Tatra Mt. Math. Publ. **39** (2008), 77–85

Tatra Mountains
Mathematical Publications

SIMULATION STUDY OF INSENSITIVITY REGIONS AND OUTLIERS

Eva Fišerová

ABSTRACT. A procedure how to identify outliers in observations in a regular linear model with constraints on mean value parameters is presented. A problem, how approximations of variance components influence the significance level of statistical test is solved by insensitivity approach. Explicit expressions of insensitivity regions are given. Behavior of insensitivity regions and the quality of the procedure for outliers identification are studied by simulations.

1. Introduction

There are several procedures how to identify outliers in observations, cf. e.g., [1]. If unknown variance components occur in a covariance matrix of an observation vector, it is of some interest to know, whether approximations of them can be used instead of their true values. Such a problem can be analyzed by the insensitivity approach.

Approximations of variance components can destroy the optimum quality of used statistical inference, namely a significance level of a statistical test for outliers identification in measurements. The main goal of the insensitivity approach is to find a set of all values of variance components which make the tolerable increase of the significance level of the test, cf. e.g., [5], [6].

Basic theoretical results on outliers and insensitivity regions in a regular linear model with outliers when mean value parameters satisfy linear constraints have been done in [3]. The aim of the paper is to study behavior of insensitivity regions and the quality of procedure for outliers identification by simulations.

2000 Mathematics Subject Classification: Primary 62J05; Secondary 62F03, 62F10, 62F30.
Keywords: regression model with constraints, variance components, outlier, insensitivity region, unbiased estimator, hypothesis testing, simulation.
The research was supported by the Council of Czech Government MSM 6 198 959 214.

2. Identification of outliers in measurement

Let us consider a linear regression model with constraints in the form

$$\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}\right), \qquad \boldsymbol{\beta} \in \mathcal{V} = \{\mathbf{u} : \mathbf{b} + \mathbf{B}\mathbf{u} = \mathbf{0}\}. \tag{1}$$

Here $\mathbf{Y}$ is an n-dimensional random vector (observation vector) which is normally distributed, its mean value is $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and the covariance matrix is $\mathrm{Var}(\mathbf{Y}) = \boldsymbol{\Sigma}$. The parametric space for $\boldsymbol{\beta}$ is $\mathcal{V}$, $\boldsymbol{\beta} \in \mathbb{R}^k$ is an unknown vector, $\mathbf{X}$ and $\mathbf{B}$ are given matrices of types $n \times k$ and $q \times k$, $\mathbf{b} \in \mathbb{R}^q$ is a given vector.

The model (1) will be supposed to be regular, i.e., the matrix $\mathbf{X}$ has full column rank, $\boldsymbol{\Sigma}$ is a positive definite and $\mathbf{B}$ has full row rank.

Let the covariance matrix $\boldsymbol{\Sigma}$ be known. The procedure for outliers identification can be done in several steps. The first step is to estimate $\boldsymbol{\beta}$ in the model (1) as the best linear unbiased estimator (BLUE) $\widehat{\boldsymbol{\beta}}$ (cf. [4], p. 80), i.e.,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{B'})^{+}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{b},$$

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{B}'(\mathbf{B}\mathbf{C}^{-1}\mathbf{B}')^{-1}\mathbf{B}\mathbf{C}^{-1} = (\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{B'})^{+}.$$

The symbol $(\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{B'})^{+}$ means the Moore–Penrose generalized inverse of the matrix $\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{B'}$ (cf. [7]), $\mathbf{C} = \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$, $\mathbf{M}_{B'} = \mathbf{I} - \mathbf{P}_{B'}$ and $\mathbf{P}_{B'} = \mathbf{B}'(\mathbf{B}')^{+}$.

The next step is to determine the residual vector

$$\mathbf{v} = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}} \sim N_n\left[\mathbf{0}, \boldsymbol{\Sigma} - \mathbf{X}(\mathbf{M}_{B'}\mathbf{C}\mathbf{M}_{B'})^{+}\mathbf{X}'\right].$$

Then we can find suspicious measurements y_i, $i = i_1, \dots, i_r$, by testing the hypothesis $H_0 : E(\mathbf{v}) = \mathbf{0}$ versus $H_a : E(\mathbf{v}) \neq \mathbf{0}$ by the test statistic

$$T = \mathbf{v}'\boldsymbol{\Sigma}^{-1}\mathbf{v} \sim \chi^2_{n+q-k}(\delta_1), \qquad \delta_1 = E(\mathbf{v})'\boldsymbol{\Sigma}^{-1}E(\mathbf{v}).$$

In view of the Scheffé theorem (cf. [8]), the ith measurement is suspicious if

$$|\{\mathbf{v}\}_i| \geq \sqrt{\chi^2_{n+q-k}(1-\alpha)}\sqrt{\{\mathrm{Var}(\mathbf{v})\}_{ii}}, \qquad i \in \{1, \dots, n\}.$$

Here $\chi^2_{n+q-k}(1-\alpha)$ is the $(1-\alpha)$-quantile of the chi-squared distribution with $n+q-k$ degrees of freedom.

If no suspicious large value $|\{\mathbf{v}\}_i|$, i.e., no suspicious measurement, is found, stop this procedure. If r suspicious measurements are found, the model (1) can be rewritten as

$$\mathbf{Y} \sim N_n\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{E}\boldsymbol{\Delta}, \boldsymbol{\Sigma}\right), \qquad \boldsymbol{\beta} \in \mathcal{V} = \{\mathbf{u} : \mathbf{b} + \mathbf{B}\mathbf{u} = \mathbf{0}\}, \qquad \boldsymbol{\Delta} \in \mathbb{R}^r, \tag{2}$$

where

$$\mathbf{E} = (\mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_r}), \quad \mathbf{e}_{i_j} \in \mathbb{R}^n, \quad j = 1, \dots, r, \quad \left\{\mathbf{e}_{i_j}\right\}_k = \begin{cases} 0, & k \neq i_j, \\ 1, & k = i_j, \end{cases}$$

and i_j is the index with suspicious large value $|\{\mathbf{v}\}_{i_j}|$. Outliers among suspicious measurements y_i, $i = i_1, \dots, i_r$, can be identified by testing the hypothesis

$$H_0 : \boldsymbol{\Delta} = \mathbf{0} \quad \text{versus} \quad H_a : \boldsymbol{\Delta} \neq \mathbf{0}. \tag{3}$$

The hypothesis (3) can be tested in the model (2) if and only if (cf. [3])

$$\mathcal{M}(\mathbf{X}\mathbf{M}_{B'}) \cap \mathcal{M}(\mathbf{E}) = \{\mathbf{0}\} \quad \Leftrightarrow \quad \mathcal{M}(\mathbf{X}', \mathbf{B}') \cap \mathcal{M}(\mathbf{E}', \mathbf{0}) = \{\mathbf{0}\}. \tag{4}$$

Here $\mathcal{M}(\mathbf{A}_{m,n}) = \{\mathbf{A}\mathbf{u} : \mathbf{u} \in \mathbb{R}^n\} \subset \mathbb{R}^m$. Under the condition (4), the BLUE $\widehat{\boldsymbol{\Delta}}$ of the parameter $\boldsymbol{\Delta}$ in the model (2) is (cf. [3])

$$\widehat{\boldsymbol{\Delta}} = \left[\mathbf{E}' \left(\mathbf{M}_{XM_{B'}} \boldsymbol{\Sigma} \mathbf{M}_{XM_{B'}}\right)^+ \mathbf{E}\right]^{-1} \mathbf{E}' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}),$$

$$\mathrm{Var}(\widehat{\boldsymbol{\Delta}}) = \left[\mathbf{E}' \left(\mathbf{M}_{XM_{B'}} \boldsymbol{\Sigma} \mathbf{M}_{XM_{B'}}\right)^+ \mathbf{E}\right]^{-1}.$$

Now the hypothesis (3) can be tested via the test statistic

$$T_{out} = \widehat{\boldsymbol{\Delta}}' [\mathrm{Var}(\widehat{\boldsymbol{\Delta}})]^{-1} \widehat{\boldsymbol{\Delta}} \sim \chi^2_r(\delta_2), \qquad \delta_2 = \boldsymbol{\Delta}' [\mathrm{Var}(\widehat{\boldsymbol{\Delta}})]^{-1} \boldsymbol{\Delta}.$$

Here it is necessary to distinguish two different significance levels. If the investigator suspects in advance in which measurements are outliers, i.e., indices of suspicious measurements are a priori known, by the Scheffé theorem, the i^*th measurement is considered to be an outlier if

$$|\{\boldsymbol{\Delta}\}_{i^*}| \geq \sqrt{\chi^2_r(1-\alpha)} \sqrt{\left\{\mathrm{Var}(\widehat{\boldsymbol{\Delta}})\right\}_{i^* i^*}}, \qquad i^* \in \{i_1, \dots, i_r\}, \tag{5}$$

where i_j is the index with suspicious measurement. Usually, the investigator has no a priori choice for outliers. Suspicious measurements are detected, e.g., from residuals. So, we are in reality performing $\binom{n}{r}$ significance tests, one for each $\binom{n}{r}$ cases. By the Bonferonni's inequality, the i^*th measurement is an outlier if

$$|\{\boldsymbol{\Delta}\}_{i^*}| \geq \sqrt{\chi^2_r \left[1 - \alpha \Big/ \binom{n}{r}\right]} \sqrt{\left\{\mathrm{Var}(\widehat{\boldsymbol{\Delta}})\right\}_{i^*, i^*}}, \qquad i^* \in \{i_1, \dots, i_r\}. \tag{6}$$

At the last step significant outliers y_{i^*} are omitted from realization of the observation vector $\mathbf{Y}$ and the whole procedure is repeated.

3. Insensitivity regions

Let the covariance matrix in models (1), (2) be in the form $\boldsymbol{\Sigma} = \sum_{i=1}^{p} \vartheta_i \mathbf{V}_i$, $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_p)' \in \underline{\vartheta} \subset \mathbb{R}^p$. Here, except $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}$, also the vector parameter $\boldsymbol{\vartheta}$ is unknown. $\mathbf{V}_1, \dots, \mathbf{V}_p$ are known symmetric matrices. The parametric space $\underline{\vartheta}$

is an open set in $\mathbb{R}^p$ with the property that if $\boldsymbol{\vartheta} \in \underline{\vartheta}$, then $\sum_{i=1}^{p} \vartheta_i \mathbf{V}_i$ is a positive definite.

If we have the approximation $\boldsymbol{\vartheta}_0$ of $\boldsymbol{\vartheta}$, outliers can be identified by the procedure given in the previous section (estimators are $\boldsymbol{\vartheta}_0$-locally BLUE only). However, the substitution of the true value $\boldsymbol{\vartheta}^*$ by its approximation $\boldsymbol{\vartheta}_0$ can destroy the optimum quality of significance levels of tests T and T_{out}.

Let $\varepsilon > 0$ be a given tolerable increase of the significance level α of the test T. Let δ_ε be given as a solution of the equation

$$\begin{aligned} &\mathcal{P}_{H_0}\left\{T(\boldsymbol{\vartheta}^*) + \delta_\varepsilon \geq \chi^2_{n+q-k}(1-\alpha)\right\} = \alpha + \varepsilon \\ &\Rightarrow \quad \delta_\varepsilon = \chi^2_{n+q-k}(1-\alpha) - \chi^2_{n+q-k}(1-\alpha-\varepsilon). \end{aligned}$$

The symbol $\mathcal{P}_{H_0}$ means the probability under the null hypothesis H_0. The insensitivity region $\mathcal{N}_\varepsilon$ for the significance level α of the test T is the set of all points $\boldsymbol{\vartheta}_0 = \boldsymbol{\vartheta}^* + \delta\boldsymbol{\vartheta}$ such that if $\boldsymbol{\vartheta}_0 \in \mathcal{N}_\varepsilon$, then the significance level is not larger than $\alpha + \varepsilon$. It can be expressed as

$$\mathcal{N}_\varepsilon = \left\{\boldsymbol{\vartheta}^* + \delta\boldsymbol{\vartheta} : \left(\delta\boldsymbol{\vartheta} - \delta_\varepsilon \mathbf{D}_t^+ \mathbf{a}\right)' \mathbf{D}_t \left(\delta\boldsymbol{\vartheta} - \delta_\varepsilon \mathbf{D}_t^+ \mathbf{a}\right) \leq \left(1 + \mathbf{a}'\mathbf{D}_t^+\mathbf{a}\right) \delta_\varepsilon^2 \right\}, \tag{7}$$

where $t > 0$ is a sufficiently large number and

$$\begin{aligned} &\mathbf{D}_t = 2t^2\mathbf{S}_K - \mathbf{a}\mathbf{a}', && \mathbf{K} = [\mathbf{M}_{XM_{B'}} \boldsymbol{\Sigma}(\boldsymbol{\vartheta}^*) \mathbf{M}_{XM_{B'}}]^+, \\ &\{\mathbf{a}\}_i = \mathrm{Tr}\{\mathbf{K}\mathbf{V}_i\}, && \{\mathbf{S}_K\}_{i,j} = \mathrm{Tr}\{\mathbf{K}\mathbf{V}_i\mathbf{K}\mathbf{V}_j\}, \ \ i,j = 1,\dots,p. \end{aligned}$$

Similarly, we have two insensitivity regions for significance levels of the test T_{out} ($\mathcal{N}^{giv}_{out,\varepsilon}$ for a priori given indices of suspicious measurements, $\mathcal{N}^{rand}_{out,\varepsilon}$ for random indices) such that

$$\begin{aligned} &\boldsymbol{\vartheta}_0 \in \mathcal{N}^{giv}_{out,\varepsilon} \quad \Rightarrow \quad \mathcal{P}_{H_0}\left\{T_{out}(\boldsymbol{\vartheta}_0) \geq \chi^2_r(1-\alpha)\right\} \leq \alpha + \varepsilon, \\ &\boldsymbol{\vartheta}_0 \in \mathcal{N}^{rand}_{out,\varepsilon} \quad \Rightarrow \quad \mathcal{P}_{H_0}\left\{T_{out}(\boldsymbol{\vartheta}_0) \geq \chi^2_r\left[1-\alpha/\tbinom{n}{r}\right]\right\} \leq (\alpha+\varepsilon)/\tbinom{n}{r}. \end{aligned}$$

The explicit expressions of $\mathcal{N}^{giv}_{out,\varepsilon}$ and $\mathcal{N}^{rand}_{out,\varepsilon}$ are given by (7) when symbols are substituted according to the scheme

$$\mathbf{a} \to \mathbf{a}_{out}, \qquad \mathbf{D}_t \to \mathbf{D}_{out,t}, \qquad \delta_\varepsilon \to \delta^{giv}_{out,\varepsilon} \text{ or } \delta^{rand}_{out,\varepsilon},$$

$$\begin{aligned} &\mathbf{D}_{out,t} = t^2\left(4\mathbf{C}_U - 2\mathbf{S}_Z\right) - \mathbf{a}_{out}\mathbf{a}'_{out}, \qquad \mathbf{Z} = \mathbf{K}\mathbf{E}\left(\mathbf{E}'\mathbf{K}\mathbf{E}\right)^{-1}\mathbf{E}'\mathbf{K}, \\ &\{\mathbf{a}_{out}\}_i = \mathrm{Tr}(\mathbf{Z}\mathbf{V}_i), \qquad \{\mathbf{C}_U\}_{i,j} = \mathrm{Tr}(\mathbf{K}\mathbf{V}_i\mathbf{Z}\mathbf{V}_j), \qquad \{\mathbf{S}_Z\}_{i,j} = \mathrm{Tr}(\mathbf{Z}\mathbf{V}_i\mathbf{Z}\mathbf{V}_j) \end{aligned}$$

for $i,j = 1,\dots,p$ and $\delta^{giv}_{out,\varepsilon}$, $\delta^{rand}_{out,\varepsilon}$ are given by

$$\begin{aligned} \delta^{giv}_{out,\varepsilon} &= \chi^2_r(1-\alpha) - \chi^2_r(1-\alpha-\varepsilon), \\ \delta^{rand}_{out,\varepsilon} &= \chi^2_r\left[1-\alpha/\tbinom{n}{r}\right] - \chi^2_r\left[1-(\alpha+\varepsilon)/\tbinom{n}{r}\right]. \end{aligned}$$

Note that the insensitivity regions aren't centered at $\boldsymbol{\vartheta}^*$. For more details about the insensitivity regions see [3]. The parameter t can be chosen in the interval $< 3, 5 >$. For the optimum value of t see [6].

4. Simulation study

EXAMPLE 4.1. In the triangle $P_1P_2P_3$ the distances $\beta_1 = \overline{P_2P_3} = 800$ m, $\beta_2 = \overline{P_3P_1} = 900$ m, $\beta_3 = \overline{P_1P_2} = 600$ m and the angles $\beta_4 = \{P_2P_1P_3\} = 60°36'36.59''$, $\beta_5 = \{P_3P_2P_1\} = 78°35'17.43''$, $\beta_6 = \{P_1P_3P_3\} = 40°48'15.98''$ are measured just twice with the accuracy σ_s (distances) and σ_ω (angles).

The problem is to identify the outlier y_1 among measurement. Using simulations we will study behavior of insensitivity regions for the significance level $\alpha = 0.05$ and its tolerable increment $\varepsilon = 0.05$ in dependence on the choice of the true (certificate) values σ_s^*, σ_ω^* and the choice of the significance level for the test T_{out} (given or random indices of suspicious measurements). The significance level for given indices is used as a heuristic approach in the sense that although suspicious measurements aren't a priori known and they are detected from residuals, the significance level is chosen as for a priori known indices. Finally, we compare the quality of the outlier identification for both types of significance levels.

The process of measurement can be modelled by (more detail cf. [2])

$$Y_{2i-1} = \beta_i + \varepsilon_{2i-1}, \qquad Y_{2i} = \beta_i + \varepsilon_{2i}, \qquad i = 1, \dots, 6,$$
$$\beta_4 + \beta_5 + \beta_6 = 180°, \qquad \beta_1 \sin\beta_5 = \beta_2 \sin\beta_4, \qquad \beta_1 \sin\beta_6 = \beta_3 \sin\beta_4.$$

The observation vector $\mathbf{Y}$ was generated in a natural way, an error term was added to the true mean $\mathbf{X}\boldsymbol{\beta}$ except $Y_1 = 800 + 0.75 + \varepsilon_1$, i.e., the observation y_1 is an outlier. The error term had the distribution $N_{12}\big[0, (\sigma_s^*)^2\mathbf{V}_s + (\sigma_\omega^*)^2\mathbf{V}_\omega\big]$, where different accuracies σ_s^*, σ_ω^* are given in Table 1, $\mathbf{V}_s = \mathrm{Diag}(\mathbf{1}_{1\times 6}, \mathbf{0}_{1\times 6})$ and $\mathbf{V}_\omega = \mathrm{Diag}(\mathbf{0}_{1\times 6}, \mathbf{1}_{1\times 6})$ are diagonal matrices.

There were done 10 000 simulations for each of five different accuracies $\boldsymbol{\vartheta}^* = \big((\sigma_s^*)^2, (\sigma_\omega^*)^2\big)'$ denoted by the case A–E. Maximum increase and decrease (in %) of $\boldsymbol{\vartheta}^*$ which makes tolerable increment $\varepsilon = 0.05$ of $\alpha = 0.05$ is given in Tables 2, 3. Changes are the same for all cases A–E. Relatively large uncertainty can be tolerated in $\boldsymbol{\vartheta}^*$ when detecting suspicious measurements (test T). If ϑ_s is true, then ϑ_ω can increase of 62% and the significance level of the test T is not larger than $\alpha + \varepsilon = 0.1$, e.g., $3.52 \le \vartheta_\omega \le 6.48$ seconds^2 in the case A. Rather smaller uncertainty occurs when searching significant outliers (test T_{out}).

The insensitivity ellipses $\mathcal{N}_{0.05}$, $\mathcal{N}_{out,0.05}^{giv}$ and $\mathcal{N}_{out,0.05}^{rand}$ for $t = 3$ are shown in Figures 1, 2. In Figure 1, they are given in dependence on the number

TABLE 1. Accuracy of measurement

case	σ_s^* (cm)	σ_ω^* (second)	ϑ_s^* (cm^2)	ϑ_ω^* (second^2)
B	1.16	1.21	1.37	1.46
C	1.11	1.68	1.23	2.83
A	1.00	2.00	1.00	4.00
D	0.84	2.21	0.71	4.90
E	0.61	2.34	0.37	5.46

TABLE 2. Tolerable changes (in %) of $\boldsymbol{\vartheta}^* - \mathcal{N}_{0.05}$.

	increase	decrease
ϑ_ω	62	12

	increase	decrease
ϑ_s	72	22

TABLE 3. Tolerable changes (in %) of $\boldsymbol{\vartheta}^* - \mathcal{N}_{out,0.05}^{giv}$, $\mathcal{N}_{out,0.05}^{rand}$.

	given indices			random indices		
	$r = 2$	$r = 3$	$r = 4$	$r = 2$	$r = 3$	$r = 4$
increase of ϑ_ω	44	41	41	44	38	36
decrease of ϑ_ω	27	20	17	27	19	14
increase of ϑ_s	44	52	59	44	48	52
decrease of ϑ_s	27	31	35	27	29	31

TABLE 4. Empirical probabilities (in %) of the number r of suspicious measurements.

r	case B	case C	case A	case D	case E
2	96.55	98.93	99.51	99.69	99.73
3	3.44	1.07	0.47	0.31	0.25
4	0.01	0	0.02	0	0.02

TABLE 5. Empirical probabilities (in %) of outliers identification.

indices	significant	case B	case C	case A	case D	case E
given	y_1 only	98.14	98.18	95.79	91.21	88.08
random	y_1 only	99.62	89.47	68.34	53.45	45.59
given	y_1 and y_j, $j \neq 1$	1.86	1.35	0.83	0.66	0.85
random	y_1 and y_j, $j \neq 1$	0	0.01	0	0.01	0.01
given	$y_j \neq y_1$ only	0	0.46	0.68	0.82	0.67
random	$y_j \neq y_1$ only	0	0	0	0	0
given	none	0	0.01	2.70	7.31	10.40
random	none	0.38	10.52	31.64	46.54	54.40

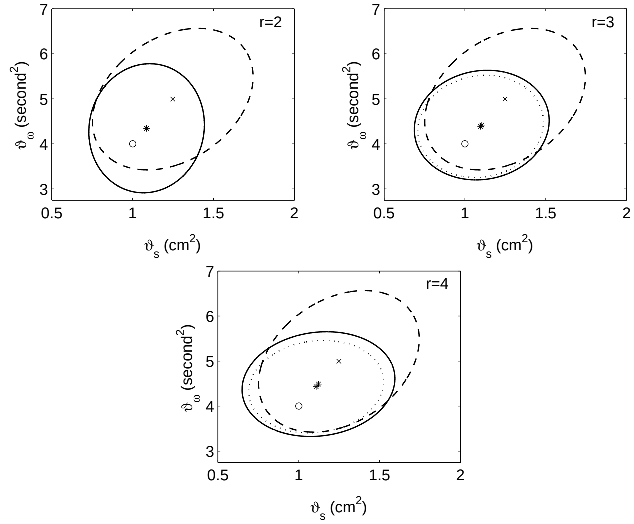

FIGURE 1. Insensitivity ellipses in dependence on the number of suspicious measurements. $\mathcal{N}_{0.05}$ by dashed line, its center at $\times$, $\mathcal{N}^{giv}_{out,0.05}$ by solid line, $\mathcal{N}^{rand}_{out,0.05}$ by dotted line. The symbol $\circ$ means $\boldsymbol{\vartheta}^*$, $*$ means the center of $\mathcal{N}^{giv}_{out,0.05}$ or $\mathcal{N}^{rand}_{out,0.05}$.

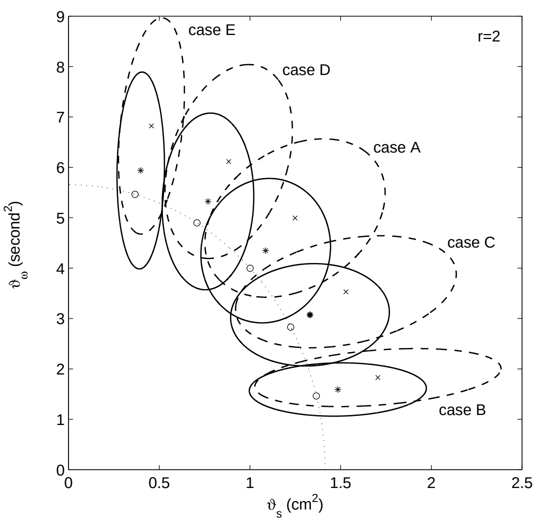

FIGURE 2. Insensitivity ellipses in dependence on $\boldsymbol{\vartheta}^*$. $\mathcal{N}_{0.05}$ by dashed line, its center at $\times$, $\mathcal{N}^{giv}_{out,0.05} = \mathcal{N}^{rand}_{out,0.05}$ by solid line, the center at $*$. The symbol $\circ$ means $\boldsymbol{\vartheta}^*$.

of suspicious measurements r. Naturally, $\mathcal{N}_{0.05}$ is independent of r. Further, $\mathcal{N}^{rand}_{out,0.05} \subset \mathcal{N}^{giv}_{out,0.05}$ for $r = 3, 4$ and $\mathcal{N}^{rand}_{out,0.05} = \mathcal{N}^{giv}_{out,0.05}$ for $r = 2$. In Figure 2, the insensitivity ellipses are presented in dependence on the choice of $\boldsymbol{\vartheta}^*$ for $r = 2$. Although ellipses $\mathcal{N}_{0.05}$, or $\mathcal{N}^{giv}_{out,0.05}$, are different for different cases A–E, tolerable changes (in %) are the same (see Tables 2, 3).

Empirical probabilities of the number of suspicious measurements r are given in Table 4. Since the distance β_1 is measured just twice, $r \geq 2$. So, y_1 and y_2 are always detected as suspicious measurements. Another one or two suspicious measurements are found very seldom. Empirical probabilities of outliers identification are given in Table 5. The most precise measurement of angles σ_ω and the worst precise measurement of distances σ_s are in the case B and vice versa in the case E. If the accuracy σ_ω is decreasing, the probability that the outlier y_1 is identified, is also decreasing. In the case B, the outlier y_1 is identified almost every time for both types of significance levels. In the case E, the probability for given indices is 88% but for random indices it is less than 50%. It seems that the outlier in measurement of distances is identified due to the accuracy of measurement of angles. The results were obtained via T, T_{out} calculated for $\boldsymbol{\vartheta}^*$.

5. Conclusion

Generally, insensitivity regions $\mathcal{N}_\varepsilon$, $\mathcal{N}^{giv}_{out,\varepsilon}$ and $\mathcal{N}^{rand}_{out,\varepsilon}$ are different. In the investigated situation, relatively large uncertainty can be tolerated in $\boldsymbol{\vartheta}^*$ when searching for suspicious measurements. Rather smaller uncertainty occurs when searching for significant outliers. Outliers identification depends on the accuracy of measurement. Heuristic approach (indices of suspicious measurements aren't a priori known but significance level is chosen as for given indices) gives better results than mathematically correct one (significance level for random indices).

REFERENCES

[1] BARNETT, V.—LEWIS, T.: *Outliers in Statistical Data*, John Wiley & Sons, Inc., Chichester, 1994.

[2] FIŠEROVÁ, E.—KUBÁČEK, L.: *Statistical problems of measurement in triangle*, Folia Fac. Sci. Natur. Univ. Masaryk. Brun., Math. **15** (2004), 77–94.

[3] FIŠEROVÁ, E.—KUBÁČEK, L.: *Insensitivity regions and outliers in mixed models with constraints*, Austrian Journal of Statistics **35** (2006), 245–252.

[4] KUBÁČEK, L.—KUBÁČKOVÁ, L.—VOLAUFOVÁ, J.: *Statistical Models with Linear Structures*, Veda, Bratislava, 1995.

[5] LEŠANSKÁ, E.: *Insensitivity regions for testing statistical hypotheses in mixed models with constraints*, Tatra Mt. Math. Publ. **22** (2001), 209–222.

[6] LEŠANSKÁ, E.: *Optimization of the size of nonsensitiveness regions*, Appl. Math., Praque **47** (2002), 9–23.
[7] RAO, C. R.—MITRA, S. K.: *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, New York etc., 1971.
[8] SCHEFFÉ, H.: *The Analysis of Variance*, John Wiley & Sons, New York, 1959.

Received September 27, 2006

Department of Mathematical Analysis
and Applied Mathematics
Faculty of Science
Palacký University
Tomkova 40
779 00 Olomouc
CZECH REPUBLIC
E-mail: fiserova@inf.upol.cz