

Analysing the firm failure process using Bayesian networks

Yuri A. Zelenkov 

E-mail: yzelenkov@hse.ru

HSE University

Address: 20, Myasnitskaya Street, Moscow 101000, Russia

Abstract

This work analyses the firm failure process stages using the Bayesian network as a modelling tool because it allows us to identify causal relationships in the firm profile. We use publicly available data on French, Italian and Russian firms containing five samples corresponding to periods from one to five years before observation. Our results confirm that there is a difference between the stages of the failure process. For firms at the beginning of a lengthy process (3–5 years before observation), cumulative profitability is the key that determines liquidity. Then, as the process develops, leverage comes to the fore in the medium term (1–2 years before observation) for economies with more uncertainty. This factor limits the opportunities for making a profit, leading to further development of the failure. There are also national specifics that are caused, firstly, by the level of economic development and, secondly, economic policy uncertainty.

Keywords: financial failure, firm failure process, Bayesian network

Citation: Zelenkov Y. (2022) Analysing the firm failure process using Bayesian networks. *Business Informatics*, vol. 16, no. 1, pp. 22–41. DOI: 10.17323/2587-814X.2022.1.22.41

Introduction

The study of firm failure is one of the key issues in business research. It can be divided into two subdomains [1]: the first is failure prediction, and the second is the-

oretical and empirical investigations of the failure process. The firm failure process allows us to consider the behaviour of failing firms in the longer perspective [2, 3], while failure prediction studies often focus on financial performance only one or few years before distress [4, 5].

However, the short-term forecasting models' main weakness is that the firms' obligations often are longer than the period during which the risk of default is estimated with perfect accuracy [6]. Thus, the company should be analysed from a longer perspective. Moreover, while some firms with a certain financial performance profile fail, others with the same profile can overcome the difficulties and return to normal operations. Therefore, many authors argue the existence of different types of failure trajectories that may lead or may not lead a firm to default depending on its prehistory and current abilities [1, 7–9].

The existence of different firm failure processes (FFP) is a well-established fact supported by much theoretical and empirical research. However, there is no consensus in the scientific community not only about the exact definition of these processes but even about the number of variants. For example, Argenti [10] detected three failure trajectories of decline in firms' financial health; Ooghe and de Prijcker [8] describe four different types of failure processes; du Jardin [11] identifies seven types of FFP.

These differences are explained, firstly, by the methodology used, for example, works [8, 10] are based on the case method, and du Jardin [11] analyses empirical data using self-organising maps. Second, the authors view the process from different angles. Papers [10, 11] focus on financial results (this approach is also used in many other works [1, 2, etc.]), while [8] considers the problem through the lens of management efficiency.

In this work, we focus on defining the specifics of the various stages of a firm's failure process. Three research questions correspond to this goal:

- ◆ *RQ1*: How do the causal relationships between the financial ratios describing the firm's state change in different periods before the default?

- ◆ *RQ2*: Are there differences in firm failure processes that are determined by country specifics?

- ◆ *RQ3*: How does the degree of economic policy uncertainty affect the firm failure process?

We use the Bayesian Network as a modelling tool because it allows us to identify causal relationships in the firm profile. We use publicly available data on French, Italian, and Russian companies which present the firms' financial ratios from one to five years before the failure.

Our results confirm that there is a difference between the stages of the failure process (*RQ1*). For firms at the beginning of a lengthy process (3–5 years before observation), cumulative profitability is the key that determines liquidity. Then, as the process develops, leverage comes to the fore in the medium term (1–2 years before observation) for economies with more uncertainty. This factor limits the opportunities for making a profit, leading to further development of the failure. There are also national specifics that are caused, firstly, by the level of economic development (*RQ2*) and, secondly, by economic policy uncertainty (*RQ3*).

The rest of the paper is organised as follows. After reviewing sources analysing FFP, we present basic concepts of Bayesian networks. Next, we describe the datasets and pre-processing operations that are necessary to prepare the data for modelling. In the last part, we analyse the network structures obtained and discuss further research to extend the proposed approach.

1. Literature review

The first research into bankruptcy forecasting began in the 1930s [12]. These studies mainly focused on comparing individual rates of successful and unsuccessful firms. However, the number of published works was rela-

tively small. The first multivariate model was presented by Altman [13], who used discriminant analysis based on five financial ratios. This model, also known as the Z-score, ushered in an era of intense research. Researchers have developed many predictive models using both statistical techniques and machine learning. It should be noted that models based on machine learning, in general, provide more excellent performance [14, 15]; however, the Z-score model with some modifications also remains relevant [4].

A common feature of predictive models is that they treat the failure prediction problem as a binary classification task. In this case, most often, the data of financial statements for a small number of periods before default is considered. In fact, these models are constructed on cross-sectional data; ratios from different time periods are often combined in one observation point; thus, the firm's individual dynamics are not taken in the account. Such an approach ignores the fact that companies change over time, all of which causes various problems and limitations [16]. In short, a firm's profile measured at time t cannot be reduced to measurements at time $t - 1$ alone, since the default, in most cases, is the result of a long process [9] and the discriminating power of ratios is unstable over time [17].

There are models based on observations of one, two or more years before failure at time t which are believed to be able to predict the state at years $t + 1$, $t + 2$, or even $t + 10$ [2, 18]. However, because they do not treat firm failure as a process, they have the same limitations as analysed in [16].

Some predictive studies use techniques that allow us to consider both the dynamics of firms' populations and their unique characteristics, such as panel regression [19] or survival analysis [20]. However, in general, the number of such works in the flow of research on predictive models is relatively small.

1.1. Firm failure process

Argenti [10] was perhaps the first who started to study the firms' failure process. He identified three patterns of decline and found that failing firms do not crash immediately after they decline. Some can delay the onset of bankruptcy for years.

D'Aveni [7] empirically tested Argenti's findings [10] that are based on case studies. According to both authors, the three failure processes are the following:

- ◆ Sudden decline, that is, the rapid collapse of the firm. This process of failure is typical of small or competitively disadvantaged firms that reoriented their strategy too boldly.
- ◆ Gradual decline, i.e., a slow and gradual process typical of bureaucratic and poorly managed firms that cannot adapt to the external environment.
- ◆ Lingering decline. This process is typical of firms that decline either rapidly or gradually but delay bankruptcy for several years. Such post-decline firms often centralise to be a threat rigid and exhibit strategic paralysis and downsizing activity.

Based on these earlier studies of failure processes, [1] postulate the existence of three types of failure process:

- ◆ short-run process when potential failure can be detected only about 1 year after the last reporting;
- ◆ mid-run process corresponds to a situation where the first signals of a potential failure can be detected 2–3 years before the default;
- ◆ long-run process when the potential failure can be detected more than 3 years before the default.

Short-term processes are more suitable for describing the situation when a firm with good performance declines suddenly. Mid-term and long-term processes, in turn, can describe two

situations: the firm never becoming successful enough or the firm becomes worse step by step.

To empirically check their proposition, the authors of [1] analysed 1234 bankrupt manufacturing SME's from different European countries. They applied four clustering methods on the eight different sets of variables (presented in [4]) over the last five years before the bankruptcy. Their results confirm the existence of three types of failure processes, differing in time scale and, therefore, in decline rate.

Summarizing their results, the following can be noted. For short-run processes, the failure risk (FR) is observable only in year $t - 1$ and negative annual profitability is the most important contributor. For mid-run processes, the failure risk can be detected in years $t - 2$ and $t - 3$, when negative annual profitability is also the most important contributor, followed by high leverage. For firms involved in a long-run process, the first signals can be detected up to $t - 3$; they are annual and cumulative profitability and leverage. These ratios contribute to FR also in years $t - 2$ and $t - 1$. Liquidity as an indicator of FR is important only for the last stages of mid-term and long-term processes.

However, this view of failure processes is not the only one. For example, authors of [8] present four different processes explained through the lens of management: (1) unsuccessful start-up due to the lack of managerial or industry-related experience, (2) ambiguous growth of firms with over-optimistic management, (3) unbalanced growth induced by management's dazzle, and (4) an apathetic mature firm managed by people lacking motivation and commitment.

In a series of works [6, 9, 11] du Jardin models various failure processes (also called trajectories or profiles) using self-organizing maps (SOM). The basic idea of the applica-

tion of SOM to study individual trajectories of firms is straightforward. Let us have panel data (observations of objects corresponding to measurements made in different time periods). If all the observations are classified on a SOM as if they were independent, it is possible to study the change of state of a given object along time [21]. To the best of our knowledge, this approach was first used in [22] to analyse the financial state of Spanish banks. The author noted that the trained SOM model groups the entities together according to their financial state similarities. Thus, new observations will be placed in a particular zone (bankrupts or non-bankrupts) according to the more activated neurons on the map. Therefore, it is possible to observe the bank's evolution using financial information from various years.

Briefly, du Jardin's methodology can be described as follows: at the first step, all firms are mapped to SOM, and the firm's observations at different points in time are considered independent. Then the trajectories of firms represented by the list of neurons that correspond to observations of one entity at sequential time points are built. In the last step, the trajectories are grouped into meta-classes, which can be viewed as processes leading or not leading to default. The author's results confirm that the generalisation error achieved with an SOM remains more stable over time than that achieved with conventional failure models (discriminant analysis, logistic regression, Cox's survival model, neural networks and ensemble methods). However, more importantly in the context of our discussion, du Jardin identifies a different number of processes that show the movement of firms between regions with different probability of failure: six in [9], seven in [11], and eight in [6]. This difference can be explained by the impact of the data used (specific years and time lag before default) and the impact of the technique of trajectories' grouping. This

may not be significant in terms of the model's performance since these works' primary goal is to improve prediction accuracy by considering the firm's prehistory. However, even if this approach improves the predictive capabilities, it does not allow us to analyse the failure processes, since it is based on the black-box model.

So, we can conclude that the existence of different firm failure processes is a well-established fact; however, there is no consensus regarding these processes' definition. To contribute to the solution of this problem we propose to use a causality modelling technique, namely Bayesian networks. Such an approach allows us to identify causalities between financial ratios at different periods before failure, which can shed light on the firm's dynamics.

1.2. Bayesian networks and causality modelling

Intuitively, causality can be defined as influence through which a cause contributes to the production of an effect, while the cause is partially responsible for the effect, and the effect is partially dependent on the cause [23]. Complex systems are characterised by the presence of multiple interrelated aspects, many of which relate to the reasoning task. Thus, one of biggest challenges is the extraction of causal relationships from empirical data and construction of models of complex systems that allow causal inference.

The declarative representation approach [24] is based on a causal model of the system about which we would like to reason. This model encodes our knowledge of how the system works and can be manipulated by various algorithms that can answer questions based on the model.

To communicate causal relationships, a causal model uses a combination of equations and graphs. Mathematical equations

that express the form of causality (e.g., linear, or non-linear) are symmetrical objects, so relationships of variables can be inverted using simple manipulations. For this reason, equations are augmented with a diagram that declares the directions of causality [25]. Such a model can be built manually based on expert knowledge or automatically using machine learning algorithms [26].

According to [27], causal inference extends predictive modelling (which involves estimating the conditional distribution $p(\mathbf{Y}|\mathbf{X})$ of the variables \mathbf{Y} and \mathbf{X} on the basis of a random sample) to causal modelling, where the model should be able to estimate the conditional distribution $p(\mathbf{Y}|\mathbf{X}||\mathbf{M})$ when manipulated \mathbf{M} .

There are several approaches to the construction of causal models, in particular structural equation modelling (SEM) and Bayesian networks (BN). The SEM [28] is limited, first, in that it requires a priori hypotheses about causality in the system. Secondly, it supposes only linear types of relationships. Therefore, in our study, we will use Bayesian networks, since they are free of such disadvantages. The structure of the network and its parameters can be extracted from data; relationships between variables are probabilistic.

A Bayesian network encodes the joint probability distribution $\mathbb{P}(\mathbf{X})$ of a set of m random categorical variables, $\mathbf{X} = (X_1, \dots, X_m)$, as a directed acyclic graph (DAG) and a set of conditional probability tables (CPT). More formally, it is a pair $\langle \mathcal{G}, \mathcal{P} \rangle$, where \mathcal{G} is the DAG whose vertices correspond to the variables in \mathbf{X} and arcs represent direct dependencies between variables, and \mathcal{P} is a collection of functions that define the behaviour of each variable in \mathbf{X} given its parents in the graph [27,29].

The representation of the full joint table $\mathbb{P}(\mathbf{X})$ takes exponential space in the number of variables m . This complexity is avoided thanks to the Markov condition, which states that in a

Bayesian network every variable is conditionally independent of its non-descendant and non-parents. Thus, for the set of random variables \mathbf{X} in \mathcal{G} , a density $\mathbb{P}(\mathbf{X})$ is

$$\mathbb{P}(\mathbf{X}) = \mathbb{P}(X_1, \dots, X_m) = \prod_{i=1}^m \mathbb{P}(X_i | \text{parents}(X_i)),$$

where $\text{parents}(X_i)$ denotes the set of variables $X_j \in \mathbf{X}$, such that there is an arc from node j to node i in the graph.

In other words, each node in the graph \mathcal{G} that corresponds to a variable has an associated CPT that contains the probability of each state of the variable given its parents in the graph. Such a presentation allows us to describe the structure of complex distribution compactly [30] and can be interpreted from two points of view [24]. First, the graph is a compact representation of a set of independencies that hold in the distribution. The other perspective is that the graph is a skeleton for factorising distribution: it breaks up the distribution into smaller factors, each over a much smaller space of possibilities.

Bayesian networks have many advantages [24]. First, this type of presentation is interpretable by a human. Second, such a structure allows us to answer queries, i.e., computing the probability of some variables given evidence of others (inference). Third, models can be constructed whether by a human expert or automatically by learning from data. In our study, we will use the latter approach – data driven learning.

From a formal perspective, the Bayesian network represents the underlying joint distribution, including probabilistic properties such as conditional independence. On the one hand, it is a more compact representation of complex multivariate distributions. On the other hand, a “good” network structure should correspond to causality, in that an edge $X \rightarrow Y$ often suggests that X “causes” Y , since each value x of X specifies a distribution over the values of Y [24, 25].

The process of construction of Bayesian networks from data \mathcal{D} includes two stages: generation of the graph representing the optimal structure of BN (structure learning), and definition of conditional probabilities (parameter learning). Many authors apply BN to problems in different domains: for example, business [31], ecology [32], healthcare [33], fault diagnosis in engineering systems [34] and many others [35].

It should also be noted that there is an extension of the BN model for longitudinal data, namely, dynamic Bayesian networks. However, dynamic models are based on the assumption that the process under study is stationary, i.e., its parameters do not change over time. According to [1] and other researchers, the firm failure process is not stationary. Therefore, in our study, we generate BN structures for different time intervals independently, supposing that comparing these structures will shed light on the peculiarities of FFP stages.

1.3. Inference and explanation in Bayesian networks

Inference is the process of computing new probabilistic information from a Bayesian network based on some evidence. It computes joint posterior probabilities for a set of variables given evidence which are the values on other variables. Inference is an NP-complete task, therefore there are algorithms that implement an exact inference but also algorithms for approximated inference that can converge slowly and even not exactly but that can in many cases be useful for applications. This capability allows us to use BN for supervised classification which aims at assigning labels to instances described by a set of predictor variables [36].

However, unlike many machine learning methods, a Bayesian network can be used not only for prediction but also for explanation [37]. Explanation tasks in Bayesian networks can be classified into three categories [38]:

- ◆ explanation of a model – presentation of the domain knowledge;
- ◆ explanation of reasoning – presentation of the results inferred and reasoning process that produced them;
- ◆ explanation of evidence, i.e., determination which values of the unobserved variables justify the available evidence.

Since our goal is to analyse the BN structures that model firms at different times before failure, explaining the model is the most important issue.

In [39] the authors give examples of model explanations. These explanations can include properties of nodes and their mutual influence that can be negative or positive. The influence of node A on node B is positive when higher values of A make high values of B more probable. The definitions of negative influence and negative link are analogous.

2. Data

Relevant data is needed to build causal models for firms that filed for default years after the measurement. In addition, an interesting question is the comparison of the FFP for different countries. Therefore, we chose three countries for analysis: France, Italy, and the Russian Federation.

To compare the economies on a macro-level, we use the Gross Domestic Product converted to constant 2017 international dollars using Purchasing Power Parity (PPP) rates and divided to the total population¹.

Figure 1 shows the change of this indicator for the selected time interval (2009–2019). France has the most stable economy; it exhibits its constant GDP per capita growth during the whole period under study. The Italian econ-

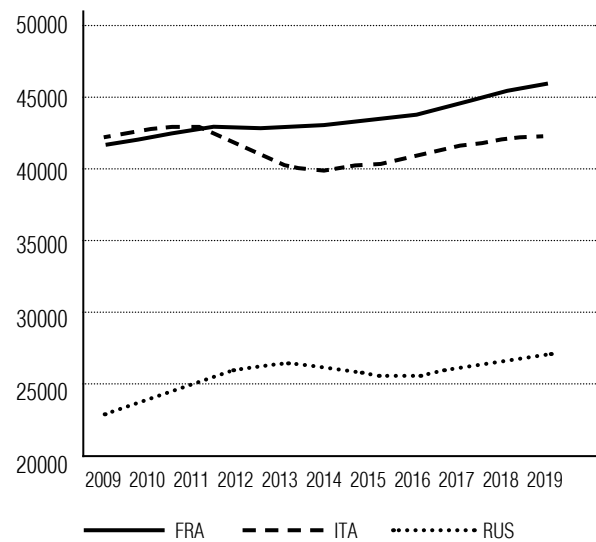


Fig. 1. Changes of GDP per capita (constant 2017 international dollars using PPP) for selected countries.

omy is more volatile; after 2010, there was a recession, and growth resumed only in 2015. The Italian economy is driven in large part by small and medium-sized enterprises, many of them family-owned. Italy also has a sizable underground economy, which is estimated as much as 17% of GDP². Based on these data, we can expect that the FFP model for French firms will be more stable than for Italian ones as they operate in a more stable environment. Note that according to the World Bank classification, both countries belong to the group of developed countries.

The Russian Federation belongs to the group of developing countries or economies in transition. The Russian economy is characterized by the significant share of the government-controlled sector and is largely regulated not by the market but by political decisions. A combination of falling oil prices, international sanctions, and structural limitations pushed Russia

¹ <https://data.worldbank.org>

² CIA (2021). The World Factbook. <https://www.cia.gov/the-world-factbook/>

into a deep recession in 2015, but GDP decline was reversed in 2017 as world oil demand picked up. All this leads to the highest growth of uncertainty. Under such conditions, it can be expected that the FFP model for Russian firms will change quite strongly at different stages.

We have collected the necessary data from the Bureau van Dijk Amadeus database³ using the following search strategy:

- ◆ Data for companies that operate in 2009–2019. We excluded 2020 data to avoid the impact of external shocks related to the COVID-19 pandemic.
- ◆ The company belongs to a small and medium sized business (SMB) – the number of employees in the last available year is limited by values $\min = 10$, $\max = 250$.
- ◆ Good companies are companies which have Active status in the last available year
- ◆ Failed companies are companies that have one of the statuses: Active (default of payment), Active (insolvency proceedings), Bankruptcy, Dissolved (liquidation), and Dissolved.

For each country under consideration, we get five samples corresponding to year $t - n$,

$n = 1, \dots, 5$ before observation in year t . Each observation has a class label that indicates the firm's state at the end of the forecast period t : failure ($Class = 1$) or non-failure ($Class = 0$). The number of observations (samples) for each time period is presented in *Table 1*; the number of failed firms is indicated in brackets. As you can see, all datasets are unbalanced. The values of the imbalance (IB) ratio computed as the ratio of negative class observations to the number of failed companies are also given in *Table 1*.

2.1. Features selection

Since our goal is to build interpretable causal models, we must reduce the number of variables in the original dataset, leaving only those that provide the optimal balance of simplicity and completeness. Therefore, we will follow [1] approach, who used four variables that included the famous Altman's Z''-score model when analysing the bankruptcy process.

In a paper presenting the initial Z-score model, Altman [13] compiled a list of 22 potentially important financial ratios, classified into five standard categories: liquidity, profitability,

Table 1.

Dataset characteristics

		$t - 5$	$t - 4$	$t - 3$	$t - 2$	$t - 1$
France	Samples	48024 (1509)	47163 (1503)	44151 (1439)	41798 (1382)	39720 (1313)
	IB ratio	30.825	30.379	29.682	29.245	29.251
Italy	Samples	55895 (5223)	56036 (5349)	56170 (5522)	56115 (5535)	55728 (5498)
	IB ratio	9.702	9.476	9.172	9.138	9.136
Russian Federation	Samples	44354 (1941)	43859 (2077)	43153 (2167)	43050 (2337)	42931 (2398)
	IB ratio	21.851	20.117	18.914	17.421	16.903

³ Bureau van Dijk. Amadeus. <https://amadeus.bvdinfo.com>

Table 2.

Financial ratios in Altman's Z-score model

Category	Financial ratio	Definition	Comments
Liquidity	WCTA	Working Capital / Total Assets	Working capital is defined as the difference between current assets and current liabilities, so this ratio is a measure of the net liquid assets of the firm relative to the total capitalisation [13]. The liquidity role is based on legal considerations, as the inability to pay the outstanding debt is a sufficient precondition for starting an official bankruptcy process [1].
Cumulative profitability	RETA	Retained Earnings / Total Assets	It is the measure of cumulative profitability over time which implicitly includes the age of a firm [13].
Annual profitability	EBITTA	Earnings before Interest and Taxes / Total Assets	It is a measure of the true productivity of the firm's assets, abstracting from any tax or leverage factors[13].
Leverage	BVETL	Book Value of Equity / Book Value of Total Liabilities	In the initial Z-score model, the Market Value of Equity was used but this approach is applicable only to publicly traded companies (Altman et al., 2017). This ratio measures the firm's ability to service liabilities using its own equity because additional debt, all other things being equal, increases bankruptcy likelihood [1].
Activity	STA (excluded)	Sales / Total Assets	Excluded from the revised Z''-score model because it is an industry-sensitive variable [4].

leverage, solvency, and activity. Only five financial ratios were included in the final discriminant function (*Table 2*). Note that higher values of all selected ratios correspond to a lower likelihood of bankruptcy.

Later, the author noted that the original model is applicable only to publicly traded companies since it includes the firm's market value [40]. For this reason, in the new version of the model, he substituted the market value of equity by book value (*Z'*-score model). The next significant improvement was the exclusion of Sales / Total assets ratio because it is an industry-sensitive variable (*Z''*-score model).

Thus, we will use four ratios (WCTA, RETA, EBITTA, and BVETL) included in Altman's *Z''*-score. One of the BN modelling preconditions is that there must be no latent variables (unobserved variables influencing the network's variables) acting as confounding factors. Based on the time-tested Altman model, we can confidently believe that this condition is met and there are no latent factors in the empirical data.

In [4], the authors tested the performance of the *Z''*-score model using a huge international dataset (more than 2.6 million observations of firms from 31 countries in the training sample). Overall, their results confirm that the model performs well despite its simplicity.

However, the authors made several important clarifications:

- ◆ the coefficients of the model must be reevaluated for each sample,
- ◆ the model based on logistic regression gives better results than the multiple discriminant analysis version.

Thus, we will use logistic regression as the basis for validating subsequent data transformations. *Table 3* presents the ROC AUC scores obtained using the logistic regression (LR) 10-fold cross-validation procedure for data that contains the above four Altman features (see the ‘Raw data’ line for each country). Note, the quality of prediction decreases as the interval between observation and evaluation increases because the classification approach ignores changes in the firm over time [16].

2.2. Discretisation

Another problem stems from the fact that the concept of a non-linear Bayesian network was developed to handle discrete or categorical data. There are three common approaches to extending the Bayesian network to continuous variables [41]. The first is to model the conditional probability density of continuous variables using parametric distributions, and then to redesign the BN learning algorithms based on the parameterisations [42]. The second approach is to use nonparametric distributions, such as Gaussian processes [43]. The third approach is discretisation, that is a process that transforms a variable, either discrete or continuous, into a finite number of intervals and associates with each interval a numerical, discrete value [44, 45].

Table 3.

ROC AUC scores (10-fold cross-validation)

Data	Model	Datasets				
		$t - 5$	$t - 4$	$t - 3$	$t - 2$	$t - 1$
France						
Raw data	LR	0.675(0.034)	0.686(0.034)	0.694(0.026)	0.705(0.024)	0.714(0.028)
Discretized data	LR	0.687(0.030)	0.696(0.029)	0.712(0.022)	0.723(0.019)	0.729(0.027)
	BN	0.686(0.028)	0.697(0.028)	0.712(0.016)	0.726(0.015)	0.727(0.032)
Italy						
Raw data	LR	0.695(0.039)	0.727(0.031)	0.744(0.026)	0.768(0.017)	0.802(0.012)
Discretized data	LR	0.716(0.027)	0.761(0.019)	0.777(0.015)	0.798(0.013)	0.826(0.009)
	BN	0.717(0.028)	0.757(0.019)	0.776(0.013)	0.809(0.021)	0.833(0.015)
Russian Federation						
Raw data	LR	0.658(0.023)	0.675(0.021)	0.691(0.014)	0.711(0.014)	0.742(0.010)
Discretized data	LR	0.676(0.024)	0.683(0.018)	0.695(0.011)	0.731(0.021)	0.757(0.020)
	BN	0.709(0.034)	0.740(0.038)	0.738(0.034)	0.743(0.030)	0.770(0.037)

The discretisation approach in the context of Bayesian networks can be divided into two parts. First, there are algorithms that discretise attributes based on interdependencies between class labels and attribute values, such as the entropy binning method [46]. These algorithms are based on classification problems. They are used to discretise all continuous variables before learning the Bayesian network structure. The next class of algorithms requires that the structure of the network be known in advance [41, 45, 47]. These algorithms start with some preliminary discretisation policy, then the structure learning algorithm is started to determine the locally optimal graph structure. The discretisation policy is then updated based on the learned network, and this cycle is repeated until convergence.

We carried out a series of experiments and found out that preliminary discretisation based on [46] allows us to learn networks with higher score. *Table 3* shows the logistic regression ROC AUC scores obtained on discretised data, which confirm the chosen approach to discretisation improves the performance of the model.

For reference, *Figs. 2* and *3* shows the distribution of raw and discretised data respectively of the Italy $t-1$ dataset. Note that the average values of the transformed ratios have changed because we now use the identification of intervals into which each variable is divided instead of the absolute values. For BN, this transformation is acceptable because the model uses joint probabilities. The number of intervals according to [46] is determined based on the joint distribution of the attribute discretised and the target variable.

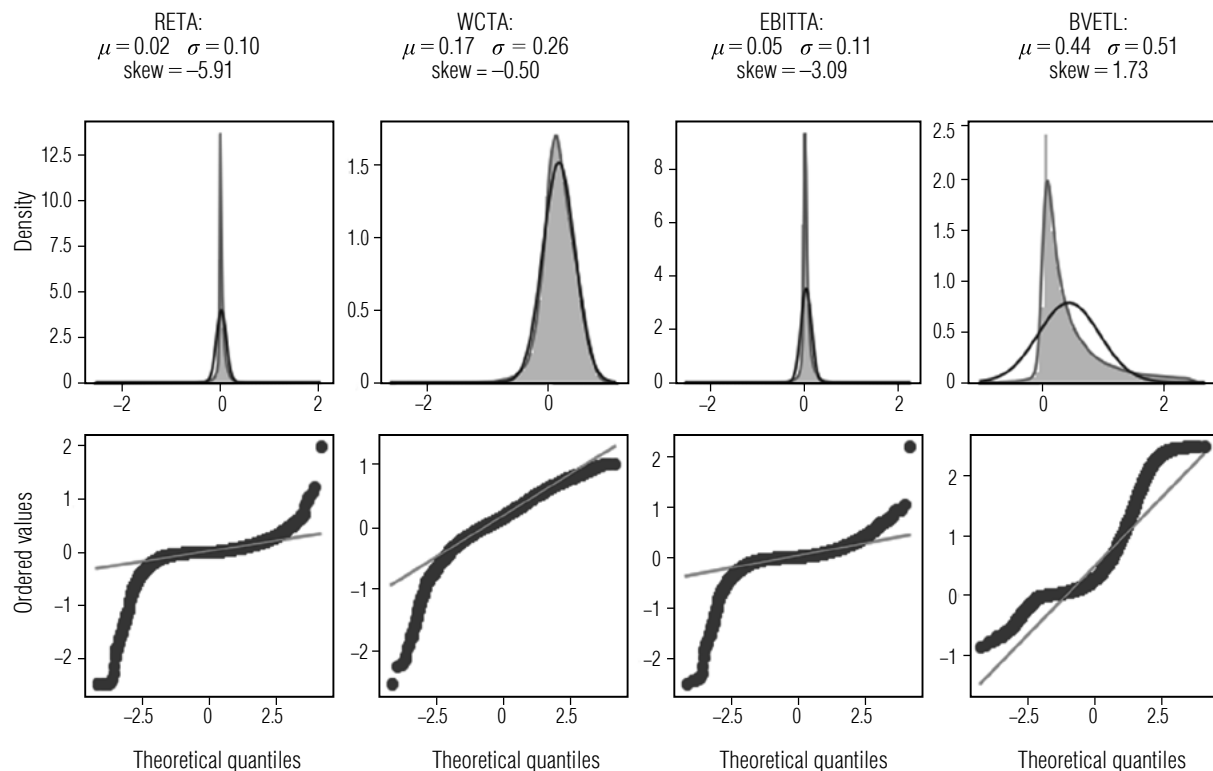


Fig. 2. Italy $t-1$ dataset: distribution of continuous data.

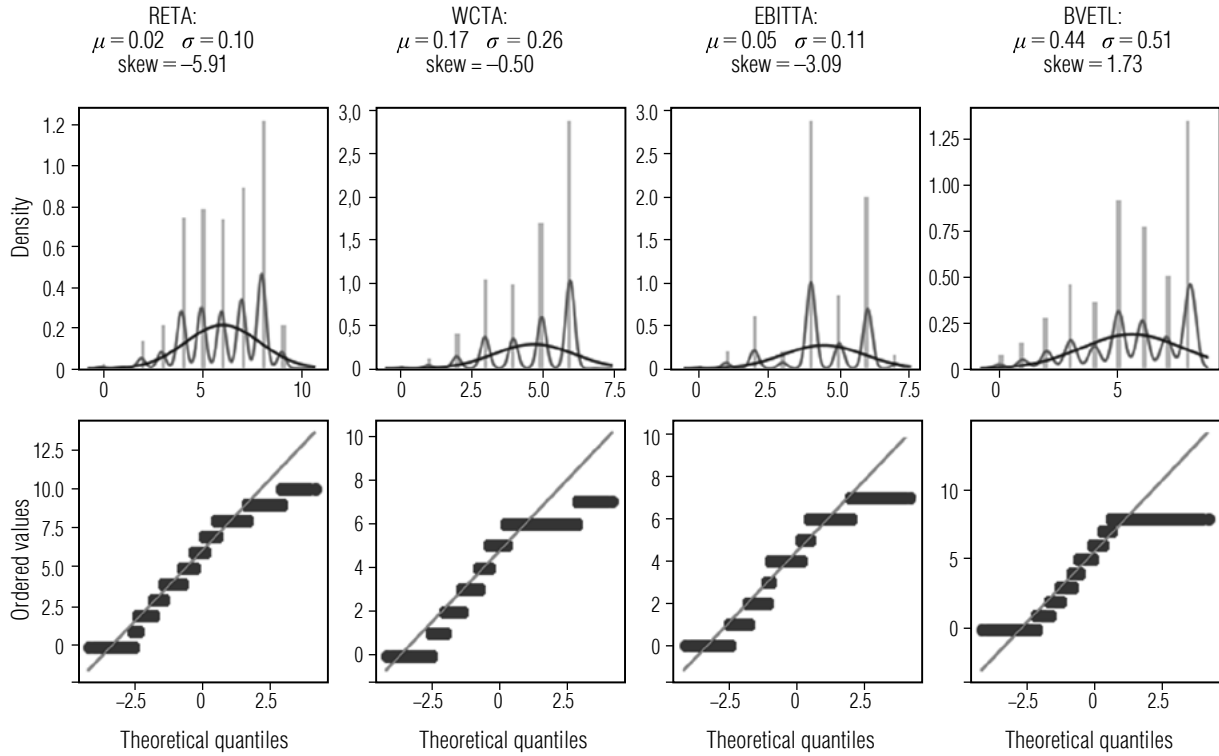


Fig. 3. Italy t – 1 dataset: distribution of discretised data.

3. Experiment and results

The process of construction of Bayesian networks from data \mathcal{D} includes two stages: first, generation of the directed acyclic graph \mathcal{G} representing the optimal structure of BN (structure learning), and next, definition of conditional probability tables \mathcal{P} for each node in the graph (parameter learning). For our research, it is most important to study the structure of Bayesian networks corresponding to different periods before default. However, we performed both stages of learning, since CPT is important for the causal inference and, therefore, use of the model as a predictive tool.

Learning the structure of Bayesian networks can be complicated for two main reasons: (1) inferring causality and (2) the super-exponential number of directed edges that could exist in a dataset. Most methods for structure learning

can be put into one of the following categories [24, 29]:

- ♦ score-based structure learning, with the goal to solve the optimisation problem

$$\operatorname{argmax}_{G \in \mathcal{G}} \operatorname{score}(G, \mathcal{D}).$$

In other words, it is the task to find the best DAG according to some score function that measures its fitness to the data. Widely adopted scores are the Bayes Dirichlet equivalent uniform (BDeu), Bayesian Information Criterion (BIC), which approximates the BDeu, and Akaike Information Criterion (AIC).

- ♦ constraint-based structure learning family of algorithms that perform a series of statistical tests to find independences among the variables and build the DAG following these constraints.

According to [24], a score-based approach evaluates the complete network structure against the null hypothesis of the empty network. Thus, it takes a more global perspective, which allows us to trade off approximations in different part of the network. Therefore, we use score-based algorithms.

To evaluate the structure, we use the BIC score. Let us have a set of random variables \mathcal{D} . Let S be a candidate Bayesian network structure and Θ_S be a vector of parameters for S . Then

$$\text{BIC} = \log P(\mathcal{D}|S, \Theta^*) - \frac{d}{2} \log N,$$

where Θ^* is the estimation of Θ_S ;
 d is the number of free parameters in S ;
 N is the dataset size.

The first term in the formula presents the logarithm of likelihood and the second one is the penalty for complexity.

BIC has two important properties that allow it to be used as a universal metric. Firstly, BIC

is an equivalence invariant, i.e., it gives the same score to equivalent models. As the number of variables grows, the number of possible network structures also grows. This property of the BIC guarantees the assignment of the same score to equivalent networks. Secondly, BIC is locally consistent when the sample size is sufficiently large.

There are many different software packages and methods that they implement (e.g., see review in [29]). We use the pomegranate, that is an open-source Python library [48] implementing few score-based methods, in particular an exact algorithm A*[49], its greedy implementation and Chow-Liu [50] algorithm.

On the first step, we tested all the algorithms in the package to find the ones that give the best results on our data. According to the tests, the exact algorithm achieves the best performance. Table 4 presents values obtained of BIC for the Bayesian Network. We also tested the Naive Bayes (NB) approach to ensure that the proba-

Table 4.

Bayesian Information Criterion (BIC) for the Bayesian Network (BN) and the Naive Bayes (NB) model

	<i>t</i> - 5	<i>t</i> - 4	<i>t</i> - 3	<i>t</i> - 2	<i>t</i> - 1
France					
NB	-201 163	-192 127	-180 853	-178 759	-159 080
BN	-171 651	-162 705	-152 862	-147 833	-129 372
Italy					
NB	-360 293	-375 556	-380 321	-391 209	-394 780
BN	-316 636	-332 979	-324 984	-326 920	-329 246
Russian Federation					
NB	-199 197	-249 535	-235 613	-240 508	-247 257
BN	-167 615	-214 696	-204 133	-206 716	-211 850

bilistic relationships between variables are beneficial. NB is the simplest form of the Bayesian network, derived from the assumption of mutual independence of exogenous variables. The results presented in *Table 4* confirm that the Naive Bayes method is inferior in accuracy to Bayesian networks. The corresponding BIC values are about 20% worse than those obtained for BN.

We also tested the performance of a classifier built based on this Bayesian network [36]. Given that the data is unbalanced, we used the decision threshold adjustment by introducing various penalties for misclassification errors [51]. So, for observation x , the predicted class label $\hat{y}(x) = 1$ if and only if $\mathbb{P}(x) \geq t$. Here the $\mathbb{P}(x)$ is an inference of BN when all variables except *Class* are known. Threshold t is computed as

$$t = C_{10} / (C_{10} + C_{01}),$$

where C_{ij} is a cost of predicting the class i when the true class is j . We set $C_{10} = 1$ and $C_{01} = IB$, where IB is the imbalance ratio of the training dataset.

Table 3 presents the ROC AUC scores obtained by 10-fold cross-validation; refer to the lines ‘Discretised dataset / BN’. As we can see, the performance of the BN classifier at least comparable with the Logistic Regression for all datasets and outperforms it in most cases especially for uncertain economies (Russia and Italy).

4. Discussion

The network structures shown in *Figs. 4–6* allow us to draw some important conclusions about the features of different stages of the firm failure process. The *Class* variable that labels the firm state (0 for healthy firms and 1 for fail companies) is on the root of graphs. This can be easily interpreted as follows. The state of the firm is the root cause that determines the values of its financial ratios. This view is consist-

ent with the problem of failure prediction when the state of the firm is computed by the values of financial ratios.

As follows from *Figs. 4–5*, for developed economies (Italy, France), the early stages of a long-term process ($t - 5$, $t - 4$) coincide. For the period $t - 5$, the cumulative profitability positively affects the difference between assets and liabilities, i.e., leverage (note, the numerator BVETL is the difference between Total Assets and Total Liabilities). Both factors then determine the firm’s current liquidity and current profitability. Note that liquidity and annual profitability are independent. However, at stage $t - 4$, annual profitability becomes a factor affecting liquidity.

For a more predictable economy (France), the model will not change during the $t - 4$, $t - 3$ and $t - 2$ periods. One year before the financial failure, the network structure for France changes and becomes like the $t - 5$ period. Overall, we can conclude that cumulative profitability is a key factor in the success of French firms.

The model representing the short-term failure process of Italian firms ($t - 2$ and $t - 1$) is changing more radically. The key factor is the difference between assets and liabilities (leverage), which determines the firm’s ability to generate profits and liquidity. Note also that the liquidity values are conditionally independent of the cumulative and annual profitability at these stages. Obviously, this is due to the higher uncertainty in the Italian economy. Firms unable to meet liabilities using their own assets cannot quickly remedy this situation by increasing productivity through borrowed resources.

For Russian companies, the key factors are leverage and cumulative profitability. Also note that in this case, the liquidity depends on all the variables under consideration (except the period $t - 2$). In general, the process can be described as follows. In the mid and long term

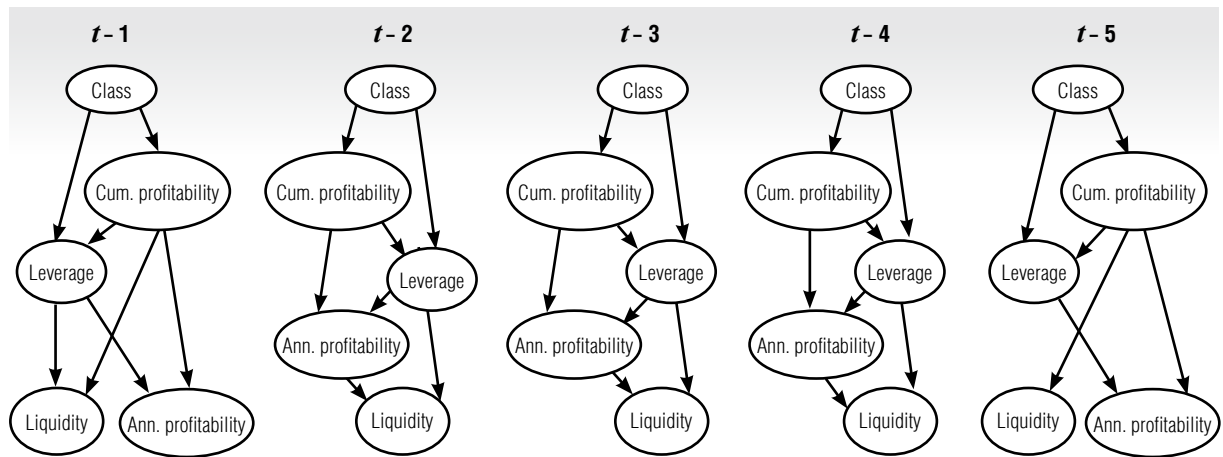


Fig. 4. France: Bayesian networks for different periods before default.

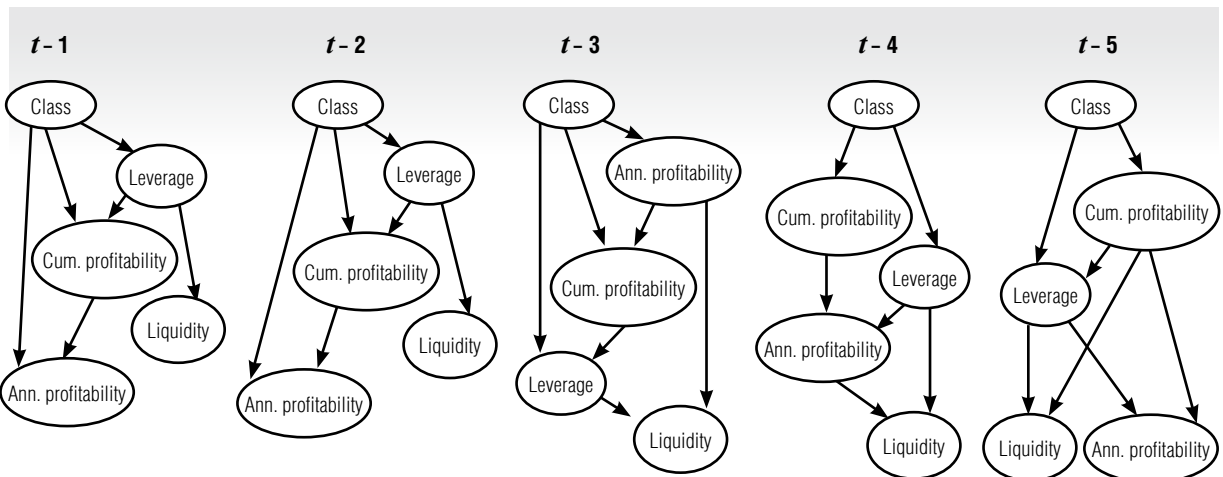


Fig. 5. Italy: Bayesian networks for different periods before default.

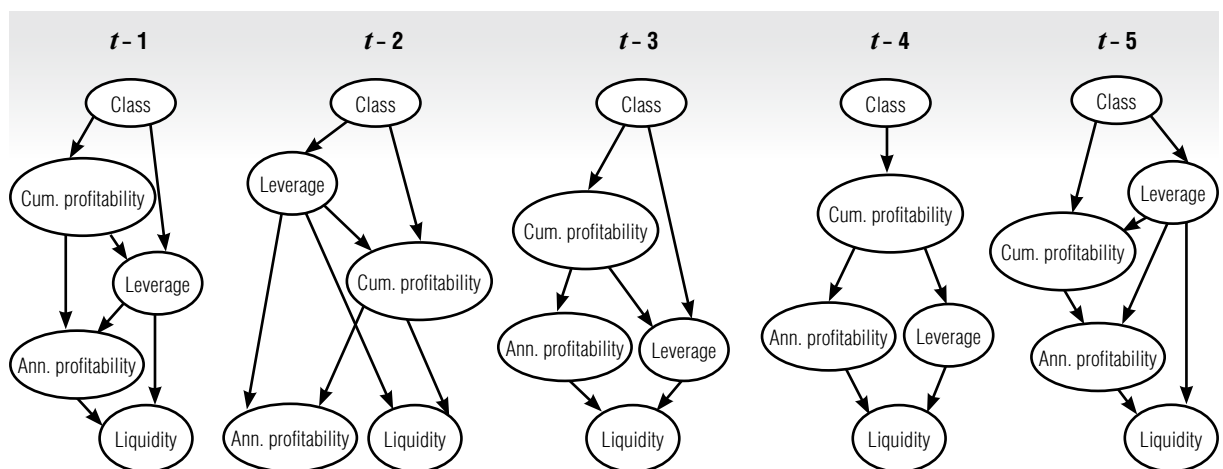


Fig 6. Russian Federation: Bayesian networks for different periods before default.

($t - 3$ and $t - 4$) cumulative profitability has a marginally positive effect on annual profitability. This can be explained by the fact that the RETA ratio implicitly reflects the firm's age [13] and its ability to generate profits sustainably. The accumulated profit also causes the number of external resources attracted. At the same time, annual profitability and leverage are conditionally independent; however, they completely determine liquidity. The conditional independence of the annual profitability and the volume of attracted resources can be explained by the fact that we are considering a fairly long process at these stages, the results of which will be evaluated in 3–4 years. Obviously, this process is more influenced by managerial decisions based on financial indicators that reflect long-term trends (cumulative profitability) than short-term results (annual profitability).

In stage $t - 2$, leverage becomes a key factor. This means that the ability to attract resources allows underperforming firms to increase profitability and increase liquidity and avoid financial disruptions in 2 years. In the year $t - 1$, cumulative profitability becomes a causal factor determining the leverage. This can be explained by the fact that potential lenders assess the firm's overall performance in the long term, which can limit the availability of borrowed resources. Annual profitability is caused by the ability to generate profit in the long term and service the debt. Leverage and annual profitability determine the current value of liquidity, which at this stage is the main indicator of potential financial failures.

The main conclusion drawn from the presented results is that the mutual influence of the factors that determine the state of the firm changes over time ($RQ1$). In general, for firms at the beginning of a lengthy process that could lead to failure, cumulative profitability is the key that determines other metrics such as liquidity and leverage. Then, as the process develops, in the medium term, the degree

of self-sufficiency, as measured by leverage, come to the fore, especially for economies with higher uncertainty. In these stages, low values of these factors limit the opportunities for making a profit. This leads to further development of the failure.

However, there are national specifics that are caused, firstly, by the level of economic development ($RQ2$) and, secondly, economic policy uncertainty ($RQ3$). This specificity is manifested both in the change in the causal relationships between factors at different stages of the firm failure process and in the rate of change of models. The most robust set of models is obtained for France, which has the lowest uncertainty. For Russia, which is characterized by the maximum growth of economic uncertainty over the past 10 years, the models change most frequently and more radically.

Thus, the resulting graphs shed light on the specifics of the various stages of the failure process. As far as we know, our paper is the first attempt to analyze FFP based on Bayesian networks. However, our research in its current form has some issues that can be possibly viewed as limitations. In particular, we can note the following:

- ◆ The sample used contains cross-sectional data for different time periods before failure. It allows us to identify differences in causal relationships at stages of FFP; but it is impossible to trace the evolution of specific firms. To solve such a problem, panel data is needed. Analysis of data containing sequential periods for good and failed firms can provide more detailed information on the causality of a firm's decline. However, solving this problem requires another tool, which can be a Dynamic Bayesian Network.
- ◆ The analysed factors are limited to only four financial ratios presented in the Altman Z''-score. We accepted this limitation based on the requirements for simplicity of the model and its further interpretation. This made

it possible to draw important conclusions about the internal dynamics of a firm. However, in further research, the financial ratios list can be extended to get more complex and detailed models. It is also necessary to study the influence of other parameters, for example, corporate governance and environmental factors.

The next issue, which is of practical interest, is the definition of the current stage of the analysed firm's process. This information can be useful for predictive model which will compute the probability of default for a few future periods. This issue is also a topic of future research.

Conclusion

Our work's main goal was to demonstrate that Bayesian networks can serve as a reliable tool for analysing the dynamics of firms and studying the firm failure process. Our results, on the one hand, highlight the specifics of stages of the failure process for different economies. On the other hand, they allow us to build predictive models that surpass Altman's Z''-score

using the same variables. As far as we know, the work presented is the first one using Bayesian networks for FFP analysis, so many issues remained outside our study's scope. Possible areas of research include:

- ◆ Building models on panel data describing the dynamics of a set of firms.
- ◆ Expansion of the number of analysed features.
- ◆ Modelling specifics of industries.
- ◆ Determining the stage of the process to predict failure in the long term.

All this opens a vast field for new studies, which, in the light of the results obtained, seem promising, since they can potentially make a significant contribution to the theoretical and empirical analysis of the firm failure process. ■

Acknowledgments

The work is a part of the project "Development of Quantitative Methods for Bankruptcy Prediction" supported by the Graduate School of Business of the HSE University.

References

1. Lukason O., Laitinen E.K. (2019) Firm failure processes and components of failure risk: An analysis of European bankrupt firms. *Journal of Business Research*, vol. 98, pp. 380–390. <https://doi.org/10.1016/j.jbusres.2018.06.025>
2. Altman E.I., Iwanicz-Drozdowska M., Laitinen E., Suvas A. (2020) A Race for Long Horizon Bankruptcy Prediction. *Applied Economics*, vol. 52, no. 37, pp. 4092–4111. <https://doi.org/10.1080/00036846.2020.1730762>
3. du Jardin P. (2021) Forecasting corporate failure using ensemble of self-organizing neural networks. *European Journal of Operational Research*, vol. 288, no. 3, pp. 869–886. <https://doi.org/10.1016/j.ejor.2020.06.020>
4. Altman E.I., Iwanicz-Drozdowska M., Laitinen E.K., Suvas A. (2017) Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-Score model. *Journal of International Financial Management & Accounting*, vol. 28, no. 2, pp. 131–171. <https://doi.org/10.1111/jifm.12053>
5. Zelenkov Y., Fedorova E., Chekrizov D. (2017) Two-step classification method based on genetic algorithm for bankruptcy forecasting. *Expert Systems with Applications*, vol. 88, pp. 393–401. <https://doi.org/10.1016/j.eswa.2017.07.025>

6. du Jardin P. (2017) Dynamics of firm financial evolution and bankruptcy prediction. *Expert Systems with Applications*, vol. 75, pp. 25–43. <https://doi.org/10.1016/j.eswa.2017.01.016>
7. D'Aveni R. (1989) The aftermath of organizational decline: A longitudinal study of the strategic and managerial characteristics of declining firms. *Academy of Management Journal*, vol. 32, no. 3, pp. 577–605. <https://doi.org/10.5465/256435>
8. Ooghe H., De Prijcker S. (2008) Failure processes and causes of company bankruptcy: A typology. *Management Decision*, vol. 46, no. 2, pp. 223–242. <https://doi.org/10.1108/00251740810854131>
9. du Jardin P., Séverin E. (2012) Forecasting financial failure using a Kohonen map: A comparative study to improve model stability over time. *European Journal of Operational Research*, vol. 221, no. 2, pp. 378–396. <https://doi.org/10.1016/j.ejor.2012.04.006>
10. Argenti J. (1976) *Corporate collapse: The causes and symptoms*. New York, NY: McGraw-Hill.
11. du Jardin P. (2015) Bankruptcy prediction using terminal failure processes. *European Journal of Operational Research*, vol. 242, no. 1, pp. 276–303. <https://doi.org/10.1016/j.ejor.2014.09.059>
12. Bellovary J.L., Giacominio D.E., Akers M.D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial Education*, vol. 33, pp. 1–42.
13. Altman E.I. (1968) Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, vol. 23, pp. 589–609.
14. Barboza F., Kimura H., Altman E. (2017) Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, vol. 83, pp. 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
15. Zelenkov Y., Volodarskiy N. (2021) Bankruptcy prediction on the base of the unbalanced data using multi-objective selection of classifiers. *Expert Systems with Applications*, vol. 185, article ID 115559. <https://doi.org/10.1016/j.eswa.2021.115559>
16. Balcaen S., Ooghe H. (2006) 35 Years of Studies on Business Failure: An Overview of the Classic Statistical Methodologies and Their Related Problems. *The British Accounting Review*, vol. 38, pp. 63–93. <https://doi.org/10.1016/j.bar.2005.09.001>
17. Bardos M. (2007) What is at stake in the construction and use of credit scores? *Computational Economics*, vol. 29, no. 2, pp. 159–172. <https://doi.org/10.1007/s10614-006-9083-x>
18. Iwanicz-Drozdowska M., Laitinen E.K., Suvas A., Altman E.I. (2016) Financial and Nonfinancial Variables as Long-horizon Predictors of Bankruptcy. *Journal of Credit Risk*, vol. 12, no. 4, pp. 49–78. <https://doi.org/10.21314/JCR.2016.216>
19. Pizzi S., Caputo F., Venturelli A. (2020) Does it pay to be an honest entrepreneur? Addressing the relationship between sustainable development and bankruptcy risk. *Corporate Social Responsibility and Environmental Management*, vol. 27, no. 3, pp. 1478–1486. <https://doi.org/10.1002/csr.1901>
20. Zelenkov Y. (2020) Bankruptcy Prediction Using Survival Analysis Technique. In: *2020 IEEE 22nd Conference on Business Informatics (CBI)*, vol. 2, pp. 141–149. IEEE. <https://doi.org/10.1109/CBI49978.2020.10071>
21. Cottrell M. (2003) Some other applications of the SOM algorithm: how to use the Kohonen algorithm for forecasting. In: *Invited lecture at the 7th International Work-Conference on Artificial Neural Networks IWANN 2003*.
22. Serrano-Cinca C. (1998) Let financial data speak for themselves. In: Deboeck, G., Kohonen, T. (eds.) *Visual Explorations in Finance with Self-Organizing Maps*. Springer, pp. 3–23.
23. Bunge M. (2017) *Causality and modern science: Fourth revised edition*. Routledge, NY.
24. Koller D., Friedman N. (2009) *Probabilistic graphical models: Principles and techniques*. MIT Press, Cambridge: MA.

25. Pearl J. (2009) *Causality*. Cambridge University Press.
26. Zhao Q., Hastie T. (2021) Causal Interpretations of Black-Box Models. *Journal of Business & Economic Statistics*, vol. 39, no. 1, pp. 272–281. <https://doi.org/10.1080/07350015.2019.1624293>
27. Spirtes P. (2010) Introduction to causal inference. *Journal of Machine Learning Research*, vol. 11, pp. 1643–1662.
28. Hair J.F., Hult G.T.M., Ringle C.M., Sarstedt M., Thiele K.O. (2017) Mirror, mirror on the wall: a comparative evaluation of composite-based structural equation modelling methods. *Journal of the Academy of Marketing Science*, vol. 45, no. 5, pp. 616–632. <https://doi.org/10.1007/s11747-017-0517-x>
29. Scanagatta M., Salmeron A., Stella F. (2019) A survey on Bayesian network structure learning from data. *Progress in Artificial Intelligence*, vol. 8, pp. 425–439. <https://doi.org/10.1007/s13748-019-00194-y>
30. Sucar L.E. (2021) *Probabilistic graphical models: Principles and applications*. Springer Nature. Cham, Switzerland.
31. Ekici A., Ekici S.O. (2021) Understanding and managing complexity through Bayesian network approach: The case of bribery in business transactions. *Journal of Business Research*, vol. 129, pp. 757–773. <https://doi.org/10.1016/j.jbusres.2019.10.024>
32. Marcot B.G., Penman T.D. (2019) Advances in Bayesian network modelling: Integration of modelling technologies. *Environmental Modelling & Software*, vol. 111, pp. 386–393. <https://doi.org/10.1016/j.envsoft.2018.09.016>
33. McLachlan S., Dube K., Hitman G.A., Fenton N.E., Kyrimi E. (2020) Bayesian networks in healthcare: Distribution by medical condition. *Artificial Intelligence in Medicine*, vol. 107, article ID 101912. <https://doi.org/10.1016/j.artmed.2020.101912>
34. Cai B., Huang L., Xie M. (2017) Bayesian networks in fault diagnosis. *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2227–2240. <https://doi.org/10.1109/TII.2017.2695583>
35. Pourret O., Naïm P., Marcot B. (2008) *Bayesian Networks: A Practical Guide to Applications*. Wiley, Hoboken.
36. Bielza C., Larranaga P. (2014) Discrete Bayesian network classifiers: A survey. *ACM Computing Surveys*, vol. 47, no. 1, article ID 5. <https://doi.org/10.1145/2576868>
37. Yuan C., Lim H., Lu T.C. (2011) Most relevant explanation in Bayesian networks. *Journal of Artificial Intelligence Research*, vol. 42, pp. 309–352. <https://doi.org/10.1613/jair.3301>
38. Lacave C., Díez F.J. (2002) A review of explanation methods for Bayesian networks. *The Knowledge Engineering Review*, vol. 17, no. 2, pp. 107–127.
39. Lacave C., Luque M., Díez F. (2007) Explanation of Bayesian networks and influence diagrams in Elvira. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 4, pp. 952–965. <https://doi.org/10.1109/TSMCB.2007.896018>
40. Altman E.I. (1983) *Corporate Financial Distress: A Complete Guide to Predicting, Avoiding, and Dealing with Bankruptcy*. Hoboken: Wiley.
41. Chen Y.C., Wheeler T.A., Kochenderfer M.J. (2017) Learning discrete Bayesian networks from continuous data. *Journal of Artificial Intelligence Research*, vol. 59, pp. 103–132. <https://doi.org/10.1613/jair.5371>
42. Weiss Y., Freeman W.T. (2001) Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, vol. 13, no. 10, pp. 2173–2200. <https://doi.org/10.1162/089976601750541769>

43. Ickstadt K., Bornkamp B., Grzegorzczak M., Wieczorek J., Sheriff M.R., Grecco H.E., Zamir E. (2010) Nonparametric Bayesian network. *Bayesian Statistics*, vol. 9, pp. 283–316.
44. Kurgan L.A., Cios K.J. (2004) CAIM discretization algorithm. *IEEE transactions on Knowledge and Data Engineering*, vol. 16, no. 2, pp. 145–153. <https://doi.org/10.1109/TKDE.2004.1269594>
45. Lustgarten J.L., Visweswaran S., Gopalakrishnan V., Cooper G.F. (2011) Application of an efficient Bayesian discretization method to biomedical data. *BMC bioinformatics*, vol. 12, no. 1, pp. 1–15.
46. Fayyad U.M., Irani K.B. (1993) Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning, In: *Proceedings of 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, pp. 1022–1027.
47. Friedman N., Goldszmidt M. (1996) Discretization of continuous attributes while learning Bayesian networks. In: *Proceedings of 13-th International Conference on Machine Learning (ICML)*, pp. 157–165.
48. Schreiber J. (2018) Pomegranate: fast and flexible probabilistic modeling in python. *Journal of Machine Learning Research*, vol. 18, no. 164, pp. 1–6.
49. Yuan C., Malone B., Wu X. (2011) Learning optimal Bayesian networks using A* search. In: *22nd International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2186–2191.
50. Chow C.K., Liu C.N. (1968) Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 462–467.
51. Elkan C. (2001) The foundations of cost-sensitive learning. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI'01)*, pp. 973–978.

About the author

Yury A. Zelenkov

Dr. Sci. (Tech.);

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

E-mail: yzelenkov@hse.ru

ORCID: 0000-0002-2248-1023