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 48, 1985

Władysław Milo*, Zbigniew Wasilewski**

SOME PROBLEMS OF ROBUST ESTIMATION IN THE CASE OF LINEAR MODELS PART 1: CHARACTERIZATION OF ROBUSTNESS

1. Introduction

We limit our attention to the robustness of point estimation methods for the following linear models

$$\mathcal{M}_0 = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi, k_0 = k, n_0 = n, \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \sigma^2 I)),$$

$$\mathcal{M}_1 = (\mathcal{R}^{n\times k}, \mathcal{S}, Y = X\beta + \Xi, k_0 = k, n_0 = n, \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \Omega)),$$

where:

$\mathcal{R}^{nxk}$ - the set of real nxk matrices,

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

$k_0 = \text{rank}(X)$, $n_0 = \text{rank}\ \mathcal{B}(Y)$, $X \in R^{nxk}$,

$\mathcal{B}(Y)$ - dispersion of random vector Y, i.e.

$\mathcal{E}$ - expectation operator,

"$\mathcal{P}_Y = \mathcal{N}_Y(X\beta,\Omega)$" - "n-dimensional random vector has n-dimensional normal distribution with $\mathcal{E}(Y) = X\beta$ and $\mathcal{B}(Y) = \Omega$".

The concept of robustness was introduced into statistics by B o x, A n d e r s o n [5] and it was concerned with the robustness of tests. A more formal presentation of the robustness

*Dr., Lecturer at the Institute of Econometrics and Statistics, University of Łódź.
**Senior Assistant at the Institute of Econometrics and Statistics, University of Łódź.

ideas was introduced by Huber [9, 10, 11], Hampel [6, 7], Bickel [4], Berger [3], Bartoszyński, Pleszczyńska [1], Zieliński [20], Huber, Carol [12], and Rousseeuw [18].

2. Robustness: Intuitive and Heuristic Meaning

The broad and intuitive meaning of robustness should be characterized more formally in order to achieve greater precision. Posing the following questions should help in it.

Q1) What is robust?

Q2) Against what something is robust? (against what kind of changes (deviation, perturbation, disturbances, transformations) this something is robust)?

Q3) In what properties of this something the robustness is manifested? (with respect to which properties the robustness is manifested?)

Q4) With respect to which qualitative measures of performance this something is robust?

Q5) How to measure robustness?

Q6) What do we loose in return to robustness?

Q7) To what degree something is robust?

Q8) Whether the concepts "stability of something" and "robustness of something" are different in meaning?

Brief qualitative answers are as follows:

Ad Q1) Robust is:

a) a statistical (econometric) method of estimation, prediction, testing (an estimator, a test, a predictor),

b) a statistical model,

c) a numerical algorithm of a given method.

Ad Q2) We postulate the robustness of a given method (or an algorithm) against the changes of:

ch1) assumptions of stochastic structure of the given statistical model,

ch2) assumptions of nonstochastic structure of the given model,

ch3) assumptions of both structures.

Possible changes (ch1) concern:

ch1.1) probability measures,

ch1.2) shifts in the parameter value of probability distribution,

ch1.3) replacing standard distribution by a superposition of distributions belonging to different classes,

ch1.4) distances between distributions.

Possible changes (ch2) concern:

ch2.1) levels of bad-conditioning of data matrix,

ch2.2) shapes of relationships (e.g. from linear into non-linear).

Possible changes (ch3) concern the elements of cross-classification of (ch1) and (ch2).

Ad Q3) We postulate that the robustness should manifest in the invariance (against changes (ch1)-(ch3)) of one or more of the following properties:

p1) unbiasedness,

p2) efficiency,

p3) precision,

p4) consistency,

p5) sufficiency,

p6) stability,

p7) shape of distribution.

Confirming or not-confirming the invariance of some enumerated properties is nothing more than a qualitative analysis. It helps to say that something is or is not robust in a sense and nothing more.

Ad Q4) We postulate that the robustness will manifest in the relative stability of the range for the following functions of statistical procedures:

f1) total bias,

f2) efficiency,

f3) total mean square error,

f4) standard error,

f5) average relative absolute prediction error,

f6) weighted total sum of prediction errors,

f7) speed of convergence of iterative statistical procedures,

f8) deviation between the p-th quantiles of distributions of statistics,

f9) distances between probability distributions of statistics (measured by distances between distribution functions or characteristic functions or moment generating functions),

f10) values of Gâteaux derivative of statistical procedure or logarithms of statistics' variance.

Ad Q5) Robustness measures will be defined by using (f1)-(f10). There are, among others, the following options:

- a robustness measure is the diameter of one of the functions (f1)-(f10) ranging for the given variant of changes (ch1)-(ch3) or their mixtures,

- a robustness measure is the ratio of the diameter of one of the ranges of functions (f1)-(f10) and changes versions (ch1)-(ch3) for a given set of standard model neighbourhoods to the diameter of the set of ranges of distances between neighbourhoods of a standard model (or a robustness measure is the limit of this ratio when the denominator approaches to zero),

- a robustness measure is the difference between the supremum of one of the ranges of functions (f1)-(f10) and the corresponding infimum defined on elements of particular neighbourhood of standard model,

- a robustness measure is the distance between a value of one of the functions (f1)-(f10) calculated for the given standard model and the range of the corresponding function calculated for the elements of particular neighbourhood of the standard model,

- a robustness measure is the distance between the set of all ranges of functions (f1)-(f10) calculated for the given standard model and the set of all ranges of functions (f1)-(f10) calculated for the set of neighbourhoods of standard model (set of supermodels).

Ad Q6) Robust methods are, in general, more numerically complex than non-robust ones. It means that the numerical cost of using them (as measured by e.g. computer time usage) is greater. We are not in a position to give reasonable comparisons of benefits and losses when using robust methods.

Ad Q7) A degree of robustness should be fixed on the grounds of indications given by the measures proposed in Ad Q5.

Ad Q8) The concept of "stability" was defined precisely within mechanics by Poincaré and next by Lapunoff. U l a m [19]

discusses it in a more general context of stability of mathematical theorems. He understands by it the property that while changing assumptions of a theorem "a little", the truth of the theorem is fully preserved or "approximately" preserved.

Interesting propositions of defining stability were given by Zolotarev [22, 23] and extended by Bednarek-Kozek, Kozek [2], and Bartoszyński, Pleszczyńska [1].

Here we are trying to make the concept of stability more concrete. Propositions of characterizations and robustness measures presented in § 3 and 4 were first formulated by Milo in the work of Milo, Wasilewski [15]. A specified kind of given method's robustness against some changes of standard model would be defined by using the concept of stability of measures belonging to the measures types (M1)-(M5).

3. A Characterization of Linear Standard Model's Neighbourhoods

First we need to define neighbourhoods of Y. They can be defined in terms of Levy's, Procherov's, Meshalkin's, Kolmogorov's or total variation metrics.

Due to the definition of random vector Y as

$$Y : (\Omega, \mathcal{F}, P) \longrightarrow (\mathcal{R}^n, \mathcal{F}_{\mathcal{R}^n}, P_Y)$$

we will use further the probability space $(\mathcal{R}^n, \mathcal{F}_{\mathcal{R}^n}, P_Y)$. We will metrize the space R^n.

After metrizing the space R^n becomes the metric space $(R^n, \varrho_{\mathcal{R}^n})$ (or shortly, if possible, writing R^n, ϱ).

Let μ_1, μ_2 be two non-negative finite measures defined on $\mathcal{F}_{\mathcal{R}^n}$ ($\mathcal{F}_{\mathcal{R}^n}$ is σ-field of Borel subsets of (R^n, ϱ)). Then we can define the above mentioned metrics as follows

a) Meshalkin's metrics [13]:

$$\varrho_M(\mu_1, \mu_2) = \sup_A \{|\mu_1(A) - \mu_2(A)|\}.$$

where: A - an intersection of at most n half-spaces, a half-space is an arbitrary set of points in R^n determined by

$$\left\{(y_1, \ldots, y_n)' \in R^n : \sum_{i=1}^{n} a_i y_i < b\right\}; \quad a_i, b \in R^1;$$

b) Levy's metrics

$$\varrho_L(\mu_1, \mu_2) = \max\left\{\varrho_L^*(\mu_1, \mu_2), \varrho_L^{**}(\mu_2, \mu_1)\right\},$$

where:

$$\varrho_L^*(\mu_1, \mu_2) = \inf_{\delta>0}\left\{\delta : \mu_1\left\{(-\infty, y]\right\} \leq \mu_2\left\{(-\infty, y]^\delta\right\} + \delta, \ \forall y \in R^n\right\},$$

$$\varrho_L^{**}(\mu_2, \mu_1) = \inf_{\delta>0}\left\{\delta : \mu_2\left\{(-\infty, y]\right\} \leq \mu_1\left\{(-\infty, y]^\delta\right\} + \delta, \ \forall y \in R^n\right\},$$

$$(-\infty, y]^\delta = \left\{z \in R^n : \varrho_{R^n}(z, (-\infty, y]) < \delta\right\}$$

c) Prochorov's metrics

$$\varrho_P(\mu_1, \mu_2) = \max\left\{\varrho_P^*(\mu_1, \mu_2), \varrho_P^{**}(\mu_2, \mu_1)\right\}$$

where

$$\varrho_P^*(\mu_1, \mu_2) = \inf_{\delta>0}\left\{\delta : \mu_1(A) \leq \mu_2(A^\delta) + \delta, \quad \forall A \in \mathcal{F}_{R^n}\right\},$$

$$\varrho_P^{**}(\mu_2, \mu_1) = \inf_{\delta>0}\left\{\delta : \mu_2(A) \leq \mu_1(A^\delta) + \delta, \ \forall A \in \mathcal{F}_{R^n}\right\},$$

$$A^\delta = \left\{y \in R^n : \varrho_{R^n}(y, A) < \delta\right\}$$

d) Kolmogorov's metrics

$$\varrho_K(\mu_1, \mu_2) = \sup_A\left\{|\mu_1(A) - \mu_2(A)| : A = (-\infty, y], y \in R^n\right\},$$

e) total variation metrics

$$\rho_{TV}(\mu_1, \mu_2) = \sup_{A \in \mathcal{F}_{R^n}} \{|\mu_1(A) - \mu_2(A)|\}.$$

In formulating experiments schemes one can replace μ_1, μ_2 with the corresponding distribution functions F_1, F_2 (or characteristic functions $\varphi_1(t)$, $\varphi_2(t)$). The enumerated metrics did not exhaust the list of all possible metrics of probability space $(R^n, \mathcal{F}_{R^n}, \mathcal{P}_Y)$. They enable us, however, to define η_{il}, i.e. the i-th neighbourhood of measure $\mu_1 = \mathcal{P}_Y = \mathcal{N}_Y$ with respect to the metrics l (where i - for a given metrics l denotes the index number attached to the given i-th value of η_l, e.g., for $i = \overline{1, 3} = 1, 2, 3$ we have $\eta_{11} = 3 \cdot 10^0$, $\eta_{21} = 3 \cdot 10^{-1}$, $\eta_{31} = 3 \cdot 10^{-2}$).
Thus,

D e f i n i t i o n 1. By η_{il} - neighbourhood of measure $\mathcal{P}_Y = \mu_1 = \mathcal{N}_Y(X\beta, \delta^2 I)$ we understand such set $U_{\eta_{il}}$ of distributions that the elements of U are distant from the measures μ_2 by the value η_{il}, i.e.

$$U_{\eta_{il}} = \{\mu_2 : \rho_l(\mu_1, \mu_2) < \eta_{il}\}, \quad l = M, L, P, K, TV.$$

Using Def. 1 we can define η_{il} - neighbourhood of the standard model $\mathcal{NM}_o$ as follows:

D e f i n i t i o n 2. By η_{il} - neighbourhood of model $\mathcal{NM}_o$ we understand such a set $\{\mathcal{NM}_o\}_{\eta_{il}}$ that $\{\mathcal{NM}_o\}_{\eta_{il}} \equiv \{R^{n \times k}, \mathcal{S}, Y = X\beta + \Xi, k_o = k, n_o = n, U_{\eta_{il}}\}$. ◆

While using the name of $\{\mathcal{NM}_o\}_{\eta_{il}}$ in defining the robustness measures is very fruitful in constructing schemes of experiments, it turns out on the other hand, that other names are also convenient (e.g. "η_{il} - set of supermodels with respect to the standard model $\mathcal{NM}_o$" or "η_{il} - set of probabilistic extensions of the standard model $\mathcal{NM}_o$").

Note: if $i = \overline{1,10}$, then under Def. 1,2 there are 10 x 5 values of η_{il} therefore it is 50 sets of supermodels for the $\mathcal{NM}_o$.

Denoting $\mu_1 = \mathcal{P}_Y = \mathcal{N}_Y(X\beta, \Omega)$ we can obtain

$$\left\{\mathcal{N}\mathcal{M}_1\right\}_{\eta_{il}} = \left(R^{n\times k}, \mathcal{B}, Y = X\beta + \Xi, k_o = k, n_o = n, U_{\eta_{il}}\right),$$

that is η_{il} - neighbourhood of model $\mathcal{N}\mathcal{M}_1$.

4. Robustness Measures of Estimators Against Changes of the Model Neighbourhoods

Let bias $(B_j) = E(B_j) - \beta$ be the bias of the j-th kind of estimator B for the parameter vector β from $\mathcal{N}\mathcal{M}_o$ or a model belonging to the set $\left\{\mathcal{N}\mathcal{M}_o\right\}_{\eta_{il}}$. Thus a total bias is tobias $(B_j) = \| E(B_j) - \beta \|$ where $\|\cdot\|$ denotes euclidean vector norm.

Let $\sigma^{(j)}_{\eta_{il}}$ denote the range of tobias (B_j) = tobias $(B_j(\eta_{il}))$ calculated for the particular distributions belonging to the $U_{\eta_{il}}$ - neighbourhood of $\mathcal{N}_Y(X\beta, \sigma^2 I)$ and the corresponding models belonging to the set $\left\{\mathcal{N}\mathcal{M}_o\right\}_{\eta_{il}}$.

In order to be more precise we need to introduce one more index "d" connected with counting the consecutive number of "pure" or "contaminated" distribution (belonging to the $U_{\eta_{il}}$) for example: $d = 1 \longrightarrow (1-\nu)\mathcal{N}_Y(X\beta, \sigma^2 I) + \nu\mathcal{N}_Y(X\beta + \sum_{m\in N_m} \mathring{V}_m j_m, \sigma^2 I)$, where $\mathring{V}_m$ is an atypicality quantity of the m-th component of vector Y, j_m denotes the unit vector of the m-th coordinate axis, N_m is the set of atypical results of observations' indices. Let the range of d be $d = 1, \ldots, d_s$. Then

$$\sigma^{(jd)}_{\eta_{il}} = \left\{\text{tobias}\left(B^{(j)}_d(\eta_{il})\right)\right\}, \qquad d = 1, \ldots, d_s$$

Under the above notations and definitions we have

D e f i n i t i o n 3. The estimator $B^{(j)}$ of β from $\mathcal{N}\mathcal{M}_o$ will be called bias-robust in the $\left\{\mathcal{N}\mathcal{M}_o\right\}_{\eta_{il}}$-neighbourhood of the standard model $\mathcal{N}\mathcal{M}_o$ if the following implication

$$\forall\varepsilon>0\ \exists\eta_{il}>0\ \forall\mu_2\ \forall d : \left(\mu_2 \in U^{(d)}_{\eta_{il}}\right) \longrightarrow \varrho\left(\sigma^{(jd)}_{\eta_{il}}, \sigma^{(j)}\right) < \varepsilon,$$

$$\sigma^{(j)} = \sigma^{(j)}(B(\eta_{il} = 0)).$$

holds. ◆

Note: the above implication can be replaced, e.g. by

$$\forall \varepsilon > 0 \; \exists \eta_{il} > 0 \; \forall \mu_2 \forall d : \left(\mu_2 \in U^{(d)}_{\eta_{il}}\right) \longrightarrow \left(\operatorname{diam}\left(\sigma^{(j)}_{\eta_{il}}\right) < \varepsilon\right),$$

for $\operatorname{diam} A = \sup_{a_1,\, a_2 \in A} \varrho(a_1, a_2)$.

As bias-robustness measures for the estimator B_j one can propose

a) $\mathrm{BROMB}_j 1 = \operatorname{diam} \sigma^{(j,d)}_{\eta_{il}}$,

b) $\mathrm{BROM}\, B_j 2 = \dfrac{\operatorname{diam} \sigma^{(j,d)}_{\eta_{il}}}{\operatorname{diam} U_{\eta_{il}}}$,

c) $\mathrm{BROM}\, B_j 3 = \lim_{\operatorname{diam} U_{\eta_{il}} \to 0} \dfrac{\operatorname{diam} \sigma^{(j,d)}_{\eta_{il}}}{\operatorname{diam} U_{\eta_{il}}}$,

d) $\mathrm{BROM}\, B_j 4 = \sup_d \left\{\mathrm{Tobias}\, (B^{(j,d)}_{\eta_{il}})\right\} - \inf_d \left\{\mathrm{Tobias}\, (B^{(j,d)}_{\eta_{il}})\right\}$,

e) $\mathrm{BROM}\, B_j 5 = \inf_d \varrho\left(\sigma^{(j,d)}_{\eta_{il}}, \sigma^{(j)}\right)$,

f) $\mathrm{BROM}\, B_j 6 = \dfrac{\sup_d \left\{\sigma^{(j,d)}_{\eta_{il}}\right\} - \inf_d \left\{\sigma^{(j,d)}_{\eta_{il}}\right\}}{\sup_d \left\{\sigma^{(j,d)}_{\eta_{il}}\right\}}$,

We say that the estimator B_1 is more bias-robust than the estimator B_2 if

$$\mathrm{BROM}\, B_1 r < \mathrm{BROM}\, B_2 r, \quad r = \overline{1,6}.$$

If someone is interested in the precision of estimators he should study the performance of B_j with respect to the mean-square-error $\mathrm{MSE}(B_j)$, that is the characterization of $\mathrm{MSE}(B_j) = \mathcal{E}\|B_j - \beta\|^2$.

We would be interested in the behaviour of $\mathrm{MSE}(B_j)$ within the probabilistic neighbourhood of $\mathcal{N}\mathcal{M}_0$, i.e. $\left\{\mathcal{N}\mathcal{M}_0\right\}_{\eta_{il}}$.

Let $\mathcal{K}^{(j)}_{\eta_{il}}$ denote the range of $\mathrm{MSE}(B^{(j)}_{\eta_{il}})$ calculated for the distributions of Y belonging to the neighbourhood $U_{\eta_{il}}$ of the distribution $\mathcal{N}_Y(X\beta, \sigma^2 I)$.

For the given distribution measure $\mu_2 \in U^{(d)}_{\eta_{il}}, d = 1, \ldots, d_a$ from this neighbourhood the estimator B_j corresponding to μ_2 will be denoted by $B^{(j,d)}_{\eta_{il}}$. Then

$$\mathcal{K}^{(j)}_{\eta_{il}} = \left\{ MSE\left(B^{(j,d)}_{\eta_{il}}\right),\quad d = 1, \ldots, d_a \right\}.$$

Thus

D e f i n i t i o n 4. The estimator B_j of vector β from $\mathcal{M}_0$ will be called MSE-robust in the neighbourhood $\{\mathcal{M}_0\}_{\eta_{il}}$ if the following implication

$$\forall \varepsilon > 0\ \exists \eta_{il} > 0 \forall \mu_2 \forall d : \left(\mu_2 \in U^{(d)}_{\eta_{il}}\right) \rightarrow \varrho\left(\mathcal{K}^{(j,d)}_{\eta_{il}}, \mathcal{K}^{(j)}\right) < \varepsilon,$$

holds. ◆

Note: another option is

$$\forall \varepsilon > 0\ \exists \eta_{il} > 0 \forall \mu_2 \forall d : \left(\mu_2 \in U^{(d)}_{\eta_{il}}\right) \longrightarrow \left(\operatorname{diam}\left(\mathcal{K}^{(j,d)}_{\eta_{il}}\right) < \varepsilon\right).$$

On the ground of Def. 4 we can define MSE-robustness measures, e. g. as follows

a1) $MSERM\ B_j1 = \operatorname{diam} \mathcal{K}^{(j)}_{\eta_{il}}$,

b1) $MSERM\ B_j2 = \dfrac{\operatorname{diam} \mathcal{K}^{(j)}_{\eta_{il}}}{\operatorname{diam} U_{\eta_{il}}}$,

c1) $MSERM\ B_j3 = \lim_{\operatorname{diam} U^{d}_{\eta_{il}} \to} \dfrac{\operatorname{diam} \mathcal{K}^{(j)}_{\eta_{il}}}{\operatorname{diam} U_{\eta_{il}}}$,

d1) $MSERM\ B_j4 = \sup_d \left\{\mathcal{K}^{(j,d)}_{\eta_{il}}\right\} - \inf_d \left\{\mathcal{K}^{(j,d)}_{\eta_{il}}\right\}\ d = 1, \ldots, d_a$,

e1) $MSERM\ B_j5 = \inf_d \varrho\left\{\mathcal{K}^{(j,d)}_{\eta_{il}}, \mathcal{K}^{(j)}\right\}$,

f1) $MSERM\ B_j6 = \dfrac{MSERM\ B_j4}{\sup_d \mathcal{K}^{(j,d)}_{\eta_{il}}}$.

A relatively synthetic robustness measure is a robustness measure of B_j against the changes of distance between the distribution $\mu_1 = \mathcal{P}_Y$ and other distributions of Y with respect to the changes of distances between corresponding distributions of B_j. We need

D e f i n i t i o n 5. By the neighbourhood $U^{(B_j)}_{\varepsilon_1}$ of $\mu_3 = \mathcal{P}_a$ we understand such a set of probability distributions of

B_j which are not more distant from φ_{B_j} than by a quantity $\bar{\varepsilon}_1$, i.e.

$$U_{\bar{\varepsilon}_1}^{(B_j)} = \left\{ \mu_4 : \varrho(\mu_3, \mu_4) < \bar{\varepsilon}_1 \right\}. \quad \blacklozenge$$

Hence

D e f i n i t i o n 6. The estimator B_j will be called distributionally-robust against changes of distributions of Y within the neighbourhood $U_{\eta_{il}}^{(Y)}$ if the implication

$$\forall \bar{\varepsilon}_1 > 0 \; \exists \eta_{il} \forall \mu_4 \; \forall d : \left(\mu_2 \in U_{\eta_{il}}^{(Y)}\right) \longrightarrow \left(\mu_4 \in U_{\bar{\varepsilon}_1}^{(B_j)}\right),$$

holds. $\blacklozenge$

Under the Def. 6 we can formulate, among others, the following distribution robustness measures:

a2) $\mathrm{DIRM}\, B_j 1 = \operatorname{diam} U_{\bar{\varepsilon}_1}^{(B_j)}$,

b2) $\mathrm{DIRM}\, B_j 2 = \dfrac{\operatorname{diam} U_{\bar{\varepsilon}_1}^{(B_j)}}{\operatorname{diam} U_{\eta_{il}}^{(Y)}}$,

c2) $\mathrm{DIRM}\, B_j 3 = \lim_{\operatorname{diam} U_{\eta_{il}}^{(Y)} \to 0} \mathrm{DIRM}\, B_j 2$,

d2) $\mathrm{DIRM}\, B_j 4 = \sup_d \left\{ \varrho(\mu_3, \mu_4) \right\} - \inf_d \left\{ d(\mu_3, \mu_{4,d}) \right\}$,

$\mu_3 = \varphi_{B_j}$, $\mu_{4,d} \neq \mu_3$, $\mu_{4,d} \in U_{\bar{\varepsilon}_1}^{(B_j)}$,

e2) $\mathrm{DIRM}\, B_j 5 = \inf_d \left(0, U_{\bar{\varepsilon}_1}^{(B_j)}\right)$,

f2) $\mathrm{DIRM}\, B_j 6 = \dfrac{\mathrm{DIRM}\, B_j 4}{\sup_d \left\{ \varrho(\mu_3, \mu_{4,d}) \right\}}$.

In the above definitions the sample size n was fixed. H a m p e l [6] proposes to vary n and consider the robustness of sequence of sample estimators with the varying sample size. Before presenting his idea let us introduce the following notation:

$\mathcal{V} = \{F, G, \ldots, P, Q\}$ - the set of all distributions measures defined on $(\mathcal{U}, \mathcal{F})$;

$\mathcal{V}_n = \{F_n, G_n, \ldots, P_n, Q_n, \ n \geq 1\}$ - the set of all distributions measures defined on $(\mathcal{R}^n, \mathcal{F}_{R^n})$;

$\{B_{(n)}, \ n \geq 1\}$ - the sequence of $B_{(n)} : (\mathcal{U}, \mathcal{F}) \to (\mathcal{R}^n, \mathcal{F}_{R^n})$.

For each sequence of n-element sample generated from $F \in \mathcal{V}$ it corresponds to the empirical distribution from $\mathcal{V}_n$.

The mapping $B_{(n)}$ induces the conditional distribution $\mathcal{L}_F(B_{(n)})$, that is, the distribution of estimator $B_{(n)}$ under the condition that the sample was generated from the distribution F.

D e f i n i t i o n 7. The sequence $\{B_{(n)}, \ n \geq 1\}$ will be called robust with respect to changes in empirical distributions if the following implication

$$\forall \varepsilon_n > 0 \, \exists \eta_n \forall G \forall n : (\varrho(F, G) < \eta_n) \longrightarrow (\varrho(\mathcal{L}_F(B_{(n)}), \mathcal{L}_G(B_{(n)})) < \varepsilon_n)$$

holds. ◆

Definitions 6 and 7 determine the robustness with respect to the estimator's distribution as such a property which causes that the conditional distribution of this estimator is changing "a little" if we change the distribution generating the sample underlying it. These facts also determine postulates that should be addressed to robust estimation methods. Due to space limitation and a homogeneity requirement we will not cover this topic.

5. Final Remarks

The paper contains some ordering of the known statistical concepts connected with the propositions of definitions of robustness measures for the estimators. We have not presented results of experiments carried out by Z. Wasilewski connected with the performance of chosen estimation methods that would be presented in the next paper. We have not described, among others, measures of:

- efficiency-robustness of estimators,
- forecasting power-robustness of estimators,
- numerical bad conditioning-robustness of estimation.

BIBLIOGRAPHY

[1] Bartoszyński R., Pleszczyńska E. (1980), Reducibility of Statistical Structures and Decision Problems, Mathematical Statistics Banach Center Publications, Vol. 6, Warsaw, PWN, p. 29-38.

[2] Bednarek-Kozek B., Kozek A. (1980), Stability, Sensitivity and Sensitivity of Characterizations, Mathematical Statistics Banach Center Publications, Vol. 6, Warsaw, PWN, p. 39-64.

[3] Berger J. (1980), Statistical Decision Theory, Berlin, Springer-Verlag.

[4] Bickel P. (1976), Another Look at Robustness, Scandin. J. of Statistics, p. 145-168.

[5] Box G., Anderson S. (1955), Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumptions, J. Roy. Statist. Soc., Ser. B, p. 1-34.

[6] Hampel F. (1971), A General Qualitative Definition of Robustness, Ann. Math. Statist., p. 1887-1896.

[7] Hampel F. (1973), Robust Estimation, Zeit. für Wahrsch. und Verwandte Gebiete, p. 87-104.

[8] Hogg R. V. (1979), An Introduction to Robust Estimation, [in:] Robustness in Statistics, ed. Launer R. L., N. Y. Academic Press, p. 1-18.

[9] Huber P. (1964), Robust Estimation of a Location Parameter, Ann. Math. Statist., p. 73-101.

[10] Huber P. (1972), Robust Statistics: a Review, Ann. Math. Statist., p. 1041-1067.

[11] Huber P. (1977), Robust Statistical Procedures, Philadelphia: Society for Industrial and Applied Mathematics.

[12] Huber-Carol C. (1978), Quelques problémes de statistique robuste et non-paramétrique. Thése le grade de docteur en sciences, Université de Paris Sud: Centre d'Orsay.

[13] Meshalkin L. (1968), On the Robustness of Some Characterizations of the Normal Distribution, Ann. Math. Statist., p. 1747-1750.

[14] Milo W., Konarzewska I. (1981), Własności

estymatorów obciążonych parametrów modeli liniowych. P.1: Estymatory regularyzujące, Paper within the Grant R.III.9.3.1.

[15] M i l o W., W a s i l e w s k i Z. (1980), Metody estymacji parametrów modeli liniowych w przypadku występowania obserwacji nietypowych, Paper within the Grant R.III.9.5.

[16] P o l l o c k K. (1978), Inference Robustness vs. Criterion Robustness: an Example, The Amer. Statist., November.

[17] R e y W. (1978), Robust Statistical Methods, Lect. Notes in Math. 690, Berlin, Springer-Verlag.

[18] R o u s s e u w P. (1981), New Inifitesimal Methods in Robust Statistics, Ph. D. Thesis, Brussels, Free University of Brussels.

[19] U l a m S. (1960), A Collection of Mathematical Problems, Vol. 8, N.Y. Interscience Publishers.

[20] Z i e l i ń s k i R. (1977), Robustness: a Quantitative Approach, Bull. Acad. Polonaise de Sciences, Serie des Sci. Math. Astr. et Phys., Vol. XXV, No. 12, p. 1281-1286.

[21] Z i e l i ń s k i R. (1980), Robustness: a Quantitative Approach, Banach Center Publications, Vol. 6, Mathematical Statistics, Warsaw, PWN, p. 353-354.

[22] Z o l o t a r e v V. (1975), O nieprerywnosti stokhasticheskikh posledovatelnostey porozhdayemykh rekurentnymi procedurami, Teoria Veroy. i yeyo Prim., Vol. 4, p. 834-847.

[23] Z o l o t a r e v V. (1976), Efekt ustoychivosti chrakterizayi raspredelenity, Zap. Nauch. Sem. LOMI AN SSSR, p. 38-55.

Władysław Milo, Zbigniew Wasilewski

WYBRANE PROBLEMY ODPORNYCH METOD ESTYMACJI PARAMETRÓW MODELI LINIOWYCH. CZ. I: CHARAKTERYSTYKA ODPORNOŚCI

Praca zawiera:

1) opis intuicyjnego i heurystycznego znaczenia odporności,

2) opis jakościowy miar odporności,

3) formalną charakterystykę otoczenia standardowego modelu liniowego,

4) definicje miar odporności estymatorów na zmiany otoczenia danego modelu standardowego (tj. miar odporności obciążenia i miar odporności błędu średniego),

5) definicje odporności w odniesieniu do empirycznego rozkładu estymatora.