

# STATISTICAL SCORING MODEL OF LITHUANIAN COMPANIES

Laima Dzidzevičiūtė\*

Vilnius University, Lithuania

---

**Abstract.** In the banking sector of Lithuania, the necessity to apply statistical scoring models has especially increased after the transposition of the New Capital Adequacy Directive into the national legal acts. According to them, banks are allowed to apply their own statistical models to calculate capital adequacy. However, banks' internal data are not always sufficient for developing internal statistical models. The need to apply statistical scoring models increases not only for banks, but also for other institutions that grant credits. Until now, only several authors in Lithuania have proposed their own statistical scoring models for corporates; however, these models were developed using very small data samples and are suitable for specific types of companies for which they were developed only. The model proposed in this article solves these problems because it is appropriate for assessment of all companies, it is not industry-specific and has been developed using a large data sample. The objective of this study was to develop a logistic regression scoring model for assessment of corporates, using data of the external register JSC Creditinfo Lietuva<sup>1</sup>. In the proposed model, there are 19 variables characterizing all the features of a company: size, locality, age, economic sector, financial condition, past due payments, negative facts and claims from external debt collection institutions.

**Key words:** statistical scoring models, logistic regression, weight of evidence, default, probability of default

---

## Introduction

In order to make a decision to grant a credit or not, banks must have in place a credit risk assessment model. During the last decades, statistical scoring models have become more and more significant in the context of all credit risk assessment models. They may be applied not only in the decision-making process, but also in other spheres of bank activities, such as the pricing process adding a higher risk premium for riskier credits, calculating specific provisions and capital adequacy, forming a bank's strategy, allocating capital, managing past due payments, identifying the clients that could be potential clients for other products, analysing risk-adjusted profitability of a bank, in management reporting systems, etc. In Lithuania, the necessity to apply statistical scoring models especially increased after the transposition of the New Capital Adequacy Directive (prepared according to the New Basel Capital Accord) into the national legal acts. According to

---

<sup>1</sup> Author of the article thanks JSC Creditinfo Lietuva for providing the data needed for modeling.

\* Corresponding author:

Doctoral student of the Quantitative Methods and Modeling Department, Faculty of Economics, Vilnius University, Saulėtekio Ave. 9–11, LT-10222 Vilnius, Lithuania;  
e-mail: dzidzevic@yahoo.com

them, banks are allowed to apply their own statistical models for calculating capital adequacy. However, internal historical data stored at banks themselves are not always sufficient for developing internal statistical scoring models. As an inquiry of Lithuanian banks<sup>2</sup> showed, only four banks apply statistical scoring models, others indicate a too short historical observation period and insufficiency of internal data. The necessity to apply statistical scoring models increases not only for banks, but also for other companies granting credits, i. e. for consumer credit, quick credit, leasing companies to assess the risk of applicants' employers.

Until now, only several authors in Lithuania have proposed their own statistical scoring models for corporates. For instance, Grigaravičius (2003) proposed a logistic regression model to forecast the bankruptcy of the companies the shares of which are sold in stock exchange, Stoškus, Beržinskienė, Virbickaitė (2007) proposed a discriminant analysis model. However, these models were developed using very small data samples and are suitable for specific types of companies only. The model proposed in this article solves these problems because it is appropriate for the assessment of all companies, it is not industry-specific and has been developed using a large data sample.

The purpose of this study was to develop a logistic regression scoring model for the assessment of corporates using data of the external register JSC Creditinfo Lietuva. Calculations were made with the SPSS program. The final result of the proposed logistic regression model is an individual probability of default (thereinafter *PD*), i. e. the probability that a concrete company will default within one year from the scoring date. The proposed model may be applied not only by banks, but also by other companies; e. g., consumer credit, quick credit and leasing companies may apply it for assessing the credit risk of clients' employers.

In the first part of the article, data used for modeling are described, and in the second part a detailed description of the modeling process comprising all the stages is given: the definition of *Bads* and the result period, segmentation of population, sampling, analysis of input variables, choosing the model form, calculation of coefficients and ex-ante validation.

## 1. Data

Data of the Lithuanian companies from all economic sectors for 2005–2008 were obtained from the external loan register JSC Creditinfo Lietuva which collects and stores companies' information about their age, locality, legal status and legal form, economic sector, annual turnover, the number of employees, managers, members of the board, subsidiaries and branches, claims, arrests and legal processes, bankruptcies,

---

<sup>2</sup> Eight commercial banks and one branch of a foreign bank participated; the inquiry was performed in 2008 by Dzidzevičiūtė (2010<sup>b</sup>)

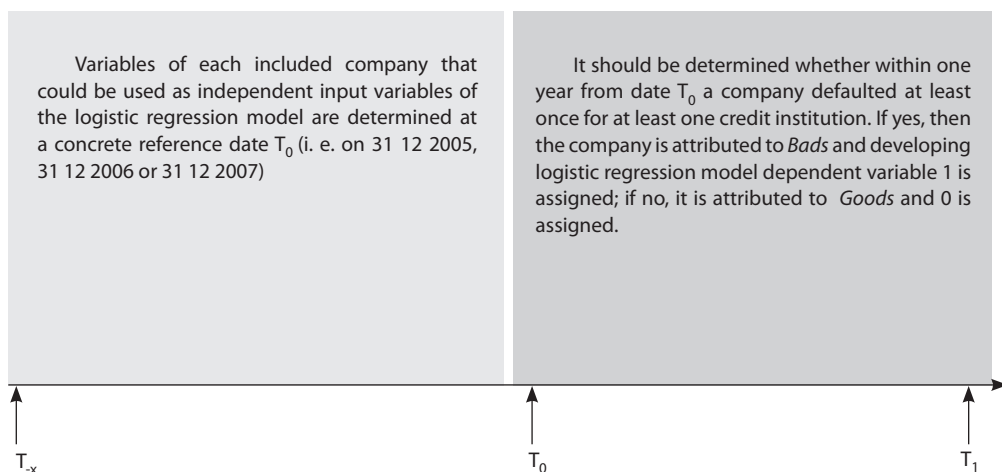


FIG. 1. Scheme of companies' data gathering

debts, changes of companies' name and address, public rating, inquiries, shares and other information from banks, leasing and telecommunication, public utility companies, public registers, etc. (<http://www.creditinfo.lt/?PageID=721>).

Each company is attributed to one of the two possible groups: to *Goods* or to *Bads*. The default criterion is used to define the status of *Bads*. Default is defined as a status of a company when payments of this company to at least one credit institution are past due more than 90 days or a bankruptcy procedure is initiated for the company<sup>3</sup>. A company is attributed to *Bads*, if it defaulted within one year from the end of a respective year, i. e. the reference date  $T_0$  (see Fig. 1).

The reference points are used: 31 December 2005, 31 December 2006 and 31 December 2007. The variables that characterize the creditworthiness of companies are taken at a concrete reference date  $T_0$ ; however, they may be calculated for the end of a year (e. g., financial ratios) or for the period  $x$  from  $T_{-x}$  to  $T_0$  (e. g., information about past due payments during two years before the reference date).

For example, variables of the company ABC are taken on 31 December 2007, i. e. the reference date is 31 December 2007. Then it is assessed whether within one year from 31 December 2007 until 31 December 2008 the ABC defaulted at least once for at least one

<sup>3</sup> The Bank of Lithuania requirements to the default definition are slightly different, i. e. *a default shall be considered to have occurred with regard to a particular obligor when either or both of the two following events have taken place: 1) the obligor is past due more than 90 days on any material credit obligation to the bank, the parent bank or any of its controlled financial undertakings, excluding the cases when the exposure amount balance does not exceed LTL 100, or another amount considered by the bank insignificant; 2) a bank considers that the obligor is unlikely to pay its credit obligations to the bank, parent bank or any of its controlled financial undertakings in full, without recourse by the bank to actions such as realising collaterals (if held)* (Bank of Lithuania, 2006). As information about unlikeliness to pay (except only the bankruptcy procedure) is not collected by JSC "Creditinfo Lietuva", the definition used in this article is narrower.

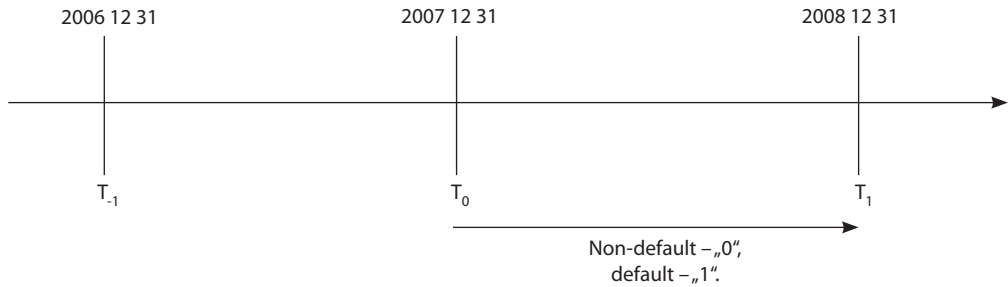


FIG. 2. Example for the ABC company

credit institution. If yes, then while forming the data array it would be attributed to *Bads* and the dependent variable 1 would be assigned. However, if ABC did not default during this one-year period, then this company would be attributed to *Goods*, and the dependent variable 0 would be assigned (see Fig. 2).

Data of each separate year were joined into one common data array and a “company-year” was used for the further analysis; e. g., if data on a concrete company are given for all three years, then the data of such company are “tripled” and used as data of three separate companies. In total, a data array of 19193 rows (“company-years”) was obtained, 376 (1.96%) of them were attributed to *Bads* and the rest 18817 (98.04%) to *Goods*.

## 2. Methodology

### 2.1. Stages of model development

The development of the logistic regression model consists of eight stages which are described in detail below (see Fig. 3).

#### Definition of “Bad” obligors and the result period

As mentioned in the first part of the paper, “Bad” is defined as a company complying with at least one of these two criteria: 1) payments of the company to

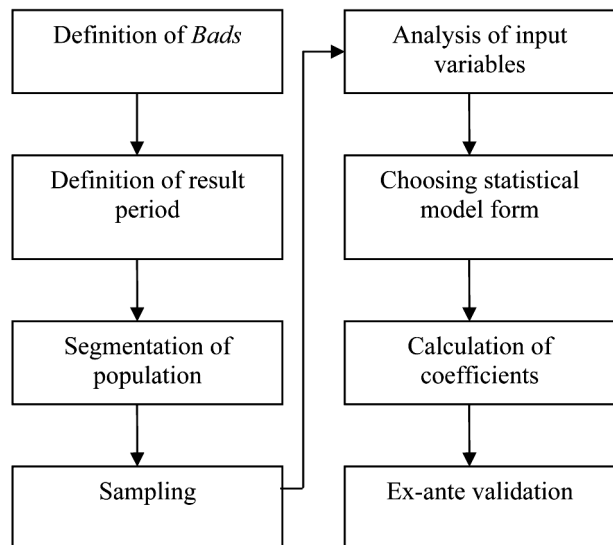


FIG. 3. Stages of statistical scoring model development

Source: L. Dzidzevičiūtė (2010<sup>6</sup>).

at least one credit institution are past due more than 90 days; 2) bankruptcy procedure is initiated for the company. The result period is equal to one year, i. e. it is assessed whether the company became “bad” within a year starting from the end of a respective year. Such duration was chosen in order to comply with the requirements of the Bank of Lithuania, prepared according to the New Capital Adequacy Directive and the New Basel Capital Accord (Bank of Lithuania, 2006; EU, 2006; BCBS, 2006).

## Segmentation of population

The proposed companies’ scoring model is generic (external) because data from an external loan register comprising information of many banks were used. As companies from all economic sectors were included, the model is recommended to assess the risk of various companies and is not industry-specific.

Also, one should notice that the model is behavioural (portfolio), i. e. it is recommended for banks to apply it for regular reassessments of already existing credit clients. JSC Creditinfo Lietuva does not gather information about credit granting date at a concrete credit institution; so, it is not possible to develop an application scoring model. The result period developing the proposed model is determined starting from the end of a respective year and not from the date of the loan granting<sup>4</sup>. However, even if the model is behavioural (and not application), it is possible to apply it even in the decision-taking process when deciding whether or not a credit should be granted.

The model was developed on a company (and not on a credit) level, i. e. it is intended for the assessment of companies and not of credits. Besides, the model may be applied for the assessment of all credit types (investment loans, working capital financing, etc.).

### 2.2. Sampling

Upon joining the data of three years into one common data array, 19193 rows (“company-years”) were obtained, of them 376 were assigned to *Bads* and 18817 to *Goods*. To adjust the initial sample several approaches were applied:

- 1) a needed sample size was calculated and compared with the initial sample size;
- 2) the structure of *Goods* and *Bads* was analyzed and the optimal structure was derived.

The following formula was applied to calculate the needed sample size (SAS, 2009; Dzidzevičiūtė, 2010<sup>a</sup>):

$$n = \left( \frac{Z_{\alpha/2} \cdot \sqrt{PD_{MAX}(1 - PD_{MAX})}}{\Delta PD} \right)^2, \quad (1)$$

---

<sup>4</sup> For the further comparison of application and behavioural scoring models, see. Dzidzevičiūtė (2010<sup>c</sup>).

where  $PD_{MAX}$  is the maximum PD that can be determined by experts analyzing the historical experience of the companies;

$\alpha$  is the significance level, i. e. 100% minus the confidence level chosen by a bank;

$Z_{\alpha/2}$  is the value of the inverse standard normal distribution function (it is possible to calculate it, e. g., applying MS Excel function NORMSINV());

$\Delta PD$  is the PD error; e. g., if the bank chooses the 95% confidence level and the 0.20% PD error, it wants to be 95% confident that the average of individual PDs calculated by the model will be no more than 20bp off  $PD_{MAX}$ .

As in the initial sample the *Bads rate* is 1.96%, in order to be conservative, a slightly higher maximum PD should be used to calculate the needed sample size (e. g., 2.4%). Suppose we want to be 95% confident that the average of individual PDs calculated by the model will be no more than 20bp off this  $PD_{MAX}$ . Then the needed sample size calculated according to formula (1) is equal to 22496. One could notice that the calculated needed sample size exceeds the initial sample, i. e. there are only 19193 rows (“company-years”) and 22496 rows are needed.

Besides, the initial proportions of *Goods* and *Bads* are 98.04% and 1.96%. Meanwhile, for logistic regression it is recommended to use 80% of *Goods* and 20% of *Bads*. To achieve such proportions, a mixture of undersampling and oversampling techniques was used, i. e. the number of *Goods* was reduced (every 26<sup>th</sup> row was deleted) and the number of *Bads* was increased (every row was repeated 13 times) to reach 20% in the total structure. After adjustment, the number of *Goods* was 18093 (79.36%) and the number of *Bads* 4706 (20.64%), in total 22799 rows.

### **2.3. Analysis of input variables, choosing statistical model form and calculation of coefficients**

The variables used in the final model were chosen in three cycles:

- 1) in the first cycle based on expert judgment, 57 variables presented in Appendix, Table A.1 were determined;
- 2) in the second cycle, 48 variables (from 57) were chosen taking into account several criteria (economic logic, monotony, individual discriminatory power of a variable);
- 3) in the third cycle, 48 variables were inputted into the SPSS program, and the final 19 variables were chosen applying the forward stepwise procedure.

#### **2.3.1. First cycle**

Initially, 57 variables characterizing all the features of a company were determined (see Appendix, Table A.1): the financial ratios, external past due payments, age, legal form,

county and economic sector of a company, information about the company's management, change of its address and name, negative facts about the company, claims from external debt collection companies, etc.

The values of all quantitative variables were joined into 10 groups by percentiles (in some cases negative values were used as a separate group, e. g., for *Total assets / Equity* because the negative values of this ratio indicate a very risky situation of a company, and small positive values, on the contrary, indicate a non-risky situation, so they cannot be mapped into the same group). For the variables *Company's group by annual turnover at the end of a year*, *Age of a company*, *Number of employees*, groups were determined based on expert judgment and not by percentiles. As all values of quantitative variables were grouped, the analysis of outliers was not made.

To code the values, the weight of evidence (thereinafter WOE) approach was applied, because applying this approach the dummies assigned accurately reflect the riskiness of a concrete group  $i$  (Dzidzevičiūtė, 2010<sup>a</sup>):

$$WOE_i = \ln\left(\frac{G_i}{B_i}\right), \quad (2)$$

where  $WOE_i$  is the WOE of the  $i$ -th group;

$G_i$  is the proportion of *Goods* in the  $i$ -th group, % from all *Goods*;

$B_i$  is the proportion of *Bads* in the  $i$ -th group, % from all *Bads*.

Table 1 provides the calculation of dummies for *County of a company*.

The higher the WOE, the lower the risk of a concrete group. When the percentage proportion of *Goods* in a respective group exceeds the percentage proportion of *Bads* in that group, WOE will be more than 0, and vice versa. As one could notice, the riskiest county is Panevėžys, as its WOE is the lowest if compared with other counties<sup>5</sup>.

The initial groups were adjusted taking into account:

- the economic logic, i. e. the risk of groups should reflect the expectations of an expert before modeling; for example, the negative values of *Total assets / Equity* should get a low WOE because they indicate a risky situation of a company, etc.;
- monotony, i. e. *Bads rate* should monotonically decrease or increase when the value of a quantitative variable increases (at least, to a certain level; for example, the distribution can be U-shaped);
- micronumerosity, i. e. if the number of values in a concrete group is very small, it is better to assign them to one of the other groups based on the similarity of *Bads rate*. For example, missing values were put into a separate group, or, in the case

---

<sup>5</sup> For the qualitative variable *County of a company*, only 10 values are possible, so WOE was calculated for each value separately. However, when there are a lot of values of a qualitative variable measured using a nominal scale, all values are sorted in ascending order of *Bads rate* and then grouped.

TABLE 1. **WOE and information value for the variable County of a company**

County	Alytus	Kaunas	Klaipėda	Marijampolė	Panevėžys	Šiauliai	Tauragė	Telšiai	Utena	Vilnius	Total
Goods	519	3.921	2.315	467	996	1.070	235	498	398	7.674	<b>18093</b>
Bads	143	1.053	533	247	546	312	52	130	156	1.534	<b>4706</b>
<b>Total</b>	<b>662</b>	<b>4.974</b>	<b>2.848</b>	<b>714</b>	<b>1.542</b>	<b>1.382</b>	<b>287</b>	<b>628</b>	<b>554</b>	<b>9.208</b>	<b>22799</b>
<b>Bads rate</b>	<b>21.60%</b>	<b>21.17%</b>	<b>18.71%</b>	<b>34.59%</b>	<b>35.41%</b>	<b>22.58%</b>	<b>18.12%</b>	<b>20.70%</b>	<b>28.16%</b>	<b>16.66%</b>	<b>20.64%</b>
<b>WOE</b>	-0.0576	-0.032	0.122	-0.7097	-0.7456	-0.1143	0.1617	-0.0036	-0.41	0.2633	
<b>Information value</b>	0.0001	0.0002	0.0018	0.0189	0.0455	0.0008	0.0003	0	0.0046	0.0258	<b>0.1</b>

Source: calculations of the author.

TABLE 2. **WOE and information value for the variable Net profit (loss) / Total assets**

Percentiles	0.1 percentile <-16.7%	0.-0.4 percentile (-16.69%-1.43%)	0.5 percentile (1.44%-3.34%)	0.6-0.7 percentile (3.35%-9.96%)	0.8 percentile (9.97%-15.52%)	0.9-1 percentile >15.52%	Missings	Total
Goods	1.384	4.788	1.749	3.897	2.036	4.157	82	<b>18093</b>
Bads	884	2.015	520	637	234	377	39	<b>4706</b>
<b>Total</b>	<b>2.268</b>	<b>6.803</b>	<b>2.269</b>	<b>4.534</b>	<b>2.270</b>	<b>4.534</b>	<b>121</b>	<b>22799</b>
<b>Bads rate</b>	<b>38.98%</b>	<b>29.62%</b>	<b>22.92%</b>	<b>14.05%</b>	<b>10.31%</b>	<b>8.31%</b>	<b>32.23%</b>	<b>20.64%</b>
<b>WOE</b>	-0.8984	-0.4812	-0.1337	0.4645	0.8167	1.0536	-0.6035	
<b>Information value</b>	0.1	0.0787	0.0018	0.0372	0.0513	0.1577	0.0023	<b>0.4290</b>

Source: calculations of the author.



of micronumerocity, assigned to one of the groups based on the similarity of *Bads rate*;

- the discriminatory power of a variable, i. e. the information value of various grouping alternatives was compared and the highest was chosen; the unproductive variables were totally excluded from the further analysis (see Appendix, Table A.1.).

Table 2 provides the adjustment of the initial grouping.

From Table 2 it is clear that some initial groups were joined (e. g., percentiles from 0.2 to 0.4) to reach the monotonously decreasing *Bads rate*, i. e. the higher the ratio, the lower the *Bads rate*. The information value for this grouping alternative was the highest.

### 2.3.2. Second cycle

From the initial 57 variables, based on their individual discriminatory power, economic logic and monotony, 48 variables were chosen and further used in the modeling. The information value was calculated using the following formula (e. g., 0.1 in Table 1 for variable *County of a company*) (SAS, 2009):

$$IV = \sum_{i=1}^n (G_i - B_i) \cdot WOE_i, \quad (3)$$

where  $IV$  is the information value of a variable.

$G_i$  is the proportion of *Goods* in  $i$ -th group, % from all *Goods*;

$B_i$  is the proportion of *Bads* in  $i$ -th group, % from all *Bads*;

$WOE_i$  is the WOE of the  $i$ -th group;

$n$  is the number of groups.

Interpreting the meaning of the information values, the following explanations were used: <0.02 – unproductive variable; 0.02–0.1 – weak predictiveness of a variable; 0.1–0.3 – medium predictiveness of a variable; >0.3 – strong predictiveness of a variable. As one could notice in Table 1, the predictiveness of the variable *County of a company* is medium, whereas the predictiveness of the variable *Net profit (loss) / Total assets* is strong. Table A.1 in Appendix provides the information values for all analyzed variables.

### 2.3.3. Third cycle

In the second cycle, 48 variables were further analyzed using the forward stepwise (Wald) procedure. The WOE values were inputted into SPSS program. Applying the forward stepwise procedure, step-by step, variables having a strong relationship with a dependent variable were included into the model, and then it was checked which variables should be excluded from the regression equation. In total, 21 steps were made; the final

model is presented in Step 21. After the procedure, 19 variables were left in the equation (Appendix, Table A.2). The PD of a company is determined applying the formulas below (Dzidzevičiūtė, 2010<sup>a</sup>):

$$PD_i = \frac{1}{1 + e^{-Z_i}}, \quad (4)$$

$$Z_i = \ln \frac{PD_i}{1 - PD_i} = b_0 + b_1 X_{1i} + \dots + b_n X_{ni}, \quad (5)$$

where  $PD_i$  is the probability that a company  $i$  will default;

$X_{1i}, \dots, X_{ni}$  are dummies of independent input variables, i. e. the WOE of a concrete group indicated in Table 3;

$b_0, b_1, \dots, b_n$  are the coefficients shown in Appendix, Table A.2, column B;

$PD_i/(1-PD_i)$  is an odd in favour of  $PD_i = 1$  (the value may vary from 0 to  $\infty$ );

$Z_i$  is a natural logarithm of the odd, also called logit.

TABLE 3. Independent input variables used in logistic regression model

Notation in regression equation	Variable	WOE*	Inclusion into regression equation
X1	<b>Company's group by annual turnover at the end of a year, thous. LTL</b>		<b>Step 12</b>
	(0–10]	0.7423	
	(10–100]	0.6225	
	(100–200]	-0.3791	
	(200–1000]	-0.2636	
	(1000–2000]	-0.1054	
	(2000–7000]	-0.0756	
	(7000–10 000]	-0.0633	
	(10 000–20 000]	0.4853	
	(20 000–100 000]	0.6751	
	>100 000 + missing values	2.1215	
X2	<b>County of a company</b>		<b>Step 5</b>
	Alytus	-0.0576	
	Kaunas	-0.0320	
	Klaipėda	0.1220	
	Marijampolė	-0.7097	
	Panevėžys	-0.7456	
	Šiauliai	-0.1143	
	Tauragė	0.1617	
	Telšiai	-0.0036	
	Utena	-0.4101	
Vilnius	0.2633		

TABLE 3 (continued)

<b>X3</b>	<b>Economic sector according to NACE 2</b>		<b>Step 3</b>
	Agriculture, forestry and fishing (Section A)	0.0755	
	Manufacturing industry, mining, quarrying and other industries (Sections B, C, D, E)	-0.3917	
	Construction (Section F)	-0.4294	
	Wholesale and retail trade, transportation and storage, accommodation and food service activities (Sections G, H, I)	0.2072	
	Information and communications (Section J)	0.4560	
	Real estate operations (Section L)	0.0533	
	Professional, scientific and technical activity, administration and services (Sections M, N)	1.1056	
	Public administration and defence, education, human health services and social work activities (Section O, P, Q)	1.2143	
	Finance and insurance activity and other services (Section R,S,T,U and K)	0.7455	
<b>X4</b>	<b>Age of a company, years</b>		<b>Step 9</b>
	<=1	0.3561	
	(1–2]	-0.3124	
	(2–3]	-0.4151	
	(3–4]	-0.6345	
	(4–5]	-0.4952	
	(5–6]	-0.2218	
	(6–7]	-0.0585	
	(7–9]	0.2315	
	(9–10]	0.2647	
	>10	0.4076	
<b>X5</b>	<b>There are / there are no records of negative information about a company during that year in external register**</b>		<b>Step 14</b>
	Yes	-2.4501	
	No	0.0414	
<b>X6</b>	<b>There are / there are no records from debt collection companies about claims to the company during that year***</b>		<b>Step 6</b>
	Yes	-2.4520	
	No	0.0602	
<b>X7</b>	<b>Number of employees</b>		<b>Step 7</b>
	<=2	0.5735	
	[3–29]	-0.0567	
	[30–39]	-0.5763	
	[40–69]	-0.0435	
	[70–99]	0.2521	
	[100–149]	0.3587	
	>150 + missing values	0.8388	
<b>X10</b>	<b>Profit (loss) before tax / Sales revenue</b>		<b>Step 20</b>
	<-15.70%	-0.8263	

TABLE 3 (continued)

	-15.69–1.24%	-0.5411	
	1.25–2.82%	0.5197	
	> 2.82%	0.5493	
	missing values	0.7843	
<b>X11</b>	<b>Net profit (loss) / Total assets</b>		<b>Step 17</b>
	<-16.70%	-0.8984	
	-16.69–1.43%	-0.4812	
	1.44–3.34%	-0.1337	
	3.35–9.96%	0.4645	
	9.97–15.52%	0.8167	
	>15.52%	1.0536	
	missing values	-0.6035	
<b>X18</b>	<b>Current assets / Amounts payable and liabilities</b>		<b>Step 18</b>
	<0.2554	-0.7301	
	0.2555–0.4460	-0.5630	
	0.4461–0.9471	-0.2989	
	0.9472–1.0741	-0.1627	
	1.0742–1.2633	0.2489	
	1.2634–1.6287	0.9182	
	1.6288–2.6630	1.1514	
	>2.6630	1.5442	
	missing values	0.1358	
<b>X20</b>	<b>Ln (Net profit (loss))</b>		<b>Step 4</b>
	<8.8968	-0.3980	
	8.8969–9.8999	-0.3597	
	9.9000–10.5798	-0.0986	
	10.5799–11.6622	0.2004	
	11.6623–12.2063	0.7272	
	12.2064–12.7400	0.8123	
	>12.7400	0.8983	
	missing values	-0.5742	
<b>X25</b>	<b>Cash and cash equivalents / Current amount payable and liabilities</b>		<b>Step 8</b>
	<0.0055	-1.0490	
	0.0056–0.0190	-0.5456	
	0.0191–0.0374	-0.4332	
	0.0375–0.1122	0.1621	
	0.1123–0.1807	0.2429	
	0.1808–1.108	0.5672	
	>1.108	0.8585	
	missing values	-0.5417	
<b>X26</b>	<b>Total assets / Equity</b>		<b>Step 2</b>
	<1.2847 + missing values	1.7303	
	1.2848–1.9445	0.8725	

TABLE 3 (continued)

	1.9446–2.9413	0.5192	
	2.9414–3.8115	0.4375	
	3.8116–7.6253	-0.2744	
	7.6254–15.3320	-0.4712	
	>15.3320 + Negative values	-0.8341	
<b>X29</b>	<b>Gross profit (loss) / Total assets</b>		<b>Step 13</b>
	<5.028%	-0.7371	
	5.029%–22.142%	-0.5062	
	22.143%–30.901%	-0.0100	
	>30.901%	0.3168	
	missing values	0.9766	
<b>X39</b>	<b>Current amount payable and liabilities / Total assets</b>		<b>Step 21</b>
	<11.2460%	0.7322	
	11.2461%–20.4192%	0.3893	
	20.4193%–28.4675%	0.1006	
	28.4676%–63.8380%	0.0566	
	63.8381%–76.5530%	-0.0404	
	76.5531%–93.2751%	-0.2669	
	>93.2751%	-0.7336	
	missing values	-0.1932	
<b>X42</b>	<b>Ln (Non-current amounts payable and liabilities)</b>		<b>Step 11</b>
	<10.3983	0.6535	
	10.3984–11.1692	0.2362	
	11.1693–11.7500	-0.1410	
	11.7501–13.3798	-0.2158	
	13.3799–13.9489	-0.2784	
	13.9490–14.6272	-0.3210	
	>14.6273	-0.3432	
	missing values	0.3342	
<b>X43</b>	<b>Sales revenue / Current assets</b>		<b>Step 16</b>
	<0.8229	-0.5463	
	0.8230–1.3839	-0.4201	
	1.3840–2.2739	-0.1449	
	>2.2740 + missing values	0.2381	
<b>X46</b>	<b>Total number of past due payments during the last year****</b>		<b>Step 15</b>
	There were no past due payments during the last year	0.3285	
	1 past due payment	-0.5268	
	2 past due payments	-0.8548	
	3–4 past due payments	-1.0295	
	5–8 past due payments	-1.4815	
	9–14 past due payments	-1.6886	
	>=15 past due payments	-1.7467	

TABLE 3 (continued)

X47	Average past due duration during the last year, days****		Step 1
	There were no past due payments during last year	0.3285	
	<6.14	0.2679	
	6.15–8.50	-0.3468	
	8.51–12.06	-0.4916	
	12.07–14.78	-0.5005	
	14.79–17.87	-1.3155	
	>17.87	-1.5708	

Source: calculations of the author.

\* WOE is multiplied by the coefficient for that variable shown in Appendix, Table A.2, column B; the lower the WOE, the riskier the group.

\*\* All negative facts about a company that are registered at JSC Creditinfo Lietuva, e. g., negative media information.

\*\*\* Only the records registered at JSC Creditinfo Lietuva are used.

\*\*\*\* Past due payments to credit institutions, leasing, telecommunication, public utility companies and other companies registered at JSC Creditinfo Lietuva.

Table 3 provides the groups of variables and their dummies (WOE) and shows the step when a concrete variable was included into the equation. One could notice that variables left in the final cycle characterize all the features of a company: *age*, *size* (group of annual turnover, number of employees and, to some extent, natural logarithms of net profit and non-current amounts payable and liabilities as bigger companies generate relatively bigger absolute amounts of net profit and take relatively bigger credits), *financial condition* (even eight financial ratios were included), *locality* (companies were grouped by counties), *economic sector* (companies were grouped according to the NACE 2 classifier), *external past due payments* (total number of all past due payments to credit institutions, leasing, telecommunication, public utility companies and other companies and the average duration of all these past due payments during the last year before the scoring date), *negative facts about a company and claims from external debt collection companies*.

For ex-ante validation, the following analyses were made:

- *analysis of the economic logic of the coefficients' mathematical signs*: the mathematical sign of a coefficient must comply with the economic logic used when developing the model. The coefficients of logistic regression equation must have a plus when the increasing value of a variable (or a dummy) indicates *ceteris paribus* an increasing risk of a company, and, on the contrary, the coefficients must have a minus when the increasing value of a variable (or a dummy) indicates *ceteris paribus* a decreasing risk of a company. In this study, groups of variables were coded with the WOE; the increasing WOE indicates *ceteris paribus* a decreasing risk of a company. Therefore, the sign of all coefficients in formulas

(4) and (5) must be a minus. As one could notice in the Appendix Table A.2, all coefficients in column B are with a minus as one could expect;

- *analysis of the significance of coefficients' inequality to 0 applying the Chi-square goodness-of-fit test*: the p values (Sig.) when applying the Chi-square goodness-of-fit test are lower than the significance level 0.05; so, the H0 hypothesis is rejected (i. e. at least one coefficient is significantly unequal to 0) (see Appendix, Table A.3).
- *analysis of the significance of coefficients' inequality to 0 applying the Wald tests*: the p values (Sig.) when applying the Wald tests prove the significance of the coefficients' inequality to 0. As one could notice in the last 21st step of the forward stepwise procedure, all Sig. values are below the significance level of 0.05, so the H0 hypothesis is rejected (i. e. all coefficients are significantly unequal to 0) (see Appendix, Table A.2).

This means that the mathematical signs of the coefficients comply with the economic logic: all coefficients are significantly unequal to 0. Besides, the overall percentage of the classification table is 83.2%<sup>6</sup>. However, a concrete institution (bank, consumer and quick credit company, leasing company), before applying the proposed model, should check its discriminatory power, the accuracy of calibration, stability, etc.<sup>7</sup> using its own data; also, a regular ex-post validation should be performed upon implementing the model.

## Conclusions

When developing the logistic regression model, the final variables were chosen in three cycles. In the first cycle, 57 variables were chosen that characterize all the features of a company: financial condition, external past due payments, age, legal form, county and economic sector, information about the company's management, change of its address and name, negative facts about the company, claims from external debt collection companies, etc. The WOE approach was applied for coding with dummies, i. e. a concrete WOE was assigned for each group of a variable's value. In the second cycle, based on the economic logic, monotony and individual discriminatory power, 48 variables were chosen for the further analysis. Then, in the third cycle, applying the forward stepwise (Wald) procedure, 19 final variables were determined. The proposed model consists of 19 variables that comprehensively characterize a company's risk. It may be applied to assess companies from all economic sectors and for all credit types (investment loans, working capital financing, etc.). The proposed model may be applied not only by banks,

---

<sup>6</sup> As data about defaults in 2009 were not received from an external register, it was impossible to make an out-of-time validation applying the validation methods presented in Dzidzevičiūtė (2010<sup>a</sup>). Besides, as the initial data sample was insufficient even for the modeling purposes, no out-of-sample validation by the above-mentioned methods was made, either.

<sup>7</sup> For validation methods, see ONB (2004), SAS (2009), Dzidzevičiūtė (2010<sup>b</sup>).

but also by other institutions that grant credits (consumer credit, quick credit, leasing companies), e. g., to assess the applicants' employers. However, before applying the proposed model, companies should validate its discriminatory power, the accuracy of calibration, stability, etc. using their own data to decide whether the model is suitable for them.

In addition to the model itself, the analysis presented in the article could be helpful for banks while developing their own models; for example, banks could choose the same or similar variables, use the results of individual discriminatory power analysis, intervals of quantitative variables, apply the proposed WOE and information value approaches, etc.

## REFERENCES

Bank of Lithuania (2006). General Regulations for the Calculation of Capital Adequacy approved by Bank of Lithuania Board Resolution No. 138 of 9 November 2006, "Valstybes ziniuos" (Official Gazette), 2006, No. 142(5442).

Basel Committee on Banking Supervision (BCBS) (2006). International Convergence of Capital Measurement and Capital Standards: Revised Structure, <http://www.bis.org/publ/bcbs128.pdf>

Dzidzevičiūtė, L. (2010<sup>a</sup>). Statistinių vertinimo balais modelių kūrimo ir taikymo ypatumai. Pinigų studijos.

Dzidzevičiūtė, L. (2010<sup>b</sup>). Statistinių vertinimo balais modelių taikymas Lietuvos bankuose. Pinigų studijos

Dzidzevičiūtė, L. (2010<sup>c</sup>). Application and behavioural statistical scoring models. Economics and Management, No. 15

European Union (2006). Directive 2006/48/EC of the European Parliament and of the Council of 14 June 2006 relating to the taking up and pursuit of the business of credit institutions (recast) (OJ 2006 L 177, p. 1)

Grigaravičius, S. (2003). Corporate Failure Diagnosis. Organizacijų vadyba: sisteminiai tyrimai, Nr. 28.

JSC Creditinfo Lietuva, <http://www.creditinfo.lt/?PageID=721>

Oesterreichische Nationalbank (ONB) (2004). Rating Models and Validation. Guidelines on Credit Risk Management.

SAS. 2009: Credit Risk Modeling Using SAS<sup>R</sup>. Course Notes.

Stoškus, S., Beržinskienė, D., Virbickaitė, R. (2007). Theoretical and Practical Decisions of Bankruptcy as one of Dynamic Alternatives in Company's Performance. Engineering Economics, No. 2 (52).



## Appendix

TABLE A.1. Analysis of variables

No.	Variable	Information value	Predictive-ness	Chosen in the 2nd cycle (Yes/No)	Variables in the final model (Yes/No)
1.	Total assets / Equity	0.528	Strong	Yes	Yes
2.	Equity / Amounts payable and liabilities	0.514	Strong	Yes	No
3.	Profit (loss) before tax / Amounts payable and liabilities	0.505	Strong	Yes	No
4.	Net profit (loss) / Amounts payable and liabilities	0.492	Strong	Yes	Yes
5.	Profit (loss) before tax / Total assets	0.443	Strong	Yes	No
6.	Net profit (loss) / Total assets	0.429	Strong	Yes	No
7.	Current assets / Amounts payable and liabilities	0.428	Strong	Yes	Yes
8.	Profit (loss) before tax / Current amount payable and liabilities	0.426	Strong	Yes	No
9.	Profit (loss) from ordinary activities / Amounts payable and liabilities	0.426	Strong	Yes	No
10.	Net profit (loss) / Current amount payable and liabilities	0.417	Strong	Yes	No
11.	Amounts payable and liabilities / Total assets	0.416	Strong	Yes	No
12.	Average past due duration during the last year, in days	0.409	Strong	Yes	Yes
13.	Total weighted past due duration during the last year, in days Comments: As reference date $T_0$ is always the last day of the year (December 31), each past due payment is weighted going backwards from the end of the respective year. Past due duration of each month is multiplied by special weight, i. e. if the beginning of the past due payment is in December, the weight is 1 (12/12), if in November, the weight is 0.917 (11/12), if in October the weight is 0.833 (10/12), etc.	0.384	Strong	Yes	No
14.	Gross profit (loss) / Amounts payable and liabilities	0.370	Strong	Yes	No
15.	Total number of past due payments during the last year	0.355	Strong	Yes	No
16.	Cash and cash equivalents / Current amount payable and liabilities	0.344	Strong	Yes	Yes
17.	Profit (loss) before tax / Sales revenue	0.339	Strong	Yes	Yes
18.	Profit (loss) from ordinary activities / Total assets	0.338	Strong	Yes	No
19.	Ln (net profit (loss))	0.335	Strong	Yes	Yes



TABLE A.1 (continued)

42.	Current amount payable and liabilities / Total assets	0.13	Medium	Yes	Yes
43.	Sales revenue / Total assets	0.115	Medium	Yes	No
44.	Ln (non-current amounts payable and liabilities)	0.102	Medium	Yes	Yes
45.	There are/there are no records of negative information about a company during that year in external register	0.101	Medium	Yes	Yes
46.	County	0.1	Medium	Yes	Yes
47.	Company's group by annual turnover at the end of a year	0.096	Weak	Yes	Yes
48.	Number of employees	0.091	Weak	Yes	Yes
49.	Sales revenue / Current assets	0.089	Weak	Yes	Yes
50.	Gross profit (loss) / Sales revenue	0.085	Weak	No	-
51.	There are/there are no records about unreliability of companies' managers during that year	0.074	Weak	Yes	No
52.	Ln (amounts payable and liabilities)	0.071	Weak	No	-
53.	Ln (amounts receivable within one year)	0.058	Weak	No	-
54.	Legal form	0.011	Unpredictive	No	-
55.	Company at least once changed the name	,01	Unpredictive	No	-
56.	Company at least once changed the name from 2000	0.009	Unpredictive	No	-
57.	Company at least once changed the address	0.002	Unpredictive	No	-

TABLE A.2. **Variables in the equation\***

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 21(t)	X1	-0.677	0.068	99.362	1	0.000	0.508
	X2	-0.958	0.060	258.601	1	0.000	0.384
	X3	-0.821	0.050	272.159	1	0.000	0.440
	X4	-0.831	0.052	256.913	1	0.000	0.435
	X5	-0.423	0.066	41.038	1	0.000	0.655
	X6	-0.755	0.056	179.921	1	0.000	0.470
	X7	-0.911	0.073	154.938	1	0.000	0.402
	X10	-0.135	0.058	5.451	1	0.020	0.874
	X11	-0.145	0.061	5.609	1	0.018	0.865
	X18	-0.164	0.048	11.707	1	0.001	0.849
	X20	-0.403	0.065	38.234	1	0.000	0.668
	X25	-0.460	0.036	162.623	1	0.000	0.631
	X26	-0.257	0.044	33.345	1	0.000	0.774
	X29	-0.336	0.048	48.871	1	0.000	0.714
	X39	-0.155	0.076	4.102	1	0.043	0.857
	X42	-0.774	0.086	80.440	1	0.000	0.461
	X43	-0.668	0.070	92.007	1	0.000	0.513
	X46	-0.330	0.058	32956	1	0.000	0.719
	X47	-0.561	0.052	114.683	1	0.000	0.571
		Constant	-1.352	0.020	4468.535	1	0.000

\*Only the last 21st step was left.

- a Variable(s) entered on step 1: X47.
- b Variable(s) entered on step 2: X26.
- c Variable(s) entered on step 3: X3.
- d Variable(s) entered on step 4: X20.
- e Variable(s) entered on step 5: X2.
- f Variable(s) entered on step 6: X6.
- g Variable(s) entered on step 7: X7.
- h Variable(s) entered on step 8: X25.
- i Variable(s) entered on step 9: X4.
- j Variable(s) entered on step 10: X30.
- k Variable(s) entered on step 11: X42.
- l Variable(s) entered on step 12: X1.
- m Variable(s) entered on step 13: X29.
- n Variable(s) entered on step 14: X5.
- o Variable(s) entered on step 15: X46.
- p Variable(s) entered on step 16: X43.
- q Variable(s) entered on step 17: X11.
- r Variable(s) entered on step 18: X18.
- s Variable(s) entered on step 20: X10.
- t Variable(s) entered on step 21: X39.

TABLE A.3. **Omnibus tests of model coefficients\***

Step 21		Chi-square	df	Sig.
Step 21	Step	4115	1	.043
	Block	6037.752	19	.000
	Model	6037.752	19	.000

\*Only the last 21st step was left.