Valuating the position of the control object based on a universal complex indicator using structured and unstructured data

Tatiana K. Bogdanova (iD)
E-mail: tanbog@hse.ru

Liudmila V. Zhukova (iD)
E-mail: lvzhukova@hse.ru

National Research University Higher School of Economics
Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

Valuating the position of a controlled object using indicators which are management and control tools is widely used in many areas of the economy. Usually such indicators are based on internal data, however, as the volume of available open information grows, algorithms for valuation of the position of certain control objects and on open structured data are appearing. The disadvantage of these models is their narrow specialization and binding only to structured, and sometimes strictly official data, which, as a rule, have a rare publication frequency. This does not allow you to track the change in the position of the object at different times. The authors have proposed a concept for constructing a universal complex indicator (UCI) for express valuation of the position of a controlled object in various types of activity: banking, educational, industrial, etc. Another difference in the construction of the UCI is that the concept presented in the article assumes, as a reference point, to take into account the requirements of regulatory authorities, while in most Russian and foreign studies, indicators are built for the needs of investors. It is also proposed to use, along with structured and unstructured data, tracking the dynamics of changes in the position of the control object. To determine the UCI values on the basis of various econometric models and methods, the components that characterize the requirements of the control bodies to the control object are calculated; using them the UCI value is determined from

the truth table. The concept proposed was tested to build an express valuation of the financial position of 108 banks for the period from 1 January 2018 to 1 February 2020. In accordance with the requirements of the Central Bank of the Russian Federation, the values of the three UCI components were obtained, and the value was calculated for each bank. The predictive ability of the constructed model, tested on three banks of the test sample, was confirmed by the consistency of the express valuation with their actual position in March 2020.

Introduction

In the current economic environment, the role of competent management in the activities of an organization of any level cannot be underestimated. Optimal management saves resources, increases the efficiency of activities, accelerates the repayment of funds, and reduces the risk of financial losses. In recent years there has been a great deal of attention to the ways and methods of management, as well as to the information on the basis of which decisions are made. The more detailed and qualitative the information used, the more accurate the management decisions made on its basis will be. Control does not exist without monitoring and controlling the state of the control object [1]. Quite a lot has been written about the problem of managing economic objects in recent years, both by Russian and foreign researchers [2].

The problems of improving the management of various objects have been touched upon in many scientific publications of the last 10 years. For example, the article [3] is focused on the issues of improving the tools for planning socio-economic development and evaluation of control objects under the conditions of active formation of the information society infrastructure. The author also assessed the impact of digital technologies on the content of control tasks — in the process of variability of control objects and subjects in large-scale organizational systems. There is an underdeveloped level of involvement of big data in management at the tactical level [4]. The author proposes the standardization of technologies for collecting, transmitting, storing and processing big data, the construction of computer mathematical models to provide a multilevel system of state strategic planning [5, 6].

In [7] the author, analyzing the research in the field of management and control, notes that under the conditions of tightening competition, an effective tool for survival and adaptation at present is internal control. At the same time, the poorly developed system of internal control of the majority of today's enterprises is considered. The author draws the conclusion that heads of the majority of firms consider introduction of internal control as expensive and labor-consuming, thus they underestimate the importance of this system.

In [1] the author identifies a number of significant management problems concerning the control functions of the organization. Thus, when carrying out control, the accent of inspection bodies is made, as a rule, only on audit of financial and economic activity; control acti-

vity turns into rechecking of accounting operations, including arithmetic recalculation. At the same time, assessment of external non-quantitative indicators, such as the level of popularity, customer loyalty, the quality of the object of management, etc. remains outside the attention of the control authorities.

From analysis of the methods and models proposed both by Russian and foreign researchers, the advantage relies on the information coming with a certain periodicity, often with a significant lag relative to the current moment. At the same time, the controlling bodies need to obtain a reliable assessment of the state of the object of management for the current moment in time, which will allow taking the necessary measures in time to prevent negative developments.

This problem can be solved by conducting an express-analysis of the current state of the control object based on the use of formalized and non-formalized, structured and unstructured information. The result of express-analysis is to obtain an express valuation of the state of the control object. Express valuation of the state of the control object is a preliminary study, which largely determines the feasibility of further, more in-depth analysis.

Considering the above, the authors of this article set the following research goal and objectives. The aim of the study is to develop a universal complex indicator for express valuation of the compliance of the economic object of management with the stated requirements on the part of regulators or relevant services on the basis of structured and unstructured data from the internet.

The main objectives of the study are:
✦ developing the concept of express valuation of the compliance of the state of the object of management with the stated requirements of current controlling;
✦ development of an algorithm for constructing a universal complex indicator for express valuation of the compliance of the state of the

control object with the stated requirements;
✦ approbation of the concept developed on the example of calculating the universal complex indicator for express valuation of the state of a Russian commercial bank.

The relevance of the study is explained by the following factors:
✦ the need to develop a tool of economic and mathematical modeling based on open data for current monitoring, analysis and forecast of the state of the control object;
✦ the need for a current assessment model, which does not depend on the moment of the release of statistical reports;
✦ conducting an express valuation of the state of the control object (for example, a bank) using open data;
✦ the need to track the dynamics of the state of the control object on the basis of structured and unstructured data.

1. Statement
of the research problem

The proposed approach will provide a solution to the problem of transformation of the existing management and control system and the development of internal control system in the context of digitalization of the economy [7, 8]. As a rule, official statistical reports are published in the public domain with a certain periodicity and an appropriate lag, which makes it difficult to respond promptly in force majeure situations.

An approach to the construction of a complex indicator based on the use of open structured data on the control object at the current moment in time to assess the degree of coastal zone deterioration is proposed in [9]. The work formulates a model for calculating a complex indicator based on structured information collected from public maps and communities about the use by the population of the territory as undeveloped wild beaches, mini hotels or places of public assembly.

One of the problems raised by the authors of this article is the problem of ongoing monitoring of control objects, with official statistical reporting coming out with great frequency and a long time lag. For example, in the banking sphere current control by the Central Bank over other commercial banks is difficult because of the delay in the receipt of information about the state of bank finances. This problem is considered in [10]. We suggest involving open sources of information and, on the basis of a mathematical model, forming a complex indicator using such data [2, 11]. The approach outlined in the article makes it possible to obtain, on the basis of open data, operational information about the state of the bank as an object of management, and to accelerate decision-making in relation to it. The result of the application of the mathematical model is the probability of license revocation (as one of the components of the reliable bank assessment model) on the basis of public financial statements and taking into account the volatility of the external environment (changes in currency exchange rates, the cost of a barrel of oil, etc.) [12–14].

A distinctive feature of this paper is the authors' proposal to use, in addition to structured and unstructured data from "open" sources of information, such as: qualitative and express valuation of the overall condition of the bank from the point of view of independent Russian rating agencies, the tone of published news mentioning the bank.

The unstructured data can be presented in different forms – text, news reports, comments, photos [15]. With the development of internet technologies, their volume is growing, and information based on such data can help in express valuation of the state of the control object.

In contrast to the works discussed above, we offer the concept of building a universal complex indicator of express valuation of the state of the control object on the basis of open structured and unstructured data obtained from official and unofficial sites. The data are collected both from statistical resources, where they are published periodically, and from internet resources regularly updated with different periodicity. In contrast to the structured data used for many years, unstructured data is an amalgamation of disparate information. This allows end-users to monitor the consequences of their decisions at an early stage, thus minimizing risks. Consideration of unstructured data in the evaluation of control objects allows for operational monitoring of the current state of the control object, taking into account qualitative characteristics (reputation, feedback, trends) based on data obtained in a non-contact method from open sources.

For processing by methods of machine learning, mathematical statistics and mathematical models, the unstructured information is pre-structured in the form of a set of some indicators by methods of mathematical statistics, machine learning or econometrics [16–18].

For example, in the conference proceedings [19] we proposed using one of the components calculated on the basis of unstructured data indicating probability of instability of a bank's position (text content of the news highlighting the negative references to the bank's name), as an attribute for constructing a complex indicator. This indicator can be used for evaluating the bank's position that is different from the stable one.

2. The concept of constructing a universal complex indicator for express valuation of the state of the control object

The proposed concept of forming a universal complex indicator (UCI) of express valuation of the state of the control object makes it possible to assess not only the current, but also to predict the future state of such a control object [19–21].

The constructed universal complex indicator (UCI) can be used for the following purposes:

✦ to assess the proximity of the state of the control object to the critical level set by the controlling body;

✦ to analyze trends in the state of the controlled object on the basis of structured and unstructured data;

✦ to collect and analyze operational information about the state of the control object;

✦ to assess the reliability of reporting data on the state of the control object;

✦ to compare the state of different-scale control objects;

✦ to take into account the external in relation to the object of management characteristics of its activity (the presence of actual mini-

hotels on the coast, the tone of the news about the director of the bank, the comfort of the infrastructure of the educational institution, the activity of the use of social networks by the respondent, etc.) [22–24].

The advantage of the use of open data is the possibility to receive information with any periodicity (not being restricted by the regularity of updating of officially published statistical information), to expand and check the correspondence of the actual state of the control object to the official data.

The conceptual scheme of building UCI express valuation of the state of the objectt for monitoring and control to prevent the development of negative events is presented in *Figure 1*.
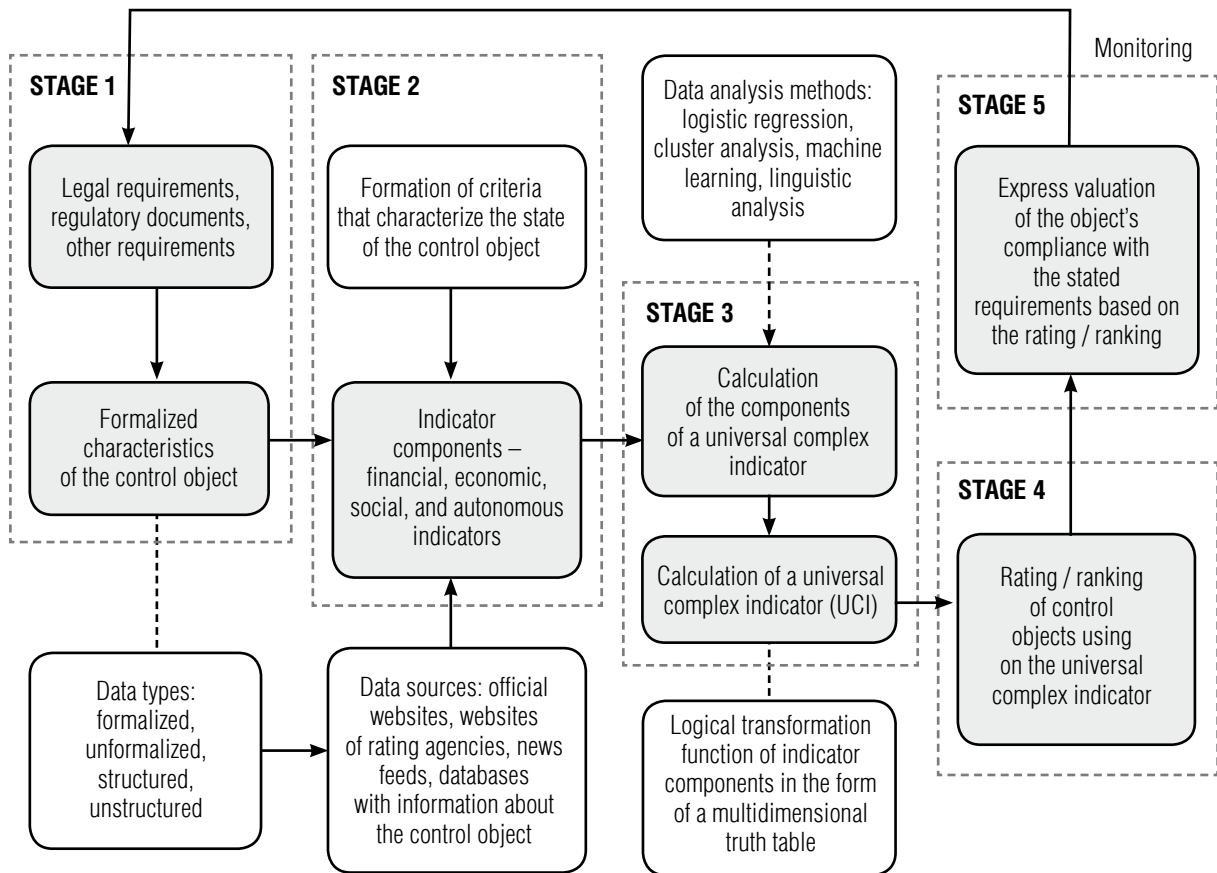

Fig. 1. Conceptual scheme of the construction of a universal complex indicator (UCI)

The construction of a universal complex indicator involves the following stages.

Stage 1. The first (initial) first stage is determining the components of UCI on the basis of information about the requirements from the supervisory body regarding the state of the control object. Such information can be obtained from decrees, orders, laws and other regulatory acts. Such a body may be either an external organization (the Central Bank, the Ministry) or an internal control body of the organization itself.

As a result of implementing the algorithm, a universal complex indicator (UCI) is formed, the values of which can be an integer representing binary or categorical values. These values reflect the state of the control object relative to the satisfactory state determined by the controlling body. So an example of binary values of the universal complex indicator is: 0 – no signs of illegal activities of the organization, 1 – there are signs of illegal activities of the organization. Most often, a categorical scale with three categories is used, for example, where: 1 means stable condition, 2 – unstable condition and 3 – condition requiring closer examination.

Stage 2. After the formation of criteria characterizing the state of the control object, which in their totality should be described by the UCI, the components of the indicator are defined. They make it possible to evaluate the state of various characteristics of the control object: economic state, the probability of an object state other than stable, media activity associated with the object, characteristics of the object's location, financial or economic state forecast of the object, etc. Then a list of UCI components is formed, and data from public sources are identified from which information can be obtained for mathematical modeling and calculation of the components. Unstructured data are pre-processed for their possible use as calculated indicators.

Stage 3. Using mathematical methods (logistic regression, cluster analysis, regression analysis, statistical analysis) based on the available data, the UCI components are calculated. They represent numerical values used to rank all the control objects.

Relying on the obtained UCI component values, the value of the universal complex indicator of express valuation of compliance of the state of the control object with the stated requirements is determined. For this purpose, the component values are substituted as arguments in the logic function that aggregates the various UCI component values.

Let us consider such a logical function. Let n requirements formulated in n criteria be received from the controlling authority and k components I_j, $j = 1, ..., k$, be formed for the universal complex indicator. The values of the components form the set X (as usual, $X \in \mathbf{R}$).

In this case, a particular control object state requirement received from a controlling entity can be described by a few UCI components, i.e. $k \geq n$. The values of the universal complex indicator (UCI) take values from some given discrete set Y, i.e. $y_i \in Y$, where $i = 1, ..., p$.

The logic function $\Psi(I_1, I_2, ..., I_k)$ is specified by the truth table (*Table 1*), based on the importance of the criteria received from the controlling body (*Figure 2*).

Each component I_j can take Q_j values. The values of the component $I_j^{q_j} \in X_j$, where $j = 1, ..., k$; $q_j = 1, ..., Q_j$; $X_j \in X$.

The total number of combinations M of all possible values of all k components is determined as follows:

$$M = \prod_{j=1}^{k} Q_j. \tag{1}$$

In the general case, the table is a matrix of the dimension $M \times k$. The first row of the matrix (combination s_1) is represented by the first values of each of the k components I_j^1. The last row of the matrix is represented by the maximum values of each of the k components $I_j^{(Q_j)}$. The rows between the first and the last rows can be formed in different ways, but so that a com-
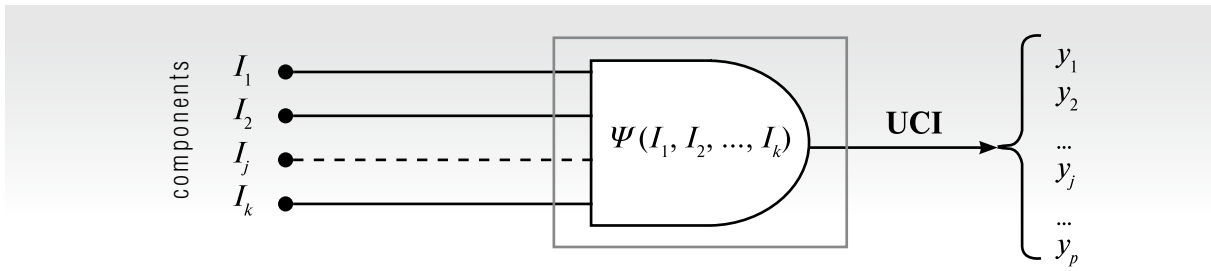
Fig. 2. UCI formation using a logic function

Table 1.

General form of the UCI multidimensional truth table

s_m − combination number, $s_m = 1, ..., M$	Components of the universal complex indicator (UCI)					The value of the universal complex indicator (UCI)
	I_1	...	I_j	...	I_k	
s_1	I_1^1	...	I_j^1	...	I_k^1	$y_{s_1} = \psi(I_1^1, ..., I_j^1, ..., I_k^1);\ y_{s_1} \in Y$
s_2	I_1^2	...	I_j^1	...	I_k^1	$y_{s_2} = \psi(I_1^2, ..., I_j^1, ..., I_k^1);\ y_{s_2} \in Y$
...	...	...	...	...	...	...
s_m	$I_1^{q_1}$	...	$I_j^{q_j}$	...	$I_k^{q_k}$	$y_{s_m} = \psi(I_1^{q_1}, ..., I_j^{q_j}, ..., I_k^{q_k});\ y_{s_m} \in Y$
...	...	...	...	...	...	...
s_M	$I_1^{Q_1}$	...	$I_j^{Q_j}$	...	$I_k^{Q_k}$	$y_{s_M} = \psi(I_1^{Q_1}, ..., I_j^{Q_j}, ..., I_k^{Q_k});\ y_{s_M} \in Y$

plete enumeration of all possible combinations of values taken by each of the I_j components is carried out.

Stage 4. In the next stage, after calculating the UCI value, the ranking of the control objects is performed.

Stage 5. After the ranking stage, an assessment is made of the compliance of the control object with the stated requirements in accord-ance with the obtained ranks (ratings). Based on the values obtained at the moment, rec-ommendations for the supervisory authori-ties regarding the control objects are devel-oped. As new information emerges during the monitoring period for a group of control facilities, data from public sources is updated regularly and the model is re-evaluated on the basis of the updated data.

3. Calculation of the universal complex indicator of express valuation of the commercial bank state

The proposed concept of constructing UCI express valuation of compliance of the state of the control object with the stated requirements on the part of regulators has been tested to obtain an express valuation of the state of a commercial bank. The controlling body for the solution of this problem is the Central Bank of Russia – the supervisory body in the banking sphere.

The construction of the UCI mathematical model of the express valuation of the commercial bank's condition in accordance with the above concept consists of five stages.

Stage 1. Formation of target measures of a bank's state based on requirements for liquidity and reliability indices derived from standards and legislative acts of the Central Bank of the Russian Federation. In accordance with the standards and requirements established by the Central Bank of the Russian Federation for commercial banks, the following characteristics were formed:

✦ probability of financial bankruptcy of the commercial bank;

✦ financial stability of the commercial bank;

✦ qualitative valuation of the commercial bank's activities by experts

✦ market participants.

To assess the formed characteristics, the values of the UCI components are calculated.

Stage 2. Determination of UCI components of the express valuation of the bank's condition:

✦ **probability of financial bankruptcy of a commercial bank (stated by the Central Bank of the Russian Federation).** To determine this characteristic, the component 1 is calculated. This component represents the forecasted probability of a commercial bank bankruptcy; it is estimated by means of logistic regression, using financial and economic factors;

✦ **financial stability of a commercial bank (stated by the Central Bank of the Russian Federation).** To determine this characteristic, the component 2 is calculated. This component is an indirect sign of instability of a commercial bank obtained on the basis of the assessment of the bank's belonging to the "troubled" group. Homogeneous groups of banks on the signs of stability are identified by the clustering method [25];

✦ **qualitative valuation of the commercial bank by experts – market participants (stated by the Central Bank of the Russian Federation).** To determine this characteristic, the component 3 is calculated. This component represents the availability of a negative valuation of the general state of the bank by a market participant – the expert agency "Expert RA". The indicator is constructed using a qualitative valuation based on semantic analysis of news texts mentioning the bank.

Stage 3. Construction of the UCI express valuation of the commercial bank's state. To construct the UCI, we used a training sample of 108 operating commercial banks. Information was obtained from eight external sources including the site of the Central Bank of Russia, rating agencies, reference books and information-analytical portals. The requirements relative to the bank's condition were taken from the norms of the regulator – the Central Bank of the Russian Federation.

To characterize the stability of the financial condition of a commercial bank and its market behavior, the significance of fluctuations in the dynamics of changes in bank rates without visible reasons was assessed. The main indicators characterizing the rates of banks on deposits were taken for the period from January 1, 2018 to February 1, 2020. Ratings agency indicators are from 2017 to 2020. General characteristics

of commercial banks according to the Bank of Russia website and other public sources were taken as of February 2020.

The UCI calculation model used more than 60 initial indicators of the following type:

✦ rates of each of the commercial banks on deposits for different terms;

✦ indicators characterizing the state of the bank: its assets and values of the statutory indicators ($H_1 - H_7$);

✦ macroeconomic indicators of the external environment: key rate, average rates for all banks of the Russian Federation and for the 30 largest banks of the Russian Federation;

✦ indicators of rating agencies for the period from 2017 to 2020;

✦ semantic (textual) information about the mention of each bank in the news feed of the rating agency.

The following central tendency and dispersion indicators were calculated based on these raw data:

✦ median, dispersion of deviations of deposit rates for each commercial bank relative to macroeconomic indicators: the key rate, the average bank rate for all banks, for the 30 largest banks in Russia;

✦ relative indicators of the rates dynamics, their medians and dispersion;

✦ relative and absolute place of the bank in the rating of banks;

✦ negative mention of the bank in the news feed on the website of the rating agency.

The criterion for checking the quality of the model was the status of the commercial bank at the time of information collection, i.e. whether it was operating or liquidated.

To obtain the UCI values, the components I_1, I_2 and I_3 were calculated as follows.

The component 1 (I_1) – predicted probability of a commercial bank going bankrupt. To obtain an estimate of the probability of revoca-

tion of the commercial bank's license, a logistic regression model was built on the basis of initial data and estimated indicators. These indicators include: the bank's position in the rating, the volatility of bank rates, a qualitative calculation indicator that characterizes the perception of the expert community of the activities of a commercial bank based on the processing of unstructured text data – mention of the downgrade of a commercial bank in the news feed on the rating agency's website. The dependent variable Y is a binary variable taking two values – 0 if the current status of the bank is "active" and 1 if it is liquidated.

The estimated logistic regression model has the form:

$$P(Y = 1) = \frac{1}{1 + e^{-z}}, \qquad (2)$$

$$z = -16.7 + 1.4955 \cdot x_1 + 0.024 \cdot x_2 + \\ + 1.128 \cdot x_3 + 0.389 \cdot x_4, \qquad (3)$$

where x_1 – descending in the ranking;

x_2 – ranking position;

x_3 – median deposit rates;

x_4 – median deviation of demand deposit rates from the market average rate.

All model coefficients are significant at the 5% level, the sensitivity (i.e. the share of liquidated banks) and specificity (the share of operating banks) indicators of the model are high, respectively 85.9% and 73.3%.

Relying on the results of the model's work, by dividing all the values obtained into three intervals, a UCI component 1 was formed. This component characterizes the probability of license revocation within six months of the time of the model's estimation:

✦ low: $P(Y = 1) \le 0.25$;

✦ medium: $0.25 < P(Y = 1) < 0.35$;

✦ high: $P(Y = 1) \ge 0.35$.

The cutoff limits were chosen based on sensitivity and specificity optimization as follows. The upper limit, equal to 0.35, was obtained

when the maximum values of sensitivity and specificity of the model were achieved simultaneously (85.9% and 73.3%). The lower limit, equal to 0.25, corresponds to a specificity value of 80%.

The component 2 (I_2) – Valuation of the bank's membership in the "disadvantaged" cluster. When clustering the training sample by the BIRCH method (two-step method) on the main financial indicators of a commercial bank, the data on the value of the seven basic standards of banks ($H_1 - H_7$), the position of a commercial bank in the banking rating developed by the rating agency and the change in this rating in the retrospective five homogeneous clusters were obtained.

The most interesting is the first cluster, accounting for 14% of the training sample. This cluster contained 30% of the liquidated banks. All banks in this cluster ranked below 200. There was also a decrease in the median of the norms of each commercial bank for the retrospective period, a decrease in the median of the dispersion of deposit rates for the previous period.

According to the results of the model, the component 2 UCI was formed – belonging to the first cluster, as a sign of "disadvantage" of a commercial bank.

The component 3 (I_3) – availability of a negative valuation of the bank's activity by a market participant – expert agency. This component is formed as a binary indicator of belonging to one of two sets. Component 3 takes value 1, if there is at least one negative reference to the name of a commercial bank in the newsfeed on the site of rating agency for the current year, and value 0, if there is no negative reference.

The values of universal complex indicator (UCI) for training and test samples were calculated from UCI components with the help of truth table (*Table 1*). The results are presented in *Table 2*.

Stage 4. Ranking the commercial banks of the training sample on the basis of the calculated UCI values. The result of this stage is a list of commercial banks sorted by decreasing UCI values.

Stage 5. Testing the developed model of UCI calculation was carried out on a test sample of three commercial banks. The results of modeling and recommendations for commercial banks from the test sample are shown in *Table 3*.

The results of approbation of the developed model on the test sample of three commercial banks, as shown in *Table 4*, are confirmed by the current status of commercial banks as of March 2020.

Thus, the concept of express valuation of commercial bank's state proposed on the basis of the analysis of open structured and unstructured data, allows us to get the probability of the bank having an unsatisfactory state. The results obtained allow us to apply preventive measures of support or protection of depositors' interests by the supervisory body (the Central Bank of the Russian Federation).

Conclusion

This paper proposes a concept and algorithm for constructing a universal complex indicator. This allows us to get an express valuation of the state of the object of control by the supervisory authorities to prevent the development of negative events.

The novelty of the proposed approach is the possibility of using the universal complex indicator (UCI) for express valuation of the control objects in various subject areas and activities using as input information structured and unstructured data from open sources.

The information base of the study for the express valuation of the state of the Russian commercial bank, including 111 banks and 79 financial and non-financial indicators from 1.01.2018 to 01.02.2020 was formed.

Table 2.

UCI values of the valuation of the bank's condition

Components of the universal complex indicator (UCI)			The value (y_j) of the universal complex indicator (UCI)
Component 1 (I_1): Probability of license revocation	**Component 2 (I_2):** Cluster membership	**Component 3 (I_3):** Availability of negative references on the rating agency's website	
High (more than 0.35)	Cluster membership is not relevant	Availability of negative references is not relevant	3: Close attention is required
Medium (0.25–0.35)	Belongs to the Cluster 1	1 (availability of negative references)	
Medium (0.25–0.35)	Belongs to the Cluster 1	0 (absence of negative references)	2: Attention is required
Medium (0.25–0.35)	Does not belong to the Cluster 1	1 (availability of negative references)	
Low (less than 0.25)	Cluster membership is not relevant	1 (availability of negative references)	
Low (less than 0.25)	Cluster membership is not relevant	0 (absence of negative references)	1: Attention is not required
Medium (0.25–0.35)	Does not belong to the Cluster 1	0 (absence of negative references)	

Table 3.

UCI values for test sample banks

Bank	**Component 1 (I_1):** Probability of license revocation	**Component 2 (I_2):** Cluster membership	**Component 3 (I_3):** Availability of negative references on the rating agency's website	The value of the universal complex indicator (UCI)
PFS–Bank	High (0.844)	Does not belong to the Cluster 1	0	3: Close attention is required
Peresvet Bank	Low (0.01)	Belongs to the Cluster 1	1	2: Attention is required
Citibank	Low (0.007)	Does not belong to the Cluster 1	0	1: Attention is not required

Table 4.

Comparison of the calculated UCI value and the current status of the commercial bank

Bank	The value of the universal complex indicator (UCI)	Current status of the bank
PFS–Bank	3: Close attention is required	The license was revoked in March 2020
Peresvet Bank	2: Attention is required	Active. Repeated violations of the regulator's mandatory ratios in the process of the bank's rehabilitation are observed
Citibank	1: Attention is not required	Active

The concept developed for building a universal complex indicator (UCI) for express valuation of the state of three Russian commercial banks was tested and a bank with a state that does not meet the requirements for banks imposed by the Central Bank of the Russian Federation was identified. Conclusions about the status of three commercial banks as of March 2020 from the test sample obtained on the basis of the analysis of UCI values were further confirmed by information on the website of the Central Bank of Russia. ∎

References

1. Mullakhmetov Kh.Sh. (2007) Problems of management control organization. *Economics of Contemporary Russia*, no 4, pp. 59−69 (in Russian).

2. Polshkov Y.N. (2019) Applied econometric approaches to risk analysis in managing the region's economic complex and the living standard of its population. *Vestnik of Institute of Economic Research*, no 1, pp. 12−18 (in Russian).

3. Pisareva O.M. (2018) Analysis of the state and characteristics of the development potential of strategic planning tools in the digital transformation conditions of the economy and Management. *MIR (Modernization. Innovation. Research)*, vol. 9, no 4, pp. 502−529 (in Russian). DOI: 10.18184/2079-4665.2018.9.4.502-529.

4. Mayer-Schönberger V., Cukier K. (2013) *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt.

5. Manyika J., Chui M., Brown B., Bughin J., Dobbs R., Roxburgh C., Byers A.H. (2011) *Big data: The next frontier for innovation, competition, and productivity*. McKinsey Global Institute.

6. Gorev A.I., Goreva E.G. (2020) About the applicability of existing data processing algorithms to big data. *Mathematical Structures and Modeling*, no 1, pp. 139−143 (in Russian). DOI: 10.24147/2222-8772.2020.1.139-143.

7. Gromova S.V. (2016) Problems of organization of internal control at the enterprise. *Young Scientist*, no 8, pp. 240−242 (in Russian).

8. Gayduk E.A. (2012) *Application of balance and optimization modeling in making management decisions at the municipal level. Doctoral thesis*. Novosibirsk (in Russian).

9. Kiryushina A.A., Zhukova L.V., Chikina L.G. (2020) Using big data in assessing the degree of pollution of the coastal zone of the seacoast of resort regions. Proceedings of the *All-Russian Scientific Conference "Russian Seas: research of coastal and shelf zones", Sevastopol, 21−25 September 2020*, pp. 409−410 (in Russian).

10. Bidzhoyan D.S., Bogdanova T.K. (2017) The concept of modeling and forecasting the probability of revoking a license of Russian banks. *Economics of Contemporary Russia*, vol. 79, no 4, pp. 88−103 (in Russian).

11. Markov S.M., Markova A.S. (2017) Key performance indicators as criteria for a comprehensive assessment of the effectiveness of the implementation of state programs to support small and medium-sized businesses. *Bulletin of Pskov State University. Series: Economics. Law. Management*, no 5, pp. 132−137 (in Russian).

12. Golovan S.V., Evdokimov M.A., Karminsky A.M., Peresetsky A.A. (2004) *Probability of default models of Russian banks II. Models and macroeconomic environment. Working Paper #WP/2004/043*. Moscow: New Economic School (in Russian).

13. Peresetsky A.A. (2010) Модели причин отзыва лицензии российских банков. *Modelling reasons for Russian bank license withdrawal. Working Paper #WP/2010/085*. Moscow: New Economic School (in Russian).

14. Malykh N.I., Prodanova N.A. (2016) Modern approaches to the assessment of the cost of commercial bank. *Statistics and Economics*, no 6, pp. 79−84 (in Russian). DOI: 10.21686/2500-3925-2016-6-79-84.

15. Borodin O.N. (2008) Extracting information from full-text data sources. *Agricultural Engineering*, no 1, pp. 42−43 (in Russian).

16. Zhang T., Ramakrishnan R., Livny M. (1996) BIRCH: An efficient data clustering method for very large databases. Proceedings of the *1996 ACM SIGMOD International Conference on Management of Data (SIGMOD '96), Montreal, Canada, 4−6 June 1996,* pp. 103−114. DOI: 10.1145/233269.233324.

17. Aivazyan S.A., Afanasiev M.Yu. (2007) Evaluation of measures aimed at managing production inefficiency factors. *Applied Econometrics*, no 4, pp. 27−41 (in Russian).

18. Peresetsky A.A. (2012) *Econometric approach to off-site analysis of Russian banks.* Moscow: HSE (in Russian).

19. Bogdanova T.K., Zhukova L.V. (2020) Building a comprehensive indicator for assessing the state of a Russian commercial bank based on structured and unstructured data. Proceedings of the *43rd Meeting of the International Scientific School-Seminar on System Modeling of Socio-Economic Processes, Moscow, 13−18 October 2020* (in Russian).

20. Bidzhoyan D., Bogdanova T. (2019) Russian banks credit risk stress-testing based on the publicly available data. *Advances in Intelligent Systems and Computing*, vol. 850, pp. 262−271. DOI: 10.1007/978-3-030-02351-5_31.

21. Sobolevskaya Yu.V. (2013) Control bodies: Competence levels. *Public Administration Issues*, no 4, pp. 113−130 (in Russian).

22. Karminsky A.M., Kostrov A.V. (2013) Modelling the default probabilities of Russian banks: Extended abilities. *Journal of the New Economic Association*, no 1, pp. 64−86 (in Russian).

23. Polyakov K.L., Zhukova L.V. (2019) Modeling the probability of credit default of clients of microfinance organizations: The case of one MFI. *HSE Economic Journal*, vol. 23, no 4, pp. 497−523. DOI: 10.17323/1813-8691-2019-23-4-497-523 (in Russian).

24. Zhukova L.V., Kiryushina A.A., Kovalchuk I.M., Ruzaeva A.V. (2018) Online education system efficiency improvement by the means of machine learning and blockchain technology. *Caspian Journal: Control and High Technologies*, no 1, pp. 56−68 (in Russian).

25. Zhukova L., Polyakov K.L. (2019) Comparative analysis of predictive analytics models in classification problems. Proceedings of the *VI International Conference on Actual Problems of Systems and Software Engineering (APSSE 2019), Moscow, Russia, 12−14 November 2019*, pp. 162−169.

About the authors

Tatiana K. Bogdanova

Cand. Sci. (Econ.);

Associate Professor, Department of Business Informatics, Graduate School of Business, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: tanbog@hse.ru

ORCID: 0000-0002-0018-2946

Liudmila V. Zhukova

Senior Lecturer, Department of Applied Economics, Faculty of Economic Sciences, National Research University Higher School of Economics, 20, Myasnitskaya Street, Moscow 101000, Russia;

E-mail: lvzhukova@hse.ru

ORCID: 0000-0003-1647-5337