Variances of Linear Regression Coefficients for Safety Margin of Technical System

Gurami Tsitsiashvili, Marina Osipova
•
Russia, 690041, Vladivostok, Radio street 7
Institute for Applied Mathematics
Far Eastern Branch of Russian Academy Sciences
guram@iam.dvo.ru, mao1975@list.ru

Abstract

In this paper, we choose such a particular formulation of the problem of calculating linear regression coefficient, when the moments of observation form an arithmetic progression or depend in proportion to some degree of the observation number and variances of the estimation of the safety margin of the technical system decrease over time. It is proved that the variance of the trend estimation in this case decreases as a certain degree of the length of the series of observations. This makes it possible to evaluate the effectiveness of non-destructive testing for the safety margin of the technical system.

Keywords: linear regression coefficient, variance calculating, independent samples of observations.

1 Introduction

The problem of studying the variance of the linear trend estimation and its dependence on the length of the time series, that this estimation is based on, is of both theoretical and practical interest. This problem is closely related to the problem of small samples in mathematical statistics. In reliability theory, this problem occurs when using linear regression analysis to predict the safety margin of a technical system (see, for example, [1], [2]). In this regard, of particular interest is the task set by O. V. Abramov to study the variance of the estimation when the intervals between successive moments of observation and the variance of the estimation of the safety margin of the technical system decrease over time.

By analogy with [3] in this paper, we choose such an honest statement of this problem, when the observation moments form an arithmetic progression or grow as less than one degree of the observation number. Similarly, it is assumed that the variance of the observation estimate also decreases in proportion to some degree of the observation number. It is proved that the variance of the trend estimation in this case decreases as a certain degree of the length of the series of observations. This makes it possible to use short series of observations to evaluate the linear trend and also to evaluate the effectiveness of non-destructive testing for the safety margin of the technical system [4] – [6].

Consider the following linear regression model $x(t) = y(t) + \varepsilon(t),\ y(t) = at + b$. Assume that at times $t_1, \dots, t_n,\ 0 \le t_1 < t_2 < \cdots < t_N$, measured values are $y(t_1), \dots, y(t_N)$ with random errors $\varepsilon_1, \dots, \varepsilon_N$. The random variables $\varepsilon_1, \dots, \varepsilon_N$ are assumed to be independent, equally distributed with zero mean and variance σ^2.

To solve this problem, replace the variable $\tilde{t} = t - T_N$, $T_N = \frac{\sum_{k=1}^{N} t_k}{N}$, and define a linear function

$$\tilde{y}(t) = y(t + T_n) = at + b + aT_N = at + \tilde{b}, \quad \sum_{k=1}^{N} \tilde{t}_k = 0, \quad \tilde{b} = b + aT_N.$$

To do this, we compute $\tilde{t}_k$, $k = 1, \dots, N$, and construct the least squares [7], [8] estimates of the coefficients a, $\tilde{b}$ of the linear regression function $\tilde{y}(t) = at + \tilde{b}$ from observations

$$\tilde{x}_1 = \tilde{y}(\tilde{t}_1) + \sigma_1 \tilde{\varepsilon}_1, \dots, \tilde{x}_N = \tilde{y}(\tilde{t}_N) + \sigma_N \tilde{\varepsilon}_N.$$

Here $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_1$ are independent random variables with zero means and unit variances.

The solution of this problem is a random vector consisting of estimates

$$\hat{a}_N = \frac{\sum_{k=1}^{N} \tilde{x}_k \tilde{t}_k}{\sum_{k=1}^{N} \tilde{t}_k^2}, \qquad \hat{b}_N = \frac{\sum_{k=1}^{N} \tilde{x}_k}{N}$$

of coefficients a, $\tilde{b}$ of linear function $\tilde{y}(t)$. The components of this vector have the following averages, variances, and covariance coefficient:

$$M\hat{a}_N = a, \qquad M\hat{b}_N = \tilde{b}, \qquad D\hat{a}_N = \frac{\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2}{(\sum_{k=1}^{N} \tilde{t}_k^2)^2}, \qquad D\hat{b}_N = \frac{\sum_{k=1}^{N} \sigma_k^2}{N^2}, \qquad cov(\hat{a}_N, \hat{b}_N) = \frac{\sum_{k=1}^{N} \tilde{t}_k \sigma_k^2}{N \sum_{k=1}^{N} \tilde{t}_k^2}. \tag{1}$$

2 Main Results

Statement 1. *The following equalities are true*

$$\sum_{k=1}^{N} \tilde{t}_k^2 = \sum_{k=1}^{N} t_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} t_k\right)^2, \tag{2}$$

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 = \sum_{k=1}^{N} \sigma_k^2 t_k^2 + \sum_{k=1}^{N} \frac{\sigma_k^2}{N^2} \left(\sum_{i=1}^{N} t_i\right)^2 - \frac{2}{N}\sum_{k=1}^{N} \sigma_k^2 t_k \sum_{i=1}^{N} t_i. \tag{3}$$

Proof. Statement 1 follows from the following equalities

$$\sum_{k=1}^{N} \tilde{t}_k^2 = \sum_{k=1}^{N} \left(t_k - \frac{1}{N}\sum_{i=1}^{N} t_i\right)^2 = \sum_{k=1}^{N} \left(t_k^2 - \frac{2t_k}{N}\sum_{i=1}^{N} t_i + \left(\frac{1}{N}\sum_{i=1}^{N} t_i\right)^2\right) =$$

$$= \sum_{k=1}^{N} t_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} t_k\right)^2,$$

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 = \sum_{k=1}^{N} \sigma_k^2 \left(t_k - \frac{1}{N}\sum_{i=1}^{N} t_i\right)^2 = \sum_{k=1}^{N} \sigma_k^2 \left[t_k^2 + \frac{1}{N^2}\left(\sum_{i=1}^{N} t_i\right)^2 - \frac{2}{N} t_k \sum_{i=1}^{N} t_i\right].$$

Equidistant series of observations. By equidistant series of observations, we will understand such a series in which the following equalities are fulfilled

$$t_k = k, \qquad \sigma_k^2 = \sigma^2, \qquad k = 1, \dots, N.$$

Statement 2. *The following relations are valid for an equidistant series of observations*

$$D\hat{a}_N = \frac{12\sigma^2}{N^3\left(1 - \frac{1}{N^2}\right)}, \qquad D\hat{b}_N = \frac{\sigma^2}{N}, \qquad cov(\hat{a}_N, \hat{b}_N) = 0, \tag{4}$$

Proof. By induction on N it is not difficult to obtain from Formula (2) the equality

$$\sum_{k=1}^{N} t_k = \frac{N(N+1)}{2}, \qquad \sum_{k=1}^{N} t_k^2 = \frac{N^3}{3} + \frac{N^2}{2} + \frac{N}{6}$$

and so it follows from the formula (3) that

$$\sum_{k=1}^{N} \tilde{t}_k^2 = \sum_{k=1}^{N} t_k^2 - \frac{1}{N}\left(\sum_{k=1}^{N} t_k\right)^2 = \frac{N^3}{12}\left(1 - \frac{1}{N^2}\right). \tag{5}$$

Therefore, from Formulas (1), (5), we get the equality (4).

Remark 1. *If independent random variables $\varepsilon_1, \dots, \varepsilon_N$ have normal distributions with zero averages and total variance σ^2, then estimates $\hat{a}_N$, $\hat{b}_N$ are independent.*

Here are the results of calculations performed using the formula (5).

N	2	3	4	5	6	7	8	9	10
$\sum_{k=1}^{N} \tilde{t}_k^2$	0.5	2	5	10	17.5	28	42	60	82.5
$12\sum_{k=1}^{N} \tilde{t}_k^2/N^3$	0.75	0.888	0.9375	0.96	0.9722	0.9795	0.9844	0.9876	0.99

N	11	12	13	14	15	16	17	18
$\sum_{k=1}^{N} \tilde{t}_k^2$	110	143	182	227.5	280	340	408	484.5
$12\sum_{k=1}^{N} \tilde{t}_k^2/N^3$	0.9917	0.9931	0.9948	0.9949	0.9955	0.9961	0.9965	0.9969

Tabl. 1. Meanings of terms $\sum_{k=1}^{N} \tilde{t}_k^2$, $12\sum_{k=1}^{N} \tilde{t}_k^2/N^3$.

From Tabl. 1 values of term $\sum_{k=1}^{N} \tilde{t}_k^2$ and its asymptotics at $N \to \infty$ it is seen that already at $N \geq 5$, the value of $\sum_{k=1}^{N} \tilde{t}_k^2$ with a five-percent and smaller relative error is approximated by the asymptotic $N^3/12$. Next, we will write this ratio as $\sum_{k=1}^{N} \tilde{t}_k^2 \approx N^3/12$.

Remark 2. *Let's now consider a series of values $t_k^{(m)}$, $k = 1, \dots, Nm$, represented as $1/m, 2/m, \dots, Nm/m$. Then approximate equalities are valid*

$$\sum_{k=1}^{Nm} \left(\tilde{t}_k^{(m)}\right)^2 \approx \frac{N^3 m}{12}, \qquad D\hat{a}_{N,m} \approx \frac{12\sigma^2}{m\,N^3}, \qquad D\hat{b}_{N,m} = \frac{\sigma^2}{Nm}.$$

Thus, replacing a series of observation moments t_k, $k = 1, \dots, N$ with a series $t_k^{(m)}$, $k = 1, \dots, Nm$, results in approximately a m reduction in the variances of the linear regression function coefficients estimates.

Series of observations with power dependencies.

Statement 3. *Now suppose that*

$$t_k = k^\alpha, \ \sigma_k = \sigma k^{-\beta}, \ 0 < \beta, \alpha. \tag{6}$$

Then the following asymptotic relations are valid

$$D\hat{a}_N \sim \frac{\sigma^2 C_*(2\alpha+1)^2(\alpha+1)^4}{N^{2\alpha+2\beta+1}\alpha^4}, \ D\hat{b}_N \sim \frac{\sigma^2}{N^{2\beta+1}(1-2\beta)}, \ 0 < 1-2\beta, \tag{7}$$

$$D\hat{a}_N \sim \frac{\sigma^2 \ln N (2\alpha+1)^2(\alpha+1)^2}{N^{2\alpha+2}\alpha^4}, \ D\hat{b}_N \sim \frac{\sigma^2 \ln N}{N^2}, \ 0 = 1-2\beta, \tag{8}$$

$$D\hat{a}_N \sim \frac{\sigma^2 C^*(2\alpha+1)^2(\alpha+1)^2}{N^{2\alpha+2}\alpha^4}, \ D\hat{b}_N \sim \frac{\sigma^2 C^*}{N^2}, \ 1-2\beta < 0, \tag{9}$$

with

$$C_* = \frac{1}{2\alpha-2\beta+1} + \frac{1}{(1-2\beta)(\alpha+1)^2} - \frac{2}{(\alpha-2\beta+1)(\alpha+1)}, \ 0 < 1-2\beta,$$

$$C^* = \sum_{k=1}^{\infty} k^{-2\beta}, \qquad 1-2\beta < 0.$$

Proof. The following inequalities are valid

$$\int_1^{N+1} \tau^\gamma d\tau \le \sum_{k=1}^{N} k^\gamma \le 1 + \int_1^{N} \tau^\gamma d\tau, \qquad \gamma < 0,$$

$$1 + \int_1^{N} \tau^\gamma d\tau \le \sum_{k=1}^{N} k^\gamma \le \int_1^{N+1} \tau^\gamma d\tau, \qquad 0 \le \gamma,$$

This is followed by asymptotic relations for $N \to \infty$

$$\sum_{k=1}^{N} k^\gamma \sim \frac{N^{\gamma+1}}{\gamma+1}, \qquad \gamma > -1; \ \sum_{k=1}^{N} k^{-1} \sim \ln N; \ \sum_{k=1}^{N} k^\gamma \to \sum_{k=1}^{\infty} k^\gamma < \infty, \qquad \gamma < -1. \tag{10}$$

and so

$$\sum_{k=1}^{N} \tilde{t}_k^2 \sim \frac{N^{2\alpha+1}\alpha^2}{(\alpha+1)^2(2\alpha+1)}, \qquad \alpha > 0. \tag{11}$$

Consider the case when $0 < 1 - 2\beta$, then from Formulas (10), (11) we get

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \frac{\sigma^2 N^{2\alpha-2\beta+1}}{2\alpha-2\beta+1}, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \sim \frac{\sigma^2 N^{\alpha-2\beta+1}}{\alpha-2\beta+1}, \qquad \sum_{k=1}^{N} \sigma_k^2 \sim \frac{\sigma^2 N^{1-2\beta}}{1-2\beta}$$

and consequently

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \sigma^2 C_* N^{2\alpha-2\beta+1},$$

$$D\hat{a}_N \sim \frac{\sigma^2 C_*(2\alpha+1)^2(\alpha+1)^4}{N^{2\alpha+2\beta+1}\alpha^4}, \ D\hat{b}_N \sim \frac{\sigma^2}{N^{2\beta+1}(1-2\beta)}. \tag{12}$$

If $1 - 2\beta = 0$, then

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \frac{\sigma^2 N^{2\alpha}}{2\alpha}, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \sim \frac{\sigma^2 N^{\alpha}}{\alpha}, \qquad \sum_{k=1}^{N} \sigma_k^2 \sim \sigma^2 \ln N$$

and so

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \frac{\sigma^2 N^{2\alpha}\ln N}{(\alpha+1)^2}, \qquad D\hat{a}_N \sim \frac{\sigma^2 \ln N(2\alpha+1)^2(\alpha+1)^2}{N^{2\alpha+2}\alpha^4}, \qquad D\hat{b}_N \sim \frac{\sigma^2 \ln N}{N^2}. \tag{13}$$

Now let's go to the case where $1 - 2\beta < 0, \ \alpha - 2\beta + 1 > 0$, then from Formulas (10), (11) we get

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \frac{\sigma^2 N^{2\alpha-2\beta+1}}{2\alpha-2\beta+1}, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \sim \frac{\sigma^2 N^{\alpha-2\beta+1}}{\alpha-2\beta+1}, \qquad \sum_{k=1}^{N} \sigma_k^2 \to \sigma^2 C^*,$$

and consequently

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \frac{\sigma^2 C^* N^{2\alpha}}{(\alpha+1)^2}, \qquad D\hat{a}_N \sim \frac{\sigma^2 C^*(2\alpha+1)^2(\alpha+1)^2}{N^{2\alpha+2}\alpha^4}, \qquad D\hat{b}_N \sim \frac{\sigma^2 C^*}{N^2}. \tag{14}$$

If $1 - 2\beta < 0, \ \alpha - 2\beta + 1 = 0$, then

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \frac{\sigma^2 N^{\alpha}}{\alpha}, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \sim \sigma^2 \ln N, \qquad \sum_{k=1}^{N} \sigma_k^2 \to \sigma^2 C^*$$

and so

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \frac{\sigma^2 C^* N^{2\alpha}}{(\alpha+1)^2}, \qquad D\hat{a}_N \sim \frac{\sigma^2 C^*(2\alpha+1)^2(\alpha+1)^2}{N^{2\alpha+2}\alpha^4}, \qquad D\hat{b}_N \sim \frac{\sigma^2 C^*}{N^2}. \tag{15}$$

Now consider the case where $\alpha - 2\beta + 1 < 0,\ 2\alpha - 2\beta + 1 > 0$, then from Formulas (10), (11) we get

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \frac{\sigma^2 N^{2\alpha-2\beta+1}}{2\alpha - 2\beta + 1}, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \to \sigma^2 \sum_{k=1}^{\infty} k^{\alpha-2\beta} < \infty, \qquad \sum_{k=1}^{N} \sigma_k^2 \to \sigma^2 C^*$$

and consequently

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \sigma^2 C^* N^{2\alpha} (\alpha+1)^2, \qquad D\hat{a}_N \sim \sigma^2 C^* (2\alpha+1)^2 (\alpha+1)^2 N^{2\alpha+2} \alpha^4, \qquad D\hat{b}_N \sim \sigma^2 C^* N^2. \tag{16}$$

If $\alpha - 2\beta + 1 < 0,\ 2\alpha - 2\beta + 1 = 0$, then

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \sim \sigma^2 \ln N, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \to \sigma^2 \sum_{k=1}^{\infty} k^{\alpha-2\beta} < \infty, \sum_{k=1}^{N} \sigma_k^2 \to \sigma^2 C^*$$

and so

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \frac{\sigma^2 C^* N^{2\alpha}}{(\alpha+1)^2}, \qquad D\hat{a}_N \sim \frac{\sigma^2 C^* (2\alpha+1)^2 (\alpha+1)^2}{N^{2\alpha+2} \alpha^4}, \qquad D\hat{b}_N \sim \frac{\sigma^2 C^*}{N^2}. \tag{17}$$

Now let's go to the case where $2\alpha - 2\beta + 1 < 0$, then from Formulas (10), (11) we get

$$\sum_{k=1}^{N} \sigma_k^2 t_k^2 \to \sigma^2 \sum_{k=1}^{\infty} k^{2\alpha-2\beta} < \infty, \qquad \sum_{k=1}^{N} \sigma_k^2 t_k \to \sigma^2 \sum_{k=1}^{\infty} k^{\alpha-2\beta} < \infty, \qquad \sum_{k=1}^{N} \sigma_k^2 \to \sigma^2 C^*$$

and consequently

$$\sum_{k=1}^{N} \sigma_k^2 \tilde{t}_k^2 \sim \frac{\sigma^2 C^* N^{2\alpha}}{(\alpha+1)^2}, \qquad D\hat{a}_N \sim \frac{\sigma^2 C^* (2\alpha+1)^2 (\alpha+1)^2}{N^{2\alpha+2} \alpha^4}, \qquad D\hat{b}_N \sim \frac{\sigma^2 C^*}{N^2}. \tag{18}$$

From Formulas (12) – (18) we get the asymptotic relations for $N \to \infty$ (7) – (9).

Remark 3. *In the first quadrant of the coefficient values $\alpha > 0, \beta > 0$ in Fig. 1 shows the areas*

$$\Gamma_1 = \{0 < 1 - 2\beta\}, \qquad \Gamma_2 = \{0 = 1 - 2\beta\}, \qquad \Gamma_3 = \{1 - 2\beta < 0, \alpha - 2\beta + 1 > 0\},$$

$$\Gamma_4 = \{1 - 2\beta < 0, \alpha - 2\beta + 1 = 0\}, \qquad \Gamma_5 = \{\alpha - 2\beta + 1 < 0, 2\alpha - 2\beta + 1 > 0\},$$

$$\Gamma_6 = \{\alpha - 2\beta + 1 < 0, 2\alpha - 2\beta + 1 = 0\}, \qquad \Gamma_7 = \{2\alpha - 2\beta + 1 < 0\}$$

for which Formulas (12), (13), (14), (15), (16), (17), (18) are obtained for the variance of estimates of coefficients of the linear regression function, respectively.

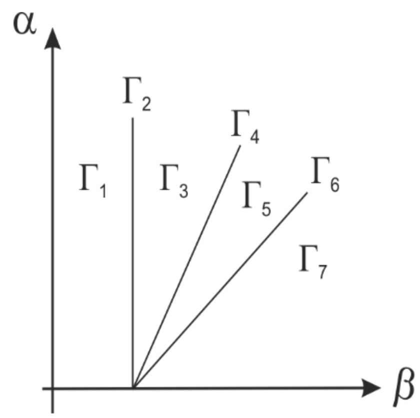

Fig 1. Domains $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5, \Gamma_6, \Gamma_7$ in the first quadrant of the coefficient values $\alpha > 0, \beta > 0$.

Remark 4. *If $\alpha < 1$ then distances between neighbour time moments t_k, t_{k+1}, $1 \le k < N$, decrease.*

3 Conclusion

The estimates of variances of linear trend coefficient estimates obtained in this work allow us to choose the moments of determining the technical system's strength reserves and their accuracy depending on the system's approximation to the rollback state.

4 Acknowledgements

The research of Gurami Tsitsiashvili is partially supported by Presidium of FEB RAS, program "Priority research for the integrated development of the far Eastern branch of RAS", project 18-5-044. The authors thank O.V. Abramov for help in setting the task in a meaningful way.

References

[1] Abramov, O. V., Tsitsiashvili, G. Sh. (2018). Prediction of failure of the controlled technical system. Informatics and control systems, 3: 42–49.
[2] Abramov, O. V., Tsitsiashvili, G. Sh. (2019). Interval estimation for the task of predicting failures of complex engineering systems. (In Russian). Proceedings of the International Symposium "Reliability and quality", 1, 113–114. (In Russian).
[3] Tsitsiashvili, G. Sh. (2019) Calculating the variance of the linear regression coefficient. (In Russian). Reliability: Theory and Applications, 14 (3): 65–68.
[4] Novokreshenov V. V. Nondestructive testing of welded joints in mechanical engineering: Textbook for academic baccalaureate /Novokreshenov V. V., Rodyakina R. V.; edited by Prokhorov N. N. 2nd ed. Moscow: Yurayt, 2018. (In Russian).
[5] Volchenko V. N., Gurvich A. K., Mayorov A. N., Kashuba L. A., Makarov E. L., Khasanov M. H. Quality control of welding /ed. - Textbook for engineering universities. Moscow: Mashinostroenie, 1975. (In Russian).
[6] GOST P 56542-2015. Non-destructive testing. Classification of types and methods. (In Russian).
[7] Rykov V. V. Mathematical statistics and experiment planning. MAKS Press, Moscow 2010. (In Russian).
[8] Borovkov A. A. Mathematical statistics and experiment planning. Additional chapters. Nauka, Moscow 1984. (In Russian).