International Research Journal of Management and Commerce
ISSN: (2348-9766)
Impact Factor- 5.564, Volume 4, Issue 12, December 2017
Website- www.aarf.asia, **Email** : editor@aarf.asia , editoraarf@gmail.com

EFFECT OF PUBLIC SECTORS BANK'S LOAN ON FINANCIAL PERFORMANCE AS WELL AS IN CAPITAL FORMATION OF MSMES IN DISTRICTS OF EAST AND WEST MEDINIPUR: DATA SCREENING AND PRELIMINARY ANALYSIS

Mr. Sajijul Islam
Research Scholar, Department of Business Administration, Vidyasagar University, West Bengal, India.

ABSTRACT

The purpose of the present study is to discuss the process of screening, editing, and preparation of initial data before going to further multivariate analysis of the study concerning the effect of Public Sector Banks' loan towards MSMEs of two selected districts of West Bengal. It is very important to conduct data screening to identify any potential violation of the basic assumptions related to the application of multivariate techniques. Moreover, initial data examination enables the researcher to gain a deeper understanding of the data collected. For this research, simple random sampling has been adopted as the sampling technique to draw a conclusion about the entire population. Samples of 360 were selected from the total population of 5973 registered entrepreneurs in the two districts for the year of 2011-12 to 2015-16. Towards the achievement of the fulfillment of the assumptions of multivariate analysis, data screening and preliminary analysis were conducted. In specifically, the study carried out response rate, non-response bias test, missing data detection and treatment, multivariate outliers detection and treatment, normality assessment, linearity assessment, common method variance assessment, multicollinearity assessment, and homoscedasticity assessment. All the assessment was conducted using IBM SPSS statistical software version 23.0 (SPSS). In brief, the data found to fulfill the requirements for further multivariate analysis.

Keywords: Public Sector Banks, loan, Micro Small and Medium Enterprises, Data Screening, Preliminary Analysis.

1. INTRODUCTION

Data screening is a fundamental part of the multivariate analysis as it helps researchers to ensure that the data underlying the analysis meet the certain requirements of the multivariate analysis in the quantitative research process (Hair, Black, Babin & Anderson, 2010). In addition, by examining and screening the collected data before going to next analysis, researchers gain a critical insight into the characteristics of the data. The first requirement is to meet the assumptions of psychometric property concerning the data, therefore making it safe to proceed to use the data for statistical analyses. Second, is the need to follow certain procedure by checking for errors and correcting the error, if any, in the data file.

However, such neglect of initial data screening may be catastrophic on the result of the multivariate analysis, as the result of the estimated standard error may be inflated (Chenick, 2008). Hence, the significance of the statistical coefficient of a regression-based path analysis or structural model are underestimated (Kura, 2014; Ringle, Sarstedt, & Straub, 2012). To meet these requirements, the present study presented the procedures of data screening and further conducted the following preliminary analysis: response rate, non-response bias test, missing data detection and treatment, multivariate outliers detection and treatment, normality assessment, linearity assessment, common method variance assessment, multicollinearity assessment, and homoscedasticity assessment (Samoel & Page, 2007; Tabanick & Fidel, 2007, Sarstedt et al. 2012, Hair et al., 2016,). The review of works of literature in this study is presented in the following section of the paper.

2. LITERATURE REVIEW

Though touch upon the availability of financial resources for their surveillance, do not penetrate much about role and effectiveness of bank finance in this respect. The commercial banks came forward and made immense help to the growth of SMEs. Here is a gap which requires analysis of the role of the banks in the post-economic reforms. It concludes the in-depth study of the banker's role in providing the credit to promote the SMEs. It does not analyse much over the effectiveness of such fund in process of growth of SMEs. Park, D. and H. A. Krishnan (2001).

The available relevant literature on MSME is reviewed with reference to their financials well as other related aspects. An attempt is made to review some of such studies. The study emphasizes more the needs and feasibilities of MSMEs and also estimates how relevantly the MSMEs emphasises the use of inherently available local resources in industrial development Kotler, P. and N. Lee (2005).

Finance always act as critical resources to improve SMEs business activities in any economy (Kelley, Singer, & Herrington, 2012; Xavier, Kelley, Kew, Herrington, & Vorderwülbecke, 2013). Therefore, access to loan fund enhance firm performance (Ayyagari, Demirgu-Kunt, & Maksimovic, 2008; Batra, Kaufmann, & Stone, 2003; Frank, Kessler, & Fink, 2010; Kyophilavong, 2011; Wiklund & Shepherd, 2005). Finally, a supportive business environment is essential principally when assessing firm performance. Therefore, past studies report roles of the different elements business environment on firm performance (Ensley, Pearce, & Hmieleski, 2006; Goll & Rasheed, 2004; Jong & Thai, 2008; Rasheed, 2005; Rueda Manzanares, Aragon Correa, & Sharma, 2008; Tang, Tang, Marino, Zhang, & Li, 2008; Tang, 2008; Tang & Hull, 2012; Wiklund & Shepherd, 2005).

3. METHODOLOGY

Technique of data analysis is a method by which researchers' analyses their collected data, and consequently delivers better understanding of the phenomenon (Pallant, 2011). In this study, data screening and pre-analysis were employed to analyse the data. Using **Cochran's formula** 360 respondents were selected among 5973 registered MSMEs of East and West Medinipur districts through simple random sampling technique. Therefore, after raw data were collected from the field, the entire questionnaires were coded and inputted into the Statistical Package for the Social Science (SPSS version 23.0). Formerly the subsequent technique of data analysis was implemented to analyse the data. Firstly, response rate, non-response bias test was conducted. Subsequently, this study adopted the approach of detection and treatment of missing values, identification of outliers for data screening (Hair et al., 2016). Lastly, normality assessment and linearity assessment, common method variance (CMV), multicollinearity assessment, and homo-scedasticity assessment analysis were conducted to meet the preliminary assumption for further multivariate analysis.

4. Data Screening

4.1 Response Rate

Table 1: Response rate of the respondents

Response	Total
No of selected respondents	360
No of met respondents	271
No of unmet respondents	89
Response rate	75.2%

In the present study, 360 respondents were selected based on Cochran's formula. The researcher has successfully collected data from 271 respondents only which makes the response rate of 75.2%. According to Sekaran and Bougie (2010), in survey studies a response rate of 30% is acceptable. Therefore, the study response rate is adequate for further analysis

4.2. Response Bias Test

The issue of non-response bias arises when there is a difference in the answers between non-response and response data (Lambert and Harrington, 1990). Non-response bias can affect the findings of the research and the generalization of the result to the population. Henceforth, there is a requirement to conduct the non-response bias test to detect this type of error before going to the main analysis. In this regards to founding out the possibility of non-response bias issue, this research followed a time-trend extrapolation method suggested by Armstrong and Overton (1977) by comparing the early and late respondents. Furthermore, according to the suggestion of Lindner and Wingenbach (2002), a minimum response rate of 50% should be achieved to minimize the issue of non-response bias.

Since, in the present study, 75.2% response rate is achieved, so the potential differences between early and late were compared using the entire study variables. Therefore, the test of response bias was carried out by dividing the respondents into two groups based on early and late respondents. Then, an independent samples t-test was conducted for the study variables to observe if there is any difference between the two groups. In view of this, the study has classified 153 respondents as early responses and 127 respondents as late responses. The researchers' was conducted an Independent-Sample t-test for both descriptive test and Levene's test for equality of variance between the early and late respondents.

Group Statistics

	Group	N	Mean	Std. Deviation	Std. Error Mean
Availability	Early	161	19.0248	2.19645	.17310
	Late	106	19.1226	2.10533	.20449
Accessibility	Early	161	19.0435	2.22021	.17498
	Late	106	19.3491	2.28053	.22150
Expected Reliance	Early	161	15.7019	1.46563	.11551
	Late	106	15.4811	1.69723	.16485
Facilities	Early	161	15.6398	1.75909	.13864
	Late	106	15.7736	1.51992	.14763
Terms & Conditions	Early	161	15.7267	1.45339	.11454
	Late	106	15.7453	1.66225	.16145
Utilization	Early	161	27.8385	2.13278	.16809
	Late	106	28.1792	2.43701	.23670
Capital Formation	Early	161	24.0745	1.56346	.12322
	Late	106	24.4245	1.67871	.16305
Financial Performance	Early	161	12.0932	1.47395	.11616
	Late	106	12.4434	1.58012	.15347

In the above table, the results of the descriptive test indicated that there were no significant statistical differences between the said two groups' mean and standard deviation.

Table3: 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
Availability	Equal variances assumed	.474	.492	-.362	265	.718	-.09780	.27027	-.62996	.43436
	Equal variances not assumed			-.365	231.419	.715	-.09780	.26792	-.62567	.43008
Accessibility	Equal variances assumed	.216	.642	-1.089	265	.277	-.30558	.28072	-.85830	.24715

	Equal variances not assumed			-1.083	220.567	.280	-.30558	.28228	-.86189	.25073
Expected Reliance	Equal variances assumed	1.502	.221	1.130	265	.259	.22073	.19531	-.16383	.60530
	Equal variances not assumed			1.097	201.532	.274	.22073	.20129	-.17617	.61763
Facilities	Equal variances assumed	2.555	.111	-.641	265	.522	-.13383	.20869	-.54473	.27707
	Equal variances not assumed			-.661	246.202	.509	-.13383	.20252	-.53272	.26506
Terms & Conditions	Equal variances assumed	.204	.652	-.096	265	.923	-.01857	.19257	-.39773	.36058
	Equal variances not assumed			-.094	203.472	.925	-.01857	.19796	-.40888	.37173
Utilization	Equal variances assumed	.551	.459	-1.206	265	.229	-.34074	.28246	-.89689	.21542
	Equal variances not assumed			-1.174	203.617	.242	-.34074	.29031	-.91314	.23167
Capital Formation	Equal variances assumed	2.123	.146	-1.738	265	.083	-.34999	.20139	-.74653	.04654
	Equal variances not assumed			-1.713	213.482	.088	-.34999	.20437	-.75284	.05285
Financial Performance	Equal variances assumed	2.939	.088	-1.846	265	.066	-.35023	.18973	-.72381	.02335
	Equal variances not assumed			-1.820	213.732	.070	-.35023	.19248	-.72963	.02917

In the above table, the results of independent-samples t-test showed that the variance between the above said two groups were same for all dimensions. The significant value of all dimensions was >0.05 significance level of Levene's test for equality of variances (Field, 2009; Pallant, 2011). Therefore, it can be concluded that this assumption has not been violated between the two groups. Therefore, no problem of response bias has occurred.

4.1.1 Missing Data

In this study, the data was collected by researcher's hand to hand from the respondents. As the researcher was checked the questionnaire and asked to refill on that time if they left any question blank without the answer to preventing missing data. So, there was no question about missing data. Though, the researcher has conducted frequency analysis after collecting all the data to check if there is any missing value or any coding mistake. After running the frequency analysis on IBM SPSS version 22, there was no missing value found.

4.1.2 Analysis of Multivariate Outliers

Outliers are the extreme scores or values of data sets. In this study, multivariate outliers were detected using Mahalanobis distance (d2) and Cooks distance statistical tool. There were four outliers with the probability of D^2 less than 0.001. And none of the respondents had a cooks distance greater than 1. Though, according to **Steavens (1984),** we found that no need to remove the outliers whose cooks distance is less than 1. But here the researcher has deleted four outliers based on Mahalanobis D^2 because they could affect distort the result of the data analysis. Mahalanobis D2 and cooks distance for all the cases are reported in Appendix. Henceforth, after removing four multivariate outliers, the final dataset in this study was 267.

4.1.3 Normality Assessment

According to the suggestions of Pallant (2001) and Hair *et al.,* (2010), in order to meet up with the assumption of a multivariate analysis, normality of the data need to be checked. There are two major ways of assessing normality: graphically and numerically. To check with the graphical method, the histogram was examined by looking at the shape of data distribution (Tabachnick and Fidell, 2007) while for numerical method, the K-S Test (sample size more than 2000) or S-W test (7<sample size<= 2000) is to be used.

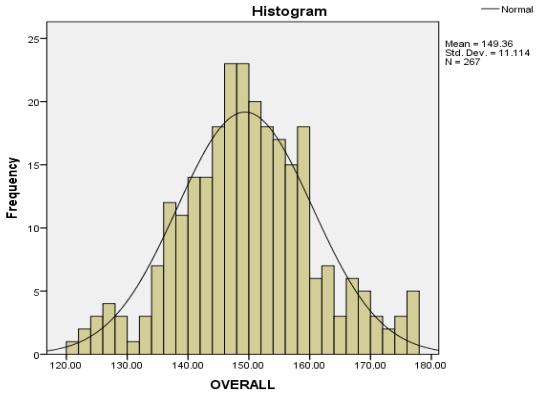


Figure 4.1: Histogram of normality distribution

Dependent variable: OVERALL

Above Figure 4.1 depicted the histogram which indicates that the normality assumption has been achieved because the histogram gave a bell shape 'normal curve'.

Table 4.1.3: Descriptives

			Statistic	Std. Error
OVERALL	Mean		149.3558	.68019
	95% Confidence Interval for Mean	Lower Bound	148.0166	
		Upper Bound	150.6951	
	5% Trimmed Mean		149.3092	
	Median		149.0000	
	Variance		123.531	
	Std. Deviation		11.11444	
	Minimum		121.00	
	Maximum		177.00	
	Range		56.00	
	Interquartile Range		14.00	

	Skewness	.110	.149
	Kurtosis	.149	.297

Table 4.1.4: Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
OVERALL	.069	267	.004	.990	267	.069

a. Lilliefors Significance Correction

Table 4.1.3 and Table 4.1.4 shows that the result of the normality test conducted for this study. In table 4.1.3, the absolute value of skewness is 0.738 (0.110/0.149) and kurtosis is 0.502 (0.149/0.297). The values of both skewness and kurtosis in this study fall within the range of $\pm$ 1.96 with the significant value of Shapiro-Wilk test greater than 0.05. So, the above tests are indicating that the data was normality distributed. Means the normality assumptions of this study were not violated.

4.1.4. Linearity Assessment

Next assumption linearity assessment has been examined through the graph-legacy diagrams-scatter/dot-simple scatter procedures with the help of SPSS 22 software. The Linearity of data could be tested by examination of scatter plots or linearity residual plot (Hair et al., 2010; Pallant, 2013).

Visual examination of the plots in this study showed a roughly straight line and not a curve. This meant that the residuals had a straight-line relationship with the predicted values of the dependent variables. Hence, linearity exists between the dependent variables and the independent variables. No serious deviations from linearity have been observed in the scatter-plots. The graphs for linearity assessment has presented in below.

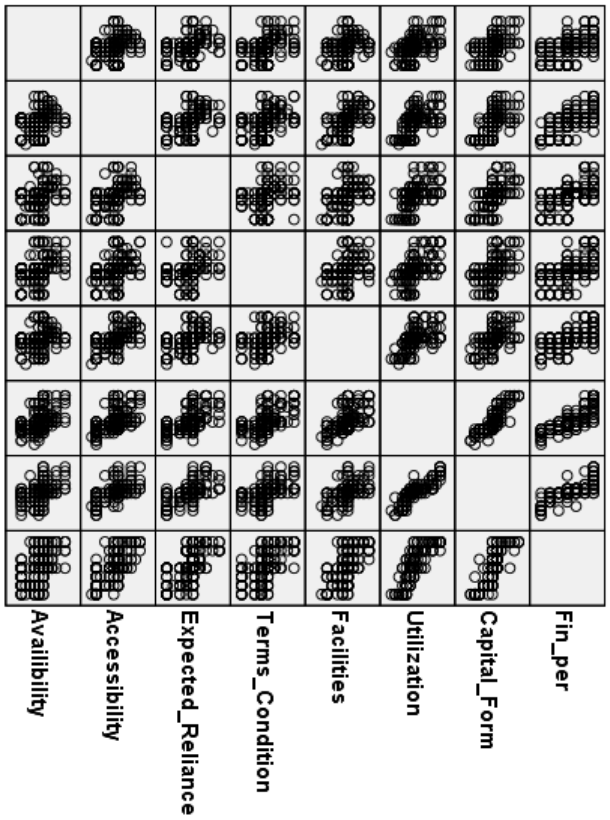


Figure 4.2: Linearity Assessment

4.1.5. CMV Assessment

Since the data collected on endogenous variables as well as exogenous variables at same time and using the same instrument, the common methods bias may exist among the collected data. Therefore, considering the potential problem caused by common method bias in behavioral studies, the researcher conducted a test to make sure that there is no variance in observed scores and the correlations are not inflated because of the CMV effect. Hence, the researcher has used a test known as Harman's one-factor test (Podsakoff et al., 2003). An un-rotated factor analysis with the entire thirty-nine variables has been conducted. The test produced that 33.723% of the total variance was accounted by the single factor which indicates the absence of common method bias in this study. According to Podsakoff et al, (2003), Kumar, (2011), and Lowry and Gaskin (2014), the common method bias is present when the single factor is explaining more than 50% of the variance.

4.1.6. Multicollinearity Assessment

Multicollinearity refers to the relationship between two or more exogenous variables, where the independent variables demonstrate little correlation with other independent variables (Hair Jr et

al. 2010). Multicollinearity problem occurs when the independent variables are highly correlated to each other (Hair Jr et al., 2010; Pallant, 2010; Tabachnick & Fidell, 2013). Therefore, when two or more variables are highly related, it means they contain unnecessary information. Therefore, all variables are not needed in the same analysis because they increase the error terms. Furthermore, when multicollinearity between variables is high, the standard error of the regression coefficient increases, so the statistical significance of these coefficients becomes less reliable. However, the most reliable statistical test of multicollinearity is an examination of tolerance and Variance Inflation Factor (VIF) with the thresholds of more than 0.1 and VIF of 10 (Hair Jr et al., 2010; Pallant, 2010). Therefore, in this study multicollinearity was tested first by examining correlation matrix and secondly by tolerance and VIF level for the independent variables. The correlation matrix of the independent variables was examined to find out if there is any indication of high correlations among the variables. According to Hair Jr et al. (2010 and Pallant, 2010), Multicollinearity exists when the correlation between independent variables is 0.9 and higher. Though, Pallant (2010) also suggested that correlation value above 0.7 as a threshold for multicollinearity among independent variables. The result in Table 4.7 shows that the correlation values are within .218 to .489 not higher than the threshold of 0.7. It is, therefore, concluded that there was no problem of high correlation among the variables.

Table 4. 7: Correlations among the Exogenous Variables

		Availability	Accessibility	Expected Reliance	Facilities	Terms & Conditions
Availability	Pearson Correlation	1	.244**	.295**	.403**	.222**
	Sig. (2-tailed)		.000	.000	.000	.000
	N	267	267	267	267	267
Accessibility	Pearson Correlation	.244**	1	.385**	.388**	.489**
	Sig. (2-tailed)	.000		.000	.000	.000
	N	267	267	267	267	267
Expected Reliance	Pearson Correlation	.295**	.385**	1	.229**	.413**

	Sig. (2-tailed)	.000	.000		.000	.000
	N	267	267	267	267	267
Facilities	Pearson Correlation	.403**	.388**	.229**	1	.273**
	Sig. (2-tailed)	.000	.000	.000		.000
	N	267	267	267	267	267
Terms & Conditions	Pearson Correlation	.222**	.489**	.413**	.273**	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	267	267	267	267	267

** Correlation is significant at the 0.01 level (2-tailed).

Table 4. 8: Multicollinearity test based on Tolerance and VIF values

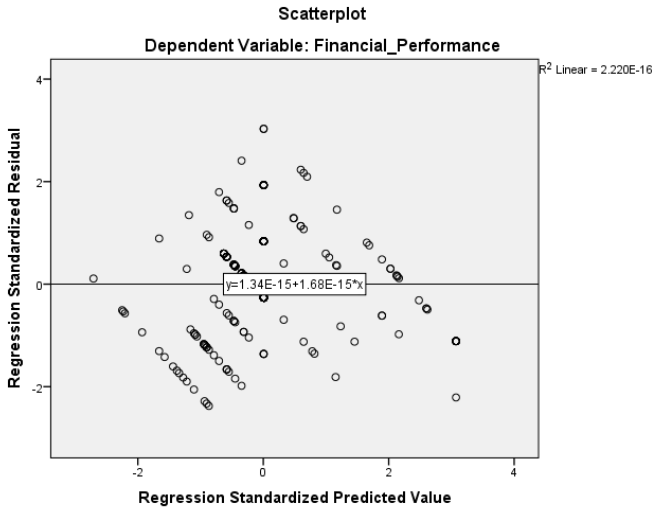
	Co-linearity Statistics	
	Tolerances	VIF
Availability	.792	1.262
Accessibility	.663	1.509
Expected Reliance	.754	1.326
Facilities	.745	1.342
Terms & Conditions	.696	1.437

Secondly, multicollinearity was tested through examination of tolerance and VIF using regression results provided by the SPSS collinearity diagnostics result. As recommended (Hair Jr et al., 2010; Pallant, 2010), this is the most important and reliable test of multicollinearity. In table 4.4 below it is clear that the tolerance ranges between 0.694 and 0.796 substantially greater than 0.1 and VIF ranges from 1.330 to 1.506, thus, is acceptable as being less than 10. In line with Hair Jr et al. (2010) and Pallant (2010), the result shows that multicollinearity does not exist in this study since tolerance values above 0.10 and VIF values are below 10. Though some researchers are also argued that if VIF is more than 3 probably have multicollinearity else if VIF is more than 5 there will be lightly multicollinearity and if it is more than 10 then there will be

definitely multicollinearity. Luckily in the present study, VIF is less than 3 so there is no question about multicollinearity.

4.1.7. Homoscedasticity Assessment

Homoscedasticity refers to the assumption that dependent variables exhibit equal levels of variance across the range of predictor variable(s) (Hair, Jr. et al., 2003, p.73). Heteroscedasticity occurs when ―the error term in a regression model does not have a constant variance (Berry & Feldman, 1985, p.73). If the assumption of homoscedasticity is unmet, the data is not appropriate for conducting a test of differences like ANOVA. In the present study, it is checked through examination of residuals which should show no pattern of increasing or decreasing residuals (Hair, Black, Babin & Anderson, 2013, p.217). It is expected to display a fairly even cigar shape along its length (Pallant, 2010). The graphs for residuals have been presented in the following figure. The figure has no pattern of increasing or decreasing residuals and it has a cigar shape. Therefore, it can be concluded that this study met the assumptions of homoscedasticity as the residuals show a cigar shape.



5. CONCLUSION

To conclude, this paper assesses the data through series of statistical techniques to ensure it fulfills the assumptions before any further multivariate analysis. Therefore, data screening, cleaning, and pre-analysis ware conducted to satisfy these assumptions. Thus, the study conducted non-response bias test, missing data analysis, multivariate outlier detection and treatment, normality assessment, linearity assessment, Common method Variance (CMV) assessment, multicollinearity assessment and homoscedasticity assessment. In brief, the data found to fulfill the multivariate analysis requirements.

References

- Park, D. and H. A. Krishnan (2001), Supplier Selection Practices among Small Firms in the United States: Testing Three Models, *Journal of Small Business Management,* 39(3) : 259 – 271.
- Kotler, P. and N. Lee (2005), CSR – Doing the Most Good for Your Company and Your Cause, New Jersey: John Wiley and Sons Inc.
- Rasheed, H. S. (2005). Foreign entry mode and performance: The moderating effects of environment. *Journal of Small Business Management*, *43*(1), 41–54.
- Wiklund, J., & Shepherd, D. (2005). Entrepreneurial orientation and small business performance: A configurational approach. *Journal of Business Venturing*, *20*(1), 71–91.
- Hair, J.F.Jr., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L. (2006), Multivariate Data Analysis. 6th ed. New Jersey: Prentice- Hall International Inc.
- Tabachnick, B.G., Fidell, L.S. (2007), Using Multivariate Statistics. 5th ed. Boston, MA: Allyn & Bacon/Pearson Education.
- Tang, J., Tang, Z., Marino, L., Zhang, Y., & Li, Q. (2008). Exploring an inverted U-shape relationship between entrepreneurial orientation and performance in Chinese ventures. *Entrepreneurship …*, *32*(1), 219–239.
- Ayyagari, M., Demirgu-Kunt, A., & Maksimovic, V. (2008). How important are financing constraints? The role of finance in the business environment. *The World Bank Economic Review*, *22*(3), 483–516.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L. (2010), Multivariate Data Analysis. 7th ed. New York: Pearson.

- Pallant, J. (2010), SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS. 4th ed. Australia: Allen & Unwin Book Publishers.
- Sekaran, U., & Bougie, R. (2010). *Research methods for business: A skill building approach* (5th ed., pp. 1–387). United Kindom: John Wiley & Sons Ltd.
- Pallant, J. (2011), SPSS Survival Manual. UK: McGraw-Hill Education.
- Tang, Z., & Hull, C. (2012). An investigation of entrepreneurial orientation, perceived environmental hostility,and strategy application among Chinese SMEs. *Journal of Small Business Management, 50*(1), 132–158.
- Tabachnick, B.G., Fidell, L.S. (2013), Using Multivariate Statistics. 6th ed. New Jersey: Pearson Education Inc. p1-983.
- Pallant, J. (2013), SPSS Survival Manual. UK: McGraw-Hill Education.
- Kraus, S. (2013). The role of entrepreneurial orientation in service firms: Empirical evidence from Austria. *TheService Industries Journal*, *33*(5), 427–444.
- Xavier, S. R., Kelley, D., Kew, J., Herrington, M., & Vorderwülbecke, A. (2013). *Global Entrepreneurship Monitor GEM 2012 Global report* (pp. 1–87). Massachusetts.