Anatol RACUL, Petru TOMIȚA

State Agrarian University of Moldova, Chisinau, Republic of Moldova
racul@rambler.ru

A PARAMETRIC APPROACH TO ASSESSING THE CREDITWORTHINESS FOR THE MOLDOVAN RURAL DEVELOPMENT NETWORK

ABSTRACT

This paper tries to adapt to the conditions that describe the Moldovan rural development network and most importantly its banking system. The investigation has two main objectives: firstly, to present the current situation of the banking sector with respect to the Moldovan agriculture and to advocate for the implementation of statistical-based credit scoring models as a mean to make the loan granting process more efficient and secondly, to present two possible statistical-based models and their implementation on a Moldovan sample of data. The choice of these two models is grounded on the fact that, from the authors' point of view, discriminant analysis and logistic regression analysis were the two models that were refined the most by a wide variety of scholars throughout the years.

Key words: rural development, creditworthiness, Moldovan agriculture.

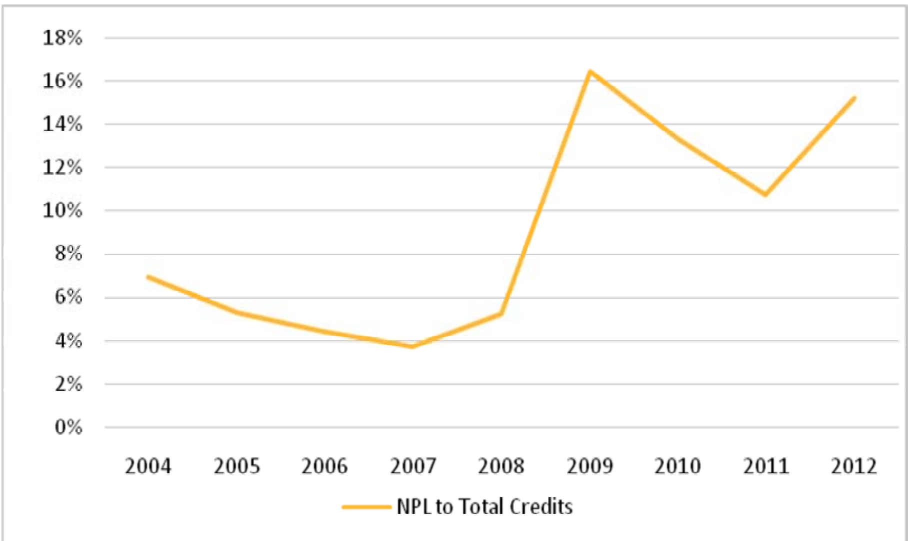
JEL Classification: G32, Q14.

1. INTRODUCTION

Since 1991, when the Republic of Moldova became independent, the performance of the banking system continuously improved in order to be successfully integrated in the global financial system. After 23 years, there are 14 licensed banks in Moldova with total assets of about 4.2 billion Euros, which represents about 86% of the Moldovan Gross Domestic Product for 2012. Some of these banks are owned by multinational financial institutions or banks, like the case of Mobiasbancă, which is owned by Groupe Société Générale since January 2007. Another example is the presence of a subsidiary of Procredit Group in Moldova.

The banking system became deeply integrated into the Moldovan economy, with an important impact on the agricultural sector, due to its role of aggregation of firms' and households' savings and that of granting credits. This integration also made the banking system rather correlated to the macroeconomic situation of the national economy and a relative high degree of pro-cyclicality can be observed. The positive trend of the Moldovan economy in the period 2000-2008 is a good example in this sense, as the share of banking credits to GDP increased from 25.2% in 2000 to 40.2% in 2007. It dropped to 39.8% in 2008 mainly due to the worsening of macroeconomic expectations related to the emerging of the international financial crisis. Taking into account the lag of the adjustment of the banks to the new macroeconomic trends, the share was 37.2% in 2010 and 42.2% in 2012. These data clearly indicate the high correlation of the penetration of the banking sector into the economy and agriculture to the expected macroeconomic trends (Mandru, 2010).

However, the effects of the recent international financial crisis on Moldovan banking system can be noticed in Graph 1. Since 2008, the ratio of the non-performing loans and provisions of loan losses to total credits (NPL) achieved record levels, to reach 16.4% in 2009. An important event that triggered this trend, besides the financial crisis, was the critical situation in the wine industry, which accounted as an important share of credit portfolios. This heavy drop in the quality of loans (almost three-fold) translated into some immediate measures from the side of the banks. The credit market collapsed and the rules for granting credits became much more severe.



Graph 1. The ratio between the non-performing loans and provisions of loan losses to total credits.

Even with this events taking place, the situation worsened in 2012. This is due to a peculiarity of the Republic of Moldova, and of most ex-Soviet Republics, namely, corruption. A more serious audit highlighted that an important part of the credit portfolio of one of the biggest banks from Moldova, the state-owned “Banca de Economii”, consists of non-performing loans. These loans were granted to some off-shore companies that were unable and, also maybe, unwilling to repay the money. A chain reaction was initiated that revealed that several other big banks, to some extent, performed similar practices in the ante-crisis period.

2. METHODOLOGY

2.1. DISCRIMINANT ANALYSIS

At its roots, discriminant analysis is a classification technique which uses data obtained from a sample of companies to draw a boundary that separates the group of reliable ones from the group of insolvent ones (De Laurentis, 2010). The discriminant function is developed in order to perform this task. If

$$Z = w_1 \bullet X_1 + w_2 \bullet X_2 + ... + w_n \bullet X_n\,, \qquad X = (X_1, X_2, ... Xn)$$

is a linear combination of the characteristics of the companies, the weights w_i have to be selected to maximize the distance between the mean values of Z for "good" and "bad" companies.

Assuming a common sample variance of the two distinct groups, the method of separation if defined as:

$$M = w^{\mathrm{T}} \cdot \frac{m_g - m_b}{(w^{\mathrm{T}} \cdot S \cdot w)^{\frac{1}{2}}},$$

where m_g represents the sample means of the "good" companies, as m_b represents the sample means of the "bad" ones. S is the common sample variance. Intuitively, M is the ratio of distance between the sample of means of the two groups and the square root of the sample variance of each group (Emel, 2003).

The value of M is maximized when

$$\frac{m_g - m_b}{(w \cdot S \cdot w^{\mathrm{T}})^{\frac{1}{2}}} - \frac{(w \cdot m_g - m_b)^{\mathrm{T}}(S \cdot w^{\mathrm{T}})}{(w \cdot S \cdot w^{\mathrm{T}})^{\frac{3}{2}}} = 0$$

which is equivalent to

$$(m_g - m_b)(w \cdot S \cdot w^{\mathrm{T}}) = (S \cdot w^{\mathrm{T}})(w \cdot m_g - m_b)^{\mathrm{T}})$$

and finally to

$$w^{\mathrm{T}} = (S^{-1}(m_g - m_b)^{\mathrm{T}})$$

The model finds the weights that applied in the initial linear combination showing the best separator of the "good" and the "bad" companies in terms of maximizing the distance between means. After the calculation of all Z values (discriminant scores), a cut-off point is selected at the average distance between the means of the two groups (Min, 2008).

2.2. LOGISTIC REGRESSION ANALYSIS

Logistic regression is one of the functional techniques used to analyze classified data. Also, logistic regression aims at solving one of the obvious flaws of linear regression approach in credit scoring. The right-hand side of the equation would be bounded to take values from 0 to 1 instead of being able to take any values from -∞ to ∞ (like in the case of linear regression). Another difference between the two methods is the way in which the coefficients are determined. Logistic regression, instead of minimizing the errors of squares, maximizes the likelihood of the occurrence of an event (Maryam, 2013). The model fitness and significance of the effects are checked through chi-square and Wald tests. The Wald test can be determined using the following equation:

$$Wald(X_i) = \left(\frac{\beta_i}{S.E.}\right)^2,$$

where β_i is the coefficient of variable *i* and S.E. is the standard error.

One of the advantages of the logistic regression analysis is the fact that it is not necessary to assume the equality of variances of the two groups and the normal distribution of the independent variables. The ratio of the likelihood of the occurrence of an event to the probability of non-occurrence of the event is defined as the odd ratio, with the following formula:

$$\frac{\pi_i}{1-\pi_i}$$

(Datoori, 2013). The following equation explains the process:

$$Logit(y) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \beta_1 \cdot X_1 + ... + \beta_n \cdot X_n,$$

where π_i is the probability of the outcome, β_i show the coefficients of logistic regression and X_i are the independent variables (Cole, 2009).

2.3. DATA DESCRIPTION USED FOR THE ANALYSIS

This analysis presented in the paper is based on secondary data obtained from Commercial Bank "Moldinconbank" SA, one of biggest banks in the Republic of Moldova. The database includes a total number of 1079 borrowers. As it is a more traditional bank, the main industries in which it operates are the agricultural sector, trading, manufacturing, services and transportation. As the bank avoids risky opportunities, the rate of its underperforming loans is around 4-5%, much lower than the national average of 10-12%. The sample contains firms well spread among the main industries, but is concentrated on companies from the agricultural sector, which is riskier. The number of non-performing loans from the sample is almost double the banks average in an attempt to find as many common characteristics of bad borrowers as possible. The number of borrowers from the sample represents around 5-6% of the total number of loans yearly issued by CB "Moldinconbank" SA.

The database includes information extracted from three financial statements of the credit applicants of the above-mentioned bank, the balance sheet, the income statement and the cash flow statement. The financial data is spread over the period from 2008 to 2012 including year-end data from the financial reports. As it can be concluded, it is the most updated data that could be obtained.

A particular useful variable that is included in the dataset is the credit rating attributed by the bank's expert and an appropriate update for every year. Originally, according to bank internal risk policies, the credit rating has only 5 possible values: 2, 5, 30, 60, and 100. Intuitively, these ratings can be interpreted as probabilities of default. The safest companies have a credit rating value of 2, representing a probability of default of approximately 2%. On the other hand, firms which have 100 as a ranking value are either nearly defaulted or have a very tough financial situation. As both credit scoring models employed in the study require that all the companies from the sample must be divided into two subcategories, a decision was taken to transform the initial credit rating into a dummy variable with only 2 possible values.

The possible values indicate whether a company is "good" or "bad". The "good" firms are the ones with sound financial situation and most likely will be repaying the loan in time with no delays. The "bad" firms should be analyzed more carefully and there is a high probability that in a short period of time they will not be able to repay the money they borrowed. Continuing the hypothesis of interpreting the initial credit ratings as probabilities of default, all the companies that initially had a credit rating of 2 or 5 were assigned as "good" and at the same time the "bad" firms were appointed as all the companies with initial credit rating of 30, 60 or 100.

As both models that are presented in this paper require inputs for only one year, initial univariate analysis was performed on separate samples, each representing one year of financial variables. As a result, the sample from the year 2010 was chosen as it presented the most relevant results.

The wide range of the size of the firms included in the sample could heavily bias the final results of the models, thus the decision was to concentrate only on SMEs, excluding large agricultural companies. On the basis of the amount of sales for 2010, about 70 companies were excluded from the analysis. As almost all of them were "good" companies, no useful information was lost in the process.

For the application of credit scoring models, financial ratios were computed using the data from the sample for the year 2010. The main purpose was to calculate some relationships between relevant sub-totals or aggregates of values, which are taken from financial statements of the firms. The main analyzed dimensions were: *Liquidity, Solvency, Profitability and Growth*.

3. DISCUSSION AND THE RESULTS OF THE ANALYSIS

Although this study aims at proposing two parametric credit scoring models for the Moldovan banking sector and its applications to the agricultural sector, there are several steps that have to be executed before the implementation of the two models. Data cleaning, univariate and bivariate analysis, transformation of indicators and defining a short list of most relevant financial ratios are crucial steps, as they have a huge impact on the final outcome of discriminant analysis or logistic regression analysis (Beaver, 2005).

From the first overview of the database, it resulted that although the total number of companies was 1079, a part of them did not contain any financial information for the year 2010, which was selected as the reference year. This was caused by two main reasons: either the companies began providing financial data to the bank from 2011 onward, in accordance to the date on which the loan was granted, or, as the final payments were due before 2010, the companies stopped providing their financial statements to the bank. Obviously these companies were excluded from the dataset. Additionally, a duplicate case analysis was performed that is necessary in order to perform the future analyses on independent observations. No duplicates were noticed in the dataset.

When performing the missing values analysis, the most alarming results were found in the variables Sales Growth and Assets Growth. Each of these variables had 14.3% missing values. This was caused by including in the ratio the values of sales and assets from the year 2009. As 14.3% of the firms began providing financial data from 2010 onward, the values for 2009 were inexistent. As these were the only ratios that quantified the growth performance of firms, the authors decided not to exclude them from the analysis. The rest of variables either did not contain missing values at all or their values were below 7%. All missing values were replaced by the respective medians of the variables, but these medians were calculated separately for the "bad" and for the "good" companies. Some of the initial ratios had to be excluded due to the excessive amount of zeroes that were contained. The variables Intangibles/Total Assets, Interest Expenses/Liabilities, Interest Expenses/EBIT were eliminated, having more that 30% of zeroes, as many of them were not true zeroes.

The last phase of data cleaning was to remove the observations that would have greatly distorted the characteristics of the financial ratios. The big companies from the sample, identified by the amount of assets and sales, were excluded. As the wide majority of them were "good" companies, the impact on the final results of the models was minor.

The univariate and bivariate analysis are vital steps in the process, as a short list of most relevant variables will be created as a result. This list is extremely useful for the two models discussed, as well as for the rest of statistical-based credit scoring models that can be applied on the same database.

The first two tasks to be performed on the financial ratios are setting a working hypothesis on the sign of the expected relation with probabilities on default (PD) and checking the structural monotonicity in regards to default risk.

Using the economic meaning of the ratios and financial knowledge, the following working hypotheses were set by the authors:

- Return on Equity – negative relationship with PD.
- Return on Assets – negative relationship with PD.

- Return on Sales – negative relationship with PD.
- Assets Turnover – negative relationship with PD.
- Inventories Turnover – negative relationship with PD.
- Receivables Turnover – negative relationship with PD.
- Receivables Period – positive relationship with PD.
- Inventories Period – positive relationship with PD.
- Payables Period – negative relationship with PD.
- Commercial Working Capital Period – positive relationship with PD.
- Cash Flow from Operations/Sales – negative relationship with PD.
- Short Term Receivables/Assets – positive relationship with PD.
- Inventories/Assets – positive relationship with PD.
- Short Term Payables/Liabilities – negative relationship with PD.
- Cash/Short Term Assets – ambiguous relationship with PD.
- Leverage – positive relationship with PD.
- Current Ratio – negative relationship with PD.
- Quick Ratio – negative relationship with PD.
- Sales Growth – negative relationship with PD.
- Assets Growth – negative relationship with PD.

One can notice the presence of the ambiguous relationship with PD of the sign of the variable Cash/Short Term Assets. It is unclear, from an economic point of view, if it is better for a company to have a high value of this ratio or not. More cash may mean the ability to produce it, but on the other hand this cash is not invested in some profitable projects and is not used for the growth of the firm. Although it is advisable to exclude this variable from the analysis, the authors decided that it should remain, but treated with great caution and with the hope that the relationship between this financial ration and PD will reveal itself during the univariate analysis.

It is also important to check if the financial indicators can assume a monotonic relationship with probability of default. As most of the ratios have necessarily positive denominators, the structural monotonicity condition is met, meaning that the value of the ratio will move in the same direction as the denominator increases or decreases. The two ratios that do not meet this requirement are Return on Equity and Leverage, due to the alternating of the sign of the Equity value. These ratios should be treated with care in the analyses and other studies should consider some structural modifications of these ratios so they would fit the requirement.

The statistical based models require that the indicators should present a decent level of discriminant power between the “good” and the “bad” borrowers. Table 1 presents the results of the independent sample t-test, which is the most suited test when dealing with one scale indicator and a nominal variable with two possible values (“good” and “bad” in our case). The results should be interpreted in accordance with the respective values of the Levene’s test.

Table 1

Independent t-test

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Net Profit/Equity	Equal variances assumed	.101	.751	.354	769	.723	2.65463	7.49833	-12.06500	17.37426
	Equal variances not assumed			.294	75.447	.770	2.65463	9.03737	-15.34699	20.65624
Operating Profit/Total Assets	Equal variances assumed	.406	.524	5.343	769	.000	12.63455	2.36483	7.99226	17.27683
	Equal variances not assumed			5.071	78.855	.000	12.63455	2.49131	7.67557	17.59352
Operating Profit/Revenues	Equal variances assumed	128.692	.000	8.208	769	.000	51.57568	6.28325	39.24133	63.91004
	Equal variances not assumed			3.681	68.400	.000	51.57568	14.01208	23.61798	79.53339
Revenues/Assets	Equal variances assumed	4.470	.035	3.752	769	.000	68.93969	18.37270	32.87309	105.00628
	Equal variances not assumed			4.537	88.982	.000	68.93969	15.19535	38.74676	99.13261
Revenues/Inventories	Equal variances assumed	2.554	.110	.884	769	.377	9735.75452	11018.74925	-11894.64131	31366.15035
	Equal variances not assumed			2.833	711.192	.005	9735.75452	3436.41137	2989.03021	16482.47882
Revenues/Short Term Receivables	Equal variances assumed	.722	.396	.554	769	.580	2758.32230	4982.33850	-7022.27543	12538.92002
	Equal variances not assumed			1.759	733.468	.079	2758.32230	1568.24232	-320.45660	5837.10120
Short Term Receivables/ (Revenues/360)	Equal variances assumed	95.863	.000	-6.590	769	.000	-1336.62391	202.82937	-1734.78884	-938.45899
	Equal variances not assumed			-2.353	67.419	.022	-1336.62391	568.01291	-2470.25277	-202.99506
Inventories/ (Revenues/360)	Equal variances assumed	31.483	.000	-4.328	769	.000	-396.98025	91.72779	-577.04682	-216.91368
	Equal variances not assumed			-2.414	69.905	.018	-396.98025	164.44906	-724.97138	-68.98912
Short Term Trade Liabilities/ (Revenues/360)	Equal variances assumed	119.446	.000	-6.191	769	.000	-1608.12378	259.77188	-2118.06990	-1098.17765
	Equal variances not assumed			-2.004	67.120	.049	-1608.12378	802.59295	-3210.05317	-6.19438
(Receivables+Inventories -Liabilities)/ (Revenues/360)	Equal variances assumed	41.297	.000	-.789	769	.430	-125.48039	158.96338	-437.53402	186.57325
	Equal variances not assumed			-.324	67.967	.747	-125.48039	386.95864	-897.65121	646.69044
Net Cash Flow from Operations/Revenues	Equal variances assumed	.004	.953	.116	769	.907	27.53790	236.78904	-437.29167	492.36748
	Equal variances not assumed			.194	121.715	.846	27.53790	141.82415	-253.22377	308.29957
Short term Receivables/Assets	Equal variances assumed	3.696	.055	-1.740	769	.082	-5.21571	2.99761	-11.10019	.66877
	Equal variances not assumed			-1.581	77.627	.118	-5.21571	3.29828	-11.78258	1.35116
Inventories/Assets	Equal variances assumed	.120	.729	.982	769	.326	2.96577	3.01970	-2.96206	8.89360
	Equal variances not assumed			.932	78.858	.354	2.96577	3.18083	-3.36569	9.29723
Short Term Trade Liabilities/Total Liabilities	Equal variances assumed	5.697	.017	1.894	769	.059	6.95710	3.67316	-.25352	14.16772
	Equal variances not assumed			2.179	86.324	.032	6.95710	3.19231	.61133	13.30287
Cash/Short Term Assets	Equal variances assumed	13.808	.000	2.861	769	.004	5.06665	1.77081	1.59046	8.54283
	Equal variances not assumed			4.103	101.882	.000	5.06665	1.23486	2.61728	7.51601
Debt/Equity	Equal variances assumed	6.218	.013	.094	769	.925	.12261	1.30341	-2.43606	2.68128
	Equal variances not assumed			.070	73.483	.944	.12261	1.74271	-3.35022	3.59545
Short Term Assets/Short Term Liabilities	Equal variances assumed	.001	.972	.198	769	.843	.12970	.65443	-1.15498	1.41437
	Equal variances not assumed			.233	87.362	.817	.12970	.55737	-.97808	1.23747
(Short Term Assets-Inventories)/Short Term Liabilities	Equal variances assumed	.613	.434	.513	769	.608	.24365	.47474	-.68830	1.17560
	Equal variances not assumed			.839	118.195	.403	.24365	.29051	-.33162	.81892
Revenues2010/Revenues2009	Equal variances assumed	.497	.481	.659	769	.510	75.13349	113.95780	-148.57179	298.83877
	Equal variances not assumed			1.379	188.111	.170	75.13349	54.48367	-32.34400	182.61099
Assets2010/Assets2009	Equal variances assumed	1.254	.263	.708	769	.479	158.86980	224.39490	-281.62943	599.36903
	Equal variances not assumed			2.180	768.948	.030	158.86980	72.87746	15.80742	301.93218

The following indicators have a statistically significant difference between the means according to the table above: Return on Assets, Return on Sales, Assets Turnover, Receivables Period, Inventories Period, Payables Period, Short Term Payables/Total Liabilities, Cash/Short Term Assets.

These are the variables that will most likely be included in the short list of variables and are most suited as inputs for the credit scoring models. Unfortunately, these variables cover two of the four dimensions that were mentioned when motivating the usefulness of financial ratios, as *Solvency* and *Growth* are not represented, and covering *Liquidity* and *Profitability*.

Another important test, which is also feasible when the independent variable has more than two possible instances, is ANOVA. It also provides the F-ratio, a statistical value that measures the strength of the discriminatory power of a financial ratio. There is a positive relationship between the value of F-ratio and the discriminatory power, as a high F-ratio means a strong discriminatory power of the indicator. As the study is performed on a single sample, the F-ratios of different financial ratios can be compared. It is obviously that the indicators that were mentioned as having a statistically significant difference between the means of the “good” and the “bad” companies have also a high value of the F-ratio. As it can be noticed, a relatively small number of variables have a high value of F-ratio, and will be reflected in the capacity of the models to separate the “good” and the “bad” companies and also on the amount of inputs that it would be feasible to be included in the models (Table 3).

The discriminatory power of the indicators can be also calculated using ROC curves and AuROC measures. The ROC curve is a plot that illustrates the performance of the indicator in separating the “bad” companies from the “good” ones. A quantifiable approach can be implemented by measuring the AuROC, the area under the ROC curve. It also has a positive relationship with strength of the discriminatory power, as the higher the AuROC, the better the discriminatory power of the indicator. An important specificity is that AuROC is calculated using the relative ranks of the observations, instead of absolute values, making it less sensible to extreme values. As the range of the values of AuROC is between 0.5 and 1, a threshold can be chosen for a variable to be suitable into entering into the short list (Table 2).

The first thing to point out is that the three ratios that were initially not aligned to the working hypothesis have an AuROC value below 0.5. At the same time, the indicators with high values of F-Ratio have the highest values for AuROC values.

The last step before making the final short list of indicators is to examine the correlations between the pairs of variables. As high correlations among variables indicate similar information, the presence of two highly correlated variables would cause more harm than good from the perspective of the scoring models. High Pearson’s Correlations, significant at 1%, can be noticed between Payables Period and Receivables Period (0.76), Quick Ratio and Current Ratio (0.894). Additionally, high Spearman's Correlations, significant at 1%, can be noticed between Inventories Period and Inventories Turnover (0.776), Receivables Period and Receivables Turnover (0.831) and between Return on Assets and Return on Sales (0.872).

As Payables Period is not aligned with the working hypothesis, but at the same time is highly correlated to Receivables Period, it can be successfully replaced in the scoring models. Also, the high correlation between the Receivables Period and Receivables Turnover makes it even clearer that the variables Receivables Period should not make it to the short list and excluded from the models when possible. Due to high correlations between Current Ratio and Quick Ratio and between Return on Assets and Return on Sales, great care should be taken when including them into the scoring models or the short list, as one should be preferred over the other.

The outliers' treatment is also a necessary step, as it is an impediment previously encountered and greatly affected the shape of the distributions of the variables. According to the approach that has been chosen, an outlier is considered an observation that has a value that is greater than the 3rd quartile plus 3xInterquartile Range or has a value that is lower than the 1st quartile minus 3xInterquartile Range. As all the outliers were counted, the results show that 3 variables have more than 10% of outliers: Inventories Turnover-13.2%, Cash Flow from Operations/Sales-20%, Leverage-13.5%. A separate treatment is necessary for these 3 ratios.

Table 2

Calculated AuROCs

Return on Equity	0.549
Return on Assets	0.741
Return on Sales	0.676
Assets Turnover	0.735
Inventories Turnover	0.675
Receivables Turnover	0.729
Receivables Period	0.71
Inventories Period	0.669
Payables Period	0.296 Non-alignment with hypothesis
Commercial Working Capital Period	0.687
Cash Flow from Operations/Sales	0.521
Short Term Receivables/Assets	0.557
Inventories/Assets	0.454 Non-alignment with hypothesis
Short Term Payables/Liabilities	0.559
Cash/Short Term Assets	0.713
Leverage	0.422 Non-alignment with hypothesis
Current Ratio	0.536
Quick Ratio	0.509
Sales Growth	0.632
Assets Growth	0.656

Table 3

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Net Profit/Equity * Risk Category Transformed	Between Groups	(Combined)	436.935	1	436.935	.125	.723
	Within Groups		2680807.764	769	3486.096		
	Total		2681244.699	770			
Operating Profit/Total Assets * Risk Category Transformed	Between Groups	(Combined)	9897.586	1	9897.586	28.544	.000
	Within Groups		266646.392	769	346.744		
	Total		276543.978	770			
Operating Profit/Revenues * Risk Category Transformed	Between Groups	(Combined)	164930.071	1	164930.071	67.379	.000
	Within Groups		1882368.068	769	2447.813		
	Total		2047298.140	770			
Revenues/Assets * Risk Category Transformed	Between Groups	(Combined)	294678.501	1	294678.501	14.080	.000
	Within Groups		16094669.20	769	20929.349		
	Total		16389347.70	770			
Revenues/Inventories * Risk Category Transformed	Between Groups	(Combined)	5876910671	1	5876910671	.781	.377
	Within Groups		5.789E+12	769	7527910719		
	Total		5.795E+12	770			
Revenues/Short Term Receivables * Risk Category Transformed	Between Groups	(Combined)	471736934.1	1	471736934.1	.306	.580
	Within Groups		1.184E+12	769	1539133601		
	Total		1.184E+12	770			
Short Term Receivables/ (Revenues/360) * Risk Category Transformed	Between Groups	(Combined)	110771570.7	1	110771570.7	43.427	.000
	Within Groups		1961543165	769	2550771.346		
	Total		2072314736	770			
Inventories/ (Revenues/360) * Risk Category Transformed	Between Groups	(Combined)	9771194.585	1	9771194.585	18.730	.000
	Within Groups		401178848.8	769	521689.010		
	Total		410950043.4	770			
Short Term Trade Liabilities/ (Revenues/360) * Risk Category Transformed	Between Groups	(Combined)	160342557.4	1	160342557.4	38.323	.000
	Within Groups		3217514103	769	4184023.541		
	Total		3377856660	770			
(Receivables+Inventories -Liabilities)/ (Revenues/360) * Risk Category Transformed	Between Groups	(Combined)	976251.162	1	976251.162	.623	.430
	Within Groups		1204842708	769	1566765.549		
	Total		1205818959	770			
Net Cash Flow from Operations/Revenues * Risk Category Transformed	Between Groups	(Combined)	47018.805	1	47018.805	.014	.907
	Within Groups		2673371880	769	3476426.372		
	Total		2673418899	770			
Short term Receivables/Assets * Risk Category Transformed	Between Groups	(Combined)	1686.695	1	1686.695	3.027	.082
	Within Groups		428437.813	769	557.136		
	Total		430124.508	770			
Inventories/Assets * Risk Category Transformed	Between Groups	(Combined)	545.362	1	545.362	.965	.326
	Within Groups		434774.323	769	565.376		
	Total		435319.685	770			
Short Term Trade Liabilities/Total Liabilities * Risk Category Transformed	Between Groups	(Combined)	3001.002	1	3001.002	3.587	.059
	Within Groups		643304.884	769	836.547		
	Total		646305.886	770			
Cash/Short Term Assets * Risk Category Transformed	Between Groups	(Combined)	1591.662	1	1591.662	8.187	.004
	Within Groups		149512.776	769	194.425		
	Total		151104.438	770			
Debt/Equity * Risk Category Transformed	Between Groups	(Combined)	.932	1	.932	.009	.925
	Within Groups		81002.807	769	105.335		
	Total		81003.739	770			
Short Term Assets/Short Term Liabilities * Risk Category Transformed	Between Groups	(Combined)	1.043	1	1.043	.039	.843
	Within Groups		20420.118	769	26.554		
	Total		20421.161	770			
(Short Term Assets-Inventories)/Short Term Liabilities * Risk Category Transformed	Between Groups	(Combined)	3.681	1	3.681	.263	.608
	Within Groups		10746.177	769	13.974		
	Total		10749.858	770			
Revenues2010/Revenues2009 * Risk Category Transformed	Between Groups	(Combined)	350007.227	1	350007.227	.435	.510
	Within Groups		619190557.2	769	805189.281		
	Total		619540564.4	770			
Assets2010/Assets2009 * Risk Category Transformed	Between Groups	(Combined)	1564921.488	1	1564921.488	.501	.479
	Within Groups		2400834267	769	3122021.153		
	Total		2402399188	770			

Also, this is another confirmation of the manipulation of the entrepreneurs of the ratio Leverage and of the indicator Cash Flow from Operations. The manipulation of the latter is due to tax reasons, as low indicators of profits translate into lower taxes. The very high number of outliers in the Cash Flow from Operations/Sales will probably be a serious limitation in considering this ratio in the short list. As the rest of the ratios contained less than 10% of outliers, these observations were substituted by the lower and upper bounds respectively. This procedure was applied to all the variables, and the three above-mentioned variables will be additionally transformed. Twenty new variables were created in result. In order to avoid the excessive accumulation of observations in the boundaries of variables, for the three problematic indicators an additional logistic transformation was performed, thus creating three new variables. It is necessary to say that, as ROC curves take into account the rank of the observations, rather than the absolute values of the differences, these transformations did not improve the discriminatory power of the variables. Also, the values of AuROC remained practically the same, due to the same reason.

Using all the information that the univariate and bivariate analysis have provided, the authors concluded that the variables having the most discriminatory power are aligned with the working hypotheses, and are most useful in separating the "good" and the "bad" borrowers are: Return on Sales, Return on Assets, Assets Turnover, Receivables Period, Cash/Short Term Assets, and Inventories Period. The low number of indicators from this short list is due to the initial low number of ratios that could be calculated and, of course, of the less than great quality of financial data.

The most tangible result of the discriminant analysis is providing canonical coefficients for the variables in the model. These coefficients, presented in Table 4, if multiplied by the respective variables and computing the sum will result in the final discriminant score of each company.

Table 4

Canonical Discriminant Function Coefficients

	Function
	1
ROS3	.010
Receivables_Period3	-.004
Inventories_Period3	-.002
Commercial_WC_Period3	.000
CashSTAssets3	.033
(Constant)	.621

Unstandardized coefficients

The score that is obtained for each firm can be computed using:

Discriminant Score = 0.621 +
+ 0.10*Return on Sales –
- 0.004*Receivables Period –
- 0.002*Inventories Period +
+ 0.000*Commercial Working Capital Period +
+ 0.33*Cash/Short Term Assets

In order to be able to rank the firms using the obtained scores, Table 5 is useful, as it compares the means of the scores of "good" and "bad" subcategories. In this case, a high score for a firm means lower probability of default.

Table 5
Comparison of the means

Functions at Group Centroids

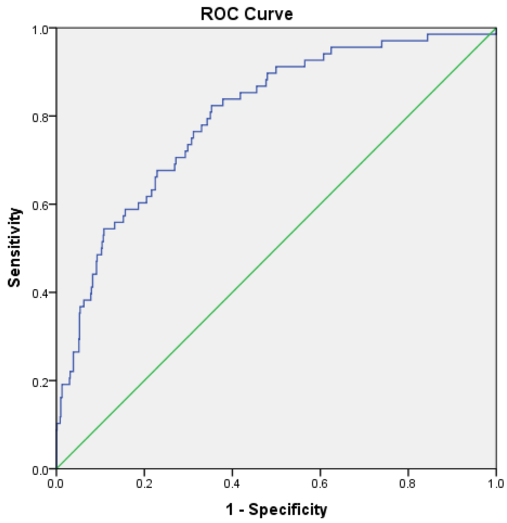
Risk Category Transformed	Function
	1
Performing	.129
Default	-1.335

Unstandardized canonical discriminant functions evaluated at group means

A very useful tool may represent the ROC curve (Graph 2) for the model and the calculation of AuROC. In this way, the model performance is measured at each possible cut-off and it successfully plots the possible tradeoff that can be made on Type 1 errors, which are represented on the Y-axis, and Type 2 errors, which are represented on X-axis.

An important observation is that the model is reconfirmed as being statistically significant at 1% (Sig. 0.00). Also, the AuROC value of 0.800 is an acceptable level of precision, suggesting a rather strong model. As the model's value of AuROC is higher than the AuROC value of every solitary variable, it can be concluded that the model "pooled" the discriminatory power of each indicator from the model into creating a score that would maximize the effectiveness. In general, a model with AuROC higher than 75% is considered rather useful in separating the two subgroups.

Some final considerations are presented in order to conclude the discussion about the discriminant analysis model. There are several strong points of this credit scoring model. Firstly, all the signs of the canonical coefficients are in order with the working hypotheses of the variables, respectively. Any contradiction from this point of view should be treated extremely cautionary and would mean that the model requires some further transformations or is of no great use at all. Additionally, the variables from the short list that was elaborated using univariate and bivariate analysis have also been selected by the model as relevant (except for Return on Assets, which is correlated to Return on Sales). At the same time, all the tests suggested that the model is statistically significant at 5%.



Area Under the Curve

Test Result Variable(s): Discriminant Scores from Function 1 for Analysis 1

Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.800	.028	.000	.745	.854

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Figure 1. ROC curve and AuROC.

At the same time, the statistical indicators that assess the capability of the model to separate the two subgroups presented rather good results. All these positive facts suggests that the discriminant analysis model can indeed be implemented on Moldovan data and provide rather accurate scores and eventually probabilities of default.

3.1. LOGISTIC REGRESSION ANALYSIS

As opposed to discriminant analysis, where the stepwise procedure excluded all unsatisfactory indicators from the model, in Logistic regression analysis (LOGIT) the expert has an important role of selecting the short list of variables that would serve as inputs for the model and also the type of the stepwise method that is to be used by the model. Including a high number of indicators would do more harm, thus making the model unstable. High correlation among inputs can also cause convergence issues of the algorithm.

When choosing the short list of indicators that would serve as inputs for the LOGIT model, the authors included four variables from the initial short list, preferring Return on Assets over Return on Sales (due to the high colinearity among them).

Table 5
Variables in the LOGIT model

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Receivables_Period3	.006	.001	52.361	1	.000	1.006
	Constant	-3.346	.224	222.518	1	.000	.035
Step 2[b]	ROA3	-.035	.008	20.778	1	.000	.966
	Receivables_Period3	.005	.001	43.328	1	.000	1.005
	Constant	-3.193	.229	195.214	1	.000	.041
Step 3[c]	ROA3	-.034	.008	17.825	1	.000	.966
	CashSTAssets3	-.097	.037	6.991	1	.008	.908
	Receivables_Period3	.005	.001	33.139	1	.000	1.005
	Constant	-2.790	.254	120.887	1	.000	.061

a. Variable(s) entered on step 1: Receivables_Period3.

b. Variable(s) entered on step 2: ROA3.

c. Variable(s) entered on step 3: CashSTAssets3.

Also the transformed versions of the indicators, without outliers, were preferred over the initially calculated ratios. The additional ratios were chosen that would maximize the explained performance of the company, in terms of the four dimensions already mentioned. The ratios should also have a significant discriminant power described in the univariate and bivariate analysis.

Another criterion was to have a model with all the variables statistically significant. This, of course, required several try-outs to be made on combinations of indicators. Also, the LOGIT model should not perform much worse than the discriminant analysis model in terms of the value of AuROC. The chosen method of iteration was “forward stepwise” as it generally presents more precise results.

The model that fitted all the criteria presented above included as initial inputs the following variables: Return on Assets, Cash/Short Term Assets, Receivables Period, Short Term Payables/Total Liabilities, Sales Growth, Assets Turnover, Inventories Turnover, and Commercial Working Capital Period.

As it can be noticed in Table 5, the LOGIT model included only three of the eight input variables. The systematic component of the model that can be calculated using these results is:

-2.790 - 0.34*ROA - 0.97*Cash/Short Term Assets + 0.05*Receivables Period

Using the following equation:

$$\pi_i = \frac{1}{1 + e^{-(systematic\ component)}} \quad i = 1, \dots, n$$

the predictive probability values of the model can be computed. These values can be interpreted as probabilities of default in our case.

Even though the model selected only three variables, the significance of Assets Turnover (p-value 0.068) and Inventories Turnover (p-value 0.076), which are not very far from the selected 5% threshold, suggest that if the LOGIT model is implemented to a more qualitative dataset, the results would be more favorable.

4. CONCLUSIONS

It is no wonder that the loan granting process for the agricultural sector of the Republic of Moldova has important flaws and inefficiencies. However, certain aspects of the risk management decision processes can be improved, thus positively impacting the assessment of credit risk. This study is performed mostly from the model developer perspective and presents two main applications.

Firstly, the study identifies the relevant financial data of a company that can be used for credit rating purposes of agricultural companies. Even in the expert-based rationale, which is currently used by the risk management divisions, these insights might prove to have significant value. Also the univariate and bivariate analysis provide very useful tools to deal with heavy discrepancies from the various financial statements of Moldovan firms.

Secondly, there are at least two statistical-based models that can successfully be used as alternatives or even substitute the current credit rating system that is has been used. Even considering their limitations, the provided results suggest a significant discriminant power of both models and a rather high level of applicability. The low quality of Moldovan data would label the LOGIT model as more appropriate, but if definitely needs some more refinement before applying it to real-life situations. The authors also hope that the results presented will contradict the prejudgment of Moldovan risk managers, which state that the specificity of the rural development network data makes the application of statistical-based models to credit risk management nearly impossible.

REFERENCES

1. De Laurentis, G., Maino, R., Molteni, L., (2010), *Developing, Validating and using Internal Ratings*, United Kingdom, John Wiley & Sons.
2. Emel, A., B., Oral, M., Reisman, A., and Yolalan, R., (2003), *A credit scoring approach for the commercial banking sector*, Socio-Economic Planning Sciences, 37(2): 103-123.
3. Min, J., H., and Lee, Y., C., (2008), *A practical approach to credit scoring*, Expert Systems with Applications, 35(4): 1762-1770.
4. Beaver, W., H., McNichols, M., F., and Rhie, J., W., (2005), *Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy*, Review of Accounting Studies, 10(1): 93-122.
5. Maryam, K., A., and Sara, M., (2013), *Comparing Logit, Probit and Multiple Discriminant Analysis Models in Predicting Bankruptcy of Companies*, Asian Journal of Finance & Accounting, 5(1).
6. Dastoori, M., and Mansouri, S., (2013), *Credit Scoring Model for Iranian Banking Customers and Forecasting Creditworthiness of Borrowers*, International Business Research, 6(10).
7. Mandru, L., Khashman, A., Carstea, C., David, N., and Patrascu, L., (2010), *The diagnosis of bankruptcy risk using score function*, in Proceeding of the WSEAS Conference Recent Advances in artificial intelligence, knowledge, engineering and data base, Cambridge.
8. Cole, R., A., and Wu, Q., (2009) *Predicting bank failures using a simple dynamic hazard model*, CFR Seminar Series Library–2009.