

Study of Economic Systems Using the Simulation-Based Statistical Modeling Method

Aleksandr Mikhaylovich Batkovskiy¹

Alina Valerevna Konovalova²

Elena Georgievna Semenova³

Valeriy Jaroslavovich Trofimets²

Alena Vladimirovna Fomina¹

¹Joint Stock Company "Central Research Institute of Economy, Management and Information Systems "Electronics",
Moscow, Russian Federation

²The Department of economic analysis and Informatics, P.G. Demidov Yaroslavl State University, Yaroslavl, Russian Federation

³The Department of Innovation and Quality Management, St. Petersburg State University of Aerospace Instrumentation,
St. Petersburg, Russian Federation

Email: batkovskiy_a@instel.ru

Doi:10.5901/mjss.2015.v6n4s4p369

Abstract

This article analyzes theoretical and methodical aspects of study of economic systems using the simulation-based statistical modeling method. Principles of simulation-based statistical modeling are demonstrated on "coin toss" and "improved Buffon experiment" simple classic models realized within the environment of MS Excel tabular processor. The theoretical basis of the method – the law of large numbers and the central limit theorem – are also demonstrated by the way of computer experiments. Setting of adequate laws of random distribution is the main condition for efficient application of simulation-based statistical modeling method. Wrong assumption regarding the type of distribution laws may result in significant errors in assessment of the output parameters of the system under consideration. Selection of a distribution law type should be based on the economic essence of the modeled random variable. The logical scheme and phases of the simulation-based statistical modeling are demonstrated by the example of investment project's risk assessment. The project cash flow model is based on known estimated NPV ratios. External and internal random factor impact model allows to set various laws of random distribution, and to take into account correlations between them. Statistical processing model, together with the module for obtaining of typical statistical estimates, includes the module for construction of empirical distribution function and selection of projects with NPV close to the modal value within the set interval. The modeling process shows that the obtained mean NPV is too optimistic in certain situations, and it would be reasonable to take the modal value of NPV as its projected value. The structure of the computer model developed using the MS Excel tabular processor environment is considered.

Keywords: economic systems, simulation-based statistical modeling, law of large numbers, MS Excel tabular processor.

1. Introduction

Most economic and mathematical application models analyzed in academic literature are deterministic. Such assumption is quite reasonable for short-term planning periods or for a sustainable economy, which is not exposed to inflation. If such terms are not met, then we may expect that the results obtained on the basis of deterministic models would have significant error. But how can we assess the consequences of an impact of random factors on the studied economic system? The main ways include:

- *Sensitivity (sustainability) analysis of the deterministic model.* In the course of such analysis, a certain set of optimization models, i.e. essentially a certain operations research model is studied. This gives certain responsiveness to the model, which allows a financial analyst to analyze the impact of possible changes of initial conditions on the optimal solution obtained before. Dynamic characteristics of models reflect the same characteristics intrinsic to real processes (Taha, H., 2010);
- *Transformation of the deterministic setting of the problem into its stochastic (probability) form.* Usually, transformation is effected in respect of two most widespread stochastic formulations known, respectively, as *M*-formulation and *P*-formulation. If just the stochastic nature of a target function is described for simplicity

purposes, then within M -formulation the task would be described as maximization (minimization) of the mathematical expectation (mean value) of a target function, and within P -formulation the task would be described as maximization of probability of obtaining the maximum (minimum) value of the target function. Solution of such class of tasks is studied within the special discipline of mathematic programming known as stochastic programming section (Trofimec, V.Y. and Mamatova, L.A., 2007);

- Use of simulation-based modeling methods, which allow to reproduce the behavior of the studied economic system under the influence of random external and internal factors, if their probability distribution is known. Simulation-based models are usually located between operations research models and statistical models, as they comprise typical characteristics of both models. The second component of these models is usually implemented with the help of statistical research also known as the Monte Carlo method. This method is widely used in different subject areas, even in economics and finance (Glasserman, P., 2003), (Dzhekel, P., 2004). When it comes to the latter, such method became widespread in inventory management problems, queueing theory in assets management tasks and investment projects.

Therefore, the natural evolution of economics and mathematical modeling, having started from deterministic models and methods, resulted in utilization of probabilistic models and methods. This article focuses on the issues of development and study of simulation-based statistics models by students with majors in finance and economics.

2. Literature Review

The idea of the statistical testing model has been known quite for a long time. For example, the π number computation method which uses random throws of a needle on a paper with parallel lines drawn has been known since 1873 (Ermakov, S.M., 1975). It is assumed that the Monte Carlo method has become the object of study starting from the work by Metropolis and Ulam (Metropolis, N. and Ulam, S., 1949).

Initially, the Monte Carlo method was used mainly for solving neutron physics tasks, where the traditional numerical methods were useless (Spanier, J. and Gelbard, E., 1969). Then it was used for solving of a wide range of gas dynamics tasks (Bird, G.A., 1976), statistical physics (Landau, D.P. and Binder, K., 2005), numerical integration and solving of the systems of non-linear equations (Halton, J.H., 2006). The Monte Carlo method quickly gained popularity in solving various applied problems, first of all the queueing theory and the games theory, communications theory etc. (Buslenko, N.P., 1970), (Gerald, H. and Shapiro, S., 1967), (Sobol, I.M., 1994), (Fishman, G.S., 1996), (Winston, W., 1998), (Kroese, D.P., Taimre, T. and Botev, Z.I., 2011), (Dick, J., Kuo, F.Y., Peters, G.W. and Sloan, I.H., 2013), (Ledder, G., 2013), (Zio, E., 2013), (Thomopoulos, N.T., 2014).

Different examples of application of the Monte Carlo method for solution of financial and economic tasks may be found in the works of Lukasevich (Lukasevich, I.Y., 1998), Watsham and Parramore (Watsham, T. and Parramore, K., 1997), Dunn and Shullis (Dunn, W.J. and Shullis, K., 2011), Chan (Chan, V., 2013), Shonkwiler (Shonkwiler, R.W., 2013), Brandimarte (Brandimarte, P., 2014), Davey (Davey, K.J., 2014).

At the present time, the scope of tasks solved using the Monte Carlo method is very wide and constantly grows. This method is used for solution of various applied scientific and engineering tasks, and respective works made at different levels are spread all over numerous information sources. We might say as well that the time, when it was easy to prepare more or less complete review of applications of the Monte Carlo method, has passed, but the time to summarize its development has not come yet.

3. Methods

3.1 *The general approach to study of economic systems using the simulation statistical modeling method*

As we mentioned above, the simulation-based statistical modeling method, also called as the statistical testing method or the Monte Carlo method (the latter name is related to the well-known city of Monte Carlo in the Principality of Monaco, famous for its gambling houses, because roulette is one of the simplest devices for generation of random numbers), is one of the most powerful methods for studying of economic systems under the impact of random factors.

In the general case, simulation is understood as the process of computer experiments with mathematical models of complex real-world systems. The goals of such experiments vary from identification of fundamental properties and patterns of the analyzed system to solution of particular practical tasks (theoretical tasks: calculation of curvilinear figures areas, calculation of constants, analysis of particle diffusion etc; practical tasks: modeling of nuclear chain reactions, population migration, behavior of red and white blood cells, option pricing etc). Development of computer hardware and

software resulted in significant expansion of the range of simulation application in economics and finance. Currently it is used both for solving of in-house management tasks and for modeling of management on the macroeconomic level.

As follows from the definition, simulation is a computer experiment. The only difference of such experiment from the real one lies in the fact that it is carried out with the system's model instead of the system itself. Real experiments with economic systems is at least irrational, requires significant costs and can be hardly put into practice. Therefore, simulation is often the only way for studying the system without real experiments.

3.1.1 Principles of simulation-based statistical modeling

Principles of simulation-based statistical modeling include:

1. Experiments with the real object (system) are replaced with other model experiments with the same (or similar) probabilistic structure;
2. Experiments are carried out quite a large number of times in order to obtain reliable statistical characteristics of random variables (relative frequencies, arithmetic mean etc.) used as approximate estimates of their probabilistic characteristics (as estimated probabilities, mathematical expectations etc.). Feasibility of such approximation is based on the law of large numbers, the main theorems of which will be described below.

The first principle of simulation-based statistical modeling can be described using the classic example of a coin toss – first of all, to carry out a physical, and then a simulation experiment. The simulation experiment of a coin toss can be easily realized in MS Excel environment by combination of IF and RAND functions: =IF(RAND()<=0,5,"heads";"tails"). "Coin toss" is effected by pushing F9 button (manual recalculation of MS Excel work sheet functions). The second principle of the simulation-based statistical modeling may be described using another classic example – calculation of π number value (introduced at information science classes in the high school). Though it seems that such mathematical problem by no means may relate to the Monte Carlo method, but as early as in 18th century French mathematician Buffon used such method for finding π number; Buffon who threw a needle on a large sheet of paper with parallel evenly spaced lines for 5,000 times, with the length of the needle the same as spacing between the lines. The probability theory suggests that in respect of n/N (where n is the number of needle hits on any line, N is the total number of throws) π number is factored, which Buffon tried to determine in such an unusual way. We would rather use another, simpler way of π number determination – throwing "stones" in a square with inscribed circle. Since the "stones" would be thrown evenly, we may suggest that n/N (where n is the number of "stones", which fall within the circle, N is the total number of thrown "stones") ratio would correspond to S_{in}/S_{sp} ratio:

$$\frac{n}{N} \approx \frac{S_{in}}{S_{sp}} = \frac{\pi R^2}{a^2}, \text{ from which } \pi \approx \frac{a^2 n}{R^2 N} \quad (1)$$

In the implemented simulation model $a = 2, R = 1$, therefore

$$\pi \approx 4 \frac{n}{N} \quad (2)$$

Figure 1 shows the visual results of one "stone" throw experiment in MS Excel, which was carried out simultaneously with 10, 100, 1000 and 10000 "stones".

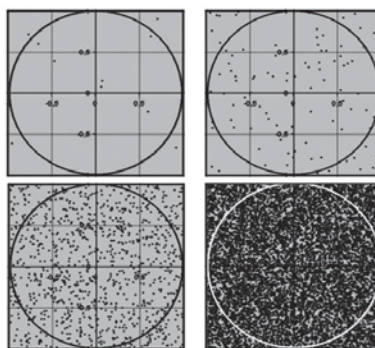


Figure 1. Result of "stone" throw computer experiment

Table 1 shows analytical results of this experiment.

Table 1. Results of the "stone" throw computer subexperiment

Result	Quantity of stones			
	10	100	1000	10000
Hit	7	76	792	7859
Missed	3	24	208	2141
Pi number (estimated)	2.8000000	3.0400000	3.1680000	3.1436000
Pi number (exact)	3.1415927	3.1415927	3.1415927	3.1415927
Absolute deviation	0.3415927	0.1015927	0.0264073	0.0020073
Relative deviation	10.87%	3.23%	0.84%	0.06%

From this very stage, intuition suggests that the accuracy of π number determination depends on the quantity of thrown "stones". Experiments with 10, 100, 1000 and 10000 "stones" prove such hypothesis, though sometimes some deviations occur, which become apparent in less exact value of π for larger amounts of "stones". Such deviations mitigate when the model is run several times and the mean value for several model runs is taken.

As shown above, quite high accuracy of task solution using the simulation-based statistical modeling method is guaranteed, generally, only when a large quantity of experiments is run; therefore, such method is utilized with the use of computer.

3.1.2 Theoretical basis of simulation-based statistical modeling

Probability limit theorems form the theoretical basis of simulation-based statistical modeling method. They may be divided into two groups, one of which is known as the *law of large numbers*, and another – as the *central limit theorem*. The law of large numbers establishes relationship between the statistical estimates of random variables and their probabilistic characteristics, while the central limit theorem establishes relationship between the law of random distribution and its limit form – the normal law of distribution (Kashyap, R.A. and Rao, A.R., 1983).

The *law of large numbers* should not be understood as one general law – it is a generic name of several theorems, from which it follows that as the number of experiments is infinitely increased, the mean values approach certain constant values. The following theorems are attributable to the law of large numbers: Bernoulli's theorem, Chebyshev's theorem, Chebyshev's generalized theorem, Markov's theorem and other theorems.

Let's briefly look at the essence of two limit theorems, which play a special role in theoretical justification of simulation-based statistical modeling – Chebyshev's theorem and Bernoulli's theorem.

Chebyshev's theorem: if in n independent tests x_1, \dots, x_n values of random variable X with mathematical expectation $M[X] = \mu_x$ are observed, then in case of an unlimited increase in the quantity of tests ($n \rightarrow \infty$) the arithmetic mean of values of a random variable converges in probability to its mathematical expectation μ_x , i.e. in any, as small as $\varepsilon > 0$ the equality holds:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n x_i - \mu_x \right| < \varepsilon \right\} = 1 \tag{3}$$

Chebyshev's theorem implies that arithmetic mean $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ of a random variable X is a natural estimate of its mathematical expectation $M[X] = \mu_x$. Chebyshev's theorem has strong mathematical proof, but it can be also demonstrated by the following simulation experiment in MS Excel. We will generate a random variable with even distribution in segment $[0, 1]$. It is known that the mathematical expectation of such random variable equals to 0.5. Let's see how the arithmetic mean of random variable values will approach its mathematical expectation with increase in the quantity of tests. Table 2 shows the results of one of the model runs.

Table 2. Results of a particular run of Chebyshev's theorem statistical check model

Indicators	Quantity of tests		
	10	100	1000
Mean	0,60567	0,53429	0,49508
Mathematical expectation	0,50000	0,50000	0,50000
Absolute deviation	0,10567	0,03429	0,00492
Relative deviation	21,13%	6,86%	0,98%

There are also some deviations, appearing in a larger deviation of the mean value from the mathematical expectation with a larger quantity of tests. Such divergences mitigate when the model is run several times and the mean value for several model runs is taken.

Bernoulli's theorem: if in any of n independent tests probability p of occurrence of event A is constant, then the probability of arbitrarily small deviation of relative frequency from probability p by the absolute value is arbitrarily close to 1, if the quantity of tests is large enough:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{m}{n} - p \right| < \varepsilon \right\} = 1 \tag{4}$$

where m is the quantity of positive outcomes (occurrence of event A); n means the total quantity of tests; $\frac{m}{n}$ means the probability estimate determined as the relative frequency of occurrence of event A .

Therefore, the value of relative frequency may be taken (with certain degree of reliability) as the probability of occurrence of an event.

Bernoulli's theorem has strong mathematical proof, but it may be also demonstrated by the following simulation experiment in MS Excel. Let's refer to the coin toss. It is known that the probability of "heads" (as well as "tails") is equal to 0.5.

Let's see how the relative frequency of "heads" falling would approach the probability of its occurrence with increase in the quantity of tests. Table 3 shows the results of one of the model runs.

Table 3. Results of a particular run of Bernoulli's theorem statistical check model

Indicators	Quantity of tosses		
	10	100	1000
"Heads"	3	54	502
"Tails"	7	46	498
Probability of "heads" (estimated)	0.300	0.540	0.502
Probability of "heads" (real)	0.500	0.500	0.500
Absolute deviation	0.200	0.040	0.002
Relative deviation	40.00%	8.00%	0.40%

There are also some deviations, appearing in a larger deviation of the relative frequency from the probability with a larger quantity of tests. Such deviations mitigate when the model is run several times and the mean value for several model runs is taken.

Therefore, the limit theorems guarantee high quality of statistical estimates obtained using the Monte Carlo method with a larger n . However, acceptable estimates of the modeled systems characteristics may often be obtained with relatively small values of n (approx. several tens and hundreds), it being proved that averaging of the values obtained from several model runs significantly increases the quality of statistical estimates (Sobol, I.M., 1994).

3.1.3 The logical scheme and phases of simulation-based statistical modeling

Simulation-based statistical modeling of economic systems is carried out according to the logical scheme shown in Figure 2 and includes the following phases:

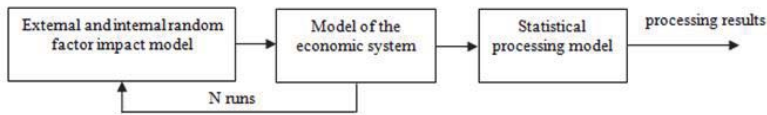


Figure 2. The logical scheme of development and analysis of applied simulation-based models

1. Development of the models of:
 - a) economic systems (establishment of relationships between the initial and output parameters of economic systems in the form of mathematical equations or inequalities);
 - b) impact of internal and external random factors (probability distributions laws for random parameters of economic system are set);
 - c) statistical processing (the set of statistical methods for calculation of the system's statistical characteristics (relative frequencies, mean arithmetic etc.) used as approximate estimates of their probabilistic characteristics (probabilities of events, mathematical expectations etc.). Feasibility of such approximation is based on the above law of large numbers;
2. Simulation-based statistical modeling of the economic system:
 - a) computer imitation of the system's random parameters;
 - b) calculation of output parameters of the economic system;
 - c) multiple repetition of previous phases a) and b) (N model runs).
3. The statistical analysis of the simulation results:
 - a) estimation of the mean values of output parameters of the economic system;
 - b) estimation of confidential intervals for mean values with set reliability level (usually 90% or 95% or 99%);
 - c) estimation of the probabilities of occurrence of relevant events;
 - d) construction of empiric distribution functions in respect of the system's output parameters;
 - e) check of the hypothesis in order to make sure that empiric distribution functions conform with the projected theoretical distribution functions.

Relying on a logical scheme and typical phases of simulation-based statistical modeling, we have developed the method and computer model of solution to the task of investment project's risks assessment.

3.2 The method of solution to the investment project's risk assessment task

From the perspective of the systemic approach, identification of significant factors and ignoring of the factors which do not affect or slightly affect the systemic characteristics, is important for construction of the economic system's model. The starting point, according to which investment attractiveness of any company may be assumed as a continuous chain of asset transformations from one form to another, is important for the investment project's risk assessment. All these transformations generate cash flows allocated in time throughout the project's life cycle. Therefore, selection of an indicator which adequately describes cash flows from the company's operations, investments and financial transactions is important for investment project's risk assessment.

3.2.1 The system's mathematical model

On the basis of analysis of indicators and methods of investment project's risk and efficiency assessment, NPV indicator was taken into account being the most widespread in investment project development practice (Kovalev, V.V., 2008), (Lipsic, I.V. and Kossov, V.V., 2011), (Moskvin, V.A., 2004). This indicator is estimated based on well-known economic ratios and estimated cash flows from the company's operations, investments and financial transaction (Vilenskij, P.L et al., 2002), (Ostrovskaya, E., 2004).

NPV indicator is estimated using the following basic formula:

$$NPV = PV - I_0, \quad (5)$$

where PV is the present value of cash flow; I_0 means the investment amount.

I_0 value is known with a high degree of accuracy, PV value is determined by formula (2):

$$PV = \frac{CF_1}{(1+r)} + \frac{CF_2}{(1+r)^2} + \dots + \frac{CF_n}{(1+r)^n} = \sum_{t=1}^n \frac{CF_t}{(1+r)^t}, \quad (6)$$

where r means the discount rate; n means the quantity of project realization periods; CF_t means net payment flow during time t .

When determining the discount rate, we proceeded from the following expression:

$$r = r_1 + r_2, \quad (7)$$

where r_1 is the long-term deposit rate in highly reliable banks; r_2 means risk premium.

In order to find CF_t , let's introduce the following designations for the project's initial parameters: Q means the production volume (in physical units); P means per unit price; FC means fixed costs; VC_1 means per unit costs; A means depreciation; T means the profit tax, %;

It is clear that the total revenue from the project's realization would amount to:

$$TR = Q \times P, \quad (8)$$

and the total costs would amount to:

$$TC = FC + VC = FC + VC_1 \times Q \quad (9)$$

Then the project's profit before taxes would amount to:

$$Pr1 = TR - TC \quad (10)$$

The tax is charged on the difference between profit and depreciation; therefore, taxable profit would amount to:

$$Pr2 = Pr1 - A \quad (11)$$

and the profit tax would amount to:

$$S = Pr2 \times T \quad (12)$$

Then the net payment flow will be represented by the cash flow from operations and calculated using the following formula:

$$CF_t = TR - TC - S = Pr1 - S \quad (13)$$

By inserting expressions (8)-(10) into formula (13) and making mathematical transformations, we obtain a resultant formula for the cash flow:

$$CF_t = [Q \times (P - VC_1) - FC - A] \times (1 - T) + A \quad (14)$$

By inserting expression (14) into formula (6), and then into formula (5), we obtain the final formula:

$$NPV = \sum_{t=1}^n \frac{[Q \times (P - VC_1) - FC - A] \times (1 - T) + A}{(1+r)^t} - I_0 \quad (15)$$

Formula (15) is a mathematical model of the studied economic system, represented here by the investment project.

Then we need to divide project parameters into deterministic and stochastic (or random) parameters. Taking into account the economic essence of parameters and the degree of random component presence therein, the following expressions were attributed to deterministic parameters: I_0 – investment amount; r – discount rate; n – number of project realization periods; A – depreciation.

The stochastic parameters include: Q – production volume, units; P – per unit price; FC – fixed costs; VC_1 – per unit costs.

3.2.2 External and internal random factor impact model

The second phase includes development of internal and external random factors impact model, for which purpose we need to set probability laws for random parameters of the studied economic system. Depending on system's peculiarities, random parameters may be subject to different distribution laws. In the developed model, we relied on the hypothesis of even or normal distribution, though the model allows to set other distribution laws.

Intrinsic function RAND, which allows to generate random value R , evenly distributed within interval $[0, 1]$, was used as a software instrument for generation of random numbers. Particular importance of R value is explained by the factor that random values of different nature can be obtained on its basis.

First of all, the analyzed model implies obtaining of a random value X with even distribution on a random segment $[a, b]$. The following formula was used for this purpose:

$$x = a + r^{[0,1]} \times (b - a), \quad (16)$$

where x means the value of a random variable X with even distribution on a segment $[a, b]$; $r^{(0,1)}$ means the value of a random variable R .

Secondly, the analyzed model implies obtaining of random variable N with standard normal distribution (mathematical expectation $\mu = 0$, standard deviation $\sigma = 1$). The following formula was used for this purpose:

$$\xi = \sum_{i=1}^{12} r_i^{(0,1)} - 12 \quad (17)$$

where ξ means the value of random variable with standard normal distribution.

Thirdly, the analyzed model implies obtaining of a random value Y with random normal distribution. The following formula was used for this purpose:

$$\eta = \mu + \sigma + \xi \quad (18)$$

where η is the normally distributed random variable with parameters μ and σ .

3.2.3 Statistical processing model

Third phase includes development of a model of statistical processing of simulation experiment results. The following statistical characteristics of a random variable NPV (used as approximate estimates of probabilistic characteristics) were applied:

- \overline{NPV} – average value of NPV ;
- $maxNPV$ – maximum value of NPV ;
- $minNPV$ – minimum value of NPV ;
- σNPV – standard deviation of NPV ;
- $varNPV$ – variation factor of NPV ;
- $NPV_{\text{see}}^{90\%}, NPV_{np}^{90\%}$ – left and right confidence interval limits with 90% reliability level;
- $NPV_{\text{see}}^{95\%}, NPV_{np}^{95\%}$ – left and right confidence interval limits with 95% reliability level;
- $NPV_{\text{see}}^{99\%}, NPV_{np}^{99\%}$ – left and right confidence interval limits with 99% reliability level;
- $P(NPV < 0)$ – the probability of project's failure to generate profits, i.e. $NPV < 0$.

In order to check the hypothesis on conformity of empiric distribution with normal distribution, the quantity of intervals is calculated by using the Sturges formula:

$$n = \text{int}(1 + 3,322 \times \lg N) \quad (19)$$

where N is the number of population units (for the analyzed method, $N = 5000$); $\text{int}(\dots)$ means taking of the integral part of a number.

Then, the range for each interval is determined by the following formula:

$$h = \frac{x_{\max} - x_{\min}}{n - 1} \quad (20)$$

where x_{\max}, x_{\min} mean, relatively, the maximum and minimum values of the statistical item in the set.

Then, the hypothesis on conformity of empiric distribution with normal distribution is checked on the basis of Pearson's goodness-of-fit test χ^2 . For this purpose, χ^2 is calculated by using the following formula:

$$\chi^2 = \sum \frac{(f_3 - f_T)}{f_T} \quad (21)$$

where f_3 and f_T mean empiric and theoretical frequencies, respectively.

Then a boundary of the right critical region $x_{np,\alpha}^{kp}$ is found and falling of χ^2 within $(x_{np,\alpha}^{kp}, +\infty)$ interval is checked. If $\chi^2 \in (x_{np,\alpha}^{kp}, +\infty)$, then the hypothesis of normality of NPV distribution is rejected at $\alpha=0,01$ confidence level (or 99% reliability level); otherwise, it is accepted.

If normality of NPV distribution is determined in a statistically significant way, then all subsequent processing of experiment results may be arranged on the basis of such fact. If the above hypothesis is not confirmed, then, on the basis of the frequency distribution, a table graphic presentation of the empiric distribution function is built, and the maximum value of NPV , i.e. NPV modal value is determined.

3.2.4 Model run

The fourth phase includes direct work with model (model run) and analysis of obtained results – statistical characteristics of a random variable. In order to achieve reliability of statistical characteristics of a random variable, tests should be run for sufficiently large number of times (second fundamental principle of simulation-based statistic modeling).

As regards the analyzed model, this principle is complied with, first of all, a sufficiently large size of the model – 5000 random variables per one model run. Second, in order to avoid possible outliers, tests are run for 10 times in the suggested model, i.e. the resultant estimate $P(NPV < 0)$ is obtained on the basis of processing of 50,000 different variations of random variables, which ensures its high statistical reliability.

4. Results

4.1 Computer model

The computer model is developed on the MS Excel platform. The work book with the model includes seven work sheets: "Mathematical model", "Simulation field", "Run results", "5 Runs chart", "Runs dispersion", "Results analysis" and "NPV distribution".

On the "Mathematical model" work sheet, the project's cash flow model is realized, and the values of its deterministic and stochastic parameters are entered. On the "Simulation field" work sheet, external and internal random factor impact model is realized. The statistical processing model is realized on "Run results", "10 Runs chart", "Runs dispersion", "Results analysis" and "NPV distribution" work sheets (Figure 3).

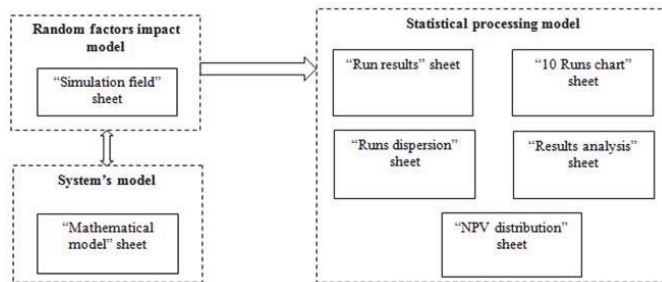


Figure 3. Computer Model's Structure in MS Excel

4.1.1 Modeling results

Let's consider as an example the results of risk simulation modeling in respect of the investment project described by Lukasevich (Lukasevich, I.Y., 1998). Table 4 shows initial parameters of the project.

Table 4. Project's parameters

Deterministic parameters of the project		Stochastic parameters of the project		
Parameter	Value	Parameter	Projected min value	Projected max value
Depreciation A , conventional units	300.00	Production volume Q , conventional units	150	300
Profit tax T , %	20%	Per unit price P , conventional units	40.00	55.00
Discount rate r , %	10%	Variable costs per product unit VC_1 , conventional units	20.00	35.00
Project period n , years	3	Fixed costs FC , conventional units	700.00	800.00
Initial investments, I_0	3,000.00			

According to the modeling results, $\overline{NPV} \approx 4600$, a $P(NPV < 0) \approx 7\%$.

In order to check the hypothesis on conformity of empiric distribution with normal distribution, the quantity of

intervals is calculated using the Sturges formula (19), and their range is calculated using formula (20).

The hypothesis on conformity of empiric distribution with normal distribution is checked on the basis of Pearson's goodness-of-fit test χ^2 using formula (21).

Then a boundary of the right critical region $x_{np,\alpha}^{kp}$ is found by using CHIINV function, and falling of χ^2 within $(x_{np,\alpha}^{kp}, +\infty)$ interval is checked. In the analyzed example, $\chi^2=144.10$; $x_{np,\alpha}^{kp}=23.21$ for confidence level $\alpha=0,01$ и $k=10$. Since $144.10 \in (23.21; +\infty)$, the hypothesis on normal distribution of *NPV* is rejected.

Further analysis based on construction of the graph of the empiric function of *NPV* distribution density function shows that distribution of *NPV* has right-sided asymmetry; therefore, the most probable (modal) value of *NPV* is less than mean value – approx. 3,700 (Figure 4).

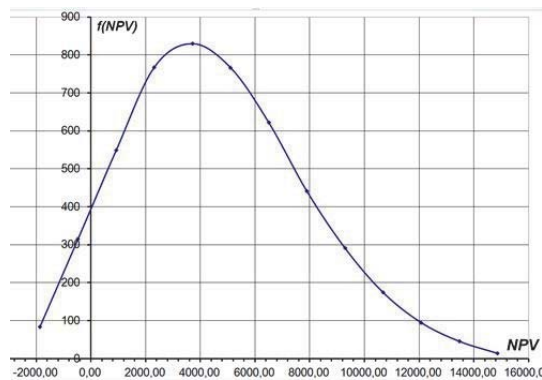


Figure 4. Density of *NPV* distribution in simulation experiment

Therefore, the mean value of *NPV* for the analyzed project is too optimistic, so further analysis should rely on the modal value of *NPV*.

In order to analyze specific parameters of the project, we may use the filter on "Simulation field" sheet within the developed computer model. For example, when selecting projects with *NPV* close to the modal value (in the analyzed example, $3690 \leq NPV \leq 3710$), we would reveal about 10-15 projects. Some of the selected projects may be excluded from further consideration as the projects which do not fully conforming with certain parameters.

5. Discussion

Depending on the applied initial data and selected distribution laws, certain situations may occur when modeling shows that *NPV* is subject to the distribution law close enough to normal distribution (approx. 30% according to model experiments statistics). Since normal distribution density has symmetrical form, the modal value of *NPV* would be equal to its mean value. Therefore, it would be reasonable to check the empiric distribution of *NPV* for normality.

In order to carry out such check, the simulation field (being a key element in external and internal random factor impact model) is used. As a result of one model run, 5,000 *NPV* values are generated simultaneously in the simulation field, which allows to make arrangements for checking of empiric distribution for normality with high level of reliability (in the analyzed model, 99% reliability level was used).

Transition from the mean to the modal value of *NPV* is conditional upon economic considerations regarding the unique nature of investment projects, which is also evidenced by the established practice of investment projects' development, requiring individual calculations for each project. In most model experiments, the modal value was, generally, lower than the mean value, as the empiric distribution function had right-sided asymmetry. Therefore, we may conclude that the obtained mean *NPV* is too optimistic, and it would be reasonable to take the modal value of *NPV* as its projected value.

Study of relationships between the key project parameters represents the crucial phase of analysis of the simulation experiment's results. Correlated variables may result in incorrect results. Therefore, correlation ratios were calculated for *Q*, *P*, *FC* and *VC*₁ variables, and they were found to be quite close to zero, prompting reasonable

assumption regarding independence of the model's key variables. It should be noticed that the correlation ratio's proximity to zero points out to absence of a linear connection between the analyzed variables, but leaves open the possibility of non-linear dependence. The value of the correlation analysis lies also in the possibility to reveal incorrect initial data. In particular, absence of the relationship between price P and per unit costs VC_1 (correlation ratio between the above values is equal to 0.001) in the analyzed example requires additional explanations, because the price should increase with costs escalation. Therefore, the determined range of changes in variable costs requires additional check and, possibly, adjustment.

6. Conclusion

Simulation-based modeling is a numerical method of computer experiments with mathematical models, which simulate behavior of real objects, processes and systems.

When complex systems exposed to random disturbances are studied, probabilistic simulation models are used where the impact of random factors is taken into account by setting probabilistic characteristics of random processes (probability distribution laws, spectral densities or correlation functions). The results, obtained from reproduction of the analyzed process in the simulation model, are random realizations. Therefore, in order to find objective and stable characteristics of the process, it should be reproduced many times with further statistical processing of the obtained data.

Development of instrumental software environments, which support the processes of preparation, translation and execution of simulation experiments, is an important factor of successful application of simulation-based statistical modeling. Advanced simulation-based modeling systems should meet the following requirements:

- 1) active interaction between the user and the model;
- 2) simple language of model description available for non-programmer user;
- 3) possibility of prompt modification of the model and customization for the user's task;
- 4) interactive control over the simulation process;
- 5) availability of an archive with typical models;
- 6) reproduction of the simulated process' and simulation results' dynamics in a form convenient for apprehension;
- 7) use of artificial intelligence methods for arrangement of conclusions.

We expect that development of the systems meeting the above requirements would allow to reduce labor costs of simulation-based modeling due to intellectualization of typical model procedures.

7. Acknowledgements

This research project was supported by the Russian Scientific Fund (RSF Project No. 14-18-00519).

References

- Bird, G.A. (1976) *Molecular Gas Dynamics*. London: Oxford University Press.
- Brandimarte, P. (2014) *Handbook in Monte Carlo Simulation: Applications in Financial Engineering, Risk Management, and Economics*. Wiley.
- Buslenko, N.P. (1970) *Metod statisticheskogo modelirovaniya [Method of Statistical Modeling]*. Moscow: Statistika [in Russian].
- Chan, V. (2013) *Theory and Applications of Monte Carlo Simulations*. InTech.
- Davey, K.J. (2014) *Building Winning Algorithmic Trading Systems: A Trader's Journey from Data Mining to Monte Carlo Simulation to Live Trading*. Wiley.
- Dick, J., Kuo, F.Y., Peters, G.W., & Sloan, I.H. (2013) *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer.
- Dunn, W.J., & Shultis, K. (2011) *Exploring Monte Carlo Methods*. Elsevier.
- Dzhekel, P. (2004) *Primenenie metodov Monte-Karlo v finansax [Application of Monte Carlo Methods in Finance]*. Moscow: Internet trejding [in Russian].
- Ermakov, S.M. (1975) *Metod Monte-Karlo i smezhnye voprosy [Monte Carlo Method and Related Matters]*. Moscow: Nauka [in Russian].
- Fishman, G.S. (1996) *Monte Carlo Concept, Algorithms and applications*. Springer.
- Gerald, H., & Shapiro, S. (1967) *Statistical Models in Engineering*. John Wiley & Sons, Inc., New York.
- Glasserman, P. (2003) *Monte Carlo Methods in Financial Engineering*. Springer-Verlag.
- Halton, J.H. (2006) Sequential Monte Carlo techniques for solving non-linear systems. *Monte Carlo Methods And Applications*. 2006, Vol. 12, 2, 113–141.
- Kashyap R.A., & Rao, A.R. (1983) *Postroenie dinamichnyx stohasticheskikh modelej po eksperimentalnym dannym [Construction of Dynamic Stochastic Models Based on Experimental Data]*. Moscow: Nauka [in Russian].

- Kovalev, V.V. (2008) *Kurs finansovogo menedzhmenta [Course of Financial Management]*. Moscow: TK Velbi [in Russian].
- Kroese, D.P., Taimre, T., & Botev, Z.I. (2011) *Handbook of Monte Carlo Methods*. Wiley.
- Landau, D.P., & Binder, K. (2005) *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press.
- Ledder, G. (2013) *Mathematics for the Life Sciences: Calculus, Modeling, Probability, and Dynamical Systems*. Springer.
- Lipsic, I.V., & Kossov V.V. (2011) *Investitsionnyj analiz. Podgotovka i oценка investitsij v realnye aktivy [Investment Analysis. Preparation and Evaluation of Investments in Real Assets]*. Moscow: INFRA-M [in Russian].
- Lukasevich, I.Y. (1998) *Analiz finansovykh operatsij. Metody, modeli, tekhnika vychislenij [Analysis of Financial Transactions. Methods, Models, Computational Technique]*. Moscow: Finansy, YuNITI [in Russian].
- Melnikov, A.V., Popova, N.V., & Skorniyakova, B.C. (2006) *Matematicheskie metody finansovogo analiza [Mathematical Methods of Financial Analysis]*. Moscow: Ankil [in Russian].
- Metropolis, N., & Ulam S. (1949) The Monte Carlo method. *J. Amer. Assoc.*, 1949, v. 44, 247, 335–341.
- Moskvin, V.A. (2004) *Upravlenie riskami pri realizatsii investitsionnykh proektov [Risk Management in the Implementation of Investment Projects]*. Moscow: Finansy i statistika [in Russian].
- Ostrovskaya, E. (2004) *Risk investitsionnykh proektov [Risks of Investment Projects]*. Moscow: Ekonomika [in Russian].
- Shonkwiler, R.W. (2013) *Finance with Monte Carlo*. Springer.
- Sobol, I.M. (1994) *A Primer for the Monte Carlo Method*. CRC Press.
- Spanier, J., & Gelbard, E. (1969) *Monte Carlo Principles and Neutron Transport Problems*. Addison-Wesley, Reading.
- Taha, H. (2010) *Operations Research: An Introduction*; 9th Edition, Printice Hall.
- Thomopoulos, N.T. (2014) *Essentials of Monte Carlo Simulation: Statistical Methods for Building Simulation Models*. Springer.
- Trofimec, V.Y., & Mamatova, L.A. (2007) *Kompyuternoe modelirovanie ekonomicheskix sistem i processov [Computer Modeling of Economic Systems and Processes]*. Yaroslavl: YarGU [in Russian].
- Vilenskij, P.L., Livshic, V.N., & Smolyak S.A. (2002) *Oценка effektivnosti investitsionnykh proektov. Teoriya i praktika [Evaluating the Effectiveness of Investment Projects. Theory and Practice]*. Moscow: Delo [in Russian].
- Watsham, T., & Parramore, K. (1997) *Quantitative Methods in Finance*. International Thomson Business Press.
- Winston, W. (1998). Financial Models Using Simulation and Optimization. *Palisade Corporation, Newfield, NY*, 107–112.
- Zio, E. (2013) *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*. Springer.