

C. PLATT and Z. PAPRZYCKI (Olsztyn)

ON THE ROBUSTNESS OF MULTIPLE REGRESSION COEFFICIENT ESTIMATORS OBTAINED BY THE p -POINT METHOD

1. Introduction. The present paper is a continuation of [4]. Our aim is to compare the robustness of the estimators of the multiple regression coefficients a_0, a_1, \dots, a_p in an extended Gauss-Markoff model, obtained by the p -point method which was described in [4] and [5] with the robustness of the estimators of the same coefficients obtained by the least squares method. The extended Gauss-Markoff model is presented in Section 2.

In [4] and [5] the following model was considered:

$$(1) \quad Y_i = a_0 + a_1(x_{i1} - \bar{x}_1) + \dots + a_p(x_{ip} - \bar{x}_p) + \varepsilon_i,$$

where

x_{i1}, \dots, x_{ip} are the constants determined by the conditions of the i -th sample,

Y_i for $i = 1, \dots, n$ are the independent random variables with the distributions

$$F_i(y) = F(y - (a_0 + a_1(x_{i1} - \bar{x}_1) + \dots + a_p(x_{ip} - \bar{x}_p))),$$

F is any distribution satisfying the conditions

$$E_F(\varepsilon_i) = 0 \quad \text{and} \quad D_F^2(\varepsilon_i) = \sigma^2.$$

In the sequel, the model (1) is called the *model* M_0 . The sample in [4] was determined as follows:

$$(2) \quad C_0 = \{(x_{i1}, \dots, x_{ip}, y_i): i = 1, \dots, n\} \subset R^p \times R.$$

The projection

$$Q(C_0) = \{(x_{i1}, \dots, x_{ip}): i = 1, \dots, n\} \subset R^p,$$

where $Q: R^p \times R \rightarrow R^p$ is a projection operation which does not change the first p coordinates. The image of this operation is called the *plan of an experiment*. In [4] the set C_0 was divided in p manners into two disjoint subsets each time, i.e., if the pairs of subsets $C_1, C'_1, \dots, C_p, C'_p$ are a result of this division, then for each pair C_k, C'_k we have

$$C_k \cup C'_k = C_0 \quad \text{and} \quad C_k \cap C'_k = \emptyset.$$

The estimators of the regression coefficients a_0, a_1, \dots, a_p , obtained by the p -point method are determined as the coefficients of a p -dimensional hyperplane passing through the gravity centers

$$g_0 = (\bar{x}_1, \dots, \bar{x}_p, \bar{y}), \quad g_r = (\bar{x}_1^r, \dots, \bar{x}_p^r, \bar{y}^r), \quad r = 1, \dots, p,$$

of the sets C_0, C_1, \dots, C_p , respectively, where $C_i \subset R^{p+1}$ and $C_i \subset C_0$ for $i = 1, \dots, p$. Assuming that the determinant (denoted by W) of the matrix

$$W = \begin{bmatrix} \bar{x}_1^1 - \bar{x}_1 & \dots & \bar{x}_p^1 - \bar{x}_p \\ \dots & \dots & \dots \\ \bar{x}_1^p - \bar{x}_1 & \dots & \bar{x}_p^p - \bar{x}_p \end{bmatrix}$$

is different from zero, the estimators of the regression coefficients are expressed by the formulas

$$(3) \quad \hat{a}_0 = \bar{y}, \quad \hat{a}_i = W_i/W, \quad i = 1, \dots, p.$$

By a simple transformation we can write (3) as follows:

$$(3') \quad \hat{a}_0 = \bar{y}, \quad \hat{a}_i = a_i + \frac{1}{W} \sum_{r=1}^p (\bar{\varepsilon}^r - \bar{\varepsilon}) W_{ri}, \quad i = 1, \dots, p.$$

The determinant W_i appearing in (3) is obtained from W by replacing the i -th column with the vector $(\bar{y}^1 - \bar{y}, \dots, \bar{y}^p - \bar{y})^T$, W_{ri} is the algebraic complement of an (r, i) entry in the matrix W , and

$$\bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i, \quad \bar{\varepsilon}^r = \frac{1}{k_r} \sum_{i \in C_r} \varepsilon_i, \quad k_r = \text{card } C_r,$$

$$C_0 = \{1, \dots, n\}, \quad C_r = \{i \in C_0: (x_{i1}, \dots, x_{ip}, y_i) \in C_r\}.$$

In [4] it is proved that under the assumptions of the model M_0 the estimators \hat{a}_i are unbiased, consistent and asymptotically normal.

2. The extended Gauss—Markoff model. Let us consider the following extension M of the model M_0 :

$$(4) \quad Y_i = a_0 + a_1(x_{i1} - \bar{x}_1) + \dots + a_p(x_{ip} - \bar{x}_p) + \varepsilon_i, \quad i = 1, \dots, n,$$

where

Y_i for $i = 1, \dots, n$ are independent random variables with the distributions

$$F_i(y) = H_i(y - (a_0 + a_1(x_{i1} - \bar{x}_1) + \dots + a_p(x_{ip} - \bar{x}_p))),$$

H_i are any distributions with the first two moments finite.

From the assumptions (4) it follows immediately that in the model M

$$E_{H_i}(\varepsilon_i) = m_i \quad \text{and} \quad D_{H_i}^2(\varepsilon_i) = \sigma_i^2.$$

THEOREM 1. For $i = 1, \dots, p$

$$(5) \quad \text{bias } \hat{a}_i = E(\hat{a}_i) - a_i = \frac{1}{W} M^T W_i,$$

where

$$\mathbf{M}^T = (M_1 - M_0, \dots, M_p - M_0), \quad \mathbf{W}_i^T = (W_{1i}, \dots, W_{pi}),$$

$$M_r = \frac{1}{k_r} \sum_{i \in C_r} m_i, \quad M_0 = \frac{1}{n} \sum_{i \in C_0} m_i.$$

The proof of Theorem 1 follows from (3) and from the assumptions of the model M .

THEOREM 2. *The risk $R(\hat{a}_i)$ for the estimators \hat{a}_i , $i = 1, \dots, p$, with square loss function is expressed by the formula*

$$(6) \quad R(\hat{a}_i) = E(\hat{a}_i - a_i)^2 = \frac{1}{W^2} \mathbf{W}_i^T \mathbf{A} \mathbf{W}_i + \frac{1}{W^2} (\mathbf{M}^T \mathbf{W}_i)^2,$$

where the vectors \mathbf{M} and \mathbf{W}_i are defined above, $\mathbf{A} = [a_{rs}]$, and

$$(7) \quad a_{rs} = \frac{1}{k_r k_s} \sum_{i \in C_r \cap C_s} \sigma_{ii} - \frac{1}{n k_r} \sum_{i \in C_0 \cap C_r} \sigma_{ii} + \frac{1}{n^2} \sum_{i \in C_0} \sigma_{ii} - \frac{1}{n k_s} \sum_{i \in C_0 \cap C_s} \sigma_{ii}, \quad r, s = 1, \dots, p.$$

The proof of Theorem 2 is based on the identity

$$(8) \quad R(\hat{\gamma}) = E(\hat{\gamma} - \gamma)^2 = \text{var } \hat{\gamma} + (\text{bias } \hat{\gamma})^2,$$

where γ is any parameter and $\hat{\gamma}$ is any estimator. From (4) it follows that the covariance matrix of the vector $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_p)^T$ is of the form

$$\text{cov } \hat{\mathbf{a}} = \mathbf{W}^{-1} \mathbf{r} (\text{cov } \boldsymbol{\varepsilon}) \mathbf{r}^T (\mathbf{W}^{-1})^T = [\mathbf{W}_i \mathbf{A} \mathbf{W}_j],$$

where

$$\mathbf{r} = \begin{bmatrix} \frac{1}{k_1} \mathbf{1}^1 - \frac{1}{n} \mathbf{1} \\ \vdots \\ \frac{1}{k_p} \mathbf{1}^p - \frac{1}{n} \mathbf{1} \end{bmatrix}, \quad \text{cov } \boldsymbol{\varepsilon} = \begin{bmatrix} \sigma_{11} & & 0 \\ & \ddots & \\ 0 & & \sigma_{nn} \end{bmatrix}.$$

$\mathbf{1}$ is a row vector with elements equal to unity, $\mathbf{1}^i$ is a row vector with k_i elements equal to unity and $n - k_i$ equal to zero. The r -th element of the vector $\mathbf{1}^i$ is 1 if the r -th observation belongs to the set C_i and 0 otherwise. An (r, s) entry in the matrix $\mathbf{A} = \mathbf{r} (\text{cov } \boldsymbol{\varepsilon}) \mathbf{r}^T$ is equal to the expression (7). The second component of the right-hand side of the expression (8) is equal to the square of (5).

3. The robustness of regression coefficients in factorial experiments. By reason of the accepted regression model, in which the values of the independent variables are appointed in advance, to compare the influence of contaminations on estimators obtained by the p -point method with the influence of these

contaminations on estimators obtained by the least squares method, we restrict ourselves to the factorial experiments based on a plan of the form

$$Q(C_0) = \sum_{j=1}^p T_{jk},$$

where

$$T_{jk} = \{\pm h/2, \pm(h/2+h), \dots, \pm(h/2+(k-1)h)\}, \quad h \in \mathbf{R}, \quad n = (2k)^p.$$

We divide the set C_0 in p manners into two subsets. For each of these divisions we define

$$C_r = \{(x_{i1}, \dots, x_{ip}, y_i): x_{ir} > 0\}, \quad r = 1, \dots, p.$$

Hence

$$\begin{aligned} \bar{x}_r &= 0 \quad \text{for } r = 1, \dots, p, \\ \bar{x}_i^r &= \begin{cases} kh/2 & \text{for } i = n, \\ 0 & \text{for } i \neq r, \end{cases} \quad i, r = 1, \dots, p, \\ W &= \begin{bmatrix} kh/2 & & 0 \\ & \ddots & \\ 0 & & kh/2 \end{bmatrix}, \end{aligned}$$

$$k_r = n/2 = (2k)^p/2 \quad \text{for } r = 1, \dots, p.$$

The formulas (5) and (6) for bias and risk can be written in the simpler form

$$(9) \quad \text{bias } \hat{a}_i = \frac{2(M_i - M_0)}{kh},$$

$$(10) \quad R(\hat{a}_i) = \frac{4((2k)^{-p} \sum_{i \in C_0} \sigma_i^2 + (M_i - M_0)^2)}{k^2 h^2}.$$

Now, we are going to obtain the same quantities for the least squares estimators of the regression coefficients in the extended model M for the same plan $\sum_{j=1}^p T_{jk}$. Let $X = [x_{ik} - \bar{x}_k] = [x_{ik}]$, $i = 1, \dots, n$, $k = 1, \dots, p$, be the matrix of the plan of the experiment and let Y be an observation vector of the random variables Y_1, \dots, Y_n . Then the vector of least squares estimators is of the form

$$\hat{b} = (X^T X)^{-1} X^T Y.$$

For the plan $\sum_{j=1}^p T_{jk}$ we have

$$X^T X = \left[\sum_{i=1}^n x_{ik} x_{ij} \delta_{jk} \right], \quad j, k = 1, \dots, p,$$

where δ_{jk} denotes the Kronecker delta and

$$\sum_{i=1}^n x_{ij}^2 = \frac{(2k)^p h^2 (4k^2 - 1)}{12}.$$

Hence

$$S = (X^T X)^{-1} X^T = \frac{12}{(2k)^p h^2 (4k^2 - 1)} X^T,$$

$$\text{bias } \hat{b} = E(\hat{b}) - a = E((X^T X)^{-1} X^T Y) - a$$

$$= E((X^T X)^{-1} X^T (Xa + \varepsilon)) - a = (X^T X)^{-1} X^T E(\varepsilon) = S \begin{bmatrix} m_1 \\ \vdots \\ m_n \end{bmatrix},$$

which implies

$$(11) \quad \begin{aligned} \text{bias } \hat{b}_r &= \frac{12 \sum_{i=1}^n x_{ir} m_i}{(2k)^p h^2 (4k^2 - 1)}, \\ \text{cov } \hat{b} &= S(\text{cov } \varepsilon) S^T, \end{aligned}$$

whence

$$\text{var } \hat{b}_r = \frac{(12)^2 \sum_{i=1}^n x_{ir}^2 \sigma_i^2}{(2k)^{2p} h^2 (4k^2 - 1)^2}.$$

By the formula (8) we get

$$(12) \quad R(\hat{b}_r) = \frac{(12)^2 \left[\sum_{i=1}^n x_{ir}^2 \sigma_i^2 + \left(\sum_{i=1}^n x_{ir} m_i \right)^2 \right]}{(2k)^{2p} (4k^2 - 1)^2}.$$

In order to compare the robustness of the p -point method estimators with the robustness of the least squares estimators we find (applying Hampel's definition [1]–[3]) the influence functions for the bias and risk of these two kinds of estimators. To do this, we assume that a contamination arises only for the i -th observation, which is identified with the point $(x_{i1}, \dots, x_{ip}) \in Q(C_0)$. This means that the following restrictions are introduced in the model M :

$$\begin{aligned} H_j &\in \{F\} & \text{for } j \neq i, \\ H_j &\in \{(1-\alpha)F + \alpha\delta_x; 0 < \alpha < 1\} & \text{for } j = i, \end{aligned}$$

where δ_x denotes the distribution concentrated at the point $x \in R$. Let us define the functional $T: \mathcal{P} \rightarrow R$, where \mathcal{P} is the family of distributions of the joint variable (Y_1, \dots, Y_n) of the form

$$\mathcal{P} = \{(F_1, \dots, F_{i-1}, (1-\alpha)F_i + \alpha\delta_x, F_{i+1}, \dots, F_n): 0 \leq \alpha \leq 1, i = 1, \dots, n\}.$$

The following expression is set up as the influence function of the functional T :

$$IC_T(x|x_{i1}, \dots, x_{ip}) = \lim_{\alpha \rightarrow 0} \frac{T(P_i(\alpha)) - T(P_i(0))}{\alpha}$$

if the above limit exists, where

$$P_i(\alpha) = (F_1, \dots, (1-\alpha)F_i + \alpha\delta_x, \dots, F_n) \in \mathcal{P}.$$

THEOREM 3. *For the bias and risk of the p -point method estimators, the influence functions are expressed by the formulas*

$$(13) \quad |IC_{\text{bias}\hat{a}_r}(x|x_{j1}, \dots, x_{jp})| = \frac{2|x|}{(2k)^p kh}, \quad r = 1, \dots, p.$$

$$(14) \quad |IC_{R(\hat{a}_r)}(x|x_{j1}, \dots, x_{jp})| = \frac{4|2x^2 - \sigma^2|}{(2k)^p k^2 h^2},$$

Proof of (13). Under the assumptions of the model M_0 it follows that

$$m_i(F_i) = 0 \quad \text{for } i = 1, \dots, p,$$

$$m_j((1-\alpha)F_j + \alpha\delta_x) = (1-\alpha)m_j(F_j) + \alpha x = \alpha x.$$

Hence, taking into consideration the formula (9) and putting in the definition of the influence function $T = \text{bias}(\hat{a}_r)$, we have

$$\begin{aligned} T(P_j(\alpha)) &= \frac{2}{kh} \left\{ \frac{1}{k_r} \left[\sum_{\substack{i \in C_r \\ i \neq j}} m_i(F_i) + m_j((1-\alpha)F_j + \alpha\delta_x) \right] \right. \\ &\quad \left. - \frac{1}{n} \left[\sum_{\substack{i \in C_0 \\ i \neq j}} m(F_i) + m_j((1-\alpha)F_j + \alpha\delta_x) \right] \right\} \\ &= \frac{2}{kh} \left(\frac{2}{n} - \frac{1}{n} \right) \alpha x = \frac{2\alpha x}{kh(2k)^p}, \end{aligned}$$

$$T(P(0)) = 0.$$

Hence we obtain the right-hand side of (13).

In the proof of (14) we use the definition of the functional T and the fact that under the assumptions of the model M_0 we have

$$\sigma_i^2(F_i) = \sigma^2, \quad \sigma_j^2((1-\alpha)F_j + \alpha\delta_x) = (1-\alpha)\sigma^2(F_j) + \alpha x^2.$$

THEOREM 4. *For the bias and risk of the least squares estimators (see (11) and (12)) the influence functions are expressed by the formulas*

$$(15) \quad |IC_{\text{bias}\hat{b}_r}(x|x_{j1}, \dots, x_{jp})| = \left| \frac{12x_{jr}x}{(2k)^p h^2 (4k^2 - 1)} \right|,$$

$$(16) \quad |IC_{R(\hat{b}_r)}(x|x_{j1}, \dots, x_{jp})| = \left| \frac{144x_{jr}^2(2x^2 - \sigma^2)}{(2k)^{2p} h^2 (4k^2 - 1)^2} \right|,$$

$$r = 1, \dots, p; j = 1, \dots, n.$$

The proof of this theorem is analogous to the proofs of (13) and (14).

In order to compare the influence functions for the estimators obtained by both considered methods we calculate the quotient

$$C_r(x|x_{j1}, \dots, x_{jp}) = \frac{|IC_{\text{bias}\hat{\delta}_r}(x|x_{j1}, \dots, x_{jp})|}{|IC_{\text{bias}\hat{\alpha}_r}(x|x_{j1}, \dots, x_{jp})|} = \frac{6|x_{jr}|k}{h(4k^2 - 1)}, \quad j = 1, \dots, n.$$

The function $C_r(x|x_{j1}, \dots, x_{jp}) = C_r(|x_{jr}|)$ is increasing with respect to $|x_{jr}|$. Hence

$$C_r(h/2) < C_r(h/2 + h) \dots C_r(h/2 + (k-1)h).$$

We put $h > 0$. If k is large enough, then the inequality

$$1 \leq C_r(x|x_{j1}, \dots, x_{jp}) \leq 3/2$$

holds. This implies the following

THEOREM 5. *If in the extended Gauss–Markoff model x_{jr} satisfies the inequality*

$$|x_{jr}| \geq h/2 + \frac{2}{3}kh,$$

then for appropriate k

$$1 \leq C_r(|x_{jr}|) \leq 3/2, \quad j = 1, \dots, n; r = 1, \dots, p.$$

COROLLARY 1. *If we take into consideration the points of factorial experiments with the plan*

$$C_0 = \sum_{j=1}^p T_{jk}$$

associated with a level greater than or equal to $\frac{2}{3}k$, then for sufficiently large k the bias-robustness of the p -point method estimators is greater than the bias-robustness of the least squares estimators.

Analogously, considering the function

$$K_r(x|x_{j1}, \dots, x_{jp}) = \frac{|IC_{R(\hat{\delta}_r)}(x|x_{j1}, \dots, x_{jp})|}{|IC_{R(\hat{\alpha}_r)}(x|x_{j1}, \dots, x_{jp})|}$$

we can formulate the following theorem:

THEOREM 6. *If in an extended Gauss–Markoff model x_{jr} satisfies the inequality*

$$|x_{jr}| \geq h/2 + \frac{2}{3}kh,$$

then for appropriate k

$$1 \leq K_r(|x_{jr}|) \leq 3/2, \quad j = 1, \dots, n; r = 1, \dots, p.$$

COROLLARY 2. *At the points of factorial experiments with plan C_0 associated with a level greater than or equal to $\frac{2}{3}k$, for sufficiently large k , the risk-robustness with respect to the square loss function of the p -point method estimators is greater than the risk-robustness of the least squares estimators.*

References

- [1] R. F. Hampel, *The influence curve and its role in robust estimation*, J. Amer. Statist. Assoc. 69 (1974), pp. 383–393.
- [2] – *A general qualitative definition of robustness*, Ann. Math. Statist. 42 (1971), pp. 1887–1896.
- [3] P. J. Huber, *Robust Statistics*, J. Wiley, New York 1981.
- [4] C. Platt and Z. Paprzycki, *On estimation of multiple regression coefficients by the p-point method*, Zastos. Mat. 20 (1988), pp. 53–65.
- [5] – *O pewnej metodzie wyznaczania estymatorów współczynników regresji wielokrotnej i jej zastosowaniach w geodezji*, Zeszyty Naukowe ART Olsztyn 1984.

DEPARTMENT OF APPLIED MATHEMATICS AND MATHEMATICAL STATISTICS
AGRICULTURAL AND TECHNICAL ACADEMY
10-740 OLSZTYN

Received on 1986.03.28
