Leonard ROZENBERG, Magdalena KIERUZEL

Wydział Informatyki ZUT w Szczecinie, ul. Żołnierska 49, 70-210Szczecin

doi:10.15199/48.2016.12.63

Using Markowitz's idea to the valuable estimation of the IT projects risk

Abstract. *Realization of information projects require planning, execution, testing, monitoring and control in all phases and aspects. This activities are made by project's members. It is also necessary to estimate the resources and costs associated with the project. The starting point is to determine the tasks structure which leads to calculating the value of the loading of work .The cost is an important information before starting the project, especially when analyzing the relationship between benefits and costs. This knowledge is also necessary because of limited budget of organization involved in the projects.*

Streszczenie. *Realizacja projektów informatycznych wymaga od osób zaangażowanych w ten proces planowania, wykonywania, sprawdzania, monitorowania oraz kontroli wszystkich jego etapów i aspektów. Na etapie przygotowania każdego projektu informatycznego do realizacji konieczne jest oszacowanie zasobów i kosztów z nim związanych. Punktem wyjścia jest określenie struktury zadań, co daje podstawę do wyliczenia nakładów pracy, czasu trwania oraz obciążenia poszczególnych osób zaangażowanych w realizację projektu. Koszt projektu jest istotną informacją przed jego rozpoczęciem, zwłaszcza przy analizie relacji korzyści-koszty, ale również w trakcie realizacji projektu, ze względu na ograniczone możliwości budżetowe każdej organizacji realizującej przedsięwzięcie. (**Zastosowanie metody Markowitza do wartościowej oceny ryzyka projektów informatycznych**)*

Keywords: Value at risk, risk management, IT project..
Słowa kluczowe: wartość ryzykowana, zarządzanie ryzykiem, projekt informatyczny.

Introduction

The budget of project may change due to a number of risk factors. It can bring unexpected costs (usually above the budget costs). The set of risk factors is inscribed in every kind of human activity that is why it is important to protect the project from the effects of risk that will occur [7]. Active risk management process requires risk assessment tools, which in valuable aspect can be an universal criteria to support the decision of starting or continuing the project.

Considering the fact that the implementation of projects is associated with the involvement of a high budget, it is reasonable to provide decision makers with tools to assess the risk of a valuable dimension of the project. The interest in the existence of valuable methods for risk assessment of the project may relate to financing institutions IT projects as well as project teams making current decisions during the execution of the project tasks.

Analysis of currently used risk assessment methods highlights the strong dependence of arbitrary knowledge of those involved in this evaluation process. This means that the fundamental role is played by the expert knowledge and experience of contractors. Outcome of the assessment is therefore a subjective relationship. Project management methodology quantify risk using only the probability and impact strength threats to the project. Commonly known today, agile IT project management methodology does not include the process's risk management, and hence do not specify how the risk would be assessed.

Project risk assessment methods such as decision trees and expected value method give an approximation of the risk of costs terms, but to estimate risk use only values from the evaluated project. Statistical analysis of historical data of other projects can significantly affect the quality of the estimation of risk in financial terms, giving understandable measures to all interested in the project. Lack of one consistent method providing a valuable risk assessment for the whole project is a basic premise for the development of this type of solution.

For the article the importance of the subject to present the Black-Scholes (1973) or the concept of pricing options for discrete variable published in 1979 in the form of the Cox-Ross-Rubinstein's model [11].

It is assumed that the project ends with success, in which the value of the maximum loss and the cost of the project will be lower than the profits and costs of the project. At the stage of defining the scope, tasks and time planned for the whole project, it is possible to establish the maximum value at risk and, consequently, to prepare an appropriate capital reserve.

The basis for determining the maximum loss in the project has components such as the number of tasks in the project and the roles assigned to specific project tasks. On the basis of the roles assigned to project tasks is to carry segregation of individual tasks according to the distribution of specific classes (conceptual).

Characteristics of the objects of research

The source of the data in the study is SourceForge Research Data Archive http://SourceForge.net platform that contains a repository of FLOSS projects types, A Repository of Free/Libre/Open Source Software Research Data. SourceForge.net is one of the largest platforms for the open source project management with registered more than 300,000 projects and nearly 3.5 million users [SourceForge.net 2011]. Data from this platform are transmitted in the form of databases to archive SRDA excluding personal data, data on the functioning of the SourceForge.net and data that can't be transferred for licensing and security reasons. The information available in the archives of SRDA is huge. The range can illustrate the complexity of the archive database SRDA based on RDBMS (Relational Database Management System) PostgreSQL, which consists of more than 100 tables.

Acquisition of data for research purposes was realized using SQL queries to the database SRDA. The vast majority of queries were made to multiple tables containing aggregate functions. Research material has been divided into the following categories:

Data on the complexity of the project - the number of tasks in a project;

- Data on the size of the project in time - project duration and duration of individual tasks in a project;
- Data concerning roles in the developing process - the number of different roles in the project fulfilled by people participating in the project;

• Data related to participation in the implementation process - the number of people participating in the project and implementation of specific actions in a project.

The test material on the basis of the state on 02.2011, distinguished 386 328 registered projects, of which 34 040 were active. Number of active projects with at least one task is 19 713, which represents 58% of the number of active projects and 5% of all projects. The maximum number of tasks in the project is 550.

Analyzing the size of the project during the 6173 distinguished projects with a duration of less than 10 days, which is a 31% share of the projects for which they were defined tasks. The maximum duration of the project is 5216 days (characterized by a single project). Figure 1 shows the number of projects depending on their execution time.

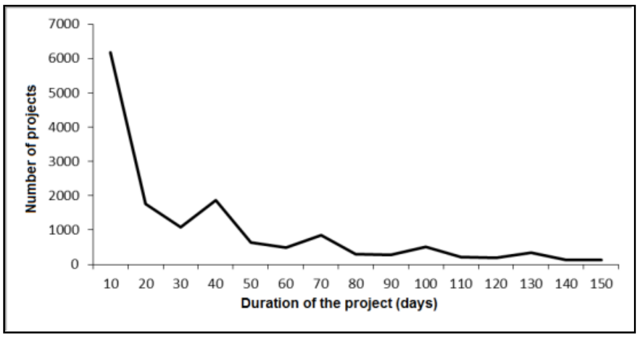


Fig. 1. The number of projects depending on their execution time.

The database http://SourceForge.net defined 29 roles and functions that can be assigned to the project participants. The largest group are the developers whose participation in IT projects accounted for 33.12% and project managers accounted for 28.16%. The smallest group consisted of technical support analysts (involved only in one project). The database distinguished roles, which are not covered by any person: compilation engineer and general manager for marketing. In the tested data, a significant majority of projects are run by one or two persons, which together represent 90% of the projects. Projects, which are assigned to 10 or more people accounted for 0.13%. Among the projects from the research base 156 projects have these characteristics. This test was used to test the valuable risk assessment. It was necessary to determine the structure of the analyzed projects. For this purpose, roles defined on the platform http://SourceForge.net were divided into 4 categories (conceptual classes):

- • Development - Developer, Tester, Cross Platform Developer, All-Hands Person, No specific roles.
- • Analysis and Design - Other Designer, Analysis / Design, Advisor / Mentor / Consultant, Requirements Engineering, Web Designer, User Interface Designer
- • Management - Project Manager.
- • Implementation and support - Unix Admin, Doc Writer, Support Manager, Editorial / Content Writer, Packager, Distributor / Promoter, Support Technician.

According to this division, for each of the 156 projects were categorization tasks - roles of people assigned to them. In each category, summed times of the tasks, which allowed to determine the structure of each project, which is the percentage of each category in the whole project.

Analysis of probability distributions for designated categories of project tasks

In order to determine the statistical measures of the composition realization projects occurring in the study sample were used principles of probability theory. Compliance testing were conducted for common distributions. Due to the fact that the categories of analysis and design as well as implementation, and the support contained a small amount of data, resampling was performed in order to maintain the relative equality of the samples with respect to the collection of development and management.

The test for normal distribution of each observation was carried out in two steps:

1 Rating of normality using probability plots.

2 Carry out compliance tests for the level of significance α = 0.05 according to the algorithms of tests:

a. Kolmogorov-Smirnov.

b. Anderson-Darling.

c. Chi-square.

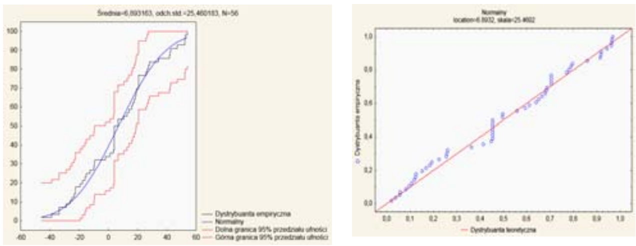
d. Shapiro-Wilk.

Verification subject to the following hypotheses:

The hypothesis H_0 : the distribution of the variable is normal.

Alternative hypothesis H_1: distribution of the test variable is not normal.

Rating of normality distribution using probability graph is to visually verify the position of the empirical points. If they fall along the line defining the theoretical distribution function, there are no grounds for rejecting the hypothesis H_0 speaking about the normal distribution. Otherwise, H_0 is rejected in favor of the H_1, so the test is not a normal distribution.



Plot of empirical distribution for category: Development	Plot type P-P for category: Development

Fig. 2. Test of probability distribution for sample Development

For the category of development on the graph the probability of empirical points is distributed along a line defining the cumulative distribution function of the normal distribution. Similarly, the case looks at the drawing of the empirical distribution function. Thus, from the observation of charts there is no reason to reject the hypothesis H_0.

Table 1. The results of the compatibility tests for: Development

Type of test							
Kołmogorow-Smirnow		Anderson-Darling		Chi-kwadrat		Shapiro-Wilk	
λ	p	A	p	χ^2	P	W	P
0,08919	0,73072	0,39081	0,85742	2,92559	0,40324	0,97478	0,28800

In all tests the compatibility, probability p is higher than the level of significance α = 0.05, which means that there is no evidence to reject H_0. Therefore, it should be assumed that the data in the research sample of development, have a normal distribution

Use of the Markowitz approach to minimize the risk of IT project

Markowitz's Portfolio Theory boils down to the creation of the so-called efficient portfolios, those that maximize revenue at a given risk and minimize risk at a given income. In case of projects that are treated as portfolios of task categories, examined in terms of cost, this approach will bring to the determination of an IT project with minimal risk and cost, i.e. minimum VaR.

After determining for the specific task categories: foot increase labor intensity, volatility and correlation coefficients, these components and the shares of each category were used to determine expected rates of increase of labor intensity and variability, for projects included in the research sample.

Expected rate of increase in workload of individual projects is determined by the formula (1). As you can see it is a weighted average of the interest rates increase the labor intensity of each category, and the weights are the shares of the category in a software project.

$$(1) \qquad r_{PR} = \sum_{i=1}^{n} w_i \bullet r_i$$

where: r_{PR} – the expected rate of growth of labor-intensity of IT project, w_i – the share of labor-intensity in each category of project tasks for which the condition is met $\sum_{i=1}^{n} w_i = 1$, r_i – the rate of growth of labor-intensity in each category of tasks, consecutive indexes i stand for respective categories W, P, Z, S, that is $n = 4$.

The variability of individual projects is determined by the formula (2). It depends on the variability of task categories that make up the project, and the correlation between them. In the case study sample derived from the platform http://SourceForge.net correlation between categories is negative. Markowitz's theory that volatility, i.e. the risk of IT project should be significantly less than the risk of each category, which is apparent from the formula (2):

$$(2) \qquad \sigma_{PR} = \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_i^2 + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_i w_j \sigma_i \sigma_j p_{ij}}$$

where: σ_{PR} – variability of IT project,, σ_i – variability of individual categories of tasks, ρ_{ij} – correlation coefficients between the different categories of tasks, in this case negative, reducting the resultant volatility, and thus reduces the risk of project implementation.

Figure 3 shows the values calculated according to the above formulas for the projects of the research sample. Brown color are selected individual projects, while the blue - a hypothetical IT projects consisting of only one category of tasks.

Taking into account all the possible values of the shares of the various categories of tasks obtaining a figure, whose interior and marked envelope, specify all the possible values of the expected rate of increase of labor intensity and variability of IT project that may occur with different proportions of each category. Defined in this way, the opportunity set.

In Figure 3, the dashed line indicated subset of projects with a minimum VaR. A subset of those IT projects, for which you can't specify the better projects, is called the efficient frontier, or efficient set. IT projects included in this set have:

• a minimum volatility at a given rate of growth of labor-intensity,

• have a minimum rate of growth of labor-intensity at a given volatility.

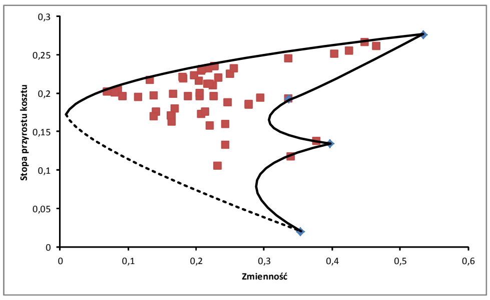


Fig. 3. Opportunity set for information project

From these considerations, project should be carried out in such a way that it could find in the set of efficiency or - at least - strove to it. Depending on the profile of risk acceptance, people preferring safer alternatives will seek to projects located in the left part of the collection efficiency, while having a tendency to risk will lead the project towards the right part of the collection efficiency.

To determine the efficient set it is necessary to use optimization methods. In general, this problem is solved using quadratic programming methods that are available in a variety of specialized software bundles. In the case of determination of an IT project with minimal risk or the project with minimal risk at a given rate of growth of labor-intensity you can use the solution based on the use of Lagrange's multipliers.

The determination of an IT project with minimal risk

To solve this problem is sought in the minimum variation of an IT project regardless of the rate of growth of labor-intensity of the project. The share of each category of tasks in a sought-after software project is defined by formula (3):

$$(3) \qquad w^* = C^{-1} \times I$$

where: w^* – vector $n+1$ – pieces, in which first n pieces are shares of each categories of tasks, and the last piece is the Lagrange's multiplier λ, C – square matrix of dimension $n+1$, which elements are defined as follows (4):

$$(4) \qquad \begin{gathered} c_{ii} = 2\sigma_i^2, i = 1,\ldots, n, \\ c_{ij} = 2\sigma_i \sigma_j p_{ij}, i, j = 1,\ldots, n, i \neq j \\ c_{i,n+1} = c_{n+1,i} = 1, i = 1,\ldots, n, \\ c_{n+1,n+1} = 0 \end{gathered}$$

C^{-1}– inverse matrix C, I – vector $n+1$ – pieces, in which first n pieces equals 0, and the last piece equals 1.

The specified set of the correlation coefficients and the volatility of the labor intensity of each category of tasks based on historical data projects managed on the platform http://SourceForge.net gave on the basis of the formula (4) vector of shares of the various categories of tasks for the IT project with minimal risk.

The result shows the relationship (5):

$$(5) \qquad \begin{vmatrix} 2{,}221 & -1{,}192 & -0{,}290 & -0{,}739 & 0{,}355 \\ -1{,}192 & 2{,}343 & 0{,}009 & -1{,}159 & 0{,}222 \\ -0{,}290 & 0{,}009 & 1{,}262 & -0{,}980 & 0{,}224 \\ -0{,}739 & -1{,}159 & -0{,}980 & 2{,}879 & 0{,}199 \\ 0{,}355 & 0{,}222 & 0{,}224 & 0{,}199 & -0{,}016 \end{vmatrix} \bullet \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{vmatrix} = \begin{vmatrix} 0{,}355 \\ 0{,}222 \\ 0{,}224 \\ 0{,}199 \\ -0{,}016 \end{vmatrix}$$

From the above figures that the IT project with minimal risk is such a project in which the shares of the individual categories of tasks are respectively (for a pre-designated volatility and correlation matrix):

• Development of 35.5%,
• Design and analysis of 22.2%,
• Management of 22.4%,
• Implementation and support of 19.9%.

The variability of such a project is 3% and the expected rate of growth of labor- intensity is 16.4%. In Figure 4, this project is designated as P1. You can see that this is a safe project, which is paid for by the high rate of growth of labor-intensity.

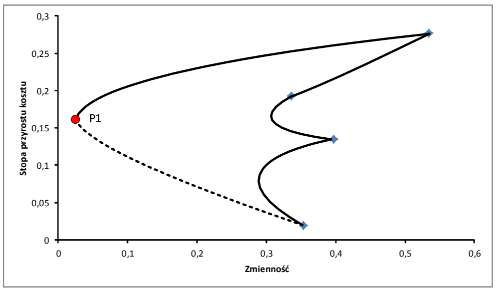


Fig4. IT project with minimal risk

The determination of an IT project with minimal risk at a given rate of growth of labor-intensive

Strategy of project management directed at increased risk will seek to reduce the rate of growth of labor- intensity at the expense of an increased risk of the project, i.e., volatility of the project. The share of each category of tasks in a sought-after software project is defined by formula (6):

$$w^{**} = D^{-1} \times I_0 \qquad (6)$$

where: w^{**} – vector $n+2$ – pieces, in which first n pieces are the shares of each categories of tasks, whereas last two pieces are Lagrange's multipliers λ i μ, D – square matrix of rank $n+2$, which elements are set out below:

$$
\begin{aligned}
&d_{ii} = 2\sigma_i^2, i = 1,...,n,\\
&d_{ij} = 2\sigma_i\sigma_j p_{ij}, i,j = 1,...,n, i \neq j\\
&d_{i,n+1} = d_{n+1,i} = 1, i = 1,...,n,\\
&d_{i,n+2} = d_{n+2,i} = r_i, i = 1,...,n,\\
&d_{n+1,n+1} = d_{n+1,n+2} = d_{n+2,n+1} = d_{n+2,n+2} = 0
\end{aligned} \qquad (7)
$$

D^{-1}– inverse matrix D, I_0 – vector $n+2$, in which n pieces equals 0, next-to-last equals 1, while the last is equal to the given rate of growth of labor-intensity of the wallet.

When solving the problem of IT project with the minimum risk at a given rate of growth of the labor intensity it is necessary, of course, to know the rate of growth of labor-intensity for each category.

Under (8) are calculations for the project at the expected rate of growth of the labor intensity of 8%.

$$
\begin{vmatrix}
1{,}946 & -1{,}296 & -0{,}729 & 0{,}080 & 0{,}094 & 1{,}595\\
-1{,}296 & 2{,}303 & -0{,}158 & 0{,}849 & 0{,}122 & 0{,}605\\
-0{,}729 & -0{,}158 & 0{,}561 & 0{,}326 & -0{,}194 & 2{,}547\\
0{,}080 & -0{,}849 & 0{,}326 & 0{,}443 & 0{,}977 & -4{,}747\\
0{,}094 & 0{,}122 & -0{,}194 & 0{,}977 & -0{,}264 & 1{,}517\\
1{,}595 & 0{,}605 & 2{,}547 & -4{,}747 & 1{,}517 & -9{,}253
\end{vmatrix}
\cdot
\begin{vmatrix} 0\\ 0\\ 0\\ 0\\ 1\\ 0{,}08 \end{vmatrix}
=
\begin{vmatrix} 0{,}221\\ 0{,}171\\ 0{,}010\\ 0{,}598\\ -0{,}143\\ 0{,}777 \end{vmatrix} \qquad (8)
$$

IT project with minimal risk at a given rate of growth of the labor intensity equal to 8% is a project in which shares of each category are as follows:
• Development of 22.1%,
• Design and analysis of 17.1%,
• Management of 1%,
• Implementation and support of 59.8%.

Variability for this project is 19.3%. On the figure the project is marked as the point P2. As you can see, the volatility of an IT project has increased much more in relation to the rate of growth of the labor intensity.

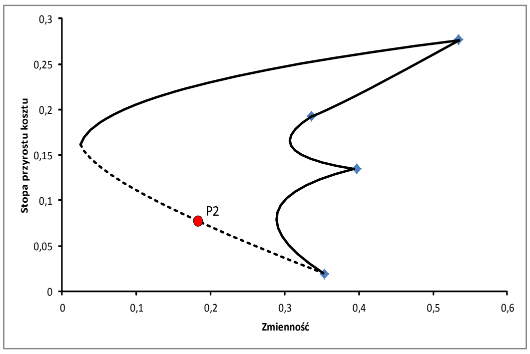


Fig. 5. The project with minimal risk for the rate of growth in labor-intensity equal to 8%

Assuming that the total labor intensity of an IT project P1 is the same as the project P2 and is, for example 4773, the value at risk (VaR) in the month period and confidence level of 95% is 384 and 1526.12. As you can see the project P2 is significantly more risky than P1.

Proposed and presented here valuable method of risk assessment of IT projects represents a new approach to risk assessment. Through the use of the concept of VaR and proposed concept is possible to get an objective assessment of the risk level of the planned or ongoing project information.

***Authors**: dr hab. inż. Leonard Rozenberg, prof. ZUT, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Informatyki Katedra Inżynierii Systemów Informacyjnych, ul. Żołnierska 49 71-210 Szczecin, e-mail:lrozenberg@wi.zut.edu.pl; dr Magdalena Kieruzel, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Informatyki Katedra Inżynierii Systemów Informacyjnych ul. Żołnierska 49 71-210 Szczecin, e-mail: mkieruzel@wi.zut.edu.pl*

REFERENCES

[1] T.Bałamut T.: Metody estymacji Value at Risk, Materiały i Studia NBP, Zeszyt 147, Warszawa 2002.
[2] Byström: Managing Extreme Market Risk using Ex-treme Value Theory, Department of Economics, Lund University, 2001.
[3] Carr M., Konda S., Monarch I., Ulrich. C: „Taxonomy-Based Risk Identification", Software Engineering Institute, Carnegie Mellon University Pittsburg, Pennsylvania, 1993
[4] Chapman R.J.: The effectiveness of working group risk identification and assessment techniques, International Journal of Project Management Vol. 16, No. 6, 1998.
[5] Christley S., Greg Madey G.:Technical Report: TR-2005-15, Dept. of Computer Science and Engineering, University of Notre Dame, October 10, 2005.
[6] Dowd K.: Beyond Value at Risk. The New science of risk management. Wiley, Chichester 1998
[7] Jajuga K.: Zarządzanie ryzykiem w przedsiębiorstwie i instytucji finansowej – metody ilościowe a wyzwania praktyki. Zeszyty naukowe Uniwersytetu Szczecińskie-go nr 394, prace Katedry Ekonometrii i Statystyki nr 15 2004.
[8] Jorion, P. Value at Risk: The new benchmark for man-aging isk, third edition, McGraw-Hill 2007.
[9] Karmańska A.: Ryzyko w rachunkowości, Wydawnic-two Diffin, Warszawa 2008.
[10] Karolak, D.W.:Software Engineering Risk Management, IEEE Computer Society Press, 1996.
[11] Pera K.: Koncepcja VaR w pomiarze ryzyka surowcowego projektu inwestycyjnego, Gospodarka Surowcami Mineralnymi. Tom 24 Zeszyt4/4 2008