International Journal of Economics, Commerce and Management
United Kingdom Vol. IV, Issue 4, April 2016
http://ijecm.co.uk/ ISSN 2348 0386

THE IMPORTANCE OF IDENTIFYING OF THE LACK OF DATA IN THE STATISTICAL EVALUATION OF FINANCIAL RISK

Zamira Veizi
"Eqrem Çabej" University of Gjirokastra, Albania
veizizamira@yahoo.com

Lorenc Koçiu
"Eqrem Çabej" University of Gjirokastra, Albania
kociulorenc@yahoo.com

Abstract

Nowadays it is very important for the business to try to identify, to assess and to determine the value of financial risk that they face. In order to business succeed in this effort, they must be supported strongly on statistical methods for accurate measurement of financial risk. For this purpose, the businesses must keep accurate statistical data and during the whole period of their activity. This means that for an accurate measurement of this business risk, one must possess time series, which in themselves can hold some lack of data. Our paper tries to highlight the fact how important is evidence of lack of data in the statistical processing models for identification, evaluation and measurement of the risk. In our paper, data is taken from about 250 SME for the period 2009-2013. From the analysis it was found that only 36.36% of variables are complete, whereas 63.64% of variables are with missing data. About cases, only 52% of cases (businesses) in the study had complete data, whereas 48% of them had missing data. Also in connection with missing data in value, was concluded that 95.75% of values were complete and 4.25 % of them were missing.

Keywords: Financial Risk, Missing Data, SME, Statistical Processing, Risk Assessment

INTRODUCTION

Lack of the data during their statistical processing is evident in many scientific papers. Researchers should be aware for the existence and their recording in scientific papers. Schlomer et al (2010), says that many researchers are unconscious for the importance of reporting and management of the lack of data and also and editors of magazines have not insisted that the authors to reflect this essential information. In the Albanian reality still not is understood the importance of reporting to the lack of data during the statistical processing of the data, quantitative or qualitative. During our paper will try to show why is important the evidence of the lack of data during the statistical processing of the data provided for the identification of financial risk. For identification of financial risk we are based on analyzing financial statements of 250 SME in the region of Gjirokastra. From the processing of these financial statements we have assessed financial risk through three groups of reports such as (1) financial ratios of risk of capital structure, (2) liquidity risk ratios and (3) risk insolvency ratios. Is important to give the meaning of SME according to Albanian legislation. In this context the definition of SME in the Republic of Albania is regulated by Law No. 8957, date 17.10.2002, "For Small and Medium Enterprises" amended. Are called micro enterprise those enterprises which employ up to 9 employees and their annual economic turnover does not exceed 10 million ALL. Small enterprises are called those enterprises which employ from 10 to 49 employees and have a turnover or total annual balance sheet less than 50 million ALL. Medium enterprises are called those enterprises which employ from 50 to 249 employees and have a turnover or total annual balance sheet until 250 million ALL.

Objectives and Research Questions of the Study

The main objective of this paper is to understand the importance of reporting and the management of the lack of data during the statistical processing of the data in their studies.

In focus of this goal are raised some research questions, as follows:

- Has affect reporting of the lack of data in identification of financial risk?
- Must be done data processing in case of absence of data largely?

LITERATURE REVIEW

The literature in support of reporting and management of the lack of data is expanding more recently. Researchers agree when are speaking for to the importance of reporting to the lack of data but they are not agree when are speaking for determining of the percentage of lack of data to be reported. Schafer (1999) has recommended that a level of 5 % of lack of data should be considered more important and certainly that should be reported during scientific papers. While

Bennett (2001) suggests that only cases when is evidenced a lack of more than 10 % of the data, statistical analyzes are likely to be biased and to put into question their reliability. While Peng et al (2006), believe in a wider interval of the lack of data, which goes up 20 %. So, according to them, up to this level (of 20%), the lack of data not damage the credibility of statistical processing of the data. Researchers have determined some kinds of lack of data and in this point contemporary literature is agree for the kind of lack of data. In this context contemporary literature recognizes two kinds of lack of data: (1) Missing Completely at Random (MCAR) and (2) Missing at Random (MAR). According to Acock (2005) and Bennet (2001) when exist MCAR data, is not possible that the lack of data and lack of values do not affect in any variable that is taken in study. Is difficult to be identified in a study the data MCAR, but Little (1998) developed an omnibus statistical test for identification if missing data are MCAR or no. Allison (2001) tried to define the data MAR, according to which the probability in this case to have the lack of data, is associated with an other variable taken in study, but it is not associated with interest variable. In this case the researcher must include observed variable in analysis in order to avoid bias.

METHODOLOGY

For processing successfully of this paper are analyzed 250 business financial statements included in the category of SMEs for the period 2009 - 2013. Integral part of these financial statements are balance sheet, income-expenses statement and cash flow. From the combined analysis of these financial statements are calculated financial reports as follows:

Financial Reports of Risk of Capital Structure

a. The ratio long-term debt / equity- this ratio shows the level of financial leverage in terms of long-term debt.
b. The ratio debt /total assets - this ratio shows the part of total assets of the company financed with debt. In this case the debt includes short-term and long term debts.
c. The ratio equity/ total assets - this ratio shows the part of assets of the company financed with own capital.
d. The ratio long-term debt / assets- this ratio shows the part of assets of the company financed with long term debt.
e. Interest coverage ratio - this ratio shows the ability of the company to cover interest payments from its profits, especially by earnings before interest and taxes.

Liquidity Risk Ratios

a. Current ratio - this ratio shows the ability of the company to cover short term debts with current assets, that is to say with 1 ALL current assets, how much money short term debts are covered.
b. Quick ratio (acid test) - this ratio which is calculated as the ratio of liquid assets to the current debts of the company. In this case is excluded the inventory by the voice of current assets, for the fact that the inventory is considered as less liquid asset.
c. Cash ratio - this ratio shows how capable is the company that with its monetary tools to cover its current liabilities as is calculated as the ratio of cash to current liabilities.

The Ratio of Insolvency Risk

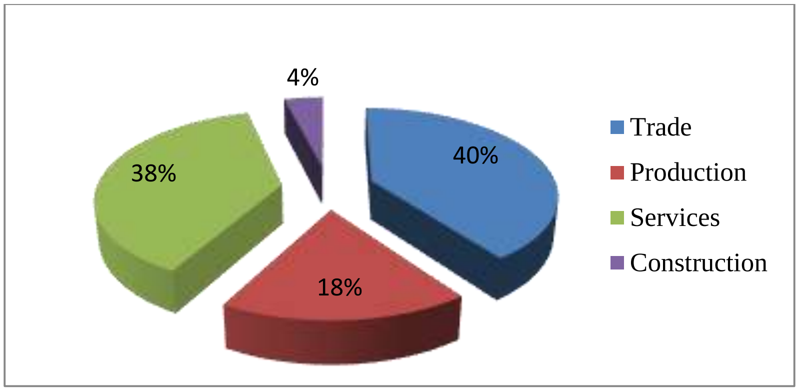
a. The ratio long - term assets /equity - this ratio shows the measure that long-term assets are financed with equity.
b. The ratio long - term /fixed equity - this ratio shows the measure that long -term assets are financed not only with equity but also with long-term debt.

The quantitative data that are obtained from these financial reports are elaborated with the help of statistical program SPSS.21, in order to obtain linear regression which will show the connection of these financial statements in the definition of financial risk. But, during processing of the data was found the lack of data in the formation of a general database. The lack of data means that not all SME that are taken in analysis have in their financial statements the same elements, also some elements of assets, obligations or the statement of income and expenditure missing. This means that when we are calculating the above financial reports is that due to the lack of elements in the financial statements will therefore bring and the absence of financial reports, which are calculated from these elements missing. Thus, this lack of data must be evidenced and to be given proper importance, in order not to deform the result of data processing. For processing of the lack of data in the statistical program SPSS.21, was used initially "Missing Value Analysis " and in the second step with the method "Multiple Imputation" (M.I.)

ANALYSIS

As we mentioned in section of "Methodology", are taken in analysis about 250 financial statements of SME that are operating in the region of Gjirokastra, which cover the period 2009-2013. The group of SMEs that are included in this study have a demographic distribution such as: 40% belonging to trade sector, 38% belonging to service sector, 18% belonging to manufacturing sector and 4% belonging to construction sector, graph 1.

Graph 1. Division according to activity

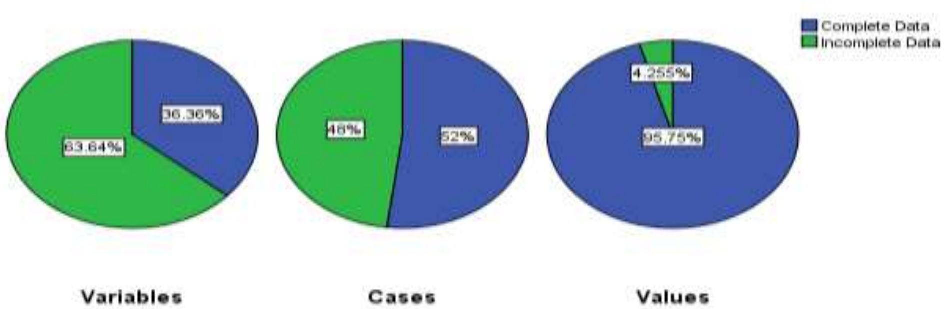


During the processing of data was found that not all variables were complete, in some of them had a lack of data. Thus, is used the method "Missing Value Analysis", in which is found that only 36.36% of variables are incomplete without lack of data, while 63.64% of variables are with the lack of data. In connection with the cases, only 52% of cases (business) that was taken in the study had complete data, while 48% of their had lack of data. Also, in connection with lack of data in the values is found that 95.75% of values were complete and 4.25% of their were with lack, the figure 2. In order to avoid spurious effects of results, in Missing Value Analysis, is used the condition that to complete the data with missing values up to the level 10%. Processing of these data is done initially with the method "Missing Value Analysis" and in the second step with the method " Multiple Imputation (M.I.)."

Multiple Imputation is more complex method of processing of missing data. The advantage of M.I. is that the final standard of their appreciate variables are based in (1) standard errors of analysis for each given data and (2) the distribution of parameters is assessed through the set data (Schlomer et al, 2010).

Through this method are improved significantly the level of significance for each group of independent variables according to respective financial reports, making more reliable the estimates of the dependent variable.

Figure 2. Overall summery of missing values



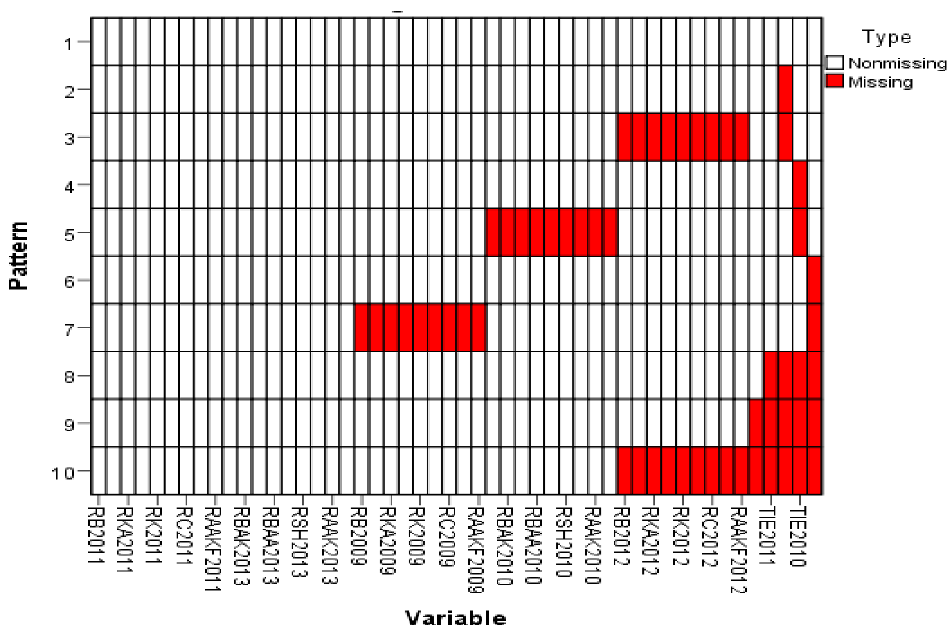
Also in the analytical way, lack of data appear through figure 3, where the red areas show financial reports that are missing. Thus for 2009, results that for some SME that are taken in analysis cannot count several reports for example; the debt ratio, capital/total assets ratio, the current ratio, the cash ratio, long-terms assets/fixed equity.

For 2010 appear to lack these reports; long-term debt/equity ratio, long-term debt/total assets ratio, quick report, long-term assets/own equity ratio, the interest coverage ratio.

For 2012 appear to lack these reports; the debt ratio, equity/total assets ratio, the current ratio, the cash ratio, long term assets /fixed equity ratio.

As a period with more complete data is presented the year 2011 and 2013 where results that those report, which were missing for years 2009, 2010, 2012, appear with the respective values.

Figure 3. Missing value patterns



If, during statistical processing of data does not reflect a lack of data and at the same time will not become their processing with the method " Missing Value Analysis ", and in the second step with the method " Multiple Imputation", then there would be not improved levels of significance. This means that the results which will conclude would be untrue and biased. But, realizing the processing of lack of data with assistance of financial program SPSS v.21, we see that are improved as levels of significance as well as and the respective standard errors.

As we see, in table 1, are reflected the analysis values before processing of the lack of data, which results very large amount of significances (>0.05), which means that these results have not statistically significant for the period 2009.

But, in the table 2 are reflected the results of the statistical analysis as are processing the lack of data with the method "Multiple Imputation". In this table result an improvement enough sensitive of the level to significance (<0.05), giving a statistically sense these variables, for the period 2009.

Table 1: Year 2009 - Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	.365	.386		.944	.355	-.436	1.165
	RB2009	.127	.438	.084	.290	.774	-.782	1.037
	RBAK2009	.059	.077	.201	.760	.455	-.101	.219
	TIE2009	.001	.002	.181	.754	.459	-.002	.005
	RKA2009	.262	.422	.199	.622	.540	-.612	1.136
	RBAA2009	.271	.938	.109	.289	.775	-1.674	2.217
	RK2009	.245	.226	1.117	1.084	.290	-.223	.713
	RSH2009	.006	.350	.032	.017	.987	-.720	.732
	RC2009	-.188	.388	-1.112	-.484	.633	-.992	.616
	RAAK2009	.143	.353	.707	.404	.690	-.590	.875
	RAAKF2009	-.110	.337	-.550	-.325	.748	-.809	.590

a. Dependent Variable: IND2009

Table 2: Year 2009 - Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
5	(Constant)	.398	.095		4.182	.000	.211	.585
	RBAA2009	.440	.133	.176	3.307	.001	.178	.701
	RKA2009	.447	.087	.339	5.164	.000	.277	.617
	RB2009	.339	.094	.223	3.612	.000	.154	.524
	RBAK2009	.057	.015	.195	3.688	.000	.027	.087
	TIE2009	.001	.000	.198	3.518	.000	.001	.002

In the table 3, are reflected the values of analysis before processing of the lack of data, which results very large values of significances (>0.05), which means that these results have not statistically significance for the period 2010.

But, in the table 4 are reflected the results of the statistical analysis as is processing the lack of data with the method "Multiple Imputation". In this table results an improvement quite sensitive of the level of significances (<0.05), giving a sense by statistically these variables for the period 2010.

Table 3: Year 2010 - Coefficients

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95.0% Confidence Interval for B**	
		B	**Std. Error**	**Beta**			**Lower Bound**	**Upper Bound**
1	(Constant)	.616	.391		1.577	.128	-.190	1.423
	RB2010	.463	.420	.265	1.104	.281	-.403	1.329
	RBAK2010	-.084	.069	-.235	-1.208	.239	-.227	.059
	TIE2010	.004	.002	.325	1.539	.137	-.001	.008
	RKA2010	.105	.415	.061	.253	.802	-.752	.963
	RBAA2010	1.227	1.100	.561	1.115	.276	-1.044	3.498
	RK2010	-.022	.033	-.191	-.688	.498	-.090	.045
	RSH2010	.075	.234	.113	.320	.752	-.409	.559
	RC2010	-.653	.534	-.444	-1.223	.233	-1.756	.449
	RAAK2010	-.198	.326	-.984	-.609	.548	-.871	.474
	RAAKF2010	.147	.317	.746	.464	.647	-.507	.801

Table 4: Year 2010 - Coefficients

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95.0% Confidence Interval for B**	
		B	**Std. Error**	**Beta**			**Lower Bound**	**Upper Bound**
7	(Constant)	.689	.063		10.925	.000	.565	.813
	RAAKF2010	-.045	.010	-.227	-4.431	.000	-.065	-.025
	TIE2010	.003	.001	.261	5.360	.000	.002	.004
	RB2010	.331	.083	.189	3.981	.000	.167	.494
	RBAK2010	-.102	.017	-.284	-6.114	.000	-.134	-.069
	RBAA2010	.574	.139	.262	4.125	.000	.300	.847
	RC2010	-.490	.125	-.333	-3.927	.000	-.735	-.245
	RSH2010	.128	.054	.193	2.372	.018	.022	.234

Table 5 reflected the values of analysis before processing of the lack of data, resulting very large amount of significances (>0.05), which means that these results have not statistical significance, for the period 2012.

But, in table 6 are reflected the results of statistical analysis after being processed the lack of data with the method "Multiple Imputation". In this table results a very sensitive improvement of the level of significances (<0.05), giving a statistical sense these variables, for the period 2012.

Table 5: Coefficients Year 2012

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.204	.276		4.364	.000	.636	1.772
	RB2012	.068	.276	.046	.245	.809	-.501	.637
	RBAK2012	.092	.038	.399	2.408	.024	.013	.171
	RKA2012	.210	.236	.184	.892	.381	-.275	.696
	RBAA2012	.273	.590	.129	.462	.648	-.943	1.489
	RK2012	-.250	.064	-.734	-3.888	.001	-.382	-.118
	RSH2012	.229	.115	.491	1.999	.057	-.007	.465
	RC2012	-.263	.251	-.311	-1.046	.306	-.780	.255
	RAAK2012	-.298	.328	-1.459	-.910	.372	-.974	.377
	RAAKF2012	.163	.310	.805	.526	.604	-.475	.801

Table 6: Year 2012 - Coefficients

Model		**Unstandardized Coefficients**		**Standardized Coefficients**	**t**	**Sig.**	**95.0% Confidence Interval for B**	
		B	**Std. Error**	**Beta**			**Lower Bound**	**Upper Bound**
8	(Constant)	**1.231**	**.030**		**41.418**	**.000**	**1.173**	**1.289**
	RK2012	**-.237**	**.014**	**-.695**	**-16.555**	**.000**	**-.265**	**-.209**
	RBAK2012	**.086**	**.009**	**.373**	**9.544**	**.000**	**.068**	**.104**
	RSH2012	**.223**	**.027**	**.478**	**8.336**	**.000**	**.171**	**.276**
	RKA2012	**.211**	**.047**	**.185**	**4.453**	**.000**	**.118**	**.305**
	RC2012	**-.250**	**.055**	**-.296**	**-4.517**	**.000**	**-.359**	**-.141**
	RAAK2012	**-.127**	**.009**	**-.620**	**-14.482**	**.000**	**-.144**	**-.110**

FINDINGS AND CONCLUSIONS

This paper showed that is very important to be given importance the identification of cases, while there is a lack of data during various statistical processing. Researchers have to be attentive that to reflect these shortages and also to be careful for the level of the absences in relation to the total data that are taken in study. Despite different opinions in contemporary literature, the most of researchers agree that up to a level of 105 of the lack of data does not compromise results and the reliability of statistical processing of the data. But, when this lack of data rise above this value, of 10%, then it can be questioned data analysis, which will require involvement of other variables in study to obtain a reliable and impartial result.

According to the analysis done by the method "Missing Value Analysis", resulted that only 36.36% of variables are complete (no lack of data), whereas 63.64% of variables are with the lack of data. About cases, only 52% of cases (business) taken in the study had complete data, whereas 48% of them had lack of data. Also in connection with the lack of data in value, is concluded that 95.75% of values were complete and 4.25 % of them were with the lack. The results showed that before processing of the lack of data, resulted very large amount of significances (>0.05), which means that these results were not statistically significant for 2009. But for this year also are reflected the results of the statistical analysis, after processing of lack of data and resulted a very sensitive improvement of the level of significances (<0.05), giving a sense statistical variables. For 2010, before processing of the lack of data result very large values of significances (>0.05), which means that these results are not statistically significant. But also the results of the statistical analysis after processing of lack of data for 2010, showed an improvement quite sensitive to the level of significances (<0.05), giving a sense statistical variables. For 2012, before processing of the lack of data result very large values of significances (>0.05), which means that these results have not statistically significant. Also for the period after processing of lack of data, resulted a very sensitive improvement of the level of significances (<0.05), giving a sense statistical variables.

REFERENCES

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25,* 464–469

Gabriel L. Schlomer, Sheri Bauman, and Noel A. Card : “Best practices for missing data management in counseling psychology”, Journal of Counseling Psychology, Vol 57, No 1, pg 1

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198–1202

Peng, C.-Y. J., Harwell, M., Liou, S.-M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), *Real data analysis* (pp. 31–78). Greenwich, CT: Information Age

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8,* 3–15

Schlomer, Gabriel L, Sheri Bauman, and Noel A Card. "Practices for Missing Data Management in Counseling Psychology, Vol 57, No 1." *Journal of Counseling Psychology*, 2010: 1-10