NASA TECHNICAL MEMORANDUM NASA TM-76218

STEADY (ROBUST) CONDITIONALLY EFFECTIVE ESTIMATION OF PARAMETERS

L. S. Gurin and K. A. Tsoy

(NASA-TM-76218) STEADY (ROBUST) CONDITIONALLY EFFECTIVE ESTIMATION OF PARAMETERS (National Aeronautics and Space Administration) 14 p HC A02/MF A01 CSCL 12A

N81-22755

Unclas

G3/64 42183

Translation of "Ustoychivyye (Robastnyye) Uslovno-Effektivnyye Otsenki Paraemtrov," Academy of Sciences USSR, Institute of Space Research, Moscow, Report Pr-382, 1977, pp. 1-14.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
WASHINGTON, D. C. 20546 JUNE 1980

STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA TM-76218		

4. Title and Subtitle

STEADY (ROBUST) CONDITIONALLY EFFECTIVE ESTIMATION OF PARAMETERS

5. Report Date

June 1980

6. Performing Organization Code

7. Author(s)

L. S. Gurin and K. A. Tsoy

8. Performing Organization Report No.

10. Work Unit No.

11. Contract or Grant No.

NASw- 3198

9. Performing Organization Name and Address

SCITRAN
Box 5456
Santa Barbara, CA 93108

13. Type of Report and Period Covered

Translation

12. Sponsoring Agency Name and Address

National Aeronautics and Space Administration
Washington, D.C. 20546

14. Sponsoring Agency Code

15. Supplementary Notes

Translation of "Ustoychivyye (Robastnyye) Uslovno-Effektivnyye Otsenki Paraemtrov," Academy of Sciences USSR, Institute of Space Research, Moscow, Report Pr-382, 1977, pp. 1-14.

16. Abstract

This article considers a generalization of the concept of conditionally-effective estimation (have the best accuracy for given limitations on the suitability of the algorithm) to the case when consideration is also given to the steadiness (robustness) of the algorithm with respect to the deviation of the law governing the error distribution from the proposed law.

17. Key Words (Selected by Author(s))	18. Distribution Statement
	Unclassified - Unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	14	

STEADY (ROBUST) CONDITIONALLY EFFECTIVE ESTIMATION OF PARAMETERS

L. S. Gurin and K. A. Tsoy

Section 1. Formulation of the Problem /3*

This study is a direct continuation of the works [1, 2], which introduced the concept of a conditionally-effective estimation and examined certain particular problems. Estimations which are optimum for a given criterion in the case of given limitations are called conditionally-effective. This approach is necessary because, among the desired properties of the estimation -- along with properties which characterize their accuracy (independence, nondisplacement, effectiveness) -- there are others, such as difficulty of the estimation algorithms, their stability with respect to deviations of the laws for error distribution from the proposed one, etc. (see, for example, [3]).

With respect to the stability of the estimation, several studies have been devoted to this, beginning with the well-known study of Huber [4]. Of the more recent works, we would only like to mention [5]. These studies, however, do not consider the difficulty of the algorithms. On the other hand, the studies [1, 2] consider only the difficulty of the algorithms, but do not consider the stability. Both limitations are considered in this article for a rather simple problem of estimation. The examined method for obtaining the best estimation is sufficiently general and can be applied to more complex problems, and the result obtained for a specific problem has already been used in practice.

---

*)Numbers in the margin indicate pagination of original foreign text.

Previously, when considering this problem, we considered the simplest case of the direct measurement, which is greatly influenced by the action of the limitations upon the selection of the conditionally-effective estimation.

Section 2. Study of the Case of the Direct Measurement /4

Let us assume there are N measured values x_i of the constant C, containing independent measurement errors of ξ_i distributed according to the Laplace law, i.e.,

$$x_i = C + \xi_i ; \quad i = 1, 2, \dots, N. \tag{1}$$

It is necessary to find the conditionally-effective estimation of the quantity C, i.e.,

$$\hat{C} = f(x_1, \dots, x_n) ; \quad n \leq N \tag{2}$$

so that $M(\hat{C}) = C$; $\sigma^2(\hat{C}) = \min$ under the condition that the computational time on a computer is $t_c \leq t_0$. Thus, we consider only the limitation on the difficulty.

If t_0 is very large as compared with N, so that the limitations on the difficulty are insignificant, then, as is known [6], we must set $n = N$ and as $\hat{C}$ we use the median of the sampling x_i , i.e., we use the method of the least moduli. If the limitation with respect to t_0 is great, then it is inadequate to consider only the estimation methods, and we must turn to specific algorithms. To determine the median, we may use several algorithms, for example, the following.

A_1. Let us formulate a shortened variational series, i.e., we find the following (in increasing order) from a part of the sampling with the volume $n \leq N$

$$x_1 = x_{i_1} = \min_i x_i ; \; x_2 = \min_{i \neq i_1} x_i ; \dots ; x_j ; \; j = \left[\frac{n}{2}\right] + 1 . \tag{3}$$

Then $\hat{\varepsilon}_1 = x_j$ (for odd n) or $x_{j-1} < \hat{\varepsilon} < x_j$ (for even n).

A_2. We use the dichotomy method described in [6].

For comparison, let us examine the method of least squares, reduced to the algorithm A_3, i.e., the determination of the arithmetic mean part of the sampling with the volume n (as is known, this corresponds to an effective estimation for a normal law governing the error distribution).

Let us use $\varphi_i(n)$ to designate the computational time on a computer for the algorithm A_1 for the volume of the sampling n. The function $\varphi_i(n)$ depends on the computer used, the language, the translator, etc. Therefore, they may be obtained by the method of statistical tests. In several cases, we may reach the conclusions analytically. Thus, let us compare the algorithms A_1 and A_2. We use $m_1(n)$ to designate the mathematical expectation of the number of comparison operations when using the algorithm A_1. /5

We have

$$m_1(n) = \frac{n(n-1)}{2} - \frac{1}{2}\left(\left[\frac{n}{2}\right]-1\right)\left(\left[\frac{n}{2}\right]-2\right) \tag{4}$$

where $m_2(n)$ satisfies the recurrence relationship

$$m_2(n) = n-1+\frac{2}{n}m_2(n-1)+\dots+\frac{2}{n}m_2\left(\left[\frac{n}{2}\right]-1\right) \tag{5}$$

Let us prove that

$$\lim_{n\to\infty}\frac{m_2(n)}{n} = 4 \tag{6}$$

Let us set

$$\frac{m_2(n)}{n} = \varphi(n) \tag{7}$$

Then we have the following from [5]

$$\varphi(n) = 1-\frac{1}{n}+\frac{2}{n}\cdot\frac{n-1}{n}\varphi(n-1)+\dots+\frac{2}{n}\,\frac{\left(\left[\frac{n}{2}\right]-1\right)}{n}\varphi\left(\left[\frac{n}{2}\right]-1\right) \tag{8}$$

We may prove by induction that $\varphi(n) < 4$. From [8], we have

$$\varphi(n) < 1-\frac{1}{n}+\frac{2}{n}\left[1-\frac{1}{n}+\dots+1-\frac{n}{2n}\right]=4-\frac{3}{n}<4.$$

Let us use $\varphi_1(n)$ to designate the following function:

$$\varphi_1(n)=\min_{\left[\frac{n}{2}\right]-1\le i\le n-1}\varphi(i). \tag{9}$$

Then we have

$$\begin{aligned}\varphi(n) &\geq 1-\frac{1}{n}+\frac{2}{n}\left[1-\frac{1}{n}+\dots+\frac{\left[\frac{n}{2}\right]-1}{n}\right]\varphi_1(n)\geq \\ &\geq 1-\frac{1}{n}+\frac{1}{4n^2}\left[3n^2-4n-3\right]\varphi_1(n).\end{aligned} \tag{10}$$

/6

If $\varphi(n)\geq\varphi_1(n)$, then $\varphi_1(n+1)\geq\varphi_1(n)$ and $\varphi_1(n)$ will be monotonically nondecreasing. But the condition $\varphi(n)\geq\varphi_1(n)$ gives (taking into account [10])

$$\varphi_1(n)<\frac{4\left(1-\frac{1}{n}\right)}{1+\frac{1}{n^2}(4n+3)}. \tag{11}$$

Let us consider two possibilities: a) beginning with a certain large n, the condition (11) is always satisfied; b) there is an arbitrarily large number of n, for which the condition (11) is not satisfied.

In the first case, beginning with any n, $\varphi_1(n)$ does not decrease monotonically, and consequently we have

$$\lim_{n\to\infty}\varphi_1(n)=\ell, \tag{12}$$

i.e., beginning with a certain $n=n_1$, $\varphi(n)\geq\ell-\varepsilon$ for an arbitrarily small $\varepsilon>0$. Using (8), for any $n_2>n_1$ (for which $\varphi(n_2)=\varphi_1(n)\leq\ell+\varepsilon$) we obtain

$$\ell\geq 1-\frac{1}{n}-\varepsilon+(\ell-\varepsilon)\frac{2}{n}\left[1-\frac{1}{n}+\dots+\frac{\left[\frac{n}{2}\right]-1}{n}\right]. \tag{13}$$

Passing to the limit at $\varepsilon\to 0$, $n\to\infty$, we obtain

$$\ell\geq 1+\frac{3}{4}\ell, \qquad \ell\geq 4. \tag{14}$$

However, we have $\varphi(n) \leq 4$; consequently,

$$l = 4 \tag{15}$$

Thus, in this case (6) is proven.

Let us consider the second case. If the condition (11) is not satisfied for any n, then for $\left[\frac{n}{2}\right]-1 \leq i \leq n-1$ we have

$$\varphi(i) \geq \frac{4\left(1-\frac{1}{n}\right)}{1+\frac{1}{n^2}(4n+3)} = f(n)\ ;\ \varphi_1(n) \geq f(n). \tag{16}$$

In addition $f(n) \to 4$ at $n \to \infty$

Thus, $\varphi_1(n)$ either lies below $f(n)$ and does not decrease monotonically, or it lies above $f(n)$. Let us consider one of the values of n_1.

$$\varphi_1(n_1) \geq f(n_1)\ ;\ \varphi_1(n_1+1) < \varphi_1(n_1). \tag{17}$$ /7

If $\varphi(n_1+1) \geq f(n_1+1)$, we may not be interested in this case. Thus, in addition to (17), let us assume we have

$$\varphi_1(n_1+1) < f(n_1+1). \tag{18}$$

Then, beginning with the value of n_1+1, $\varphi_1(n)$ again increases. Let us estimate the difference $\varphi_1(n_1)-\varphi_1(n_1+1)=\Delta(n_1)>0$. It is clear that $\varphi_1(n_1) \leq \varphi(n_1-1)$ and $\varphi_1(n_1+1)$, if it does not equal $\varphi_1(n_1)$, but equals $\varphi(n_1)$ (according to the definition of the function $\varphi_1(n)$) Thus,

$$\Delta(n_1) < \varphi(n_1-1)-\varphi(n_1)\ ;\ \ \varphi(n_1-1)-\varphi(n_1)>0. \tag{19}$$

However,

$$\varphi(n_1-1)-\varphi(n_1) = -\frac{1}{n_1(n_1-1)} - \frac{2}{n_1}\varphi(n_1-1)\frac{n_1-1}{n_1} + 2\varphi(n_1-2)(n_1-2)\left[\frac{1}{(n_1-1)^2}-\frac{1}{n_1^2}\right]+\cdots$$
$$+\frac{2}{n_1-1}\varphi\left(\left[\frac{n_1}{2}\right]-1\right)\frac{\left(\left[\frac{n_1}{2}\right]-1\right)}{n_1-1}. \tag{20}$$

We should note that, depending on the evenness of n_1, the form of

(20) changes, but we make the estimation for the worst case. Considering that the left part is positive, we may again discard the negative terms and we obtain the estimate (we should recall that $\varphi(n) < 4$):

$$\varphi(n_1-1)-\varphi(n_1)<\frac{4}{n_1-1}+8(n_1-2)\left[\frac{2n_1-1}{n_1^2(n_1-1)^2}\right]+8(n_1-3)\left[\frac{2n_1-1}{n_1^2(n_1-1)^2}\right]+\dots< \tag{21}$$
$$<\frac{4}{n_1-1}+\frac{8}{n_1}<\frac{12}{n_1-1}.$$

Thus, in every case

$$\varphi_1(n_1+1)>f(n_1)-\frac{12}{n_1-1} \tag{22}$$

We thus find that in this case (6) holds.

A comparison of (4) and (6) shows that for rather large n, the algorithm A_2 is best, even if the noncomputational operations (which are basically connected with the organization of the cycles) /8 comprise a larger part in the algorithm A_2 than in the algorithm A_1.

When comparing the algorithms A_2 and A_3, we must consider that the algorithm A_3 is best under the condition

$$\varphi_3(2n)<\varphi_2(n) \tag{23}$$

(since the dispersions of the estimations are equal, if the algorithm A_3 uses a sampling volume which is twice as large [6]).

The algorithm A_3 for the sampling volume 2n includes 2n-1 additions and one division, and A_2 includes 4n comparisons for the sampling volume n. Considering the identical order of these quantities and the dependence of the computational time on several factors which are not considered, we may see that the final conclusion may be reached only on the basis of a numerical experiment with a computer. This experiment was done on the BESM-6 computer, and for the corresponding sampling volumes, gave the ratio of the computational time of 2.8, in favor of the algorithm A_3. Thus, although the method of least moduli gives an effective estimation, a conditionally-effective estimation is obtained with the use of an algorithm employing the method of least squares.

However, if we consider the stability of the estimation, then the picture greatly changes. It is known that the median is much less sensitive to "lost" points than the arithmetic mean (see, for example, [5]). If we pass from the simplest problem to a more complex one, even to problems of linear regression analysis, then the picture is greatly complicated. However, the statements made in this section show that to obtain stable, conditionally-effective estimations, we must use combined algorithms which use both the method of least squares and the method of least moduli. Let us consider the corresponding problem.

Section 3. Study of the Problem of Linear Regression /9

We shall use the model

$$y_i = a + bx_i + \xi_i , \quad i = 1, 2, \ldots, N \tag{24}$$

where y_i are the measurement results, x_i -- the values of the independent variable, and the measurement errors ξ_i are independent and have a distribution which is similar to the normal distribution, more precisely, the density of the distribution probability ξ_i has the form:

$$f(x) = \alpha f_1(x) + (1-\alpha) f_2(x), \quad 0 \leq \alpha < 1 \tag{25}$$

In formula (25)

$$f_2(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}; \tag{26}$$

$$f_1(x) = \begin{cases} \frac{1}{2a} & -a \leq x \leq a; \\ 0 & |x| > a. \end{cases}$$

At $\alpha = 0$, we obtain the normal law for the error distribution.

The following algorithms are considered (estimations).

A_1. The customary algorithm for the method of least squares (MLS).

A_2. The algorithm for the method of the least squares with preliminary grouping. Thus, N values of the independent variable x (we assume that the measurements are performed for equal values of x) form n groups with respect to $m = N/n$ points. We have

$$\bar{x}_j = \frac{1}{m}(x_{m(j-1)+1} + \dots + x_{mj}); \quad \bar{y}_j = \frac{1}{m}(y_{m(j-1)+1} + \dots + y_{mj}) \tag{27}$$

It is clear that, instead of (24), we obtain the problem

$$\bar{y}_j = c + d\bar{x}_j + \eta_i \quad , j = 1, 2, \dots, n, \tag{28}$$

and the estimations $\hat{c}$, $\hat{d}$, obtained by the regular MLS, may be used as the estimation for a and b, i.e.,

$$\hat{a}_2 = \hat{c} \; ; \; \hat{b}_2 = \hat{d}$$

Below, we shall use $\hat{a}_\nu$, $\hat{b}_\nu$ to designate the estimations of a and b obtained using the algorithm A_ν. For greater determinacy, /10
we introduce the second index m for estimations with grouping. This index designates the number of points in the group, for example, we shall designate definite estimations as follows

$$\hat{a}_{2,m} = \hat{c} \; ; \; \hat{b}_{2,m} = \hat{d}. \tag{29}$$

A_3, Combined algorithm with preliminary grouping. In this case, in contrast to A_2, in (27) $\bar{y}_j$ is replaced by the median of the corresponding group of values y_i, and we shall designate it by $\tilde{y}_j$. Then, instead of (28), we obtain

$$\tilde{y}_j = e + f\bar{x}_j + \mu_j \quad , j = 1, 2, \dots, n \tag{30}$$

and, consequently,

$$\hat{a}_{3,m} = \hat{e}\ ;\ \hat{b}_{3,m} = \hat{f}, \tag{31}$$

where $\hat{e}$ and $\hat{f}$ -- estimations of e and f from (30), obtained by the regular MLS.

Thus, we have the following set of estimations: $\hat{a}_1, \hat{b}_1, \hat{a}_{2,m}, \hat{b}_{2,m}, \hat{a}_{3,m}, \hat{b}_{3,m}$ (the second index assumes several values).

To compare the estimations, the method of statistical modeling was used. We set $a = 0;\ b = 0{,}1$ in the model of (24). In addition, the case of $\alpha = 0;\ 0.05;\ 0.25$ was considered in (25). In formula (26), we have $a = 30$. In addition, at $N = 15$, we assume $m = 3;\ 5$ and at $N = 45$ $m = 5,\ 5,\ 15$. For each estimation, using the results of $M = 100$ realizations, we obtain the average computational time on the computer $\bar{t}$ (in seconds per 100 realizations), the average displacement $\bar{\Delta}$, and the average dispersion $\bar{\sigma}^2$. The results of the experiment are given in Table 1-2.

Section 4. Conclusions and Problems of Further Research

1. At $\alpha = 0$, grouping is not advantageous, as would be expected from theoretical considerations.

2. For the linear model, the individual algorithms differ very little based on the computational time. /13

3. For deviations of the normal distribution ($\alpha > 0$), the algorithm A_3 is more advantageous than A_2, since it leads to increased accuracy. Thus, for large values of α, it is better to use algorithms with large values of m (for great deviations from normal distribution, the estimation in the form of the median is more effective than large groups).

Considering that in real problems we may expect smaller values of α, on the basis of the conclusions given above, we may recommend the algorithm A_3 with the value $m \approx N^{2/3}$.

TABLE 1

QUALITY INDICES OF ESTIMATIONS AT N = 15

Algorithm (estimation)		A_2			A_3		
α		0	0,05	0,25	0	0,05	0,25
	$\bar{t}$		7,I				
m=1 (A_1)	$\bar{\Delta} a$	0,0987	-0,0773	0,7307			
	$\bar{\Delta} b$	0,0942	0,II03	-0,0030			
	$\bar{\sigma}^2(a)$	0,2335	4,996I	22,2970			
	$\bar{\sigma}^2(b)$	0,0037	0,0789	0,3558			
	$\bar{t}$		7,52			7,89	
m=3	$\bar{\Delta} a$	0,0892	-0,I080	0,0667	0,I596	-0,0367	0,9I33
	$\bar{\Delta} b$	0,0957	0,II47	0,006I	0,0867	0,I033	-0,0067
	$\bar{\sigma}^2(a)$	0,2423	4,7742	24,8740	0,3062	0,822I	9,9670
	$\bar{\sigma}^2(b)$	0,0037	0,076I	0,406I	0,0049	0,0088	0,I454
	$\bar{t}$		7,2I			7,39	
m=5	$\bar{\Delta} a$	0,0869	-0,0738	0,7059	0,0674	0,0I43	0,5064
	$\bar{\Delta} b$	0,0959	0,I098	0,0005	0,0950	0,0999	0,0365
	$\bar{\sigma}^2(a)$	0,2520	5,8354	26,3240	0,4267	0,4754	7,5002
	$\bar{\sigma}^2(b)$	0,0040	0,0900	0,4520	0,0063	0,007I	0,I440

TABLE 2

QUALITY INDICES OF ESTIMATIONS AT N = 45

Algorithm (estimation)		A_2			A_3		
α		0	0.05	0,25	0	0.05	0.25
m = 1 (A_1)	$\bar{t}$		7.19				
	$\bar{\Delta} a$	0,0890	0,0072	-0,0805			
	$\bar{\Delta} b$	0,0978	0,I006	0,I037			
	$\bar{\sigma}^2(a)$	0,0798	I,6594	7,8093			
	$\bar{\sigma}^2(b)$	0,000I	0,0023	0,0II0			
m = 5	$\bar{t}$		7,39			8,69	
	$\bar{\Delta} a$	0,089I	-0,0026	-0,09I4	0,0722	-0,0083	0,I343
	$\bar{\Delta} b$	0,I0I0	0,I042	0,0976	0,0980	0,I008	0,0955
	$\bar{\sigma}^2(a)$	0,0805	I,7025	7,9594	0,I296	0,I926	2,686I
	$\bar{\sigma}^2(b)$	0,000I	0,0024	0,0II3	0,0002	0,0003	0,0032
m = 9	$\bar{t}$		7,22			7,93	
	$\bar{\Delta} a$	0,0828	-0,0I56	-0,II07	0,I2I7	-0,0086	0,0I95
	$\bar{\Delta} b$	0,0979	0,I0I6	0,I050	0,0963	[illegible]	0,0967
	$\bar{\sigma}^2(a)$	0,0838	I,7037	7,9738	0,I55I	0,2I38	I,4372
	$\bar{\sigma}^2(b)$	0,000I	0,0025	0,0II6	0,0002	0,0003	0,0023
m = 15	$\bar{t}$		7,06			7,6I	
	$\bar{\Delta} a$	0,0739	0,0399	-0,I3I0	0,0373	[illegible]	0,0502
	$\bar{\Delta} b$	0,0988	0,099I	0,I060	0,0998	0,0092	0,I007
	$\bar{\sigma}^2(a)$	0,0854	I,843I	8,483I	0,2390	0,2692	0,6096
	$\bar{\sigma}^2(b)$	0,000I	[illegible]	0,0I25	0,0004	0,0004	0,00I2

ORIGINAL PAGE IS
OF POOR QUALITY

The conclusions obtained may be refined and expanded as the result of future research, which is presently being carried out in the following directions:

1. Estimations of nonlinear models are considered. In this case, there must be a greater difference between individual algorithms in terms of difficulty.

2. For the problem considered of linear regression, algorithms are additionally studied which are based on excluding the lost points [7, 8].

REFERENCES /14

1. Gurin, L. S., Tsoy, K. A. Use of statistical modeling in the development and comparative analysis of methods for estimating parameters. In the Collection: Obrabotka kosmicheskoy informatsii (Processing of space information). Moscow, Nauka press, 1976, pp. 36-57.

2. Gurin, L. S., Tsoy, K. A. Conditionally-effective estimations of parameters of nonlinear models. Preprint. Moscow, IKI, AN SSSR, Pr-257, 1975, 34 pages.

3. Statisticheskiye metody - eksperimental'noy fizike (Statistical methods - experimental physics). Moscow, Atomizdat, 1976, 336 pp.

4. Huber, P. J. Robust estimation of a location parameter. Ann. Math. Stat., 1964, 35, No. 1, 73-101.

5. Polyak, B. T., Tsypkin, Ya. Z. Noise-resistant identification. In the Collection: Reports of the IV Symposium of the IFAK "Identification and estimation of system parameters," Part I, Tbilisi, Izdatel'stvo Metsniereba, 1976, pp. 190-213.

6. Mudrov, V. I., Kushko, V. L. Metody obrabotki izmereniy (Methods of processing measurements), Moscow, Sovetskoye radio, 1976, 192 pp.

7. Thompson, W. R. Ann. Math. Stat., 1962, 6, 214.

8. Dixon, W. J. Biometrics, 1953, 9, 74.