Modeling Risks on Licensed Markets:
on the Example of the Russian Alcohol Market

Olga M. Pisareva[1] and Anna I. Denisova[2]

State University of Management, Moscow, Russia
[1] o.m.pisareva@gmail.com, [2] a.i.denisova@inbox.ru

Abstract. The methodology of mathematical and computer modeling of risks of licensed commodity markets of the Russian Federation on the example of the alcohol market is presented in the article. The illustrative example shows the possibility of assessing the probability of a risky event using the logistic regression model. The methodology for the assessment of the expected damage cost from non-payment of taxes and fees on illegally circulating products to the budget is considered. A simulation approach to risk modeling and its implementation in relation to the alcohol market is presented.

Keywords: decision making, mathematical modeling, risk modeling, logistic regression, risk prediction, alcohol market

1 Introduction

Traditionally markets associated with the production and sale of potentially hazardous products (for example, tobacco market, alcohol market, pharma market and so on) are the sources of close attention and concern on the part of state control bodies. Negative events in these markets can lead to damage to the national economy, life and health of people, the environment, etc. Therefore, it is important to assess the risks of such events in a timely manner in order to prevent or reduce their dramatic consequences.

One of the forms of state regulation of such "potentially dangerous" markets is licensing. In the Russian Federation, this is fixed in Federal Law No 99-FZ of 04.05.2011 "On licensing of certain types of activities". The specific state control bodies work in such markets involves the use of the so-called risk-based management model. This is due to the need to timely identify the possibility of the emergence of a risk source and assess its possible negative consequences.
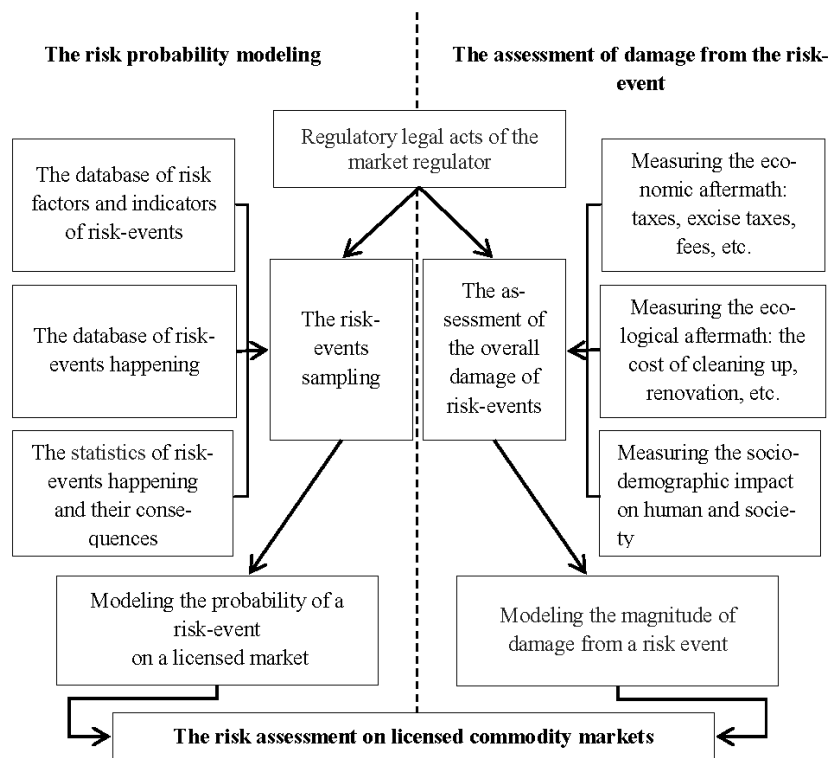
The problems of identifying, assessing and managing risks have been the subject of a large number of studies by such well-known scientists as Knight F. [5], Senchagov V. [12], Purdy G. [11], Kachalov R. [4] and others. The methodology of risk assessment and its consequences on the basis of mathematical and computer modeling methods is considered in the works of such researchers as Ayvazyan S. [2,3], Varshavskiy A. [13], Makarov V. and Bakhtizin A. [7], Mkhitaryan V. [3,8], etc. We used the general methodological approaches presented in the works of these scientists to model risks in licensed commodity markets [10].

We will present the experience of constructing a model for assessing the risks of shortfall in the consolidated state budget of funds from taxes and other charges that must be paid for illegally circulating products on the example of the alcohol market. The calculations were based on the open data from the Federal Service of State Statistics of the Russian Federation (www.gks.ru), Treasury of the Russian Federation (www.roskazna.ru) and Federal Service for Alcohol Market Regulation (www.fsrar.ru, AMR).

In accordance with ISO 31000 [1] international standard, it is customary to determine the risk (R) as a combination of the likelihood of occurrence of a risk event (P) and its associated consequences (C):

$$R = P \times C \tag{1}$$

In accordance with this, we will present two stages of the model implementation: modeling the probability of a risk event occurrence and estimating the amount of damage from its occurrence, which is reflected in general in Fig. 1.

The risk probability modeling

The assessment of damage from the risk-event

Regulatory legal acts of the market regulator

The database of risk factors and indicators of risk-events

The database of risk-events happening

The statistics of risk-events happening and their consequences

The risk-events sampling

The assessment of the overall damage of risk-events

Measuring the economic aftermath: taxes, excise taxes, fees, etc.

Measuring the ecological aftermath: the cost of cleaning up, renovation, etc.

Measuring the socio-demographic impact on human and society

Modeling the probability of a risk-event on a licensed market

Modeling the magnitude of damage from a risk event

The risk assessment on licensed commodity markets

Fig. 1. The scheme of the risk assessment on licensed commodity markets

2 Description of the risk probability modeling

By "risky events" (risk-events) here we mean, first of all, the fixed offenses of the current regulatory legal acts (RLA). Risk-indicators (risk-factors, signs of risk) are facts that in general do not indicate the realization of a risk-event, but because of their presence or absence, especially in combination, it becomes possible to judge the expectation of a risk event.

Suppose we can establish a relationship between the risky event and the set of attendant risk-indicators in the form of mathematical model, enabling to measure the risk probability according to risk signs.

Let there be the sampling of factors, presented as the matrix of independent variables $X_{n \times m}$ and the result vector $Y_{n \times 1}$, n – the sample size, m – number of independent variables. Let Y be a discrete binary variable: value "1" means an event happening, and value "0" means the contrary – non-occurrence of event. Values of risk-indicators (independent variables) can be both discrete and continuous.

Then the source data can be represented as follows:

$$Y = (y_i); \quad y_i = \begin{cases} 1; \\ 0; \end{cases}$$

$X = (x_{ij})$ – matrix of risk-factors, here x_{ij} is the j-th indicator value of i-th object, $i = 1, \ldots, n$; $j = 1, \ldots, m$.

In general, the task is to estimate the probability of the state of the system on the basis of a set of external factors:

$$P = F(X, Y). \tag{2}$$

Here P is the estimated probability of event happening; $F(X, Y)$ is the function of estimation; X is the risk characteristics matrix and Y is the events vector.

The next necessary step is to collect statistical information about the occurrence of risk events and the impact of the accompanying indicators on them, as well as the choice of the dominant indicators (factors) among them.
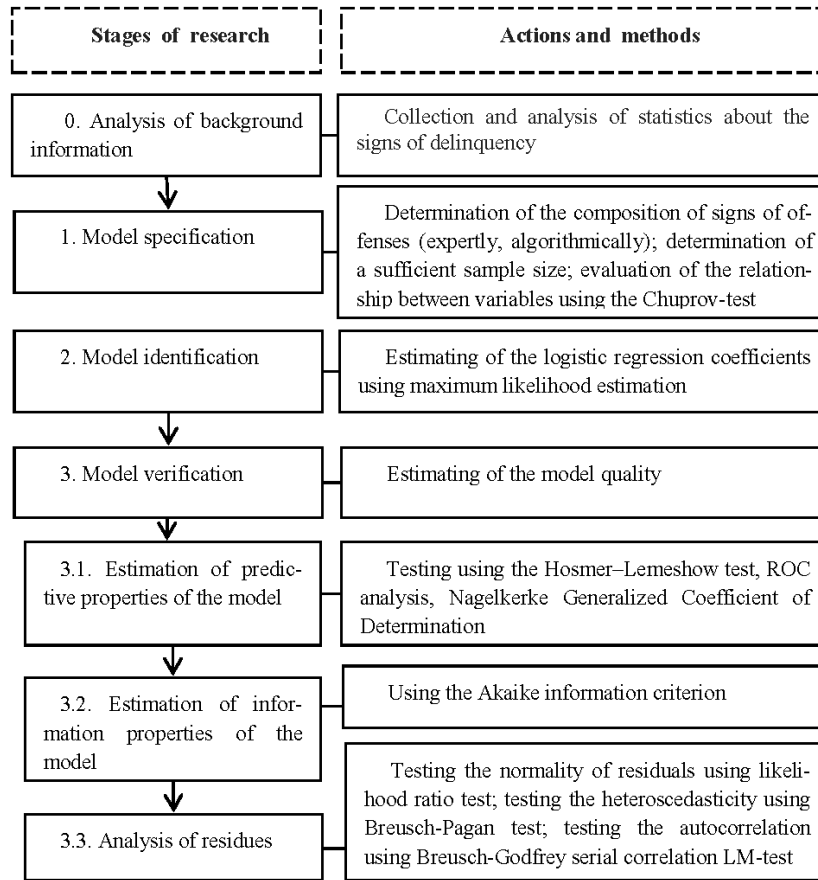
Taking into account the nature of the initial data and the experience of previous studies [2, 3, 7–9, 13], we can designate the following groups of methods used in risk modeling: artificial intelligence methods (neural modeling, logical-probabilistic method, etc.), methods of operations research (linear programming, nonlinear optimization, etc.), econometric modeling and machine learning methods (decision tree, regression of binary choice, analysis of discriminant functions, etc.).

In practice, the binary logistic regression method has proved to be the most satisfactory in terms of accuracy and ease of implementation. Here the probability of Y can be estimated as:

$$P = F(z) = \frac{e^z}{1 + e^z}, \tag{3}$$

where $z = b_0 + b_1 x_1 + \cdots + b_n x_n$; $b_0, \ldots, \beta_m$ – the regression parameters; x_0, $\ldots, x_m$ – indicators of risk-event (cut-off value is not less than 0.75).

The coefficients of the model (3) are estimated using the maximum likelihood estimator. The assumption of the lack of normality in the residues distribution using MLE is an advantage of logistic regression either. The scheme for logistic regression constructing is shown in Fig. 2. As an illustrative example of risk factors there are 8 unlawful act indicators as shown in Table 1. Verification of the adequacy of the results of model calculations to actual observations was carried out according to the data 2016.

Stages of research	Actions and methods
0. Analysis of background information	Collection and analysis of statistics about the signs of delinquency
1. Model specification	Determination of the composition of signs of offenses (expertly, algorithmically); determination of a sufficient sample size; evaluation of the relationship between variables using the Chuprov-test
2. Model identification	Estimating of the logistic regression coefficients using maximum likelihood estimation
3. Model verification	Estimating of the model quality
3.1. Estimation of predictive properties of the model	Testing using the Hosmer–Lemeshow test, ROC analysis, Nagelkerke Generalized Coefficient of Determination
3.2. Estimation of information properties of the model	Using the Akaike information criterion
3.3. Analysis of residues	Testing the normality of residuals using likelihood ratio test; testing the heteroscedasticity using Breusch-Pagan test; testing the autocorrelation using Breusch-Godfrey serial correlation LM-test

Fig. 2. The relationship of the stages and methods of logistic regression evaluation

By forming a representative, reliable and statistically stable sample for building a model, it is necessary to know the ratio of the number of "good" and the number of "bad" outcomes in the general population. The minimum allowable sample size (n), based on the known ratio of the number of realized and unre-

Table 1. Example of a list of independent variables (risk-indicators)

No	Variable	Interpretation
1	x_1	Returns to the supplier of large volumes of products (including repeatedly)
2	x_2	The names and brands of products that were noticed in illegal traffic earlier
3	x_3	Absence of GPS data on transportation of products
4	x_4	Excess of volumes of wholesale and retail turnover over its production
5	x_5	The organization being inspected had a violation earlier
6	x_6	The discrepancy between the volumes of purchased and used raw materials with the volumes of products produced
7	x_7	The presence of visual signs of non-compliance of products with state standards
8	x_8	Visually determined signs of forgery of documents

alized events, on the number of "bad" outcomes, was found on the basis of the ratio [3]:

$$n = \frac{z_\gamma^2 \times q \times (1 - q)}{\alpha^2}. \tag{4}$$

Here q is share of "bad" outcomes, α is maximum reasonable error of share estimate, z_γ is the value of standardized normal distribution in reliability level γ.

According to AMR statistics, the share of established unlawful acts inspections was 86%, that is $q = 0.86$. If $\gamma = 0.95$; $z_\gamma = 1.96$; $\alpha = 0.05$ (hereinafter), the minimal allowable sample size is 186 ($n = \frac{1.96^2 \times 0.86 \times (1-0.86)}{0.05^2} = 185.01$).

The choice of a set of the most important risk indicators was carried out in two stages. At first, experts reduced the number of risk factors from 252 to 8. The Delphi procedure was used here, three of its rounds were conducted. In this procedure, a group of regional representatives of the Federal Service for Alcohol Market Regulation participated. Each of them had the same official powers. Secondly, we additionally evaluated the relationship between the factors and the result variable.

Since the estimation of the ratio of non-numerical variables did not give clear results, it was decided to consider their different combinations. Due to a large number of factors, the task becomes computationally complex. For this reason, a software algorithm was implemented in the programming language R. It allowed formulating a list of models with significant coefficients by the method of automatic recalculation of various combinations of independent variables X.

For each generated combination, a binary logistic model was constructed and the significance of its coefficients was estimated. The observed insignificant coefficients were consistently eliminated. Then in a new iteration of the cycle a new set of factors was generated again. As a result of the program, the list of 5 models containing more than three significant coefficients was formed. Number

of iterations is $k = C_8^8 + C_8^7 + C_8^6 + C_8^5 = 93$. We limited the combinations of up to five in order to avoid unnecessary recounting. This approach is quite flexible and does not limit the specification of the model. We can also use the possibilities of computer modeling.

Based on a comparison of their quality characteristics, a choice was made in favor of the model:

$$P\{y = 1 \mid x\} = \frac{e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6}}{1 + e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6}}, \tag{5}$$

here $z = 1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6$, the coefficients are significant.
$\quad\quad\quad\;\;(0.012)\quad\;\;(0.003)\quad\;\;(0.002)\quad\;\;(0.004)$

The model (5) has minimum value of the Akaike information criterion ($AIC = 139.3$); maximum value of Nagelkerke coefficint of determination ($R_{Ng}^2 = 0.31$); as Hosmer–Lemeshow test $[\chi^2(\alpha, n + 1) = \chi^2(0.05, 5) = 11.07] > [HL = 4.592]$, $p(HL) = 0.71$ is value high enough, indicating that the model values meet the observed. That means the model (5) is well-calibrated. The area under the ROC curve (Receiver Operating Characteristic curve, shown in Fig. 3) is 0.82, so the probability of a correct model definition is rather high. The testing residuals on the normality of the distribution by the likelihood ratio criterion and on heteroscedasticity by the Breusch–Pagan test did not reveal anomalies that forbade the use of the model (5). In general, the results obtained are characteristic for models of this type. The model has a high level of quality, which is effective to solve practical problems.
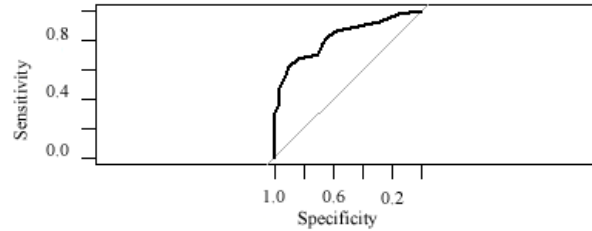
Fig. 3. ROC curve of model (5)

Further, the limiting effects of risk factors were evaluated. These effects mean change in the probability when the factor changes by one and signal change in the uncertainty of the binary choice situation. In the logistic model, the small change Δx_k of the k-th independent variable lead to the probability change $P\{y = 1 \mid x\}$: $P\{y = 1 \mid x\} \cong \frac{\partial P\{y=1|x\}}{\partial x_k}\Delta x_k = \frac{\partial F(x^T\beta)}{\partial x_k}\Delta x_k = \frac{\beta_k e^z}{(1+e^z)^2}\Delta x_k$. Here $F(\cdot)$ is the logistic model (3), $z = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$, x_i, ..., x_m are the values of m independent variables, β_0, β_1, ..., β_m are the coefficients of regression [9].

If $G = \frac{e^z}{(1+e^z)^2} = \frac{e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6}}{(1+e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6})^2}$ the probability changes will be when the new indicator is identified: $P\{y = 1 \mid x\} = 1.19 \times G \times \Delta x_1$;

$P\{y = 1 \mid x\} = 1.41 \times G \times \Delta x_4$; $P\{y = 1 \mid x\} = 1.42 \times G \times \Delta x_5$; $P\{y = 1 \mid x\} = 1.36 \times G \times \Delta x_6$.

The x_5 and x_4 factors have approximately the greatest effect on the result variable, and are interpreted as "The organization being inspected had a violation earlier" and "Excess of volumes of wholesale and retail turnover over its production". Variable x_1 has the smallest effect.

3 An Assessment of the cost of damage and modeling of risks

It was decided, as a characteristic of the consequences of direct economic damage (directly affecting the economy) to take an estimation of the amount of taxes and charges unpaid to the state budget because of illegally circulating products in the licensed market. A similar approach to assessing risk as underfunding was mentioned in [6]. Note that the transfer from rubles to dollars is based on the weighted average rate for 2016: 1 dollar was equal 67.04 rubles.

So the total damage from non-payment of excises taxes and VAT for the year (C, dollars) will be equal to:

$$C = 0.18 \times S_0 \times Q + S_0 \times A_0, \tag{6}$$

here S_0 – annual volume of unaccounted products in total sales volume (dekaliters, dl); A_0 – averaged excise rate calculated on the basis of official rates specified in the Tax Code for all types of market products:

$$A_0 = \sum_{j=1}^{l} k_j a_j g_j, \tag{7}$$

a_j – excise rate for the j-th type of products; l – count of type of products; g_j – object of taxation (for example, excise on alcohol is paid for the amount of anhydrous ethyl alcohol in the product); k_j – market share of the j-th type of product; Q – average price per unit of sold products (dollar):

$$Q = \sum_{j=1}^{l} k_j q_j$$

q_j – price for the j-th type of products (dollar/liter);

$$S_0 = \frac{\Delta A \times S^2}{\Delta A \times S + B}, \tag{8}$$

here the total volume of product S (dl) is the sum of the volume of product with which taxes are paid (S_1) and the volume of unaccounted product (S_0): $S = S_1 + S_0$; B – amount of excise taxes paid to the budget for the year (dollar). In addition, we can note that

$$A_0 = \frac{B}{S_1}.$$

Harmonizing the units of measurement, we get the difference between A_0 and averaged "actual" excise rate for all types of market products:

$$\Delta A = \frac{B}{S_1} - \frac{B}{(S_1 + S_0)}.$$

Dividing C by the number of risk-events (N) detected in the year we get the average annual amount of damage from one risky event:

$$\bar{C} = \frac{C}{N}.$$

However, one cannot expect a constant value of damage in the event of any risk event and nor should it be expected that $\bar{C}$ will be the maximum value of the damage. Suppose, that $\bar{C}$ is the value corresponding to the middle of the interval at which the amount of damage varies randomly from event to event. Hence, we can suppose that $2 \times \bar{C}$ is a maximum value of probable damage.

Obviously, the damage cannot be negative, in the case of the equality 0 – it is about the absence of damage. So, for example, let C be a random variable varying in accordance with the uniform distribution law in the interval $[0; 2 \times \bar{C}]$: $C = R[0; 1] \times 2 \times \bar{C}$. Here $R[0; 1]$ is a random uniformly distributed quantity on the interval $[0; 1]$. We use this quantity to generate pseudo random numbers in the process of computer simulation. We also considered $2 \times \bar{C}$ as the upper bound to take into account the possible variations in the amount of damage not related to the budget risk.

Then as a method for predicting the magnitude of the risk, we use computer simulation based on the Monte–Carlo method: we perform a large number of simulations (N_{MC}) for random variable values. As a result, we have not the only value of the magnitude of the risk, but its probability distribution.

Finally, we estimate the risk value in accordance with (1) and taking into account the described methodology:

$$R = \frac{e^z}{1 + e^z} \times C. \tag{9}$$

In addition, based on the obtained values, it is possible to calculate some characteristics: the expected value of risk

$$\bar{R} = \frac{\sum_{i=1}^{N_{MC}} R_i}{N_{MC}};$$

the variance and standard deviation of risk

$$S^2 = \frac{\sum_{i=1}^{N_{MC}} (R_i - \bar{R})^2}{N_{MC} - 1}, \quad s = \sqrt{S^2};$$

the coefficient of risk variation

$$cv = \frac{s}{R};$$

here N_{MC} – number of simulated Monte–Carlo runs, R_i – risk value in the i-th run.

4 Results of computer modeling of the risk value

Let us illustrate the possibility of applying the proposed methodology for risk modeling on licensed commodity markets by the example of the taxes and fees non-payment for illegally circulating products in the Russian alcohol market.

According to the Federal State Statistics Service, the volume of alcohol sales (S) in 2016 was 974.5 million decaliters (here 106.9 million dl containing more than 9% alcohol; 8.8 million dl with content less than 9%; 51.9 million dl wine, excluding sparkling wines; 22.0 million dl sparkling wines; 780.6 million dl beer; 4.3 million dl other). The corresponding shares of the retail sales market by kinds of said alcohol products are: 1) containing more than 9% alcohol – 0.11; 2) with content less than 9% – 0.009; 3) wine, excluding sparkling wines – 0.053; 4) sparkling wines – 0.023; 5) beer – 0.801; 6) other (in particular, various fruit wines) – 0.04.

The Tax Code (came into force in 2016) specifies such excise rates: 1) 500 rubles per liter of anhydrous ethyl alcohol; 2) 400 rubles per liter of anhydrous ethyl alcohol; 3) 9 rubles per liter of products; 4) 26 rubles/liter; 5) 20 rubles per liter; 6) 9 rubles per liter.

Then the average value of the excise rate on the basis of official rates approved in the Tax Code and taking into account the market conditions calculated by the formula (8) is $A_0 = 0.63$ dollar. The amount of budget revenues from excise duties on alcoholic products is 5.2 billion dollars.

Proceeding from (8) it is possible to estimate the volume of illegal products in the total volume of the sold goods: $S_0 = 1447.5$ million liters. So the damage from non-payment of excises is $S_0 \times A_0 = 0.91$ billion dollars per year. And the share of illegal products in the volume sold is $1447.5/9745 = 0.149$.

According to calculations, the weighted average price (Q) for alcohol products calculated on the basis of data on average prices for different types of alcoholic beverages (data of the Federal State Statistics Service) amounted to 3.15 dollar per liter in 2016.

Then, the damage from VAT non-payment (18%) is $0.18 \times S_0 \times Q = 0.82$ billion dollar. Based on (6) we can estimate the total damage in 2016: $C = 0.18 \times S_0 \times Q + S_0 \times A_0 = 1.73$ billion dollars.

The rounded average number of violations found on the alcohol market (related to non-compliance with tax, customs legislation, with the sale of products with counterfeit excise or special marks or unmarked, violation of license conditions and illegal manufacture at legal enterprises) was $N = 22,200$ for the last 5 years (by Federal Service of State Statistics). So, we can estimate the average "cost" of one risk event taking n as the number of risky events: $\bar{C} = \frac{1.73}{22200} = 0.078$ million dollars.

Finally, using formula (9), we obtain a model for estimating the magnitude of the risk:

$$R = P \times C = \frac{e^z}{1 + e^z} \times C = \frac{e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6}}{1 + e^{1.19x_1 + 1.41x_4 + 1.42x_5 + 1.36x_6}} \times R[0; 1] \times 2 \times 0.078.$$

Then a simulation model for estimating the value of the risk was implemented using the Monte–Carlo method. 22,200 tests were conducted in accordance with the average number of violations in the alcohol market over the last 5 years. The fragment of the results is shown in Table 2.

Table 2. Fragment of computational values of runs of the simulation model of risk assessment

No	x_1	x_4	x_5	x_6	P	C, million dollar	R, million dollar
1	1	0	1	0	0.931502	0.034834	0.032448
2	0	1	0	0	0.803766	0.132731	0.106685
3	0	0	1	1	0.941585	0.123103	0.115912
4	1	1	1	1	0.995413	0.027095	0.026971
5	1	0	1	1	0.981476	0.089822	0.088158
...							

Based on the results of the experiments, the maximum value of the risk of one unlawful act is 0.1552 million dollar. The expected value of the risk is 0.0697 million dollar, the coefficient of variation is $cv = 0.59$, which indicates about the average probability of happening of a risk event.

It is useful to assess the probability of a risky event happaning with a sufficiently high level of reliability, for example, in 80%. We divide the number of cases in which the probability of unlawful act was above 0.8, by the total number of runs: $P = P_{0.8}/N_{MC} = 0.81$ (sufficiently large value in practical problems). Note that this value is pretty close to 0.86, the specified AMR in 2016, as the proportion of checks in which violations were found.

5   Conclusion

We note that the calculated share of illegal products among those sold on the market (equal to 0.149) corresponds to the official data during the last years submitted by the AMR. According to their information, the average share of illegal products among those audited in 2013, 2014, 2015, 2016 is estimated at 0.23. The discrepancy is explained by the fact that some illegal products can be identified before the sales stage, and because the fact of sale was not also reflected in the documents. In addition, we should expect improved modeling results due to the identification and accumulation of additional information on inspections of market participants. The assessment of arrears for excise duties for wine, beer, and products with alcoholic content above and below 9% totaled 66.102 billion rubles (0.986 billion dollar) (by Report on taxes and tees, fines and tax sanctions in 2016, the Federal Tax Service of the Russian Federation). The amount of non-payments calculated in the work amounted to 0.91 billion dollar. Such a discrepancy is primarily because the report does not include arrears

for all types of products, and the averaging of some indicators present in the calculations may affect accuracy.

The presented arguments of checking the quality of the event risks modeling in the markets of licensed products testify to the possibility and the need for further improvement in order to increase the possibility of using simulation results in the practice of managing markets for potentially hazardous products.

References

1. Risk management – risk assessment techniques, iSO/IEC 31010:2009
2. Ayvazyan, S., all: Modeling risk patterns of russian systemically important financial institutions. Review of Applied Socio-Economic Research 1(1), 70–80 (2011)
3. Ayvazyan, S., Mkhitaryan, V.: Applied Statistics. Fundamentals of Econometrics. Unity, Moscow (2001)
4. Kachalov, R.: Economical Risk Management: theory and applications. Nestor–Historia, Moscow, Saint Petersburg (2012)
5. Knight, F.: Risk, Uncertainty and Profit. Hart, Schaffner & Marx; Houghton Mifflin Co., Boston, MA (1921)
6. Kovaleva, T.: Organization of budget management in the subject of the russian federation. Financy and Credit 5 (2003)
7. Makarov, V., Bakhtizin, A.: Social modeling – new computer breakthrough (agent-oriented models). Economika, Moscow (2013)
8. Mkhitaryan, V., Karelina, M.: Mathematical modeling of integration policy risks based on the method of fuzzy performance. In: Information Technologies in Economics and Management. Dagestan State Technical University. pp. 60–64 (2006)
9. Nosko, V.: Econometrics for Beginners (Additional chapters). IETP, Moscow (2005)
10. Pisareva, O.: Methodological and applied aspects of modeling risks of unlawful events in the field of licensed economic activity. Economics and management: problems, solutions 3(66)(6), 159–167 (2017), annual international round table "System Economics, Social and Economic Cybernetics, Soft Measurements in the Economy – 2017", Moscow, June 8, 2017, Financial University under the Government of the Russian Federation
11. Purdy, G.: Raising the standard – the new iso risk management standard (2009)
12. Senchagov, V.: Economic security: geopolitics, globalization, self-preservation and development. Institute of Economics of the Russian Academy of Sciences. Moscow. ZAO Finstatinform (2002)
13. Varshavskiy, A.: Challenging innovations: risks and responsibilities (on the example of food products of domestic production). CEMI RAS, Moscow (2009)