Project risk management in the construction of high-rise buildings

Boris Titarenko[1*] , *Amir* Hasnaoui[2], *Roman* Titarenko[3], and *Liliya* Buzuk[4]

[1]Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia
[2]The Finance and Economics Department, La Rochelle Business School, France
[3]Stockholm School of Economics, Russia
[4]Russian State Social University, Moscow

Abstract. This paper shows the project risk management methods, which allow to better identify risks in the construction of high-rise buildings and to manage them throughout the life cycle of the project. One of the project risk management processes is a quantitative analysis of risks. The quantitative analysis usually includes the assessment of the potential impact of project risks and their probabilities. This paper shows the most popular methods of risk probability assessment and tries to indicate the advantages of the robust approach over the traditional methods. Within the framework of the project risk management model a robust approach of P. Huber is applied and expanded for the tasks of regression analysis of project data. The suggested algorithms used to assess the parameters in statistical models allow to obtain reliable estimates. A review of the theoretical problems of the development of robust models built on the methodology of the minimax estimates was done and the algorithm for the situation of asymmetric "contamination" was developed.

1 Introduction

One of the project risk management processes in the construction of high-rise buildings is a quantitative analysis of risks. The quantitative analysis usually includes the assessment of the potential impact of project risks and their probabilities. If to solve the first problem, it is sufficient to attract a qualified specialist in this field and provide him or her with necessary statistical information and tools to estimate the potential damage from the risk occurrence, then to estimate the probability of risks it's necessary to apply more sophisticated and specialized methods.

The main problem at the conceptual stage of project management is the selection of approaches and methods that will be used in project risk management. The traditional meth-

* Corresponding author: boristitarenko@mail.ru

ods are usually based on explicit or implicit assumptions about the normality of the probabilistic laws of indicators distribution (see, e.g., Carvalho and Rabechini [1]; Mbaidheen and Alawneh [2]).

In the case when these distribution laws are far from normal (Gaussian), the application of standard methods can lead to a substantial distortion of the results as was shown by Tukey [3]. He showed that even at small deviations of the model from Gaussian, the traditional estimates quickly lose their optimal features.

The robust methods were developed which allow to find stable estimates in situations of inhomogeneous populations (Huber [3]).

According to Huber "small deviations from the model assumptions should impair the performance only slightly, that is, the latter (in terms of the asymptotic variance of an estimate) should be close to the nominal value calculated at the model". And "for statistics representable as a functional T of the empirical distribution, qualitative robustness is essentially equivalent to weak continuity of T" (Huber and Ronchetti [3]).

At present, a robust approach is widely used in various branches of mathematical statistics: factor analysis (Yuan and Zhong [4, 5]; Toman [6]), discriminant analysis (Todorov [7]), claster analysis (Nevolainen [8]), regression models and multivariate analysis (Cizak [9]; Agostinelli [10]), general problems (Morgenthaler [11]; Toman, [6]; Mbaidheen and Alawneh [2]).

At the end of the 20th century, it was first suggested to use the robust methods for project risk management. In recent years, studies in this field have been continued (Titarenko [12,13]; Bubnov [14]; Volkov et al. [15]).

Following Huber's approach, these authors have considered not only the estimates with the smallest asymptotic variance, but the estimates with minimum asymptotic bias and suggested to use in the robust algorithms the procedures of both types at the same time, for example, to obtain a robust estimate of location parameter by minimizing the asymptotic variance and a scale parameter by minimizing the asymptotic bias.

The purpose of this paper is to develop a system model of project risk management and to form on its basis such approaches that combine the traditional methods of project risk analysis with the robust methods to ensure stable estimates of risk parameters.

This paper has the following structure. Section 2 applies and extends the robust approach of P. Huber for the tasks of regression analysis of project data. Section 3 shows theoretical problems of the robust models development. Section 4 develops the algorithm for the situation of asymmetric "contamination". Section 5 concludes the paper.

2 Robust risk management approaches

Standard approaches of economic information processing: definition of averages, calculations of regression dependencies, etc. are based on explicit or implicit assumptions that the distribution laws of economic indicators are normal.

In case these distributions are far from normal, the use of standard methods can lead to substantial distortion of the results. Lack of information in most cases does not allow to establish reliably the law of its distribution. Therefore there is a problem to create such methods of information processing that would be less sensitive to the kind of its distribution law and, in particular, would not so much respond to large random deviations. In the 1960-1970s J. Tukey and P. Huber developed an entire branch of statistics dealing with the construction and the study of "robust" procedures.

Probabilistic models used in project management are quite diverse due to a variety of project types. However, the applied classical assessment methods of models parameters, methods of finding links between technical and economical these models. Application of the robust methods allows to obtain sufficiently reliable (robust to possible inconsistencies of the models and the real phenomena) estimates of the models parameters. indexes, methods of experiments planning and etc. are sensitive to possible deviations from the hypothetical models. This dramatically decreases the reliability of the estimates and results received from these models. Application of the robust methods allows to obtain sufficiently reliable (robust to possible inconsistencies of the models and the real phenomena) estimates of the models parameters.

In initiating to establish a robust procedure it's necessary to consider three interrelated aspects of the problem:
1. Description of a supermodel, reflecting any possible deviations from a "true" model.
2. Establishment of procedures among which will be chosen the best.
3. Defining the purpose of the robust procedure, i.e. the definition of "optimal" procedure. The robustness assessment will be in the fact that its characteristics (bias, efficiency, etc.) change very little within the supermodel.

Deviations from the theoretical model caused by the "gross" errors or other reasons lead to the loss of efficiency. Therefore, the most common optimality criterion is connected with a characteristic of the estimates efficiency. The estimate should have a high efficiency for the "theoretical" model to maintain a sufficiently high efficiency within the framework of the supermodel, i.e. within the theoretical model.

In economic research it is often necessary to study the relations between different economic indicators. The dependence between some indicators z_1, ..., z_n and linked with them by another x indicator can be usually submitted in the form of a linear regression equation

$$x = \theta_1 z_1 + ... + \theta_n z_n + \xi \,. \tag{1}$$

Classical method of solving the problem of finding estimates of the parameters θ_1, ..., θ_n in (1) is a method of least squares that consists in solving the minimization problem

$$\sum_{i=1}^{N} (x_i - \theta_1 z_{1i} - ... - \theta_n z_{ni})^2 = \min_{\theta_1, ..., \theta_n} \tag{2}$$

i.e. in choosing such θ_1, ..., θ_n that N observed sets $(x, z_1, ..., z_n)$ give the least deviation in terms of (2).

Solution of the problem (2) is equivalent to solving a set n of linear equations

$$\sum_{i=1}^{N} (x_i - \theta_1 z_{1i} - ... - \theta_n z_{ni}) z_{ji} = 0 \quad (j = 1, ..., n) \tag{3}$$

Possible way to obtain estimates that are resistant to gross errors is to replace the quadratic function in (2) to another that is less sensitive to large deviations of x_i.

Assume that a random variable ξ in (1) has a distribution $P_\varepsilon(y) = (1 - \varepsilon)P(y) + \varepsilon H(y)$ where theoretical and "contaminating" distributions of $P(y)$ and $H(y)$ are symmetric, i.e. $P(y) = 1 - P(-y)$, $H(y) = 1 - H(-y)$. All distribution functions P_ε with such a view let's combine into a single class K_ε. If $P(y)$ is a function of the normal distribution, this

model of gross errors shows the situation when the approximately $(1-\varepsilon)N$ deviations ξ_i are the subject to the normal law (as in the classical case), and some εN of them are gross errors. Value ε (intensity of "contamination" or the probability of gross errors occurrence) let consider as a known number.

For such a case P. Huber [3] suggested to use developed by him a common approach for the estimation of the location parameter to obtain sustainable estimates θ_1, ..., θ_n and instead of (2) to solve the problem

$$\sum_{i=1}^{N} F(x_i - \theta_1 z_{1i} - ... - \theta_n z_{ni}) = \min_{\theta_1, ..., \theta_n} \qquad (4)$$

with some properly chosen function F. This problem is reduced to solving a set of equations (as a rule, nonlinear already)

$$\sum_{i=1}^{N} f(x_i - \theta_1 z_{1i} - ... - \theta_n z_{ni}) z_{ji} = 0 \quad (j = 1, ..., n) \qquad (5)$$

where $f(u) = F'(u)$.

3 Theoretical problems of the robust models development

P. Huber managed to prove the consistency, nonbias and asymptotic normality of the estimates resulting from (5), and B. Polyak proved the minimaxity of their asymptotic dispersions.

P. Huber [3] proved the following statement. Let

(a)   values ξ_i the realization of the random component ξ are independent and equally distributed with the symmetric distribution function $P_\varepsilon(y) \in K_\varepsilon$;

(b)   known vectors of the input data $\mathbf{z_i} = (z_{1i}, ..., z_{ni})$ are random, independent from ξ and have the same distribution with a finite non-degenerate covariance matrix $\mathbf{B}$, so that $b_{rs} = E(z_{ri} z_{si})$;

(c) $F(y)$ is an arbitrary non-negative symmetric, convex or monotone with y positive function, with some A and any y_1 and y_2 meeting the condition $|F(y_2) - F(y_1)| < A|y_2 - y_1|$;

(d)   in class K_ε there is a distribution P_ε^0 with a density $p_\varepsilon^0(y)$ for which the integral of Fisher information $I(p) = \int_{-\infty}^{\infty} \left(\frac{p'(y)}{p(y)} \right)^2 p(y) dy$ takes the lowest value, and thus $F_0(y) = -\ln p_\varepsilon^0(y)$ is a convex function of y.

Then

(a)   estimates $\hat{\theta}_i$ resulting from the solution of the problem (4) when $N \to \infty$ are consistent, unbiased and have an asymptotically normal distribution with covariance matrix

$$S(F, p_\varepsilon) = \mathbf{B}^{-1} \frac{\int\limits_{-\infty}^{\infty} (F'(y))^2 p_\varepsilon(y) dy}{\left(\int\limits_{-\infty}^{\infty} p_\varepsilon(y) dF'(y)\right)^2} ; \qquad (6)$$

(b)   for all $p_\varepsilon \in K_\varepsilon$

$$\mathbf{S}(F_0, p_\varepsilon) \leq \mathbf{S}(F_0, p_\varepsilon^0) = \mathbf{B}^{-1}(I(p_\varepsilon^0))^{-1} \leq \mathbf{S}(F, p_\varepsilon^0) \qquad (7)$$

where the matrix inequality $\mathbf{A} \geq \mathbf{B}$ means a non-negative certainty of the quadratic form $y^T(\mathbf{A} - \mathbf{B})y$.

Thus, the statement implies that the estimates obtained from (4) with the function $F(y) = F_0(y)$ the largest possible in class K_ε of covariance matrix is the minimum of all possible.

The problem of finding the Huber's robust estimates for the parameters θ_1, ..., θ_n of a linear regression is to solve the problem

$$\sup_{p_\varepsilon} \mathbf{S}(F, p_\varepsilon) = \min_F , \qquad (8)$$

which having due regard to (6) is reduced to the problem

$$\sup_{p_\varepsilon} \frac{\int\limits_{-\infty}^{\infty} f^2(y) p_\varepsilon(y) dy}{\left(\int\limits_{-\infty}^{\infty} p_\varepsilon(y) f'(y) dy\right)^2} = \min_f \qquad (9)$$

$$f(x) = \begin{cases} -\dfrac{p'(x)}{p(x)}, & x \in G = \left\{x : \left|\dfrac{p'(x)}{p(x)}\right| < k\right\}; \\[3mm] k, & x \in G_+ = \left\{x : -\dfrac{p'(x)}{p(x)} \geq k\right\}; \\[3mm] -k, & x \in G_- = \left\{x : -\dfrac{p'(x)}{p(x)} \leq -k\right\}. \end{cases}$$

$$\frac{\varepsilon k}{1-\varepsilon} = -\int\limits_{G_+} (kp(x) + p'(x)) dx - \int\limits_{G_-} (kp(x) - p'(x)) dx . \qquad (10)$$

Minimization of the convex function (4) is a simple computationally problem. For its solution well convergent procedures can be applied.

Consider the most common variant of the gross errors model when $P(y)$ is the normal distribution with zero mathematical expectation and σ^2 variation. In this case, the Huber's model is significantly different from the commonly used multivariate regression model.

Firstly, independent variables z_i in the model are considered not as the known constants, but as the random ones. In other words, the obtaining estimates $\hat{\theta}_j$ will be the most accurate

not for the given specified z_i, but only for the average for the plurality of different possible combinations of such z_i. Therefore, there is no guarantee that for the specific values z_i the obtaining estimates will be good enough.

Secondly, the Huber's model is adapted only for estimating the regression coefficients $\hat{\theta}_j$ but it is not applicable for the variation estimate σ^2. Moreover, as shown in the above formulas for obtaining estimates $\hat{\theta}_j$ the variation σ^2 should be known in advance.

Let's take the quite simple procedure allowing to estimate θ_j jointly with σ^2. To that end, as a function f take

$$f(y) = \max(\min(k, y), -k) \tag{10}$$

where $k = k(\varepsilon)$ is defined as

$$\frac{\varepsilon k}{1-\varepsilon} = \frac{2}{\sqrt{2\pi}}\left(e^{-\frac{-k^2}{2}} - k\int_k^\infty e^{-\frac{x^2}{2}} dx\right).$$

The solution of problem (2) by the transition to solve the following set of equations:

$$\sum_{i=1}^N f\left(\frac{x_i - \sum_{m=1}^n \theta_m z_{mi}}{\sigma}\right) z_{ij} = 0 \ \ (j = 1, ..., n); \quad \sum_{i=1}^N f^2\left(\frac{x_i - \sum_{m=1}^n \theta_m z_{mi}}{\sigma}\right) = \beta \tag{11}$$

where to obtain more accurate and nonbias without "contamination" estimates σ it's possible to choose

$$\beta = (N-n)\int_{-\infty}^\infty f^2(y)p(y)dy$$

The procedure is iterative and based on the piecewise linear behavior function $f(y)$.

Let $\theta^{(r)}, \sigma^{(r)}$ be the estimates for $\theta = (\theta_1, ..., \theta_n)$ and σ received at r-th step of the procedure. Let divide all possible i values into three sets:

$$H = \left\{i : x_i - \sum_m \theta_m z_{mi} < -k\sigma^{(r)}\right\}; \quad C = \left\{i : \left|x_i - \sum_m \theta_m z_{mi}\right| \le k\sigma^{(r)}\right\};$$

$$B = \left\{i : x_i - \sum_m \theta_m z_{mi} > k\sigma^{(r)}\right\}.$$

Now the set of equations (11) can be written as

$$\sum_C \left(x_i - \sum_m \theta_m z_{mi}\right) z_{ji} + \left(\sum_B z_{ji} - \sum_H z_{ji}\right) k\sigma = 0, \ \ j = 1, ... n; \tag{12}$$

$$\sum_C \left(x_i - \sum_m \theta_m z_{mi} \right)^2 + \left(\sum_B 1 - \sum_H 1 \right) k\sigma^2 = \beta\sigma^2 \tag{13}$$

where the letters standing under the index of summation denotes the set of values *i* which is summed.

4 Computational algorithm

1. Choose the initial values $\theta^{(0)}$ and $\sigma^{(0)}$. Suppose $r = 0$.
2. Get sets *H*, *C* and *B*.
3. Calculate matrix **W**, vectors *y* and *d* as well as scalar quantities *Q* and *M* as follows

$$w_{mi} = \sum_C z_{ji} z_{mi} \; ; \quad y_m = \sum_C x_i z_{mi} \; ; \quad d_m = \sum_B z_{mi} - \sum_H z_{mi} \; ;$$

$$Q = \sum_C x_i^2; \quad M = \sum_B 1 + \sum_H 1$$

4. Solve sets of equations

$$\sum_m w_{mj}\eta_m = y_j, \quad j = 1, \dots n \; ; \qquad \sum_m w_{mj}\delta_m = d_j, \quad j = 1, \dots n$$

relatively to values $\eta_1, \dots, \eta_n$ and $\delta_1, \dots, \delta_n$.

5. Values $\theta_m(\sigma) = \eta_m + k\sigma\delta_m$ meet (12) for all σ. If to substitute them into (13), from the obtained equation it is possible to define σ

$$\sigma^{(r+1)} = \sqrt{\frac{Q - \sum_m \eta_m y_m}{\beta - k^2 \left[\sum_m \delta_m d_m + M \right]}} \; .$$

6. Suppose $\theta_m^{(r+1)} = \eta_m + k\sigma^{(r+1)}\delta_m$ and find new sets of *H*, *C* and *B*. If they are the same as the sets found at the previous iteration, then $\theta_m^{(r+1)}, \sigma_m^{(r+1)}$ is the final solution of the problem (11). Otherwise, let's change *r* to $r+1$ and go to step 3 of the algorithm. Since there is only a finite number of combinations for sets *H*, *C* and *B*, the procedure either ends in a finite number of steps or starts periodically to repeat itself.

To compile the program it is necessary to follow some cautions. So, if *C* contains very few elements, the denominator in the expression for $\sigma^{(r+1)}$ can become very low or even negative. In this case, it is necessary to increase $\sigma^{(r)}$ and repeat step 2 of the algorithm. It is necessary to limit the total number of iterations to avoid infinite loops.

The procedure gives estimates for $\theta_1, \dots, \theta_n$ and σ. The covariance matrix for $\theta_1, \dots, \theta_n$ is determined approximately by

$$\frac{N\beta\sigma^2}{(N-n)(N-M)}\left(1 + \frac{nM}{N(N-M)} \right)\mathbf{W}^{-1} \tag{14}$$

In conclusion, it's necessary to emphasize that this procedure despite the simplicity is heuristic and its optimality is not proved.

We consider that it's more efficient to combine Huber's approach to estimate parameters $\theta_1, \ldots, \theta_n$ with the sustainable assessment method σ given above. In this case, the second equation (11) should be replaced by

$$\sigma^2 = Med\left(\frac{\left(x_i - \sum_{m=1}^{n} \theta_m z_{mi}\right)^2}{a^2}\right) \tag{15}$$

where *Med* is a median and $a \approx 0,675$ is a 75% quantile of the standard normal distribution.

5 Discussion

Currently the robust methods cover the following range of tasks: regression and correlation analysis, dispersion analysis, factor analysis, planning of experiments, simulation, statistical estimation of models parameters, estimation of systems reliability.

This paper applies and expands the robust approach of P. Huber for the tasks of regression analysis of project data. The suggested algorithms used to estimate the parameters in statistical models allow to obtain more reliable estimates. Theoretical problems of the robust models development built on the methodology of minimax estimates and proposed by P. Huber are shown as well as the algorithm for the situation asymmetric "contamination" is developed. Thus, this paper shows that the optimal solutions largely depend on the methods of information processing on which basis these solutions are taken. Reliability of the solutions can be reached through the use of special robust procedures for the analysis of management information. An adequate model and a robust information technology decision-making represent the necessary and sufficient conditions for effective management.

References

1. M. Carvalho, R.J. Rabechini, *Impact of risk management on project performance: the importance of soft skills*. International Journal of Production Research **53**(2), 321–340 (2015)

2. M. Mbaidheen, A. Alawneh, *Assessing the risk of corporate strategic change*. International Journal of Information, Business and Management **9(**1), 45–56 (2017)

3. P. Huber, E.Ronchetti, *Robust statistics*, second ed. J. Wiley, New Jersey.(2009)

4. K. Yuan, W. Chan, Y. Tian, *Expectation–robust algorithm and estimation equations for means and dispersion matrix with missing data*. Ann. Inst. Math. **68**, 329–351.(2016)

5. K. Yuan,. X. Zhong, *Outliers, leverage observations and influential cases in factor analysis: using robust procedures to minimize their effect*. Sociological Methodology **38**, 329–368(2008)

6. A. Toman, *Robust confirmatory factor analysis based on the forward search algorithm*. Stat. Papers **55**, 233–252 (2014)

7. V. Todorov, *Robust selection of variables in linear discriminant analysis.* Stat. Math. & Appl **15**, 395–407 (2007)

8. J. Nevolainen, S. Datta, H. Ojc*, Inference on the marginal distribution of clustered data with informative cluster size*. Stat. Papers **55**, 71–92 (2014)

9. P. Cizak, *Efficient robust estimation of regression models*. Comput. Stat. Data Anal **55**, 774–788 (2011)

10. C.Agostinelli, A. Leung, V.Yohair, R. Zomer, *Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination*. TEST **24**, 441-461 (2015)

11. S. Morgenthaler*, A survey of robust statistics*. Stat. Math. & Appl. **15**, 271–293(2007)

12. B Titarenko, System approach to the project risk management. Vestnik MGSU **5**, 340–343 (2011)

13. B.Titarenko, S.Titov, R. Titarenko, Applied Mechanics and Materials **638–640,** 2338–2341 (2014)

14. G. Bubnov, B. Titarenko, S.Titov. R. Titarenko, *Isorisk curves as a tool for increasing flexibility of risk management in engineering projects*. Contemporary Engineering Sciences **8(21)**, 991–999 (2015)

15. A. Volkov, A Sedov, P Chelyshkov., B.Titarenko,., G.Malyha, , E. Krylov, Research Journal of Pharmaceutical, Biological and Chemical Sciences **7(3),** 2416–2420 (2016)