gamma_j |\beta_j|, \quad \Phi^{(o)}_{XXXII} = \left[\sum_{j=1}^{k} \gamma_j |\beta_j|^2 \right]^{1/2},$$

(19b) $$\Phi^{(o)}_{XXXIII} = \left[\sum_{j=1}^{k} \gamma_j |\beta_j|^{3/2} \right]^{2/3}, \quad \Phi^{(o)}_{XXXIV} = \Phi^{(o)}_{XXXI} + \Phi^{(o)}_{XXXII},$$

(19c) $$\Phi^{(o)}_{XXXV} = \max_{1 \leq j \leq k} \gamma_j |\beta_j|, \quad \Phi^{(o)}_{XXXVI} = \frac{\Phi^{(o)}_{XXXI}}{E'_{XXXI}\mathbb{1} + \Phi^{(o)}_{XXXI}},$$

$$(19d)\quad \Phi^{(o)}_{XXXVII} = \frac{\Phi^{(o)}_{XXXII}}{E'_{XXXII}\mathbf{1} + \Phi^{(o)}_{XXXII}}, \quad \Phi^{(o)}_{XXXVIII} = \frac{\Phi^{(o)}_{XXXIII}}{E'_{XXXIII}\mathbf{1} + \Phi^{(o)}_{XXXIII}},$$

$$\Phi^{(o)}_{XXXIX} = \frac{\Phi^{(o)}_{XXXIV}}{E'_{XXXIV}\mathbf{1} + \Phi^{(o)}_{XXXIV}}, \quad \Phi^{(o)}_{XL} = \frac{\Phi^{(o)}_{XXXV}}{E'_{XXXV}\mathbf{1} + \Phi^{(o)}_{XXXV}},$$

E_j, j = XXXI, ..., XXXV are corresponding residuals vectors.

N o t i c e 1. Possible options for $B^{(i)}$ are:

$$B^{(i)} = \left[X^{(i)\prime}X^{i}\right]^{-1}X^{(i)\prime}Y^{(i)}, \quad N_{(i)} = \{i_1, \ldots, i_{ni}\},$$

n is the size of sample "i", $N_{(i)}$ is the set of sample "i" indices: $Y^{(i)} = (Y_{i_1}, Y_{i_2}, \ldots, Y_{i_{n_i}})$, $X^{(i)} = (X_1^{(i)} \mid X_2^{i} \mid \ldots X_k^{(i)})$, $X^{(i)}_{\cdot j} = (X^{(i)}_{i_1,j}, \ldots, X^{(i)}_{i_{n_i},j})'$, $j = 1, \ldots, k$; $b^{(i)}$ is therefore the vector of values of estimator $B^{(i)}$ obtained by using sample $(X^{(i)}, Y^{(i)})$, $i = i_1, \ldots, i_{n_i}$, where i_i is the number of data samples which are at our disposal a priori.

N o t i c e 2. Possible choices for $Y^{(p)}$ are:

a) $Y^{(p)}: = Y^{(i)}$, $Y^{(i)} = (Y_{i_1}, \ldots Y_{i_{n_i}})'$, $i = i_1, \ldots, i_{n_i}$ are other a priori n-size samples of Y obtained out of model-like modes,

b) $Y^{(p)}: = Y^{(p)}$, $p \neq i$, $p = p_1, \ldots, p_p$ are other a priori n-size sample values of Y obtained by using model-like modes (not necessary linear models and not necessary with the same number of explanatory variables); in this case $Y^{(p)}$ can be interpreted as the p-th kind predictor of Y, which by definition, is a function of the assumed form of estimator $B^{(p)}$ for the parameters of p-th kind model that approximates Y.

The two kinds of functionals (distinguished by Arabic and Roman numerals as subscripts) introduced above may be matched with each other. At the outset of this matching we obtain:

(20a) $\Phi^{(o,o)}_{l,m} := \Phi^{(o)}_{l} + \Phi^{(o)}_{m}$, $l = 1, \ldots, 14;$ $m = I, \ldots, XL,$

(20b) $\Phi^{(o,o)}_{l,q,i} := \Phi^{(o)}_{l} + \Phi^{(o)}_{q,i}$, $\begin{cases} l = 1, \ldots, 14; & q = I, \ldots, XXIII, \\ i = i_1, \ldots, i_{n_i}, \end{cases}$

(20c) $\Phi^{(2,o)}_{\nu,m} := \Phi^{(2)}_{l} + \Phi^{(o)}_{m}$, $\nu = 1, \ldots, 13,$ $m = I, \ldots, XL,$

(20d) $\Phi^{(2,o)}_{\nu,q,i} := \Phi^{(2)}_{\nu} + \Phi^{(o)}_{q,i}$, $\begin{cases} \nu = 1, \ldots, 13, & q = I, \ldots, XXIII, \\ i = i_1, \ldots, i_{n_i}. \end{cases}$

The result of matching is, therefore, a great abundance of estimator quality functionals. To each functional corresponds an estimation problem (cf. for instance, (5) in § 3 and next sections) and, as a solution of this problem some estimator. It is easy to see how many different estimators will be obtained on the grounds of the proposed functionals. In the next section we show a few examples how to do it.

5. SOME REGULARIZING ESTIMATORS DERIVED FROM QUADRATIC FUNCTIONALS

For the functional $\Phi^{(o)}_{2,II}$ the estimation problem is

(21) $$\min_{\beta} \Phi^{(o)}_{2,II}(\beta).$$

Using differential calculus we have its intermediate solution in the form

$$\frac{\partial\Phi^{(o)}_{2,II}(\beta)}{\partial\beta} = \frac{\partial\Phi^{(o)}_{2}}{\partial\beta} + \frac{\partial\Phi^{(o)}_{II}}{\partial\beta} =$$

$$= \frac{1}{2}2(-X'Y + X'X\beta) + \frac{1}{2}2\gamma[-X'Y^{(p)} + X'X\beta] = 0,$$

and the final solution is

$$\beta: = B^{(o)}_{2,II,p} = \frac{1}{1+\gamma} B^{(o)}_{o} + \frac{\gamma}{1+\gamma}(X'X)^{-1}X'Y^{(p)}, \quad p = p_1, \ldots, p_p.$$

Under the functional $\Phi^{(o)}_{2,XXII}$ the estimation problem is

(22) $$\min_{\beta} \Phi^{(o)}_{2,XXII}(\beta).$$

Its solution may be obtained by solving the following system (with respect to β)

$$\frac{\partial\Phi^{(o)}_{2,XXII}}{\partial\beta} = \frac{2}{2}(-X'Y + X'X\beta) + \frac{2}{2}\gamma\beta = 0.$$

The solution is just Hoerl-Kennard's ridge estimator $\beta: = B^{(o)}_{2,XXII} = (X'X + \gamma I)^{-1}X'Y$.

Under $\Phi^{(o)}_{2,XXXII}$ we shall find the solution of the problem

(23) $$\min_{\beta} \Phi^{(o)}_{2,XXXII}(\beta).$$

Because

$$\frac{\partial\Phi^{(o)}_{2,XXXII}}{\partial\beta} = \frac{2}{2}(-X'Y + X'X\beta) + \frac{2}{2}\Gamma\beta,$$

therefore $\partial\Phi^{(o)}_{2,XXXII}/\partial\beta = 0$ iff

$$\beta: = B^{(o)}_{2,XXXII} = (X'X + \Gamma)^{-1}X'Y.$$

In the case of functional $\Phi^{(o)}_{2,II,i}$ the estimation problem is

(24) $$\min_{\beta} \Phi^{(o)}_{2,II,i}(\beta).$$

Since

$$\partial\Phi^{(o)}_{2,II,i}/\partial\beta = \frac{2}{2}(-X'Y + X'X\beta) + \frac{2}{2}\gamma\beta - \frac{2}{2}\gamma b^{(i)},$$

therefore $\partial\Phi^{(o)}_{2,II,i}/\partial\beta = 0$ iff

$$\beta := B^{(o)}_{2,II,i} = (X'X + \gamma I)^{-1}(X'Y + \gamma b^{(i)}), \quad i = i_1, \dots, i_{n_i}.$$

In the case of $\Phi^{(o)}_{2,XV,i}$ we search for

$$(25) \qquad \min_{\beta} \Phi^{(o)}_{2,XV,i}(\beta).$$

Because $\partial\Phi^{(o)}_{2,XV,i}/\partial\beta = \frac{2}{2}(-X'Y + X'X\beta) + \frac{2}{2}(\Gamma\beta - \Gamma b^{(i)})$, therefore $\partial\Phi^{(o)}_{2,XV,i}/\partial\beta = 0$ iff

$$\beta := B^{(o)}_{2,XV,i} = (X'X + \Gamma)^{-1}(X'Y + b^{(i)}).$$

All these quadratic metric estimators are, more or less, robust against strong correlation between explanatory variables in the model $\mathcal{NM}_o$ and anomal results of observations for Y.

Under the assumptions of model $\mathcal{NM}_o$ and definition of $\mathcal{E}, \mathcal{D}$, MSE we have

$$\mathcal{E}(B^{(o)}_{2,II,p}) = \frac{1}{1+\gamma}\beta + \frac{\gamma}{1+\gamma}(X'X)^{-1}X'\mathcal{E}(Y^{(p)}) \neq \beta, \quad p = p_1, \dots, p_p,$$

$$\mathcal{E}(B^{(o)}_{2,XXII}) = (X'X + \gamma I)^{-1} X'X\beta \neq \beta,$$

$$\mathcal{E}(B^{(o)}_{2,XXXII}) = (X'X + \Gamma)^{-1}X'X\beta \neq \beta,$$

$$\mathcal{E}(B^{(o)}_{2,II,i}) = (X'X + \gamma I)^{-1}(X'X\beta + \gamma b^{(i)}) \neq \beta, \quad i = i_1, \dots, i_{n_i},$$

$$\mathcal{E}(B^{(o)}_{2,XV,i}) = (X'X + \Gamma)^{-1} (X'X\beta + \Gamma b^{(i)}) \neq \beta, \quad i = i_1, \dots, i_{n_i}.$$

The relations confirm biasedness of all chosen quadratic metric regularizing estimators.

If $\operatorname{cov}(B^{(o)}_o, Y^{(p)}) = 0$, then

$$\mathcal{D}(B^{(o)}_{2,II,p}) = \mathcal{E}(B^{(o)}_{2,II,p} - \mathcal{E}(B^{(o)}_{2,II,p}))(B^{(o)}_{2,II,p} - \mathcal{E}(B^{(o)}_{2,II,p}))'$$

$$\mathcal{D}(B^{(o)}_{2,II,p}) = \left(\frac{1}{1+\gamma}\right)^2 \sigma^2 (X'X)^{-1} + \left(\frac{\gamma}{1+\gamma}\right)^2 (X'X)^{-1} X'\mathcal{D}(Y^{(p)}) X (X'X)^{-1},$$

$$\mathcal{D}(B^{(o)}_{2,XXII}) = \sigma^2 (X'X + \gamma I)^{-1} X'X (X'X + \gamma I)^{-1},$$

$$\mathcal{D}(B^{(o)}_{2,XXXII}) = \sigma^2 (X'X + \Gamma)^{-1} X'X (X'X + \Gamma)^{-1},$$

$$\mathcal{D}(B^{(o)}_{2,II,i}) = \mathcal{D}(B^{(o)}_{2,XXII}),$$

$$\mathcal{D}(B^{(o)}_{2,XV,i}) = \mathcal{D}(B^{(o)}_{2,XXXII}).$$

By definition $MSE(B^{(o)}_j) = \mathcal{E}((B_j - \beta)'(B^{(o)}_j - \beta))$. For biased estimators (see T h e o b a l d [7]) we have

$$MSE(B^{(o)}_j) = tr\, \mathcal{D}(B^{(o)}_j) + tr\, (bias\, B^{(o)}_j))(bias\, (B^{(o)}_j))'.$$

Using this relation it is easy to derive mean square errors for our chosen estimators. For example,

$$MSE(B^{(o}_{2,XXII}) = tr\, \mathcal{D}(B^{(o)}_{2,XXII}) + tr\, ((X'X + \gamma I)^{-1} X'X -$$

$$- I)\beta\beta'((X'X + \gamma I)^{-1} X'X - I)'.$$

Because $tr\, A = tr\, P'AP$, $P'P = PP' = I$, therefore,

$$MSE(B^{(o)}_{2,XXII} = \sigma^2 tr\, (\Lambda + \gamma I)^{-1} \Lambda (\Lambda + \gamma I)^{-1} +$$

$$+ tr\, ((\Lambda + \gamma I)^{-1} \Lambda - I)\alpha\alpha'(\Lambda + \gamma I)^{-1} \Lambda - I)',$$

where $\Lambda = diag(\lambda_1, \ldots, \lambda_k)$, is the diagonal matrix of eigen values for $X'X$.

$$MSE(B^{(o)}_{2,XXII} = \sigma^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + \gamma)^2} + \gamma^2 \sum_{i=1}^{k} \frac{\alpha_i^2}{(\lambda_i + \gamma)^2}, \quad \alpha = P'\beta.$$

Regularizing estimator $B^{(o)}_{2,XXII}$ dominates $B^{(o)}_2 = B^{(o)}_o$ in the sense of MSE iff $MSE(B^{(o)}_{2,XXII}) < MSE(B^{(o)}_{2})$. The latter relation holds iff

$$\frac{\gamma^2 \sum_i \lambda_i \alpha_i^2}{2\gamma \sum_{i=1}^{k} \lambda_i + k\gamma^2} < \sigma^2. \tag{26}$$

Applying similar reasoning for the estimator $B^{(o)}_{2,XXXII}$ we have

$$MSE(B^{(o)}_{2,XXXII}) = \sigma^2 \sum_{i=1}^{k} \frac{\lambda_i}{(\lambda_i + \gamma_i)^2} + \sum_{i=1}^{k} \frac{\alpha_i^2}{\left(1 + \frac{\lambda_i}{\gamma_i}\right)^2}.$$

Hence

$$MSE(B^{(o)}_{2,XXXII}) < MSE(B^{(o)}_{2}) \quad \text{iff}$$

$$\frac{\sum_{i=1}^{k} \gamma_i 2\lambda_i \alpha_i^2}{2 \sum_{i=1}^{k} \lambda_i \gamma_i + \sum_{i=1}^{k} \gamma_i^2} < \sigma^2. \tag{27}$$

For $(B^{(o)}_{2,II,i})$ we have under $\underline{b}^{(i)} = P'b^{(i)}$

$$MSE(B^{(o)}_{2,II,i}) = \sigma^2 \sum_{j=1}^{k} \frac{\lambda_j}{(\lambda_j + \gamma)^2} + \sum_{j=1}^{k} \frac{\alpha_j^2}{(1 + \gamma^{-1}\lambda_j)^2} -$$

$$- 2\gamma^2 \sum_{j=1}^{k} \frac{\underline{b}_j^{(i)}\alpha_j}{(\lambda_j + \gamma)^2} + \gamma^2 \sum_{j=1}^{k} \frac{(\underline{b}_j^{(i)})^2}{(\lambda_j + \gamma)^2}$$

and hence

$$MSE(B^{(o)}_{2,II,i}) < MSE(B^{(o)}_{2}) \quad \text{iff}$$

$$(28)\qquad \frac{\gamma^2 \sum_{j=1}^{k} \lambda_j \alpha_j^2 - 2\gamma^2 \sum_{j=1}^{k} \underline{b}_j^{(i)} \lambda_j \alpha_j + \gamma^2 \sum_{j=1}^{k} \lambda_j (\underline{b}_j^{(i)})^2}{2\gamma \sum_{j=1}^{k} \lambda_j + k\gamma^2} < \sigma^2.$$

From relations (26), (27), (28) which determine the domination of estimators $B^{(o)}_{2,XXII}$, $B^{(o)}_{2,XXXII}$, $B^{(o)}_{2,II,i}$ (domination in the MSE sense) over the Gauss-Legendre estimator $B^{(o)}_{2}$. it is easy to find conditions on the regularization parameters $\gamma(\gamma_1)$ which assure the truth of these relations. The calculation of values for $\gamma(\gamma_1)$ may be done by solving the equation $\| Y - X B^{(o)}_{2,m} \| = \delta^2$, where δ^2 is such that $\| Y - XB \| \leqslant \delta^2$, $m = XXII, XXXII$ or solving the equation $\| Y - XB^{(o)}_{2,II,i} \| = \delta^2$ (solving all the two equations with respect to γ, γ_1).

6. FINAL REMARKS

We have proposed an unified approach to the analysis of regularizing estimators. Our analysis is by no means complete. The careful reader of this paper may easily see that the presented classification of regularizing estimators would be extended by introducing:

- different assumptions concerning the form of estimators for γ and γ_1,
- assumptions that we do not know a priori the vector $b^{(i)}$ but we shall estimate it by using some estimators $B^{(i)}$,
- some additional weighting of regularizing distance between B and $B^{(i)}$. These weights should tell us about dispersion in $B^{(i)}$, and/or dispersion in Y.

In writing this text we were mainly inspired by the works of Tikhonov, Arshenin [8], Lanczos [3], Hoerl, Kennard [1, 2], Morozov [5].

The text is the extension of Milo's [4] work done within the contract R.III.9.5.7. More detailed presentation of the above results can be found first in our 1981, 1982 works under R.III.9..

In this paper we have derived, as an illustration, only five different estimators corresponding to five different estimation quality functionals. The detailed analysis of their properties as well as the derivation of other estimators will be presented in the subsequent paper.

REFERENCES

[1] Hoerl A., Kennard R. (1970a): *Ridge Regression. Biased Estimation for Non-Orthogonal Problems*, "Technometrics", 12, p. 55-67.

[2] Hoerl A., Kennard R. (1970b): *Ridge Regression. Applications to Non-Orthogonal Problems*, "Technometrics", 12, p. 69-82.

[3] Lanczos C. (1958): *Iterative Solution of Large-Scale Linear Systems*, J.SIAM, 1, p. 91.

[4] Milo W., (1978): *Estymacja parametrów ogólnych modeli liniowych. Cz. III*, work within the contract R.III.9.5.7, p. 1-39.

[5] Morozov V. (1966): *O reshenyi funkcyonalnykh uravnenij metodom regularyzacji*, DAN SSSR, 166(3), p. 510-512.

[6] Ronner A. (1977): *P-norm Estimators in a Linear Regression Model*, Groningen VRB Drukkerijen bv.

[7] Theobald C. (1974): *Generalizations of Mean Square Errors Applied to Ridge Regression*, J. Roy. Statist. Soc., Ser. B, 36, p. 103--106.

[8] Tikhonov A., Arshenin V. (1974): *Metody reshenya nekorektnykh zadach*, Moskva, Nauka.

Władysław Milo

O OBCIĄŻONYCH ESTYMATORACH REGULARYZUJACYCH. CZĘŚĆ I

Celem artykułu jest przedstawienie:

a) numerycznej analizy konsekwencji złego uwarunkowania macierzy X'X,

b) propozycji definicji źle postawionych zadań estymacji parametrów ogólnych modeli liniowych,

c) zunifikowanej statystycznej analizy estymatorów regularyzujących.

Dla ilustracji wprowadzono konkretne postacie estymatorów z kwadratowego funkcjonału jakości estymacji oraz podano warunki dominacji tych estymatorów nad estymatorem Gaussa-Legendre'a w sensie błędu średniokwadratowego.