

Statistical Estimation of Corruption Indicators in the Firm

Yu. A. Pichugin

Saint-Petersburg State University
of Aerospace Instrumentation
67, Bolshaya Morskaya str.
Saint-Petersburg, 190000, Russia

O. A. Malafeyev

Saint-Petersburg State University
7-9, Universitetskaya nab.
Saint-Petersburg, 199034, Russia

Copyright © 2016 Yu. A. Pichugin and O. A. Malafeyev. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

A new statistical procedure of anti-corruption control of economic activity is proposed in the paper. The task of the firm's corruption checking is formulated in terms of statistical hypothesis testing. To make checking procedure more rigorous, it is proposed to formulate a hypothesis regarding the current, not the average data, as is customary in classical statistics.

Mathematics Subject Classification: 62P20

Keywords: corruption, statistical estimation

1 Introduction

In the last decades mathematical modeling of socioeconomic systems and processes have received definite status of the legal existence (e.g. [1-21] and references therein) . Here we demonstrate how mathematical statistics methods

help to reveal corrupt activity in firms. The ability to counteract corrupt activities of firms is limited due to its hidden nature, extent of its dissemination, and, accordingly, a lack of resources to identify these ones. In this regard, a sufficiently reliable criteria for the evaluation of the presence of indirect indicators of corruption in firms is of particular relevance. Several studies have shown [21] that the circumstantial evidence of corruption in the sphere of economic activity includes: the short time of tendering, the inflated volume of the tender, the violation of the procedure of filing the application, the violation of the deadline for the submission of the application, the filing of a complaint from losing bidders, changes in the contract compared with the bid, the winning bidder has not declared all its participants, and so on.

2 The estimation of the corruption area boundary

Suppose that k indirect corruption indicators $\{x_1, x_2, \dots, x_k\}$ are selected as components of the vector $\mathbf{X} = (x_1, x_2, \dots, x_k)^T$ (T is a transposition operator). Let $R_{\mathbf{C}} \subset R^k$ be the corruption area. Assume that the vector \mathbf{X} obeys a multivariate normal distribution. This assumption is valid when the distribution is sufficiently localized in the area of admissible values. The boundary of the region $R_{\mathbf{C}}$ can be determined statistically by the data sample values $\{\mathbf{X}_j, j = 1, 2, \dots, n\}$, obtained on the basis of the accounting statements of firms whose economic activity is of a corrupt nature.

Compute the estimate vector of the average values

$$\Theta_{\mathbf{C}} = n^{-1} \sum_{j=1, \dots, n} (\mathbf{X}_j) \quad (1)$$

and the matrix of cross-covariances

$$\mathbf{V}_{\mathbf{C}} = (n - 1)^{-1} \sum_{j=1, \dots, n} (\mathbf{X}_j - \Theta_{\mathbf{C}})(\mathbf{X}_j - \Theta_{\mathbf{C}})^T. \quad (2)$$

The degeneracy of the matrix $\mathbf{V}_{\mathbf{C}}$ means the dependence of selected indirect corruption indicators and points to the necessity to reduce their list. The boundary of the corruption area $R_{\mathbf{C}}$, given that, by assumption, \mathbf{X} obeys a multivariate normal distribution, i.e. $\mathbf{X} \sim N(\Theta_{\mathbf{C}}, \mathbf{V}_{\mathbf{C}})$, can be defined in two different ways: as elliptic hypersurface, or a hyperplane. In the first case we have

$$W_{\mathbf{C}} = \{\mathbf{X} : f_{\mathbf{C}}(\mathbf{X}) = c_{\alpha}\}$$

Here, the probability density function $f_{\mathbf{C}}(\mathbf{X}) = f(\mathbf{X}, \Theta_{\mathbf{C}}, \mathbf{V}_{\mathbf{C}})$ is defined by the formula

$$f_{\mathbf{C}}(\mathbf{X}) = (2\pi)^{-k/2} (\det \mathbf{V}_{\mathbf{C}})^{-1/2} \exp(-0.5(\mathbf{X} - \Theta_{\mathbf{C}})^T \mathbf{V}_{\mathbf{C}}^{-1} (\mathbf{X} - \Theta_{\mathbf{C}})),$$

. The value of the constant c_α is such that the probability measure for the distribution $N(\Theta_{\mathbf{C}}, \mathbf{V}_{\mathbf{C}})$ inside the ellipsoid $W_{\mathbf{C}}$, i.e. $R_{\mathbf{C}}$, is equal to $1 - \alpha$. Note that the ellipsoid $W_{\mathbf{C}}$, might go beyond the boundary of the admissible values of the vector \mathbf{X} region . Then a probability estimation should be calculated in terms of conditional probabilities. This remark also applies to probability estimation discussed below. The second approach to evaluating the boundaries of the $R_{\mathbf{C}}$ is similar to E. Altman’s method [1], [17]. Let the orthogonal matrix \mathbf{P} be computed so that

$$\mathbf{P}^T \mathbf{V}_{\mathbf{C}} \mathbf{P} = \Lambda_{\mathbf{C}} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k).$$

Columns of the matrix \mathbf{P} can be placed in such a manner as to satisfy the condition $\lambda_1 \geq \lambda_2 \geq \dots \lambda_k$. Then one can take a hyperplane $\phi_{\mathbf{C}} = \{\mathbf{X} : \mathbf{P}_k^T \mathbf{X} = \delta_\alpha\}$, as the boundary of $R_{\mathbf{C}}$ where \mathbf{P}_k is the last column of the matrix \mathbf{P} corresponding to the smallest eigenvalue (λ_k). The sum of the squared deviations of all values \mathbf{X}_j from hyperplane $\phi_0 = \{\mathbf{X} : \mathbf{P}_k^T \mathbf{X} = \mathbf{P}_k^T \Theta_{\mathbf{C}}\}$ equals to the minimum value $(n - 1)\lambda_k$, i.e. the hyperplane ϕ_0 reproduces the original sample with minimal quadratic error. The value of constant δ_α should be calculated based on the selected level of confidence $(1 - \alpha)$, as the boundary of the left-sided critical region for Student random variable $(\lambda_k)^{-1/2} \mathbf{P}_k^T (\mathbf{X} - \Theta_{\mathbf{C}})$, i.e.

$$\delta_\alpha = \mathbf{P}_k^T \Theta_{\mathbf{C}} - \lambda_k^{1/2} t_{n-1}^{\alpha/2}$$

($t_{n-1}^{\alpha/2}$ is quantile of the Student distribution). This means that for corrupt firms the probability to be occurred below the level δ_α is equal to $\alpha/2$, since the Student test is the two-sided one.

3 Probability estimation of corruption absence in the firm analysed

Suppose that for some firm B "analyzed for corruption" there observations $\{\mathbf{X}_t, t = 1, 2, \dots, T_B\}$, are given for some time period T_B . Assume that the values of \mathbf{X}_t aren't included in $R_{\mathbf{C}}$, because otherwise the economic activity of the firm would be considered as corrupt. Making use of formulas (1) and (2), we calculate similar estimates for the parameters Θ_B and \mathbf{V}_B , assuming that the vector of indices \mathbf{X} of the estimated firms obeys a multivariate normal distribution $\mathbf{X} \sim N(\Theta_B, \mathbf{V}_B)$, which indirectly corresponds to the assumption of corruption absence.

The probability measure of corruption absence may be defined as a probability measure (distribution $N(\Theta_B, \mathbf{V}_B)$) of the interior of the ellipsoid $W_B = \{\mathbf{X} : f_B(\mathbf{X}, \Theta_B, \mathbf{V}_B) = f_*, \text{ where } f_B(\mathbf{X}_*) = f_*\}$. The value of \mathbf{X}_* is calculated

as the extremum point of the Lagrange function

$$L(\mathbf{X}, \lambda) = \ln f_B(\mathbf{X}) - \lambda(\ln f_{\mathbf{C}}(\mathbf{X}) - \ln c_\alpha) \quad (3)$$

when the boundary of $R_{\mathbf{C}}$ is defined as the ellipsoid $W_{\mathbf{C}}$ (the logarithm in (3) is used to simplify the differentiation) and

$$L(\mathbf{X}, \lambda) = \ln f_B(\mathbf{X}) - \lambda(\mathbf{P}_k^T \mathbf{X} - \delta_\alpha),$$

when the boundary $R_{\mathbf{C}}$ is defined as the hyperplane $\phi_{\mathbf{C}}$. In the latter case, the solution is quite simple:

$$\lambda = (\delta_\alpha - \mathbf{P}_k^T \Theta_B)(\mathbf{P}_k^T \mathbf{V}_B \mathbf{P}_k)^{-1},$$

$$\mathbf{X}_* = \Theta_B + \lambda \mathbf{V}_B \mathbf{P}_k.$$

Geometrically, the point \mathbf{X}_* is the point of contact of the ellipsoid W_B with boundary $R_{\mathbf{C}}$ [17], [18]. Now, taking a confidence level $1 - \alpha$ (here the choice does not depend on the construction of the boundary $R_{\mathbf{C}}$), compare with it the probability measure $P(\text{Int}W_B)$ of the interior of the ellipsoid W_B . If $P(\text{Int}W_B) > 1 - \alpha$, the economic activity of the firm is not heavily corrupt, if $P(\text{Int}W_B) \leq 1 - \alpha$, this should be treated as a signal for more rigid control.

4 Check for corruption in terms of statistical hypotheses

According to the classical theory of hypotheses testing here would be to test the hypothesis $H_{\mathbf{C}} : \Theta_B \in R_{\mathbf{C}}$. As shown in [17], it is enough to verify the hypothesis $H_* : \Theta_B = \mathbf{X}_*$. However, the scale of variation of the average values (Θ_B) for the analyzed firm will be, as is known, the $T_B^{1/2}$ times smaller than the scale of variation of the immediate values (\mathbf{X}) , since under the correct hypothesis H_* estimating of Θ_B obeys the distribution $N(\mathbf{X}_*, T_B^{-1} \mathbf{V}_B)$. This makes it relatively non-sensitive to checking of the firm, whose economic corrupt action was of an episodic nature. In this regard, we should test the hypothesis $H'_{\mathbf{C}} : X \in R_{\mathbf{C}}(H'_* : \mathbf{X} = \mathbf{X}_*)$. This hypothesis uses the distribution of indirect corruption indicators (components of \mathbf{X}). It is interpreted as abuse of firm economic actions of corruption character. This approach is consistent with probabilistic estimates of section 2. Appropriate statistics under these assumptions are components of the vector

$$s = (s_1, s_2, \dots, s_k)^T = (\Lambda_B)^{-1/2} \mathbf{Q}^T (\mathbf{X}_* - \Theta_B),$$

which has Student’s distribution with number of freedom degrees $(T_B - 1)$. Here \mathbf{Q} is an orthogonal matrix such that the matrix $\mathbf{Q}^T \mathbf{V}_B \mathbf{Q} = \Lambda_B$ has diagonal form. In addition, the value

$$r = (\mathbf{X}_* - \Theta_B)^T \mathbf{V}_B^{-1} (\mathbf{X}_* - \Theta_B)$$

with the increase of the sample size T_B asymptotically approaches the distribution χ^2 with the number of freedom degrees k . In the case where the hypothesis $H'_C : \mathbf{X} \in R_C(H'_* : \mathbf{X} = \mathbf{X}_*)$ is not rejected by statistical test, the company falls into the category of suspected of corrupt activities.

5 Example

Let’s consider a numerical example. Let’s assume for simplicity of exposition that the number of indirect corruption indicators is reduced to two ($k=2$), and the boundary of the corruption region is a straight line. Table 1 shows a sample of values for the company’s activities within a certain period of observation. This table represents the calculated average values, which are the coordinates of the test point.

Table 1: Simulated data of test results

n	x_1	x_2	n	x_1	x_2
1	24	1	11	38	3
2	32	41	12	19	11
3	35	38	13	29	29
4	36	39	14	39	3
5	24	12	15	19	8
6	34	39	16	21	15
7	38	41	17	3	21
8	17	8	18	37	32
9	29	7	19	16	12
10	2	13	20	33	3
Average value				26,25	18,8

Making use of the table 1 data calculate matrices $V_B = \begin{pmatrix} 123,250 & 53,842 \\ 53,842 & 215,432 \end{pmatrix}$, $Q = \begin{pmatrix} 0,418147 & 0,90838 \\ -0,90838 & 0,418147 \end{pmatrix}$, $\Lambda_B = \text{diag}(240,216; 98,466)$.

For the illustrative purpose, we assume that the corruption area lies on the straight line $0,75x_1 + 0,25x_2 = 52$ (Altman’s boundary). In this case, the test point falls in the safe region.

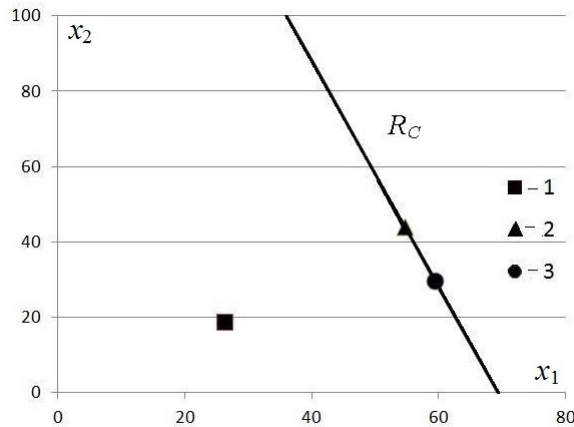


Figure 1: The arrangement of the elements of statistical analysis:1 – test point; 2 – nearest boundary point;3 – boundary point of maximum likelihood

Table 2 shows the coordinates of boundary points nearest to the test point and the coordinates of the boundary points of the maximum likelihood (all three points are presented in Fig.1). This table contains the values of the test according to Student criterion. The sample size is not so big to consider the distribution of the magnitude (5) sufficiently similar χ^2 .

Table 2: The results of the statistical test

Boundary points	x_1	x_2	t	
			$ s_1 $	$ s_2 $
Closest point	59,385	29,845	1,541	2,568
The point of maximum likelihood	54,644	44,068	2,247	1,534
Critical value $\alpha = 0,05$	$t_{19}^{0,025} = 2,433$			

As can be seen from the table 2, for the nearest boundary points one of the Student statistics exceeds the critical value, so it can give a false confidence of the tested company to be free of corruption. However, for the point of maximum likelihood, both statistics are in the confidence region. Thus, despite the fact that, on average, the tested firm is estimated as free of corrupt, in principle this one may commit acts of corruption character. Such firm must be subject to appropriate additional monitoring.

6 Conclusion

The statistical estimation procedure proposed in the paper significantly expands the possibilities of control of firms economic activity with regards to its corruption activity. However, the estimation procedure is becoming more rigorous and more sensitive to changes of indirect corruption indicators in firm economic activity. The procedure can be used also in other fields of mathematical modeling. It is shown that the corrupt checking of the system with a large number of parameters may need to test a complicated hypothesis that arises from restrictions on the parameters. It is shown that the verification of this complex hypothesis is reduced to the verification of simple hypotheses about boundary point, which gives the maximum likelihood. This fact and the algorithm proposed here significantly expands the possibilities of adequate testing for the corruption in the complex multiparameter structures and systems, increasing the reliability of findings and conclusions.

Acknowledgements. The work was supported by RFBR, project No. 14-06-00326.

References

- [1] E. Altman et al., *Managing Credit Risk*, 2nd Edition, John Wiley and Sons, 2008. <http://dx.doi.org/10.1002/9781118266236>
- [2] L.A. Bondarenko, A.V. Zubov, A.F. Zubova, S.V. Zubov, V.B. Orlov, Stability of quasilinear dynamic systems with after effect, *Biosciences Biotechnology Research Asia*, **12** (2015), no. 1, 779-788. <http://dx.doi.org/10.13005/bbra/1723>
- [3] L.A. Bondarenko, A.V. Zubov, V.B. Orlov, V.A. Petrova, N.S. Ugegov, Application in practice and optimization of industrial information systems, *Journal of Theoretical and Applied Information Technology*, **85** (2016), no. 3, 305-308.
- [4] V.V. Dikusar, A.V. Zubov, N.V. Zubov, Structural minimization of stationary control and observation systems, *Journal of Computer and Systems Sciences International*, **49** (2010), no. 4, 524-528. <http://dx.doi.org/10.1134/s1064230710040027>
- [5] X. Grigorieva, O.A. Malafeev, A competitive many-period postman problem with varying parameters, *Applied Mathematical Sciences*, **8** (2014), no. 146, 7249-7258. <http://dx.doi.org/10.12988/ams.2014.47591>

- [6] V.N. Kolokoltsov, O.A. Malafeyev, Mean-Field-Game Model of Corruption, *Dynamic Games and Applications*, (2015), 1-14.
<http://dx.doi.org/10.1007/s13235-015-0175-x>
- [7] E.P. Kolpak, L.S. Maltseva, S.E. Ivanov, On the stability of compressed plate, *Contemporary Engineering Sciences*, **8** (2015), no. 20, 933-942.
<http://dx.doi.org/10.12988/ces.2015.57213>
- [8] E.P. Kolpak, I.S. Frantsuzova, K.S. Alexandrovich, A mathematical model of thyroid tumor, *Global Journal of Pure and Applied Mathematics*, **12** (2016), no. 1, 55-66.
- [9] O.A. Malafeyev, N.D. Redinskikh, G.V. Alferov, Electric circuits analogies in economics modeling: Corruption networks, *2nd International Conference on Emission Electronics*, (2014), 28-32.
<http://dx.doi.org/10.1109/Emission.2014.6893965>
- [10] O.A. Malafeev, Stationary strategies in differential games, *USSR Computational Mathematics and Mathematical Physics*, **17** (1977), no. 1, 37-46.
[http://dx.doi.org/10.1016/0041-5553\(77\)90067-2](http://dx.doi.org/10.1016/0041-5553(77)90067-2)
- [11] O.A. Malafeev, Equilibrium situations in dynamic games, *Cybernetics and System Analysis*, **10** (1974), no. 3, 504-513.
<http://dx.doi.org/10.1007/bf01068721>
- [12] O.A. Malafeyev, L.A. Petrosyan, Differential search games - Dynamic-games with complete information, *Vestnik Leningradskogo Universiteta, Seriya Matematika. Mekhanika. Astronomiya*, **2** (1983), 26-30.
- [13] O.A. Malafeev, Essential non-cooperative n-person games, *Vestnik Leningradskogo Universiteta, Seriya Matematika. Mekhanika. Astronomiya*, **1** (1978), 50-53.
- [14] O.A. Malafeev, Natural metric and points of equilibrium in non-cooperative games, *Vestnik Leningradskogo Universiteta, Seriya Matematika. Mekhanika. Astronomiya*, **4** (1978), 143-145.
- [15] O.A. Malafeev, Dynamic games with dependent motions, *Doklady Akademii Nauk SSSR*, **23** (1973), 783-786.
- [16] Y.B. Mindlin, E.P. Kolpak, N.A. Gasratova, Clusters in system of instruments of territorial development of the Russian Federation, *International Review of Management and Marketing*, **6** (2016), no. 1S, 245-249.

- [17] Yu.A. Pichugin, Geometrical aspects of testing complex statistical hypotheses in mathematical simulation, *St. Petersburg Polytechnic University Journal. Physics and Mathematics*, **218** (2015), no. 2, 123-137, (In Russian). <http://dx.doi.org/10.5862/jpm.218.13>
- [18] Yu. Pichugin, G. Alferov, O. Malafeyev, Parameters estimation in mechanism design, *Contemporary Engineering Sciences*, **9** (2016), no. 4, 175-185. <http://dx.doi.org/10.12988/ces.2016.512319>
- [19] T. Widiharih, S. Haryatmi, Gunardi, D-optimal designs for weighted exponential and generalized exponential models, *Applied Mathematical Sciences*, **7** (2013), no. 22, 1067-1079.
- [20] W. Zhu, Y. Song, Solvability of system of generalized vector quasi-equilibrium problems, *Applied Mathematical Sciences*, **8** (2014), no. 53, 2627-2633. <http://dx.doi.org/10.12988/ams.2014.43183>
- [21] Identifying and Reducing Corruption in Public Procurement in the EU, https://ec.europa.eu/anti-fraud/sites/antifraud/files/docs/body/identifying-reducing_corruption_in_public_procurement_en.pdf

Received: April 25, 2016; Published: June 16, 2016