ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
21,4 (1993), pp. 461–479

C. PLATT and Z. PAPRZYCKI (Olsztyn)

THE GENERALIZED p-POINT METHOD OF ESTIMATION OF REGRESSION COEFFICIENTS

0. Introduction. This work is a generalization of papers [5] and [6]. A multiple regression model in which p independent variables are arbitrary random variables is considered. In order to define a generalized sample division in such a model, multifunction notation is introduced. Divisions used in [5] and [6] are particular cases of the very general one presented in the paper. On the basis of this division the estimators of regression coefficients are determined.

1. Assumptions on the regression model. Let $\mathbf{Z} = (X_1, \dots, X_p, Y)$ be a random vector defined on a probability space (Ω, B_Ω, P) with values in a measure space $(\mathfrak{R}_\mathbf{Z}, B_{\mathfrak{R}_\mathbf{Z}}, \nu_\mathbf{Z})$.

We say that the random vector $\mathbf{Z}$ *corresponds to a regression model* if it satisfies the following assumptions:

ASSUMPTION 1.1. $(\mathfrak{R}_\mathbf{Z}, B_{\mathfrak{R}_\mathbf{Z}}, \nu_\mathbf{Z})$ is a product space, i.e.

(1.1) $\mathfrak{R}_\mathbf{Z} = \mathfrak{M}_\mathbf{Z} \times Q_\mathbf{Z}$ where $\mathfrak{M}_\mathbf{Z} \subset \mathbb{R}^p$ and $Q_\mathbf{Z} \subset \mathbb{R}$ are Borel subsets,
$B_{\mathfrak{R}_\mathbf{Z}} = B_{\mathfrak{M}_\mathbf{Z}} \times B_{Q_\mathbf{Z}}$ is a σ-field which is the product of the Borel σ-fields $B_{\mathfrak{M}_\mathbf{Z}}$ and $B_{Q_\mathbf{Z}}$ of $\mathfrak{M}_\mathbf{Z}$ and $Q_\mathbf{Z}$,
$\nu_\mathbf{Z} = \mu_\mathbf{Z} \times \pi_\mathbf{Z}$ is a product measure where the measures $\mu_\mathbf{Z}$ and $\pi_\mathbf{Z}$ are defined on $(\mathfrak{M}_\mathbf{Z}, B_{\mathfrak{M}_\mathbf{Z}})$ and $(Q_\mathbf{Z}, B_{Q_\mathbf{Z}})$, respectively.

ASSUMPTION 1.2. There exists a density $f_\mathbf{Z}$ of the random vector $\mathbf{Z} = (X_1, \dots, X_p, Y)$ with respect to the measure $\nu_\mathbf{Z}$ such that

$$\widetilde{f}_\mathbf{Z}(x_1, \dots, x_p) = \int_{Q_\mathbf{Z}} f_\mathbf{Z}(x_1, \dots, x_p, y)\, d\pi_\mathbf{Z} > 0 \tag{1.2}$$

1991 *Mathematics Subject Classification*: Primary 62J05.
Key words and phrases: efficiency, estimator, influence function, multiple regression model, robustness, sample division.

for each $(x_1, \ldots, x_p) \in \mathfrak{M}_{\mathbf{Z}}$. The function $\widetilde{f}_{\mathbf{Z}}(x_1, \ldots, x_p)$ defined by (1.2) is the marginal density function of the vector $(X_1, \ldots, X_p)$ with respect to the measure $\mu_{\mathbf{Z}}$.

ASSUMPTION 1.3. The random vector $\mathbf{Z} = (X_1, \ldots, X_p, Y)$ satisfies the conditions:

$$
\begin{aligned}
&E(Y \mid X_1, \ldots, X_p) = a_0 + a_1 X_1 + \ldots + a_p X_p\,,\\
&E(X_1) = \ldots = E(X_p) = 0\,,\\
&E((Y - a_0 - a_1 X_1 - \ldots - a_p X_p)^2 \mid X_1, \ldots, X_p) = \sigma^2 < \infty\,,
\end{aligned}
\tag{1.3}
$$

and X_i and $Y - a_0 - a_1 X_1 - \ldots - a_p X_p$ are independent for $i = 1, \ldots, p$, where $(a_0, \ldots, a_p) \in \mathbb{R}^{p+1}$ is a vector of constants, and $\sigma^2 \in \mathbb{R}$.

2. The method of sample division

2a. *The sample.* A *sample of size n* is a random vector $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ in which $\mathbf{Z}_i$, $i = 1, \ldots, n$, are independent random variables of the form $\mathbf{Z}_i = (X_{i1}, \ldots, X_{ip}, Y_i)$ and

$$\mathbf{Z}_i : (\Omega, B_\Omega, P) \to (\mathfrak{R}_{\mathbf{Z}_i}, B_{\mathfrak{R}_{\mathbf{Z}_i}}, \nu_{\mathbf{Z}_i})\,.$$

Moreover, we assume that each $\mathbf{Z}_i$, $i = 1, \ldots, n$, satisfies assumptions 1.1–1.3. It is clear that this definition admits different distributions of the random vectors $\mathbf{Z}_i$. This is an essential generalization of the usual definition.

For convenience, the following notations are introduced:

$(\mathfrak{R}_i, B_{\mathfrak{R}_i}, \nu_i)$ for $(\mathfrak{R}_{\mathbf{Z}_i}, B_{\mathfrak{R}_{\mathbf{Z}_i}}, \nu_{\mathbf{Z}_i})$,
$(\mathfrak{R}, B_{\mathfrak{R}}, \nu)$ — product of the measure spaces $(\mathfrak{R}_i, B_{\mathfrak{R}_i}, \nu_i)$, $i = 1, \ldots, n$,
$(\mathfrak{M}, B_{\mathfrak{M}}, \mu)$ — product of the spaces $(\mathfrak{M}_i, B_{\mathfrak{M}_i}, \mu_i)$, $i = 1, \ldots, n$,
$(\mathbf{X}_i, Y_i)$ — a component of $\mathbf{Z}$.

The space $(\mathfrak{R}, B_{\mathfrak{R}}, \nu)$ is called a *sample space* and $(\mathfrak{M}, B_{\mathfrak{M}}, \mu)$ a *sample space plan*, or briefly, a *plan*.

2b. *Sample division.* We now consider multifunctions φ^* on (Ω, B_Ω) with values in $2^{\mathbb{R}^p}$ such that there exists a multifunction $\varphi : \mathfrak{M} \to 2^{\mathbb{R}^p}$ for which

$$\varphi^*(\omega) = \varphi(\mathbf{X}_1(\omega), \ldots, \mathbf{X}_n(\omega)), \qquad \omega \in \Omega\,.$$

Here $2^{\mathbb{R}^p}$ is the family of all subsets of $\mathbb{R}^p$. The multifunction φ is required to satisfy the conditions:

$$
\begin{aligned}
&\{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{M} : \mathbf{x}_i \in \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n)\} \in B_{\mathfrak{M}}, \quad i = 1, \ldots, n\,,\\
&P^{\mathbf{Q}\circ\mathbf{Z}}\Big(\bigcap_{i=1}^{n} \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{M} : \mathbf{x}_i \notin \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n)\}\Big) = 0\,,
\end{aligned}
\tag{2.1}
$$

where

$$Q(x_1, \ldots, x_p, y) = (x_1, \ldots, x_p) \quad \text{for } (x_1, \ldots, x_p, y) \in \mathbb{R}^p \times \mathbb{R}$$

and $P^{\mathbf{Q}\circ\mathbf{Z}}$ is the probability induced on $\mathfrak{M}$ by the variable $(\mathbf{Q} \circ \mathbf{Z})(\omega) = (Q(\mathbf{Z}_1(\omega)), \ldots, Q(\mathbf{Z}_n(\omega)))$.

Each such multifunction φ^* induces a multifunction on (Ω, B_Ω), called a *sample division*, defined by $\mathbf{C}^*(\omega) = \{\mathbf{Z}_i(\omega) : \mathbf{X}_i(\omega) \in \varphi^*(\omega)\}$. Alternatively, we may also define a multifunction $\mathbf{C} : \mathfrak{R} \to 2^{\mathbb{R}^{p+1}}$ by $\mathbf{C}(\mathbf{z}) = \{\mathbf{z}_i : \mathbf{x}_i \in \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n)\}$. Then

$$\begin{aligned}(\mathbf{C} \circ \mathbf{Z})(\omega) &= \{\mathbf{Z}_i(\omega) : \mathbf{X}_i(\omega) \in \varphi(\mathbf{X}_1(\omega), \ldots, \mathbf{X}_n(\omega))\} \\ &= \{\mathbf{Z}_i(\omega) : \mathbf{X}_i(\omega) \in \varphi^*(\omega)\} = \mathbf{C}^*(\omega)\,.\end{aligned}$$

Hence $\mathbf{C} \circ \mathbf{Z} = \mathbf{C}^*$. Therefore we do not differentiate between $\mathbf{C}^*$ and $\mathbf{C}$. It will always be clear from the context which multifunction is discussed. Hence, further we write $\mathbf{C}$ in both cases.

Notice that the size k of $\mathbf{C}(\omega)$ is a random variable: $k = \operatorname{card}\mathbf{C} : (\Omega, B_\Omega) \to \{0, 1, \ldots, n\}$.

LEMMA 2.1. *The size $k(\omega)$ of an arbitrary division $\mathbf{C}(\omega)$ is greater than zero with probability* 1.

Proof.

$$\begin{aligned}P(\{\omega : k(\omega) = 0\}) &= P(\{\omega : \mathbf{C}(\omega) = 0\}) \\ &= P^{\mathbf{Q}\circ\mathbf{Z}}\Big(\bigcap_{i=1}^{n}\{(\mathbf{x}_1, \ldots, \mathbf{x}_n) : \mathbf{x}_i \notin \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n)\}\Big) = 0\,,\end{aligned}$$

which follows directly from assumption (2.1).

3. Estimation of regression parameters. In order to obtain estimators of the regression parameters $a_0, \ldots, a_p$ in the model (1.3) we carry out $p+1$ divisions $\mathbf{C}_0, \ldots, \mathbf{C}_p$ in the sample $\mathbf{Z}$ by choosing $p+1$ multifunctions $\varphi_0, \ldots, \varphi_p$. To each division $\mathbf{C}_r$ there corresponds a random variable $\mathbf{g}_r : \Omega \to \mathbb{R}^p \times \mathbb{R}$ defined by

$$\mathbf{g}_r(\omega) = (\overline{X}_1^r(\omega), \ldots, \overline{X}_p^r(\omega), \overline{Y}^r(\omega))$$

where

$$\overline{X}_j^r(\omega) = \frac{1}{k_r(\omega)} \sum_{i \in K_r(\omega)} X_{ij}(\omega)\,, \qquad \overline{Y}^r(\omega) = \frac{1}{k_r(\omega)} \sum_{i \in K_r(\omega)} Y_i(\omega)\,,$$

with $k_r(\omega) = \operatorname{card}\mathbf{C}_r(\omega)$ and $K_r(\omega) = \{i : \mathbf{Z}_i(\omega) \in \mathbf{C}_r(\omega)\}$ for $j = 1, \ldots, p$, $r = 0, 1, \ldots, p$. We assume that $\mathbf{C}_0$ is equal to the sample $\mathbf{Z}$ and $k_0 = n$.

LEMMA 3.1. *For fixed $\omega \in \Omega$ the set of the points $\mathbf{g}_0(\omega), \mathbf{g}_1(\omega), \ldots, \mathbf{g}_p(\omega)$ uniquely determines a p-dimensional hyperplane if and only if the matrix*

$$\mathbf{W}(\omega) = \begin{bmatrix} \overline{X}_1^1(\omega) - \overline{X}_1^0(\omega) & \ldots & \overline{X}_p^1(\omega) - \overline{X}_p^0(\omega) \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \overline{X}_1^p(\omega) - \overline{X}_1^0(\omega) & \ldots & \overline{X}_p^p(\omega) - \overline{X}_p^0(\omega) \end{bmatrix}. \tag{3.1}$$

is nonsingular (see [5]).

We assume that $P(\{\omega : \det \mathbf{W}(\omega) = 0\}) = 0$, $E(\det \mathbf{W}(\omega)) \neq 0$ and $E(\overline{X}_j^r)$ exists for $0 \leq r \leq p$, $1 \leq j \leq p$.

As in [5] we propose

$$\widehat{a}_0 = \overline{Y}^0, \quad \widehat{a}_r = \frac{W_r(\omega)}{W(\omega)}, \quad r = 1, \ldots, p, \tag{3.2}$$

as estimators of the regression coefficients in assumption 1.3, where $W(\omega) = \det \mathbf{W}(\omega)$ and $W_i(\omega)$ is the determinant obtained from $W(\omega)$ by replacing the ith column with the vector $(\overline{Y}^1 - \overline{Y}^0, \ldots, \overline{Y}^p - \overline{Y}^0)^T$.

4. Properties of the estimators obtained by the p-point method

LEMMA 4.1. *For each vector $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{M}$ and $i = 1, \ldots, n$*

$$E(Y_i \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) = a_0 + \sum_{j=1}^{p} a_i x_{ij}$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$.

The lemma follows from the conditional expectation formula and from assumption (1.3).

LEMMA 4.2. *For $r = 0, 1, \ldots, p$*

$$E(\overline{Y}^r) = a_0 + \sum_{j=1}^{p} a_j E(\overline{X}_j^r).$$

Proof. Let us introduce the following notations:

$$M_{i_1,\ldots,i_k} = \{(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{M} : \mathbf{x}_{i_s} \in \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \text{ for } s \in \{1, \ldots, k\} \text{ and } \mathbf{x}_i \notin \varphi(\mathbf{x}_1, \ldots, \mathbf{x}_n) \text{ for } i \notin \{i_1, \ldots, i_k\}\}$$

where φ is the multifunction representing the division $\mathbf{C}_r$, $k = 0, 1, \ldots, n$ and $1 \leq i_1 < \ldots < i_k \leq n$. It results directly from assumption (2.1) that the sets $M_{i_1,\ldots,i_k}$ are measurable and disjoint, i.e.

$$M_{i_1,\ldots,i_k} \cap M_{j_1,\ldots,j_l} = \emptyset \quad \text{for } \{i_1, \ldots, i_k\} \neq \{j_1, \ldots, j_l\}$$

and

$$P\Big(\bigcup \{M_{i_1,\ldots,i_k} : 1 \leq i_1 < \ldots < i_k \leq n,\ k \geq 1\}\Big) = 1.$$

We can write

$$E\left(\frac{Y_{i_1}+\ldots+Y_{i_k}}{k}\,\middle|\,\mathbf{X}_1=\mathbf{x}_1,\ldots,\mathbf{X}_n=\mathbf{x}_n\right)$$

$$=\frac{1}{k}\sum_{s=1}^{k}E(Y_{i_s}\mid\mathbf{X}_1=\mathbf{x}_1,\ldots,\mathbf{X}_n=\mathbf{x}_n)=\frac{1}{k}\sum_{s=1}^{k}\Big(a_0+\sum_{j=1}^{p}a_j x_{i_s j}\Big)$$

$$=a_0+\sum_{j=1}^{p}a_j\frac{x_{i_1 j}+\ldots+x_{i_k j}}{k}=a_0+\sum_{j=1}^{p}a_j\overline{x}_j^r$$

where $\overline{x}_j^r$ is the arithmetical mean of $x_{i_1 j},\ldots,x_{i_k j}$. Hence

$$E(\overline{Y}^r)=\int_{\mathfrak{M}}E(\overline{Y}^r\mid\mathbf{X}_1=\mathbf{x}_1,\ldots,\mathbf{X}_n=\mathbf{x}_n)\widetilde{f}_1(\mathbf{x}_1)\ldots\widetilde{f}_n(\mathbf{x}_n)d(\mu_1\times\ldots\times\mu_n)$$

$$=\sum_{k=1}^{n}\sum_{1\le i_1<\ldots<i_k\le n}\int_{M_{i_1,\ldots,i_k}}E(\overline{Y}^r\mid\mathbf{X}_1=\mathbf{x}_1,\ldots,\mathbf{X}_n=\mathbf{x}_n)\widetilde{f}_1(\mathbf{x}_1)\ldots$$

$$\ldots\widetilde{f}_n(\mathbf{x}_n)d(\mu_1\times\ldots\times\mu_n)$$

$$=\sum_{k=1}^{n}\sum_{1\le i_1<\ldots<i_k\le n}\int_{M_{i_1,\ldots,i_k}}\Big(a_0+\sum_{j=1}^{p}a_j\overline{x}_j^r\Big)\widetilde{f}_1(\mathbf{x}_1)\ldots\widetilde{f}_n(\mathbf{x}_n)d(\mu_1\times\ldots\times\mu_n)$$

$$=\int_{\mathfrak{M}}a_0\widetilde{f}_1(x_1)\ldots\widetilde{f}_n(x_n)d(\mu_1\times\ldots\times\mu_n)$$

$$+\sum_{j=1}^{p}a_j\int_{\mathfrak{M}}\overline{x}_j^r\widetilde{f}_1(\mathbf{x}_1)\ldots\widetilde{f}_n(\mathbf{x}_n)d(\mu_1\times\ldots\times\mu_n)=a_0+\sum_{j=1}^{p}a_jE(\overline{X}_j^r)$$

and the lemma is proved.

LEMMA 4.3. *If, for some* $r=0,1,\ldots,p$, *there exists* $0<c_r\le 1$ *such that*

$$\forall_{\varepsilon>0}\quad\lim_{n\to\infty}P(\{\omega:|k_r(\omega)/n-c_r|<\varepsilon\})=1$$

then for each $m\ge 1$, $\lim_{n\to\infty}(P(k_r=1)+\ldots+P(k_r=m))=0$.

Proof. By assumption, for every $\varepsilon>0$ there exists n_0 such that for $n>n_0$

$$P(\{\omega:|k_r(\omega)/n-c_r|<\varepsilon\})>1-\varepsilon,$$

i.e.

$$P(\{\omega:k_r(\omega)<n(c_r-\varepsilon)\}\cup\{\omega:k_r(\omega)>n(c_r+\varepsilon)\})<\varepsilon.$$

Without loss of generality, we may assume that $c_r-\varepsilon>0$. Then $n(c_r-\varepsilon)>m$ for n sufficiently large, hence

$$P(\{\omega:k_r(\omega)\le m)\}<P(\{\omega:k_r(\omega)<n(c_r-\varepsilon)\})<\varepsilon,$$

which completes the proof.

Now set

$$D^j_{rn}(\varepsilon) = \{\omega \in \Omega : |\overline{X}^r_j(\omega) - E(\overline{X}^r_j)| \geq \varepsilon\}, \quad j = 1, \ldots, p,$$
$$B^k_{rn} = \{\omega \in \Omega : k_r(\omega) = k\}$$

where the subscript n denotes the sample size.

THEOREM 4.1. *If*

$$\underset{\varepsilon>0}{\forall} \underset{r=0,\ldots,p}{\forall} \underset{0<c_r\leq 1}{\exists} \lim_{n\to\infty} P(\{\omega : |k_r(\omega)/n - c_r| < \varepsilon\}| = 1$$

and

$$\underset{\varepsilon>0}{\forall} \lim_{k\to\infty} \sup_{n\geq k} P(D^j_{rn}(\varepsilon) \mid B^k_{rn}) = 0 \quad \text{for } j = 1, \ldots, p$$

then

$$\underset{\varepsilon>0}{\forall} \lim_{n\to\infty} P(\{\omega : |\widehat{a}_r(\omega) - a_r| > \varepsilon\}) = 0 \quad \text{for } r = 0, 1, \ldots, p.$$

Proof. First we must show that for fixed $r = 0, 1, \ldots, p$ and $j = 1, \ldots, p$, $\overline{Y}^r$ and $\overline{X}^r_j$ tend stochastically to the constants $E(\overline{Y}^r)$ and $E(\overline{X}^r_j)$, respectively.

By the total probability formula we obtain

$$P(D^j_{rn}(\varepsilon)) = \sum_{k=1}^{n} P(B^k_{rn}) P(D^j_{rn}(\varepsilon) \mid B^k_{rn}).$$

For $m \geq 1$, define

$$P'(D^j_{rn}(\varepsilon)) = \sum_{k=m+1}^{n} P(B^k_{rn}) P(D^j_{rn}(\varepsilon) \mid B^k_{rn}).$$

Since

$$P(D^j_{rn}(\varepsilon)) \leq |P(D^j_{rn}(\varepsilon)) - P'(D^j_{rn}(\varepsilon))| + |P'(D^j_{rn}(\varepsilon))|$$

it is sufficient to show that for any $\varepsilon > 0$ there exist n_0 such that for $n > n_0$

$$|P(D^j_{rn}(\varepsilon)) - P'(D^j_{rn}(\varepsilon))| < -\varepsilon/2, \tag{4.1}$$

$$P'(D^j_{rn}(\varepsilon)) < \varepsilon/2. \tag{4.2}$$

Indeed,

$$|P(D^j_{rn}(\varepsilon)) - P'(D^j_{rn}(\varepsilon))| = \sum_{k=1}^{m} P(B^k_{rn}(\varepsilon)) P(D^j_{rn}(\varepsilon) \mid B^k_{rn}) \leq \sum_{k=1}^{m} P(B^k_{rn}).$$

By Lemma 4.3 the right side tends to zero for any m. Therefore there exists n' such that (4.1) holds for $n > n'$. Also under the assumptions of Theorem 4.1, for sufficiently large m, if $m < k \leq n$ then

$$P(D^j_{rn}(\varepsilon) \mid B^k_{rn}) = P(|\overline{X}^r_j(\omega) - E(\overline{X}^r_j)| \geq \varepsilon \mid k_r(\omega) = k) < \varepsilon/2.$$

Hence

$$P'(D_{rn}^j(\varepsilon)) = \sum_{k=m+1}^{n} P(B_{rn}^k)P(D_{rn}^j(\varepsilon) \mid B_{rn}^k) < \sum_{k=m+1}^{n} P(B_{rn}^k)\varepsilon/2 < \varepsilon/2$$

and (4.2) is satisfied. This ends the proof of the stochastic convergence of $\overline{X}_j^r$ to $E(\overline{X}_j^r)$.

To prove the stochastic convergence of $\overline{Y}^r$ to $E(\overline{Y}^r)$ notice that by assumption (1.3) we can write

$$Y_i = a_0 + a_1 X_{i1} + \ldots + a_p X_{ip} + U_i, \quad i = 1, \ldots, n,$$

where

$$U_i = Y_i - E(Y_i \mid X_{i1}, \ldots, X_{ip})$$

are independent random variables. Hence $\overline{Y}^r = a_0 + a_1\overline{X}_1^r + \ldots + a_p\overline{X}_p^r + \overline{U}^r$ where

$$\overline{U}^r = \overline{Y}^r - E(Y_i \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n)\,.$$

Since $E(\overline{U}^r) = 0$, by Lemma 4.2 we have

$$\overline{Y}^r - E(\overline{Y}^r) = a_1(\overline{X}_1^r - E(\overline{X}_1^r)) + \ldots + a_p(\overline{X}_p^r - E(\overline{X}_p^r)) + \overline{U}^r\,.$$

We have to show the stochastic convergence of $\overline{U}^r$ to 0 for $r = 1, \ldots, p$. It is enough to notice that by (1.3) the variables U_i and $\mathbf{X}_i$ are independent. Then

$$P(|\overline{U}^r(\omega)| \geq \varepsilon \mid k_r(\omega) = k) = P(|\overline{U}^r(\omega)| \geq \varepsilon)$$

and by the weak law of large numbers

$$\lim_{n\to\infty} \sup_{n\geq k} P(|\overline{U}^r(\omega)| \geq \varepsilon \mid k_r(\omega) = k) = 0\,.$$

The estimators $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_p)$ of the regression coefficients are the solution of a system of p equations which in matrix notation has the form $\mathbf{W}(\omega)\widehat{\mathbf{a}} = \mathbf{h}$ where $\mathbf{h} = (\overline{Y}^1 - \overline{Y}^0, \ldots, \overline{Y}^p - \overline{Y}^0)^T$. Then

$$\widehat{\mathbf{a}} = \mathbf{W}^{-1}(\omega)\mathbf{h}\,. \tag{4.3}$$

From Lemma 4.2 it follows that

$$E(\overline{Y}^r) - E(\overline{Y}^0) = \sum_{j=1}^{p} a_j(E(\overline{X}_j^r) - E(\overline{X}_j^0)), \quad r = 1, \ldots, p\,. \tag{4.4}$$

Set

$$\mathbf{h}_1 = (E(\overline{Y}^1) - E(\overline{Y}^0), \ldots, E(\overline{Y}^p) - E(\overline{Y}^0))^T\,,$$
$$\mathbf{W}_1 = [E(\overline{X}_j^r) - E(\overline{X}_j^0)]_{j,r=1,\ldots,p}\,.$$

Then from (4.4), $\mathbf{h}_1 = \mathbf{W}_1\mathbf{a}$. By Slutsky's theorem the matrix $\mathbf{W}(\omega)$ tends stochastically to $\mathbf{W}_1$, the vector $\mathbf{h}$ tends to $\mathbf{h}_1$, and by (4.3) the vector $\widehat{\mathbf{a}}$ of

estimators tends to $\mathbf{W}_1^{-1}\mathbf{W}_1\mathbf{a} = \mathbf{a}$. This completes the proof of Theorem 4.1, since the convergence of the estimator $\widehat{a}_0 = n^{-1}\sum_{i=1}^{n} Y_i$ to a_0 is evident.

It is easy to notice that for all terms of the matrix $\mathbf{W}(\omega)$ the following relations hold:

$$E(\overline{X}_j^r \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) = \overline{x}_j^r,$$

$$E(\overline{Y}^r \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) = a_0 + \sum_{j=1}^{p} a_j \overline{x}_j^r,$$

where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$ are fixed values and $\overline{x}_j^r$ is the arithmetical mean of the quantities x_{ij} for i such that $\mathbf{x}_i \in \mathbf{C}_r$. The first of these relations is obvious. The second comes from Lemma 4.1.

THEOREM 4.2. *The vector* $\widehat{\mathbf{a}} = (\widehat{a}_1, \ldots, \widehat{a}_p)^T$ *is an unbiased estimator of the vector* $\mathbf{a} = (a_1, \ldots, a_p)^T$.

Proof. As in [5] we can show that

$$E(\widehat{a}_r \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) = a_r, \quad r = 1, \ldots, p.$$

The conditional expectation does not depend on the values of the random variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$, which completes the proof.

For fixed $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathfrak{M}$ we set

$$\begin{aligned} E(U_i \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) &= 0, \\ E(U_i^2 \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n) &= \sigma^2, \end{aligned} \tag{4.5}$$

$i = 1, \ldots, n$.

We assume that $\mathbf{a}_0 = (a_0, \ldots, a_p)^T$ is the vector of parameters in the model (1.3) and $\widehat{\mathbf{a}}_0 = (\widehat{a}_0, \ldots, \widehat{a}_p)$ is the vector of estimators of these parameters.

THEOREM 4.3.

$$\begin{aligned} \mathbf{\Sigma}_0 &= E[(\widehat{\mathbf{a}}_0 - \mathbf{a}_0)(\widehat{\mathbf{a}}_0 - \mathbf{a}_0)^T \mid \mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n] \\ &= \begin{bmatrix} \sigma^2/n & 0 \\ 0 & \sigma^2 \mathbf{W}_0^{-1} \mathbf{A} (\mathbf{W}_0^{-1})^T \end{bmatrix} \end{aligned}$$

where

$$\mathbf{A} = \left[\frac{k_{rs}}{k_r k_s} - \frac{k_{r0}}{k_r k_0} - \frac{k_{s0}}{k_s k_0} + \frac{k_{00}}{k_0 k_0} \right]_{r,s=1,\ldots,p}, \tag{4.6}$$

$k_{rs} = \operatorname{card}(\mathbf{C}_r \cap \mathbf{C}_s)$, *and* $\mathbf{W}_0$ *is the matrix obtained from* $\mathbf{W}(\omega)$ *by substituting* $\overline{x}_j^r$ *for* $\overline{X}_j^r$.

The proof is analogous to one presented in [5].

LEMMA 4.4. *Let $\{F_n\}_{\in\mathbb{N}}$ be a sequence of real-valued functions defined on $\mathbb{R}^p \times \mathbb{R}^m$. If*

$$\underset{\{S_n\}_{n\in\mathbb{N}}}{\forall} \Big(\lim_{n\to\infty} S_n = S_0,\ S_n, S_0 \in \mathbb{R}^m \ \Rightarrow \underset{x\in\mathbb{R}^p}{\forall} \lim_{n\to\infty} F_n(x, S_n) = F(x, S_0) \Big)$$

then

$$\underset{x\in\mathbb{R}^p}{\forall}\ \underset{\varepsilon>0}{\forall}\ \underset{n_0\in\mathbb{N}}{\exists}\ \underset{\delta>0}{\exists}\ \underset{n>n_0}{\forall}\ \underset{y\in\mathbb{R}^m}{\forall} \quad (\|y - S_0\| < \delta \ \Rightarrow\ |F_n(x, y) - F(x, S_0)| < \varepsilon)$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^m$.

Since the proof of this lemma is evident we omit it.

LEMMA 4.5. *Let $\{T_n\}_{n\in\mathbb{N}}$ be a sequence of random vectors on $\mathbb{R}^m$ with values in $\mathbb{R}^p$ and $\{S_n\}_{n\in\mathbb{N}}$ a sequence of random vectors on Ω with values in $\mathbb{R}^m$. Let T and S be random vectors of the same type as T_n and S_n respectively. If for each sequence $\{s_n\}_{n\in\mathbb{N}}$ in $\mathbb{R}^m$,*

$$\lim_{n\to\infty} s_n = s_0 \ \Rightarrow\ \underset{t\in\mathbb{R}}{\forall} \lim_{n\to\infty} P(T_n < t \mid S_n = s_n) = P(T < t \mid S = s_0) \tag{4.7}$$

and

$$\underset{\delta>0}{\forall} \quad \lim_{n\to\infty} P(\|S_n - s_0\| < \delta) > 0$$

then

$$\underset{t\in\mathbb{R}^p}{\forall}\ \underset{\varepsilon>0}{\forall}\ \underset{\delta>0}{\exists}\ \underset{n_0\in\mathbb{N}}{\exists}\ \underset{n\geq n_0}{\forall} \quad |P(T_n < t \mid \|S_n - s_0\| < \delta) - P(T < t \mid S = s_0)| < \varepsilon\,.$$

Proof. Defining $P(T_n < t|S_n = s_n) = F_n(t, s_n)$ and $P(T < t|S = s_0) = F(t, s_0)$ we obtain from Lemma 4.4

$$\underset{t\in\mathbb{R}^p}{\forall}\ \underset{\varepsilon>0}{\forall}\ \underset{\delta>0}{\exists}\ \underset{n_0\in\mathbb{N}}{\exists}\ \underset{n\geq n_0}{\forall}\ \underset{y\in\mathbb{R}^m}{\forall}$$
$$(\|y - s_0\| < \varepsilon \ \Rightarrow\ |P(T_n < t \mid S_n = y) - P(T < t \mid S = s_0)| < \varepsilon)\,.$$

The right-hand side of the above implication may be written as

$$-\varepsilon + P(T < t \mid S = s_0) < P(T_n < t \mid S_n = y) < \varepsilon + P(T < t \mid S = s_0)\,,$$

and after simple calculations

$$-\varepsilon + P(T < t \mid S = s_0) < \frac{\displaystyle\int_{\|y-s_0\|<\delta} P(T_n < t \mid S = y) f_{s_n}(y)\, dy}{\displaystyle\int_{\|y-s_0\|<\delta} f_{s_n}(y)\, dy} < \varepsilon + P(T < t \mid S = s_0)\,.$$

This gives

$$-\varepsilon + P(T < t \mid S = s_0) < P(T_n < t \mid \|S_n - s_0\| < \delta) < \varepsilon + P(T < t \mid S = s_0)\,,$$

and hence the assertion of the lemma.

LEMMA 4.6. *If sequences of random vectors* $\{T_n\}$ *and* $\{S_n\}$ *satisfy assumption* (4.7) *of Lemma* 4.5 *and if for some* s_0

$$\forall_{\varepsilon>0} \lim_{n\to\infty} P(\|S_n - s_0\| < \varepsilon) = 1$$

then

$$\forall_{t\in\mathbb{R}^p} \lim_{n\to\infty} P(T_n < t) = P(T < t \mid S = s_0).$$

Proof. By the total probability formula,

$$P(T_n < t) = P(\|S_n - s_0\| \geq \delta)P(T_n < t \mid \|S_n - s_0\| > \delta) + P(\|S_n - s_0\| < \delta)P(T_n < t \mid \|S_n - s_0\| < \delta), \tag{4.8}$$

providing that $P(\|S_n - s_0\| \geq \delta) > 0$ and $P(\|S_n - s_0\| < \delta) > 0$. Then by Lemma 4.5 for fixed t and $\varepsilon > 0$ there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for $n \geq n_0$

$$|P(T_n < t \mid \|S_n - s_0\| < \delta) - P(T < t \mid S = s_0)| < \varepsilon/2. \tag{4.9}$$

By assumption, there exists $n_1 \in \mathbb{N}$ such that for $n \geq n_1$

$$P(\|S_n - s_0\| < \delta) > 1 - \varepsilon/2. \tag{4.10}$$

Hence for $n \geq \max(n_0, n_1)$, from (4.8)–(4.10) we obtain

$$\begin{aligned} P(T_n < t) &> \left(1 - \frac{\varepsilon}{2}\right) P(T_n < t \mid \|S_n - s_0\| < \delta) \\ &> \left(1 - \frac{\varepsilon}{2}\right)\left(P(T < t \mid S = s_0) - \frac{\varepsilon}{2}\right) \\ &= -\left(\frac{\varepsilon}{2} + \frac{\varepsilon}{2}P(T < t \mid S = s_0) - \frac{\varepsilon^2}{4}\right) + P(T < t \mid S = s_0) \\ &> -\varepsilon + P(T < t \mid S = s_0) \end{aligned}$$

since

$$\frac{\varepsilon}{2} + \frac{\varepsilon}{2}P(T < t \mid S = s_0) - \frac{\varepsilon^2}{4} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} - \frac{\varepsilon^2}{4} < \varepsilon.$$

In a similar way we get $P(T_n < t) < \varepsilon + P(T < t \mid S = s_0)$, which yields

$$|P(T_n < t) - P(T < t \mid S = s_0)| < \varepsilon.$$

Finally, notice that if the conditions $P(\|S_n - s_0\| \geq \delta) > 0$ and $P(\|S_n - s_0\| < \delta) > 0$ are not both satisfied then under our assumptions the only possibility is $P(\|S_n - s_0\| \geq \delta) = 0$ and $P(\|S_n - s_0\| < \delta) = 1$, which does not affect the proof.

THEOREM 4.4. *If*

(i) $$\frac{1}{n} \operatorname{card} \mathbf{C}_r \xrightarrow{P} c_r > 0 \quad \text{for } r = 0, 1, \ldots, p,$$

(ii) $$n\mathbf{A} \xrightarrow{P} \mathbf{A}_0 \quad \text{where } \mathbf{A}_0 \text{ is a positive definite matrix,}$$

(iii) $$\mathbf{W} \xrightarrow{P} \mathbf{W}_0 \quad \textit{where } |\det \mathbf{W}_0| > 0,$$

then the joint distribution of the random vector $\mathbf{T}^{(n)} = (T_1^{(n)}, \ldots, T_p^{(n)})$ *where*

$$T_j^{(n)} = \frac{\widehat{a}_j^{(n)} - a_j}{\sqrt{D^2(\widehat{a}_j^{(n)})}}, \qquad j = 1, \ldots, p,$$

is asymptotically normal $N(0, \mathbf{\Sigma}_n)$ *where*

$$\mathbf{\Sigma}_n = \left[\frac{\mathbf{W}_r^t \mathbf{A} \mathbf{W}_s}{\sqrt{\mathbf{W}_r^T \mathbf{A} \mathbf{W}_r} \sqrt{\mathbf{W}_s^T \mathbf{A} \mathbf{W}_s}} \right].$$

Proof. First we define a sequence of random vectors S_n whose consecutive components are $n^{-1} \operatorname{card} \mathbf{C}_0, \ldots, n^{-1} \operatorname{card} \mathbf{C}_p$ followed by all elements of the matrices $n\mathbf{A}$ and $\mathbf{W}$. Then each S_n has values in $\mathbb{R}^m$ where $m = (p+1) + 2p^2$.

Under assumptions (i)–(iii) these vectors satisfy the condition

$$\bigvee_{\varepsilon > 0} \lim_{n \to \infty} P(\|S_n - s_0\| < \varepsilon) = 1$$

where s_0 is the vector whose components are $c_0, \ldots, c_p$ followed by all elements of the matrices $\mathbf{A}_0$ and $\mathbf{W}_0$. Furthermore, as a sequence of random vectors $\{\mathbf{T}^{(n)}\}$ we take the statistics

$$\mathbf{T}^{(n)} = (T_1^{(n)}, \ldots, T_p^{(n)})$$

depending on the sample size n. Using Lemma 4.6, it is enough to prove the convergence of the conditional distributions of the vectors $\mathbf{T}^{(n)}$ given that $\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_n = \mathbf{x}_n$ to the distribution $N(0, \mathbf{\Sigma}_n)$ under the assumptions:

(i′) $$\lim_{n \to \infty} k_r/n = c_r > 0,$$

(ii′) $$\lim_{n \to \infty} n\mathbf{A} = \mathbf{A}_0 \quad \text{where } \mathbf{A}_0 \text{ is a positive definite matrix},$$

(iii′) $$\lim_{n \to \infty} \mathbf{W} = \mathbf{W}_0 \quad \text{where } |\det \mathbf{W}_0| > 0.$$

In cases (i′)–(iii″) we have limits of sequences whose terms are not random. The proof of the above statement is presented in [5].

5. The influence of the way of sample division on the efficiency of estimators. In the class of linear estimators of unknown regression parameters, the least squares estimators are the best. Therefore the efficiency of estimators obtained by other methods is defined as the ratio of variances or generalized variances of estimators obtained by the considered method and the least squares method.

In the case of estimation of regression coefficients in the model $E(Y \mid X = x) = a_0 + a_1 x$, Wald [9] proposes to divide the sample into two groups taking to one group those values of the sample which are less than the

median, and to the other the remaining ones. This method, however, does not lead to high efficiency.

Nair and Shrivastava [4] and later Bartlett [1] and Theil and Yzeren [8] showed that better efficiency can be attained if the sample is divided into three groups in the following way. The first group consists of those observations X_i which are less than some fixed quantity Q_1 and the second consists of those X_i which are greater than some given Q_2 ($Q_2 > Q_1$). The remaining observations do not take part in the estimation.

The authors investigate the efficiency of the estimators obtained by this method using samples from different distributions. In particular, if the variable X has uniform distribution the efficiency of the considered estimators is 0.89.

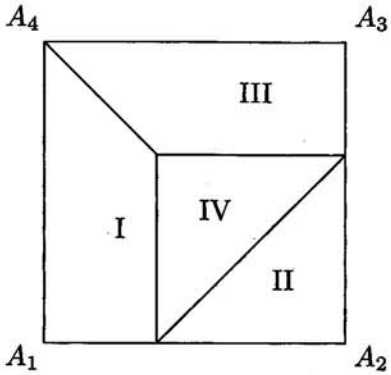

Fig. 1

As a generalization of this method, to estimate the regression coefficients in the model $E(Y \mid X_1, X_2) = a_0 + a_1 x_1 + a_2 x_2$ Theil and Hooper divide the sample into four groups. In particular, they consider the case when the bivariate random variable (X_1, X_2) has uniform distribution on the square $A_1A_2A_3A_4$. The optimal sample division is presented in Fig. 1. The part of the square labelled IV corresponds to that part of the sample which does not take part in the estimation of the parameters.

The efficiency of the estimators in that case is 0.61. This is significantly less than for the division of the sample into three groups for $p = 1$. It should be mentioned that both estimation methods presented above are troublesome because of the very complicated divisions.

6. Divisions by the p-point method. On an n-dimensional plan of some space we define multifunctions $\varphi_r : \mathfrak{M} \to 2^{\mathbb{R}^p}$, $r = 0, 1, \ldots, p$, by

$$\varphi_0(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathbb{R}^p,$$

$$\varphi_j(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \left\{(\mathbf{x}_1, \ldots, \mathbf{x}_p) \in \mathbb{R}^p : x_j \leq \frac{x_{1j} + \ldots + x_{nj}}{n}\right\}$$

for $j = 1, \ldots, p$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip}) \in \mathbb{R}^p$ for $i = 1, \ldots, n$. Conditions (2.1) are satisfied, so the above multifunctions induce the following divisions $\mathbf{C}_0, \ldots, \mathbf{C}_p$:

$$\mathbf{C}_0 = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)\}, \quad \mathbf{C}_j = \{(\mathbf{x}_i, y_i) : x_{ij} < \overline{x}_j\}, \tag{6.1}$$

where

$$\overline{x}_j = \frac{x_{1j} + \ldots + x_{nj}}{n}.$$

Such divisions are considered in [5], [6].

The matrix $\mathbf{W}$ defined by (3.1) is the matrix of a system of equations whose solution is the vector of estimators $(\widehat{a}_1, \ldots, \widehat{a}_p)$. In order to examine the behaviour of $\mathbf{W}$ as $n \to \infty$ when the division (6.1) is used, we must investigate the stochastic convergence of $\overline{X}_j^r - \overline{X}_j^0$.

For simplicity we assume that the considered random vectors $\mathbf{X}_1, \ldots, \mathbf{X}_n$ have the same distribution as a conventional vector $\mathbf{X} = (X_1, \ldots, X_p)$.

LEMMA 6.1. *For* $j = 1, \ldots, p$

$$\overline{X}_j^j - \overline{X}_j^0 = \frac{n}{2k_j} d_n^{(j)} \quad \textit{where} \quad d_n^{(j)} = \frac{1}{n} \sum_{i=1}^{n} |\overline{X}_{ij} - \overline{X}_j^0|.$$

The proof of this lemma can be found in [7].

It follows directly from the above that

$$\overline{X}_j^j - \overline{X}_j^0 \xrightarrow{P} \frac{d_{x_j}}{2p_j}$$

where

$$d_{x_j} = \int_{\mathbb{R}} |x - m| \widetilde{f}_j(x)\, d\mu, \quad p_j = \int_{\{X_j < m\}} \widetilde{f}_j(x)\, d\mu, \tag{6.2}$$

$\widetilde{f}_j$ is a marginal density function of the variable X_{ij} and m is its expectation.

LEMMA 6.2. *For* $j, l = 1, \ldots, p$, *if* $p_l > 0$, *then*

$$\overline{X}_l^j - \overline{X}_l^0 \xrightarrow{P} \frac{d_{x_l}}{2p_l} \frac{S_{x_j x_l}}{S_{x_l}^2}$$

where $S_{x_j x_l} = \operatorname{Cov}(X_j, X_l)$ *and* $S_{x_l}^2 = D^2(X_l)$.

Proof. It is enough to notice that

$$\overline{X}_l^j - \overline{X}_l^0 = (\overline{X}_j^j - \overline{X}_j^0) \frac{\overline{X}_l^j - \overline{X}_l^0}{\overline{X}_j^j - \overline{X}_j^0}$$

and that $(\overline{X}_l^j - \overline{X}_l^0)/(\overline{X}_j^j - \overline{X}_j^0)$ is an estimator of the slope of a straight line regression describing the dependence of X_l upon X_j. Hence by Theorem 4.1 the above estimator is stochastically convergent to the estimated

coefficient which can be written as $S_{x_j x_l}/S^2_{x_l}$. The conclusion now follows from Lemma 6.1.

Lemma 6.2 shows that

$$\mathbf{W} \xrightarrow{P} \operatorname{diag}\left[\frac{d_{x_1}}{2p_1 S^2_{x_1}}, \frac{d_{x_2}}{2p_2 S^2_{x_2}}, \ldots, \frac{d_{x_p}}{2p_p S^2_{x_p}}\right] \cdot \Sigma_X \tag{6.3}$$

where Σ_X is the covariance matrix of the vector X.

To compare the p-point method with the four groups method let us consider the bivariate distribution of a random variable (X_1, X_2) used by Theil and Hooper to investigate the efficiency of the estimators obtained by the four groups method. Therefore we assume that the vector (X_1, X_2) has uniform distribution on the unit square with vertices $A_1(0,0)$, $A_2(1,0)$, $A_3(1,1)$, $A_4(0,1)$. From (6.2)

$$p_1 = p_2 = 1/2, \quad d_{x_1} = d_{x_2} = 1/4,$$

and hence taking into account (6.3), the generalized variance $\det \Sigma$ of the vector of estimators satisfies

$$n^2 \det \Sigma = \frac{n^2 \det \mathbf{A}}{\det^2 \mathbf{W}} \xrightarrow{P} (16\sigma^2)^2$$

where the matrix $\mathbf{A}$ is defined by (4.6).

Calculating similarly the generalized variances for the least squares estimators we obtain an expression $(12\sigma^2)^2$ and hence

$$e = \sqrt{\frac{(12\sigma^2)^2}{(16\sigma^2)^2}} = 0.75.$$

Applying the divisions (6.1) we reached a considerable improvement of efficiency from 0.61 up to 0.75.

In the methods presented in [1], [4], [7], [8] one of the groups does not take part in the estimation. Moreover, when p increases, the efficiency decreases and the divisions are more and more complicated. Despite this the authors of those methods maintain that their methods are the best. But for the sample division (6.1) the efficiency of the estimators is independent of p and equal to 0.75.

7. The extended Gauss–Markov model and examination of an influence function. The properties of the conditional expectation and variance of the variable $U_i = Y_i - a_0 - a_1 X_{i1} - \ldots - a_p X_{ip}$ defined by (4.5) follow from assumptions 1.1–1.3 on the regression model.

Now we generalize the previous assumptions by assuming that the distributions of the variables U_i are characterized as follows:

$$
\begin{aligned}
E(U_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_n) &= m_i(\mathbf{X}_1, \ldots, \mathbf{X}_n), \\
D^2(U_i \mid \mathbf{X}_1, \ldots, \mathbf{X}_n) &= \sigma_i^2(\mathbf{X}_1, \ldots, \mathbf{X}_n)
\end{aligned} \tag{7.1}
$$

for $i = 1, \ldots, n$, where m_i are not necessarily zero and $\sigma_i^2 = \sigma^2$ need not hold almost everywhere as is the case in (4.5).

For simplicity of notation we fix the variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and assume that they are equal to some constants, say $\mathbf{x}_1, \ldots, \mathbf{x}_n$, where $\mathbf{x}_i \in \mathbb{R}^p$. Under this assumption we can omit in (7.1) the symbols $\mathbf{X}_1, \ldots, \mathbf{X}_n$ writing briefly m_i and σ_i for the conditional expectation and variance of U_i respectively.

THEOREM 7.1. *For $r = 1, \ldots, p$ the bias and risk of the considered estimators are*

(i) $$\operatorname{bias}(\widehat{a}_r) = E(\widehat{a}_r) - a_r = \frac{1}{\det \mathbf{W}} \mathbf{M}^T \mathbf{W}_r$$

where

$$\mathbf{M}^T = (M_1 - M_0, \ldots, M_p - M_0), \; M_j = \frac{1}{k_j} \sum_{i \in C_j} m_i, \; \mathbf{W}_r = (W_{1r}, \ldots, W_{pr})^T$$

and W_{jr} is the algebraic complement of the element ω_{jr} of the matrix $\mathbf{W}$, and

(ii) $$R(\widehat{a}_r) = E(\widehat{a}_r - a_r)^2 = \frac{1}{\det^2 \mathbf{W}} \mathbf{W}_r^T \mathbf{D} \mathbf{W}_r + \frac{1}{\det^2 \mathbf{W}} (\mathbf{M}^T \mathbf{W}_r)^2$$

where $\mathbf{D}$ is the matrix with elements

$$d_{ls} = \frac{1}{k_l k_s} \sum_{i \in \mathbf{C}_l \cap \mathbf{C}_s} \sigma_i^2 - \frac{1}{k_0 k_l} \sum_{i \in \mathbf{C}_0 \cap \mathbf{C}_l} \sigma_i^2 - \frac{1}{k_0 k_s} \sum_{i \in \mathbf{C}_0 \cap \mathbf{C}_s} \sigma_i^2 + \frac{1}{k_0^2} \sum_{i \in \mathbf{C}_0} \sigma_i^2 .$$

The proof can be found in [6].

As in [6], we now assume that a contamination happened for the ith observation $(\mathbf{X}_i, Y_i)$ only (recall that by assumption $\mathbf{X}_i$ has some fixed value $\mathbf{x}_i$).

An *influence function* for a functional $T : \mathcal{P} \to \mathbb{R}$ is defined to be the expression

$$IC_T(x \mid x_{i1}, \ldots, x_{ip}) = \lim_{\alpha \to 0^+} \frac{T[P_i(\alpha)] - T[P_i(0)]}{\alpha},$$

provided that the limit exists. Here

$$
\begin{aligned}
P_i(\alpha) &= (F_1, \ldots, F_{i-1}, (1 - \alpha) F_i + \alpha \delta_x, F_{i+1}, \ldots, F_n) \in \mathcal{P}, \\
\mathcal{P} &= \{P_i(\alpha) : 0 \leq \alpha \leq 1, \; i = 1, \ldots, n\},
\end{aligned}
$$

for the distribution F_i we assume

$$(7.2) \qquad m(F_i) = \int_{\mathbb{R}} y\,dF_i = 0, \quad \sigma^2(F_i) = \int_{\mathbb{R}} (y - m(F_i))^2\,dF_i = \sigma^2$$

and δ_x is the distribution concentrated at $x \in \mathbb{R}$.

Taking for T bias and risk of estimators we obtain the influence functions of interest to us. First notice that as a consequence of assumptions (7.2) we have

$$m_{i'} = 0, \qquad \sigma_{i'}^2 = \sigma^2 \quad \text{for } i' \neq i,\ i = 1, \dots, n,$$
$$m_i = \alpha x, \qquad \sigma_i^2 = (1-\alpha)\sigma^2 + \alpha(1-\alpha)x^2 .$$

THEOREM 7.2. *For the divisions* (6.1) *the influence functions for bias and risk are*

$$\text{(i)} \qquad IC_{\operatorname{bias} \hat{a}_r}(x \mid x_{i1}, \dots, x_{ip}) = \frac{1}{\det \mathbf{W}} \mathbf{L}^T \mathbf{W}_r$$

where $\mathbf{L}^T = (l_1, \dots, l_p)$ *and*

$$l_s = \begin{cases} \left(\dfrac{1}{k_s} - \dfrac{1}{n}\right) x & \text{if } x_i \in \mathbf{C}_s , \\ -\dfrac{1}{n} x & \text{if } x_i \in \mathbf{C}_s' , \end{cases}$$

$$\text{(ii)} \qquad IC_{R(\hat{a}_r)}(x \mid x_{i1}, \dots, x_{ip}) = \frac{1}{\det^2 \mathbf{W}} \mathbf{W}_r^T \mathbf{U} \mathbf{W}_r$$

where $\mathbf{U} = [u_{ls}]_{l,s=1,\dots,p}$, *with*

$$u_{ls} = \begin{cases} \left(\dfrac{1}{k_l k_s} - \dfrac{1}{n k_l} - \dfrac{1}{n k_s} + \dfrac{1}{n^2}\right)(x^2 - \sigma^2) & \text{if } x_i \in \mathbf{C}_l \cap \mathbf{C}_s , \\ \left(-\dfrac{1}{n k_s} + \dfrac{1}{n^2}\right)(x^2 - \sigma^2) & \text{if } x_i \in \mathbf{C}_l' \cap \mathbf{C}_s , \\ \left(-\dfrac{1}{n k_l} + \dfrac{1}{n^2}\right)(x^2 - \sigma^2) & \text{if } x_i \in \mathbf{C}_l \cap \mathbf{C}_s' , \\ \dfrac{1}{n^2}(x^2 - \sigma^2) & \text{if } x_i \in \mathbf{C}_l' \cap \mathbf{C}_s' , \end{cases}$$

Proof. (i) Taking for T the bias of the estimator $\widehat{a}_r$ in the definition of the influence function and using Theorem 7.1 we obtain

$$IC_{\operatorname{bias} \hat{a}_r}(x \mid x_{i1}, \dots, x_{ip}) = \lim_{\alpha \to 0^+} \frac{\frac{1}{\det \mathbf{W}} \mathbf{M}^T(P_i(\alpha)) \mathbf{W}_r - \frac{1}{\det \mathbf{W}} \mathbf{M}^T(P(\alpha)) \mathbf{W}_r}{\alpha} .$$

By (7.2), $\mathbf{M}^T(P_i(0)) = (0, \dots, 0)$ and, for $\alpha \neq 0$, $\mathbf{M}^T(P_i(\alpha)) = (l_1, \dots, l_p) = \mathbf{L}^T$. Hence we have the assertion of (i). The more complicated calculations leading to (ii) are omitted.

Remember that the above results have been achieved under the assumption that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are fixed. Now we no longer assume the variables to be fixed and consider the influence function in two cases: when the variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ have multidimensional uniform and normal distribution.

7a. *Influence functions in the case of uniform distribution.* In this case we assume that $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent random variables with the same uniform distribution on a p-dimensional polyhedron. This distribution has density

$$f(x_1, \ldots, x_p) = \begin{cases} 1 & \text{if } -1/2 \le x_j \le 1/2 \text{ for all } j, \\ 0 & \text{otherwise} \end{cases} \tag{7.3}$$

with respect to p-dimensional Lebesgue measure. Hence we immediately obtain $\mathbf{\Sigma} = \frac{1}{12}\mathbf{I}$ where $\mathbf{I}$ is the unit matrix. By (6.2) we have

$$d_{x_j} = \int_{-1/2}^{1/2} |x|\, dx = 1/4, \qquad p_j = 1/2,$$

hence by (6.3)

$$\mathbf{W} \xrightarrow{P} \tfrac{1}{4}\mathbf{I}$$

and the Slutsky theorem yields

$$\mathbf{W}_r^T \xrightarrow{P} (0, \ldots, 0, (1/4)^{p-1}, 0, \ldots, 0)$$

where $(1/4)^{p-1}$ is in the rth place. Using the Slutsky theorem once again we obtain

$$\det \mathbf{W} \xrightarrow{P} (1/4)^p.$$

Now we turn to examining the influence functions. Using Theorem 7.2 we obtain

$$\begin{aligned} &nIC_{\text{bias}\,\hat{a}_r}(x \mid x_{i1}, \ldots, x_{ip}) \\ &\quad = \frac{1}{\det \mathbf{W}} n\mathbf{L}^T \mathbf{W}_r \xrightarrow{P} \frac{1}{(1/4)^p}(nl_1, \ldots, nl_p)(0, \ldots, (1/4)^p, \ldots, 0)^T \\ &\quad = \frac{nl_r}{(1/4)^p}(1/4)^{p-1} = 4nl_r, \end{aligned}$$

and since $|nl_r| \xrightarrow{P} x$ we get

$$|nIC_{\text{bias}\,\hat{a}_r}(x \mid x_{i1}, \ldots, x_{ip})| \xrightarrow{P} 4|x|. \tag{7.4}$$

Next, by the same theorem $n^2 u_{ls} \xrightarrow{P} x^2 - \sigma^2$ and after simple calculations we have

$$|n^2 IC_{R(\hat{a}_r)}(x \mid x_{i1}, \ldots, x_{ip}) \xrightarrow{P} 16|x^2 - \sigma^2|. \tag{7.5}$$

To compare the above results we give the similar relations for the influence functions of the estimators $\widehat{b}_r$ obtained by the least squares method:

$$
\begin{aligned}
&|nIC_{\text{bias}\,\hat{b}_r}(x \mid x_{i1}, \ldots, x_{ip})| \xrightarrow{P} 12|x_{ir}||x|\,, \\
&|n^2 IC_{R(\hat{b}_r)}(x \mid x_{i1}, \ldots, x_{ip})| \xrightarrow{P} 144|x_{ir}||x^2 - \sigma^2|\,.
\end{aligned} \tag{7.6}
$$

Defining the influence coefficients

$$
\begin{aligned}
e_{\text{bias}} &= \frac{IC_{\text{bias}\,b_r}(x \mid x_{i1}, \ldots, x_{ip})}{IC_{\text{bias}\,\hat{a}_r}(x \mid x_{i1}, \ldots, x_{ip})}\,, \\
e_R &= \frac{IC_{R(b_r)}(x \mid x_{i1}, \ldots, x_{ip})}{IC_{R(\hat{a}_r)}(x \mid x_{i1}, \ldots, x_{ip})}\,,
\end{aligned} \tag{7.7}
$$

we can formulate the following theorem.

THEOREM 7.3. *For independent random variables* $\mathbf{X}_1, \ldots, \mathbf{X}_n$ *with distributions determined by* (7.3),

$$
|e_{\text{bias}}| \xrightarrow{P} 3|x_{ir}|, \qquad |e_R| \xrightarrow{P} 9x_{ir}^2\,.
$$

The proof is immediate from (7.4)–(7.6).

It can be seen that the p-point method is more robust than the least squares method with regard to both bias and risk for $|x_{ij}| > 1/3$, because for such $|x_{ij}|$, the values of the influence function for the p-point method are less than for the least squares method. Notice also that for the p-point method the influence of a disturbance is constant no matter for which $\mathbf{X}_i$ it happened.

7b. *Influence functions in the case of normal distribution.* Now we assume that the variables $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are independent with the same multivariate normal distribution $N(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a matrix of the form

$$
\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)\,.
$$

By assumption,

$$
d_{x_j} = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(\frac{-x^2}{2\sigma_j^2}\right) dx = \sqrt{\frac{2}{\pi}}\sigma_j, \qquad p_j = \frac{1}{2}\,.
$$

THEOREM 7.4. *For independent random variables with the same distribution* $N(0, \boldsymbol{\Sigma})$, *by* (6.3) *and* (7.7),

$$
|e_{\text{bias}}| \xrightarrow{P} \sqrt{\frac{2}{\pi}} \frac{|x_{ir}|}{\sigma_r}, \qquad |e_R| \xrightarrow{P} \frac{2}{\pi} \frac{x_{ir}^2}{\sigma_r^2}\,.
$$

References

[1] M. S. Bartlett, *Fitting a straight line where both variables are subject to error*, Biometrics 5 (1949), 207–212.

[2] R. F. Hampel, *The influence curve and role in robust estimation*, J. Amer. Statist. Assoc. 69 (1974), 383–393.

[3] Z. Hellwig, *Determining the parameters of a linear regression by the 2-point method*, Zastos. Mat. 3 (1956), 66–81 (in Polish).

[4] K. R. Nair and M. P. Shrivastava, *On a simple method of curve fitting*, Sankhyā 6 (1942), 121–132.

[5] C. Platt and Z. Paprzycki, *On estimation of multiple regression coefficients by the p-point method*, Zastos. Mat. 20 (1) (1988), 53–65.

[6] —, —, *On robustness of multiple regression coefficients estimators obtained by the p-point method*, ibid. 20 (2) (1990), 203–210.

[7] H. Theil and J. W. Hooper, *The extension of Wald's method of fitting straight lines*, Rev. Inst. Internat. Statist. 26 (1956), 37–47.

[8] H. Theil and J. van Yzeren, *On the efficiency of Wald's method of fitting straight lines*, ibid. 24 (1956), 17–26.

[9] A. Wald, *The fitting of straight lines if both variables are subject to error*, Ann. Math. Statist. 11 (1940), 284–300.

CZESŁAW PLATT
ZBIGNIEW PAPRZYCKI
DEPT. OF APPLICATION OF MATHEMATICS
AND MATHEMATICAL STATISTICS
OLSZTYN UNIVERSITY OF AGRICULTURE AND TECHNOLOGY
10-740 OLSZTYN, POLAND

Received on 30.3.1990;
revised version on 6.8.1992