



¹ Rocsana ȚONIȘ, ² Radu BUCEA-MANEA

ANALYSIS MODELS FOR THE ECONOMICAL-FINANCIAL ACTIVITY OF THE ROMANIAN SME'S FROM A REPRESENTATIVE SAMPLE

^{1,2} CENTER FOR EUROPEAN STUDIES AND MOBILITY, ROMANIA

ABSTRACT: Romanian SMEs activity analysis represents an area of strategic interest to the national economy, having in mind that official statistics of the Ministry for SMEs, Commerce, Tourism and Liberal Professions (MIMMCTPL) show that most Romanian active companies are SMEs. SMEs produce a percentage of 70 % of country GDP and ensure approximately 70% of jobs [1]. In this connection I create a representative sample for with the Romanian SMEs for the years 2008-2010. I spread this sample in four clusters, using TwoStepCluster method and I have accomplished one linear regression model for each cluster. Thus it can be analyzed the SMEs collar and establish different strategies for future development.

KEYWORDS: linear regression, clusters, SMEs, economical-financial analysis

JEL Classification: C5 - Econometric Modeling

INTRODUCTION

The data were taken from the CD list of Romanian companies [2] which has inscribed a database on companies registered in Romania and from the database of the Ministry of Finance. From the above mentioned CD we have been taken attributes such as: name, address, phone, phone number, web address, number of employees, turnover and profit. We selected only companies with the number of employees under 250, while turnover less than 50.000.000 Euros.

From the Ministry of Finance database we have been selected data regarding the national activity classification description (CAEN), companies' assets and liabilities by introducing the company CUI and the online database release data from companies' balance sheet for years in which they have been active. We had chosen only dates for the past three years, 2008-2010. [3]

The lack of time determined us to choose a representative sample. In further research we may develop data source as to contain all SMEs in Romania. We calculated the number of companies needed to ensure that sample to be representative using formula 1. [4]:

$$n = t^2 * p * (1-p) / e^2, \quad (1)$$

where: n = sample size,

t = the value of theoretical accepted probability (take usually $t=1.96$ for a level of confidence by 95 percent),

p = the percentage in population that has the sample characteristic (usually=0.50),

e = the limit error admitted for representativeness (accepts values between 1 percent și maximum 5%)

After applying this formula, the sample got to be corrected in terms of total population size as follows (formula 2):

$$n_1 = n / [1 + (n-1)/N] \quad (2)$$

where: n_1 = the size of corrected sample

n = the size of the sample obtained above

N = total population

Thus for a confidence level of 95% and a sample error of +/-5% from the total number of 374210 SMEs are enough 384 companies that the sample to be representative. The level of confidence expressed likelihood that the real value of an indicator to be in the range of confidence. We had chosen the value of 95%. The error of sampling (confidence interval) represents the area values calculated indicator which can have a sample. Total population (optional) represents the total number of individuals, households and firms etc. for the sample to be representative. The sample size represents the number of questionnaires, records, balance sheets, needed to have representative answers for total population.

Decision models for each cluster

Data taken from the Ministry of Finance site were then introduced in SPSS and split in four clusters using Two Step Cluster method depending on the net profit, turnover, the number of employees and year, as noted in Figure 1. As evaluation fields were chosen total debts, capital and CAEN. [5]

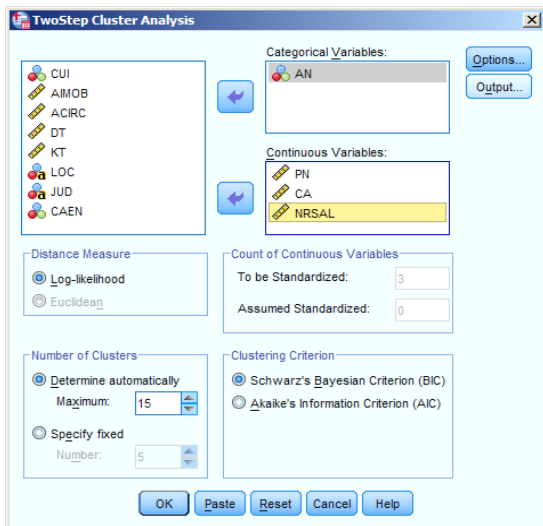


Figure 1. Criteria of splitting sample in clusters

Applying this method results four clusters whose extent of cohesion and separation scattering is good (almost 1) as it can be noted in the figure 2.

Initial data base had contained 1797 records, relating to 599 SMEs for years 2008-2010, but some of these fields were null. Finally database contains 1533 records relating to 511 companies for years 2008-2010.

Notice in figure 3 that the first cluster contains 534 records (34.8 % of total records) in particular from 2008 companies which have an average of 7 employees, with a turnover average of 1010895 RON and a net profit average of 48643 RON, most of them being small companies.

The second cluster consists of 100 records (6,5%) from 2008 (38%), 2009 and 2010, companies that have on average of 66 employees, with a turnover average of 20826339 RON and a net profit average of by 285195 RON, most of them being medium-sized enterprises.

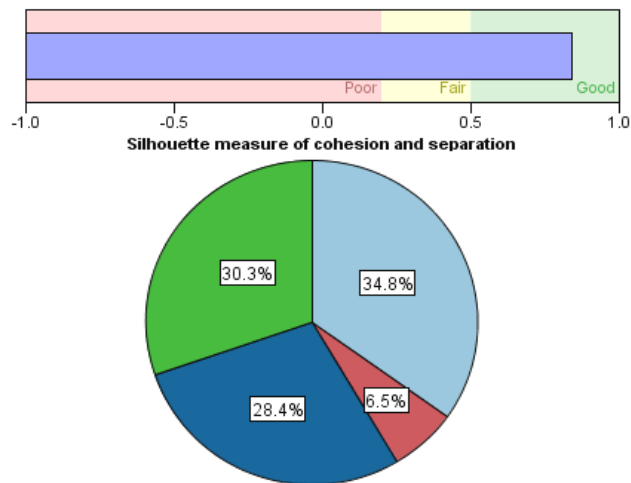


Figure 2. The Two Step Cluster model representatives and clusters size

Cluster	1	4	3	2
Size	34.8% (534)	30.3% (464)	28.4% (435)	6.5% (100)
Inputs	AN 2008 (100.0%) NRSAL 7.22 CA 1010895.39 PN 48643.51	AN 2009 (100.0%) NRSAL 6.63 CA 986969.68 PN 27453.12	AN 2010 (100.0%) NRSAL 7.38 CA 1132991.33 PN 15131.19	AN 2008 (38.0%) NRSAL 65.95 CA 20826339.24 PN 285195.47
Evaluation Fields	DT 461,649.81 KT 237,794.71 CAEN 22 (5.2%)	DT 440,366.77 KT 274,212.92 CAEN 22 (5.2%)	DT 663,438.11 KT 345,018.23 CAEN 22 (5.5%)	DT 6,330,549.66 KT 3,855,311.41 CAEN 28 (6.0%)

Figure 3. Each cluster characteristics

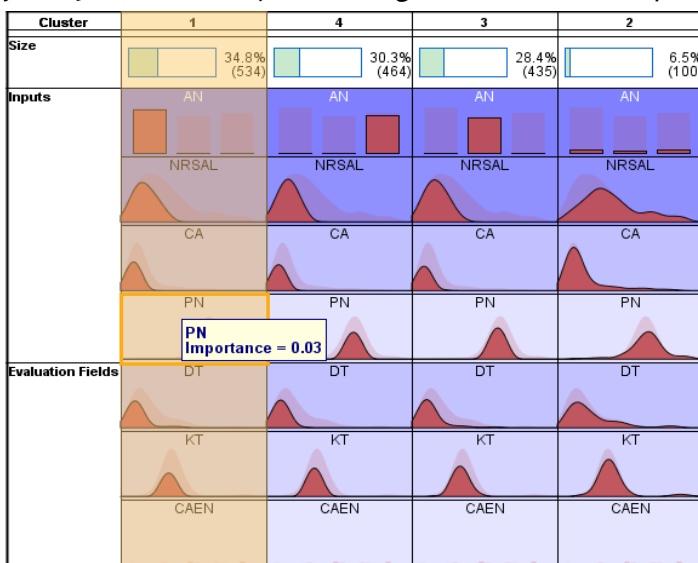


Figure 4. Graphic representation of clusters characteristics

The third cluster consists of 435 records (28.4%), in particular from 2010, companies which have on average of 7 employees, with a smaller turnover average of 1132991 RON and net profit average of 15131 RON. They have a turnover average bigger than the turnover average from the fourth cluster, but companies in the forth cluster are more profitable.

The fourth cluster consists of 464 records (30.3%), in particular from 2009, companies which have on average of 6 employees, with a smaller turnover average of 986969 RON and a net profit average of 27453 RON, being in particular small enterprises.

These characteristics are better represented in the figure 4.

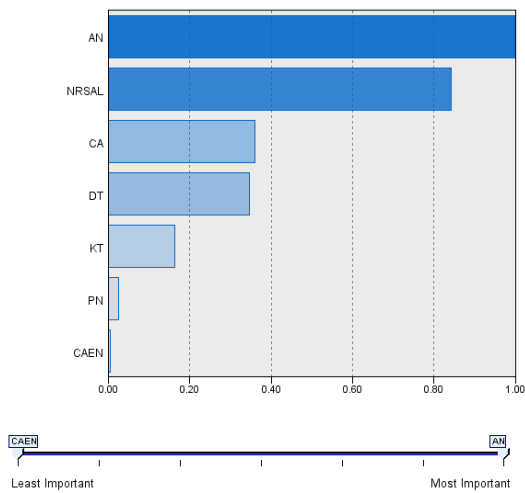


Figure 5. The weight of splitting clusters criteria

The weight of factors taken into account was established by SPSS: the most important factors are year and the number of employees, factors with average importance are turnover, debts total and capital and less important factors are net profit and CAEN. (Figure 5)

For each cluster we design a linear regression model, using SPSS. In the figure 6 we may observe the dependent (net profit) and independent (total expenditure, total revenues and number of employees) variable that form the model equation.

For the first cluster we design a linear regression equation that calculates the Net profit depending on total expenditure, total revenues and number of employees.

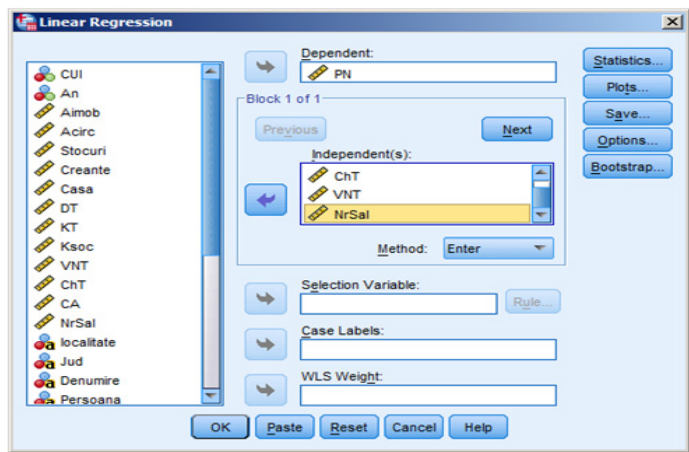


Figure 6. Dependent and independent variables for linear regression model

Table 1. Model summary for the first cluster

R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
.978	.956	.955	30794.887	1.935

We start from the null hypothesis H_0 : the equation coefficients are 0 with the alternative hypothesis H_1 : the equation coefficients are different from 0. The model rejected the null hypothesis.

- a. Predictors: (Constant), NrSal, VNT, ChT
- b. Dependent Variable: PN

Model summary contains information concerning the coefficient of correlation and standard error of estimation. It has to be noted the determination coefficient R Square (R^2) expressing what percent of dependent variable is explained by independent variable, by the regression equation.

The data series is well represented by the model because R^2 is 0,956. R^2 takes values within the range 0 and 1. As R^2 takes values closest to 1, the regression model adjusts better data from sample. It is our case, too. (Table 1)

R^2 adjusted indicates that 95.5% of dependent variable (profit net) variance was explained by the independent variables (total revenues, total expenses, number of employees). The rest of 4.5 % is explained by residual values. [6]

Table 2. Anova model for the first cluster

Model	Sum of Squares	Df	Mean Square	F	Sig.	%
Regression	10856503574481.700	3	3618834524827.230	3816.028	.000	95.55
Residual	505457250183.095	533	948325047.248			4.449
Total	11361960824664.800	536				

- a. Predictors: (Constant), NrSal, VNT, ChT
- b. Dependent Variable: PN

In the Anova table, important information is F statistics, used to test the independent variables significance (only one in simple regression case, the procedure gets substance in the case of multiple regression). The column Sig. is displayed the likelihood critical test, so that if $Sig < 0.05$ (probability of 95 %) or 0.001 (probability of 99 %) rejects the null hypothesis of significance lack (H_0) of the independent variables in favor of alternative hypothesis that the regression model is one significant (H_1).

Table 3. Coefficients for the first cluster

Model	Unstandardized Coefficients		Standardized Coefficients	T	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	1371.323	1684.824		.814	.416	-1938.386	4681.033
VNT	.822	.009	12.029	93.721	.000	.805	.839
ChT	-.821	.009	-11.509	-89.544	.000	-.839	-.803
NrSal	-489.636	171.179	-.032	-2.860	.004	-825.905	-153.368

a. Dependent Variable: PN

Coefficients table contains information regarding the equation coefficients: Column B presents the coefficient value, Std. Error present the standard error of the coefficient (standard deviation in the distribution of the poll coefficient), Beta presents the standardized coefficient value (shows with how many deviations Y changes if X is changed by a standard deviation), T statistics - presents the coefficient significance test, Sig. Presents the critical test likelihood. Therefore, a coefficient is significant (different from zero in the regression equation) if Sig < 0.05, or 0.001.

With a likelihood of 95 % the regression equation constant is covered by the interval [-1938.386 ; 4681.033], and equation parameters are covered by the intervals: [0.805 ; 0.839] for total revenues, [-0.839 ; 0.803] for total expenses and [-825.905 ; 153.368] for the number of employees. Notice that the value of significance coefficient t (.Sig) is less than 0.001, for all parameters, which confirms the validity of all coefficients of the equation parameters. Thus the model adjusts correctly the data series. But the value of significance coefficient t (.Sig) is 0.4 for the equation constant and the confidence interval is large. Thus the constant was not very accurate calculated. (table 3)

The Durbin- Watson coefficient (1.93) is close to the value 2, so autocorrelation between variables is very small and the model errors are very small. Now it can be applied the least squares method, to analyze residues. The model variance is not constant, so not all parameters are zero. We accept the alternative hypothesis of heteroscedasticity, so we can design a regression equation. (table 1)

For the first cluster the curve regression equation is (3) :

$$PN=0.822*VNT-0.822*ChT-489.636*NrSal+1371 \text{ (RON)}, \tag{3}$$

where: PN = net profit

VNT = total revenue

ChT = total expenditure

NrSal = number of employees

This equation allow SMEs from the cluster one to observe to what extent the specified parameters (total revenue, total expenses, number of employees) influences the value of net profit. One may replace parameters with different desirable values to do What IF analysis. For example: how is changing net profit if total revenue increases by 5000 and the number of employees increases by 2.

The values and interpretations are similar for the second, third and forth cluster but regression equation differs. We may notice, however that the first two clusters are more homogeneous and are better adjusted by the curve equation regression.

Table 4. Model summary for the second cluster

R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
				R Square Change	F Change	df1	df2	Sig. F Change	
.995	.991	.991	138994.887	.991	3514.176	3	96	.000	1.292

a. Predictors: (Constant), NrSal, VNT, ChT

b. Dependent Variable: PN

Table 5. Anova model for the second cluster

Model	Sum of Squares	df	Mean Square	F	Sig.	%
Regression	203677185207941.000	3	67892395069313.700	3514.176	.000	99.09762
Residual	1854679551565.870	96	19319578662.145			0.902381
Total	205531864759507.000	99				

a. Predictors: (Constant), NrSal, VNT, ChT

b. Dependent Variable: PN

Table 6. Model coefficients for the second cluster

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
(Constant)	-134419.546	28183.454		-4.769	.000	-190363.260	-78475.831
VNT	.904	.009	18.878	101.140	.000	.886	.922
ChT	-.905	.009	-18.827	-100.331	.000	-.922	-.887
NrSal	864.657	404.629	.024	2.137	.035	61.474	1667.840

Dependent Variable: PN

The parameters values for the curve regression equation for the second cluster and the model significance can be found in the Tables 4, 5, and 6. The equation is (4):

$$PN=0.904*VNT-0.905*ChT+864.657*NrSal-134419.546 \text{ (RON)} \tag{4}$$

Table 7. Model summary for the third cluster

R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin-Watson
				R Square Change	F Change	df1	df2	Sig. F Change	
.941	.885	.885	51101.228	.885	1109.157	3	431	.000	2.020

a. Predictors: (Constant), NrSal, VNT, ChT

b. Dependent Variable: PN

Table 8. Model coefficients for the third cluster

Model	Unstandardized Coefficients		Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
(Constant)	-1280.259	3112.894		-.411	.681	-7398.599	4838.082
VNT	.863	.015	13.626	56.567	.000	.833	.893
ChT	-.865	.016	-13.418	-55.788	.000	-.896	-.835
NrSal	5.338	315.765	.000	.017	.987	-615.293	625.969

a. Dependent Variable: PN

The curve regression equation for the third cluster is (5):

$$PN = 0.863 * VNT - 0.865 * ChT + 5338 * NrSal - 1280.259 \quad (RON) \quad (5)$$

This model doesn't adjusted very well the data series - R^2 is $0.88 < 0.95$, in particular because of the number of employees variable for which the T test is insignificant (Sig=0.987) and the constant (Sig.=0.681). But the coefficients of the total revenue and total expenditure parameters are very well calculated (Sig.=0.000) (Tables 7, 8).

Table 9. Model summary for the fourth cluster

R	R Square	Adjusted R Square	Std. Error of the estimate	Change Statistics					Durbin-Watson
				R Square Change	F Change	df1	df2	Sig. F Change	
.902	.814	.813	52461.293	.814	676.041	3	462	.000	2.027

a. Predictors: (Constant), NrSal, VNT, ChT

b. Dependent Variable: PN

Table 10. Model coefficients for the fourth cluster

Model	Unstandardized Coefficients		Standardized Coefficients Beta	T	Sig.	95.0% Confidence Interval for B	
	B	Std. Error				Lower Bound	Upper Bound
(Constant)	915.226	3124.712		.293	.770	-5225.183	7055.635
VNT	.778	.019	12.780	41.518	.000	.741	.815
ChT	-.779	.019	-12.432	-40.578	.000	-.817	-.742
NrSal	252.407	374.989	.017	.673	.501	-484.489	989.303

The curve regression equation for the fourth cluster is (6):

$$PN = 0.778 * VNT - 0.779 * ChT + 252.407 * NrSal + 915.226 \quad (RON) \quad (6)$$

This model doesn't adjusted very well the data series - R^2 is $0.81 < 0.95$, in particular because of the number of employees variable for which the T test is insignificant (Sig=0.501) and the constant (Sig.=0.770). But the coefficients of the total revenue and total expenditure parameters are very well calculated (Sig.=0.000) (Tables 9, 10).

CONCLUSIONS

Regression models designed allow SMEs from different clusters to observe to what extent the specified parameters (total revenue, total expenses, and number of employees) influence the value of net profit. It also allow What If analyses, replacing parameters with different desirable values. Thus a SMEs may create strategy of how increase the revenue for the future, how to reduce the expenditure or to alter the number of employees as to reach the proposed net profit value.

REFERENCES

- [1.] <http://www.startups.ro/analize/ce-sunt-imm-urile>
- [2.] www.listafirme.ro
- [3.] <http://www.mfinante.ro/agentiCod.html?pagina=domenii>
- [4.] <http://www.smarquest.ro/en/resources.html>
- [5.] Spiricu, L., Bădin, L., Ciomara, R., Mitruț D., Eficiență și productivitate - Tehnici de măsurare software și aplicații economice, București, Ed. Economică, 2001
- [6.] Tănăsioiu, O., Iacob A., Ilieanu, B. Introduction a L'Econometrie, București, Ed A.S.E., 2008

