Exploring Management Capability in SMEs Using Transactional Data

Citation for published version:
Ma, Y, Andreeva, G & Ansell, J 2016, 'Exploring Management Capability in SMEs Using Transactional Data', *Journal of the Operational Research Society*, pp. 1-8. https://doi.org/10.1057/jors.2014.90

Digital Object Identifier (DOI):
10.1057/jors.2014.90

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
Journal of the Operational Research Society

Exploring Management Capability in SMEs Using Transactional Data

By Yigui Ma, Jake Ansell and Galina Andreeva

The University of Edinburgh Business School

Abstract

Small and Medium Sized Enterprises (SMEs) have become very important in most world economies. Governments have developed policies to support them and within UK the government has encouraged lending to SMEs with policies such as treating SMEs as similar to retail banking. Most SMEs credit relationships are organized through relationship banking and aspect of which is the confidence a banker may have in the SME's management. Management Capability is therefore critical to the success of a SME and its access to finance. The paper reports findings from work on determining Management Capability from qualitative and quantitative measures. In the study we have used Principal Component Analysis and Partial Least Squares regression to elicit measures for Management Capability. The results indicate some success in determining measures for Management Capability.

Key words: Credit risk assessment, Management capability, Small and Medium Sized Enterprises

Introduction

Lending to Small and Medium Sized Enterprises (SMEs) by Banks has attracted considerable attention in recent years, especially after the 'Credit Crunch'. This interest is driven in part by the fact that SMEs account for the majority of firms in the economy and represent a significant share of employment (Ma and Lin, 2010). Furthermore, most large companies usually start as small enterprises, so the ability of SMEs to develop and invest becomes crucial to any economy wishing to prosper, (De la Torre et al, 2010). The problem for the Banks is related to some difficulties specific to this sector, such as informational opacity. The need is to improve assessment modelling of SMEs by taking into account available information.

Within banking over the recent past credit managers have attempted to overcome this issue via relationship lending, and aspect of which is the confidence a banker may have in the SME's management. Management Capability is therefore critical to the success of a SME and its access to finance.

Managerial capability is commonly an important aspect of small business success. Research in this area mainly focus on qualitative description, often cited examples indicating management capability are planning sophistication, (Carter & Auken, 2006; Gaskill et al, 1993; Lussier, 1995; Maes et al, 2005; Perry, 2001); use of information system, record keeping (Lussier, 1995; Maes et al., 2005) and the use of professional advice (Gaskill et al., 1993; Lussier, 1995; Maes et al., 2005) These qualitative information requires judgments and is not easily assessed. Quantitative hard information has the advantage of being verifiable and comparable. In addition, the collection method of hard information need not be personal. However, there is little reported research connecting qualitative research on managerial capability with quantitative information.

The aim of the current paper is to address this deficit with an explorative study to quantify management capability using companies' quantitative transactional characteristics. This may allow banks to assess SMEs management capability, and therefore improve assessment accuracy on SMEs credit risk. It will also relate the findings to the previous work, with further contributions to the SMEs credit risk modelling taking in soft measures.

This paper is organized as follows. The next section covers previous research on managerial capability on the corporate financial performance. The third section is methodology, including principal component analysis and partial least squares regression. The following section describes data sample obtained from a UK bank. Then the results of principal component analysis and the results of partial least squares regression are presented. The subsequent section compares the predictive power of principal component analysis, partial least squares regression and often used logistic regression model. The final section is the conclusion.

Related Literature Review

Managerial capability is often an important aspect of small business success. It affects the planning process, financial conditions, the competitive environment, and decisions on growth and expansion.

Haswel and Holmes (1989) report that managerial inadequacy, incompetence, inefficiency, and inexperience are a consistent theme explaining small business failure. Poor management is interrelated with several issues such as poor financial conditions, inadequate accounting records, limited access to necessary information and lack of good managerial advice.

Keats and Bracket (1988) propose a conceptual model of small firm performance, based on strategic entrepreneurship and organizational theory. It provides a basis for explaining how owner characteristics, behaviour and contextual factors relate to small firm performance.

One of the earliest empirical studies of exploring owner-manager and firm characteristics to explain business failure is by Larson and Clute (1979). The numerous characteristics shared by failed firms are directly related to personal decision-based characteristics of the owner (lack of insight, inflexibility, emphasis on technical skills, etc.), managerial deficiencies (lack of management skills and appropriate managerial training, etc.) and financial shortcomings (no accounting background, poor cash flow analysis and financial records, etc.). However, Maes et al. (2005) found that owner-manager and company characteristics have no direct significant impact on financial performance. Significant paths have been found between these two aspects and management practices.

Gaskill et al. (1993), limit to apparel and accessory industrial sector, find that four factors that play important role in explaining small enterprise failure, are managerial and planning functions, working capital management, competitive environment, and growth and expansion. These four factors and their assocaited variables are not independent.

In a comparative study between key factors influencing SMEs' failure between the UK and Nigeria, it is found that internal factors such as management are the most significant factors for UK SMEs, while external factors such as economic condition and infrastructure play an important role for Nigerian SMEs, (Ihua, 2009).

The findings of Isachenkova and Weeks (2009) are congruent with the requirements of the Basel II Accord that encouraged banks to develop internal credit rating systems by taking into account relevant non-financial soft information. They investigate the importance of managerial capital to involuntary insolvency and acquisition in UK SMEs. The owner-manager's human capital was cast in terms of firm-specific, professional-specific, and generic components, measured respectively by tenure, education and age. Additional information such as previous experience of employement, and intentions about future growth are included. Their result indicate that firms run by managers with higher human capital and intentions to pursue a strategy of growth have greater survival propects and are less likely to be forced into insolvency or become acquired.

Much of this previous research mainly focuses on the qualitative aspects; there is no exploration of the relationship between qualitative based managerial capability and hard quantitative information. This will be explored within this paper.

Modelling Methodology

The information from the bank, which will be described in more detail later, consist of two elements: a set of predictive variables X and outcome variables Y. Principal Component Analysis (PCA) is a traditional multivariate statistical method commonly used to reduce the number of predictive variables and resolve the multi-colinearity problem in regression modelling. It aims to produce a limited number of linear combinations of the original set variables which are orthogonal and encapsulate a high proportion of the original variability. One drawback of PCA technique is that it derives the linear combinations only based on the characteristics of X, the predictive variables. It does not take account of the outcome variables Y. It is an unsupervised dimension reduction technique. Partial Least Squares regression (PLS) allows the user to capture the information in the predictive variables (X) as well as in the relation between the predictive (X) and outcome variables (Y). It provides an alternative approach to PCA technique. It is developed by Wold in the late 1960s for econometrics, and is very popular in area such as chemical engineering (Geladi and Kowalski, 1986, Kleinbaum et al, 1998, Mevik and Wehrens, 2007). An introduction to the technique and a statistical account of it can be found in Geladi and Kowalski (1986) and Tobias (2003). The comparison of principal component analysis and partial least squares regression can be found in Maitra and Yan (2008). The following two parts summarize the modelling methodology of principal component analysis and partial least squares regression.

Principal Component Analysis

Principal component analysis is employed to reduce the number of variables, usually in an exploratory manner to investigate the underlying relationships among a set of variables. It transforms a set of correlated response variables into a smaller set of uncorrelated variables called the principal component. Each principal component comprise of linear combinations of the original variables calculated usually using the correlation matrix of the original variables. PCA achieves the reduction by finding the eigen values and vectors of the correlation matrix, though, sometimes the variance covariance matrix may be used. The eigen values reflect the shares of the total variation and the eigen vectors are the weighted linear combinations of the variables, often referred to as scores. Finding the PCA is equivalent to finding a rotation of the original axes to form new axes so that the 1st principal component contains the highest amount of variation and 2nd principal component contains the highest amount of variation perpendicular to the first, and so on for all principal components. The i^{th} component can be described as

$$PC_i = W_{i1}X_1 + W_{i2}X_2 + \ldots + W_{ip}X_p$$

The principal components are extracted in decreasing order of importance, and the important principal components are those that express and contain more variance of dataset. Often only the first few principal components account for the majority of the variance accounted for the original set of variables. (Stevens 1992)

Partial least squares regression

Partial least squares regression (PLS) is a supervised factor extraction method in a way that captures as much information in the raw predictive variables as well as in the relation between the predictive and target variables, see Maitra and Yan (2008),

Assume that X is a $n \times p$ matrix and Y is a $n \times q$ matrix. PLS tries to find a linear decomposition of X and Y such that

$$X = TP' + E$$
$$Y = UQ' + F$$

where X and Y are the matrices of predictors and outcomes. The matrices on the right hand side of this model are defined by

T=X-scores   U=Y-scores

P=X-loadings Q=Y-loadings

E=X-residuals F=Y-residuals

PLS algorithms works by successively extracting factors from both X and Y such that covariance between X-score and Y-score is maximized. For a good PLS model, the first few factors show a high correlation between the X-scores and Y-scores. PLS has the flexibility to extract orthogonal X factors while not restricting themselves to the original model of PLS. Therefore, in this study, X-scores from the PLS decomposition is used separately for a regression to predict DBI, in order to compare the results with logistic regression.

To determine which predictors to eliminate from the analysis, Variable Importance for the Projection (VIP) is used. VIP coefficient reflects the relative importance of each X variable for each X factor in the prediction model. VIP coefficients thus represent the importance of each X variable in fitting both the X- and Y-scores since the Y-scores are predicted from the X-scores. If a predictor has a small value of VIP, then it is a prime candidate for deletion.

Data Sample

The data used in this study is from the SME portfolio of a leading UK bank. 35,000 observations on SMEs have been selected. They are randomly chosen from each of the five selected regions of the bank.

The predictive variables consist of 72 quantitative variables and 5 qualitative variables. The variables are transactional characteristics of SMEs which reflect their credit behaviour, such as repayment history and account usage behaviour over the past month, the past three months, the past six months, and the past twelve months. A large number of characteristics have high correlations.

The outcome variable, Y, is measured after 12 months of observation, from October 2007 to October 2008. This is regarded as the norm for the development of a application scoring system, see Thomas et al (2001). Two dependent variables are used. One is DBI, which has three values, which are good, bad and indeterminate. The bad observations account only 2% of the total sample, large amount of observations are good with 87%, indeterminate observations account 11% of total observations. This variable is used to assess the prediction power of models based on extracted principal components and latent variables, and compare the findings with the results of logistic regression.

The other dependent variable is a risk indicator. The risk indicator (RI) is a measure of the risk of the customer, ranging from 0 to 9 (0=low risk, 9=high risk). RI will be used in partial least squares as dependent variable to extract latent variables.

Results of Principal Component Analysis

The number of components to be extracted is usually determined by considering the change in the amount of variation accounted for by a component, often through the Scree plot (Rencher, 2002). In this case a major drop in the variation occurred between the 4^{th} and 5^{th} component and therefore only the first four components are considered. The first 4 components account for 50% of the variation of the original variables. It is possible to further rotate the variables to produce more interpretable components (Rencher, 2002).

Figure 1 shows the variables composition of the first four principal components. The variables with loadings above 0.5 are retained for further analysis. First component (PC1) does seem to be a measure of size with many variables contributing. The other components do seem to have fewer major variables, and may yield insight into management capability. After exploring the major contributing variables (with loadings above .5) to the principal components, it is found:

Second components (PC2) relates to the credit turnover and debit turnover, e.g. average monthly customer generated credit turnover of last three months, ratio of total debit to total credit on cheque and deposit account of last three months. This reflects transactional accounting information. Good accounting control will not yield bad credit turnover and debit turnover. PC2 can be attributed to poor management ability.

Third component (PC3) consists predominantly of variables describing delinquency in payment, for example, average number of days in excess of last three months, the worst consecutive Days Past Due during this month. Delinquency in payment is considered to be connected with management capability, as good control of the account will not yield days in excess in payment. It is consistent with findings of Wichmann (1983).

Fourth component (PC4) comprises the age of the account and hard-core balance, such as age of all account entity, time associated with the bank, trend in hard-core balance.

Results of Partial Least Squares regression

The optimal number of factors is determined by the predicted residual sum of squares (PRESS). The PRESS for cross-valiation is computed for successive factors, and the minimum PRESS point is selected as the basis for identifying the corresponding number of dimensions for the optimal model. Six latent factors would seem recommended by the PRESS result. According to Table 1, these six latent factors explain 53% of variance in the X variables; explain 59% of variance in the Y variables. The more a factor explains of the variation in the Y variables, the more powerful it will be in explaining the variation in the new sample of dependent values. The more a factor explains in the variation of the X variables, the more it well reflects the observed values of the set of independent variables. Increasing the number of latent factors to 15, the variation in X variables increased to 68%, but variance of Y stayed the same at 59%.

The plot of the X- and Y-scores for the first four latent factors is shown in Figure 2. In the plot, the X score is plotted on the X axis, and the Y score is plotted on the Y axis. These plots show the direction toward which each PLS X-factor is projecting with respect to the Y-factor. It is found that high correlation between X- and Y-scores for the first component, X-score is associated with increasing value of Y-score. A looser correlation exists for the second components. But the relationships for the third and fourth components are not strong. Recalled in Table 1, first two components contribute significantly, more than 5% of original variance, in the response variables. If ‘significant’ components are retained according to the criteria used by Carrascal et al (2009), which are those explaining more than 5% of original variance in the response variable, then the first two components were selected.

The meaning of each latent factor is understood by exploring the significant weights of predictors contributing to latent factors. One rule of thumb in confirmatory PLS factor analysis is loadings should be .7 or higher to confirm that independent variables identified a priori are represented by a particular factor (Hulland, 1999), however, the .7 standard is a high one and real-life data may well not meet this criterion, will use a lower level such as .4 for the central factor and .25 for other factors (Raubenheimer, 2004). This study chooses the variables which are higher than .4 as contributing variables for latent factor.

There is no obvious cluster pattern in latent factor 6, therefore the composition of first five latent variables is considered only. The interpretation of the remaining latent factors is:

First factor (LF1) represents current account and deposit account balance with one major predictor contributing.

Second factor (LF2) relates to soft features of the company, which include age of the account, time associated with the bank, region at which the customer relationship is managed, and industry type the company belongs to.

Third factor (LF3) is a management score.

Fourth factor (LF4) consists of customer generated credit turnover, such as trend in customer generated credit turnover for quarter 1 compared to the average of the other three quarters.

Fifth factor (LF5) is delinquency in payment, similar explanation as principal components 3 in PCA.

The extracted latent factors from PLS are close to the extracted principal components from PCA. Factor four (LF4) and five (LF5) from PLS are similar to second factor (PC2) and third factor (PC3) from PCA. It is concluded that they both reflect management capability of the company. In addition, PLS gives more priority to soft features of the company. This could be the reason that PLS captures response information. It confirms the importance of soft feature in prediction of credit risk of corporate. In the next part, the prediction using latent factors based on PLS is compared with PCA and stepwise logistic regression.

A comparative analysis of prediction power across models

The prediction power of partial least squares regression and principal component analysis can be compared with traditional logistic regression model. Adjusted R-square is used to test fitness of the model. Partial least square regression shows the best model fitness among the three models, see Table 2. In terms of prediction ability, we have compared the outcome of first four principal components from PCA, first six latent factors from PLS, with the traditional stepwise logistic regression. To specify, the X-scores from the PLS decomposition are used separately for a regression to predict three states response variable DBI.

Area under the Receiver Operating Curve (AUROC) provides a measure of discrimination which the likelihood that a subject with probability of default than a subject with probability of non-default for an entire range of possible cutoff points. Considering there are three statuses of response, 'indeterminate' observations are considered to combine with either 'good' or 'bad'. Industry habit is to combine 'indeterminate' with 'good'. However, Ma (2011) employed cluster analysis, and notably found that most 'indeterminate' observations show similar pattern with 'bad'. Therefore, 'indeterminate' is grouped with 'bad' as one group while leave 'good' when setting binary outcomes in this study.

From Table 2, partial least squares regression performs better than principal component analysis according to the value of AUROC. The PLS is more efficient than the PCA due the supervised nature of its algorithm. Logistic regression seems perform best, it is the reason that more variables contributing in the regression model. But when the predictive variables are highly correlated, the tests of statistical significance that the stepwise method is based on are not sound, as independence is one the primary assumptions of these tests.

Conclusion

This study elicits the concept of management capability from quantitative transactional characteristics, using principal component analysis and partial least squares regression. It is the first study of investigating the quantitative reflection of qualitative concept of management capability.

The findings of both modelling framework are consistent in terms of management capability. It is found that financial measure (credit turnover and debit turnover) and the performance measure (number of days in excess of the account) could be considered as reflecting management capability. Drawing upon this finding, credit managers might be suggested to emphasize the quantitative measures when assessing SMEs management.

In addition to the same latent factors of PLS as PCA, soft features of the company are extracted as a significant factor in PLS. The contributing soft features are age of the account, region where the customer relationship is managed, and the industry the company belongs to. This finding confirms the importance of companies' soft feature in determining the credit risk of the company, as PLS has the advantage of capturing the relationship between predictor variables and outcome variable. It is suggested that this soft information should be included when predicting SME credit risk.

The performance of PCA and PLS are compared with traditional logistic regression. Across the three models, PLS has the best model fit. Partial least squares regression has better predictive accuracy than principal component analysis. Logistic regression shows the highest predictive power, but not distinct, it is the reason that stepwise method allows more variables to contribute.

Figure 1: Contributing variables of first four components of principal component analysis

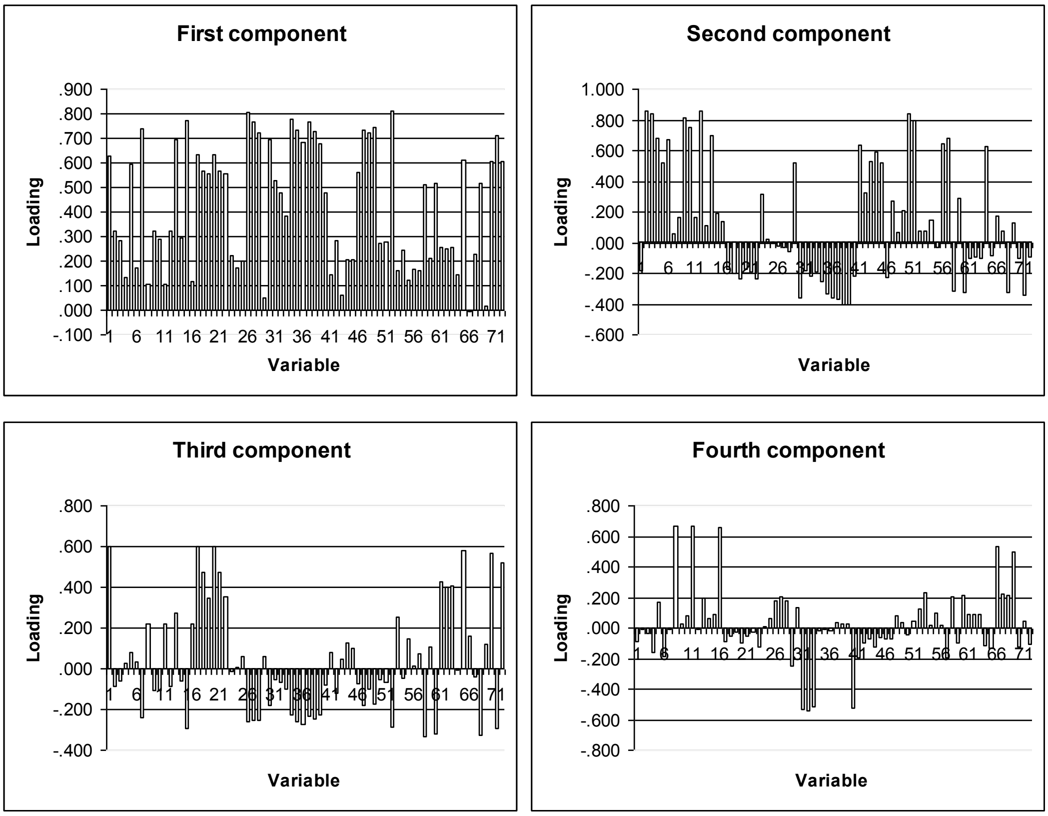
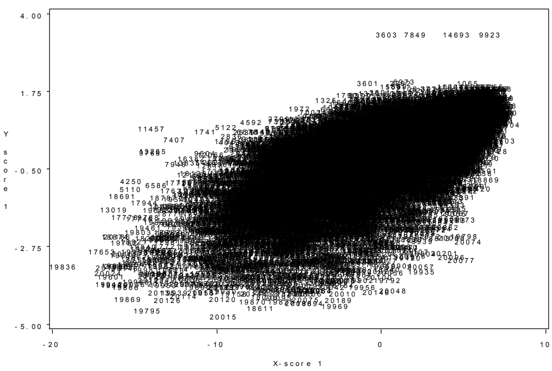
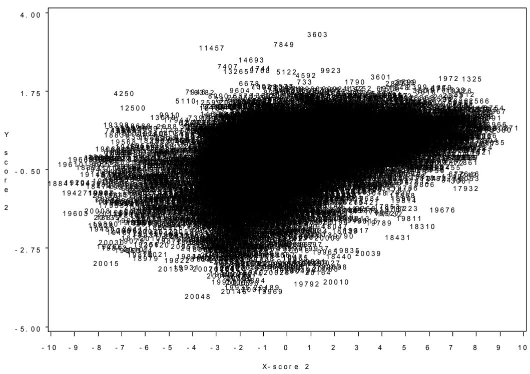


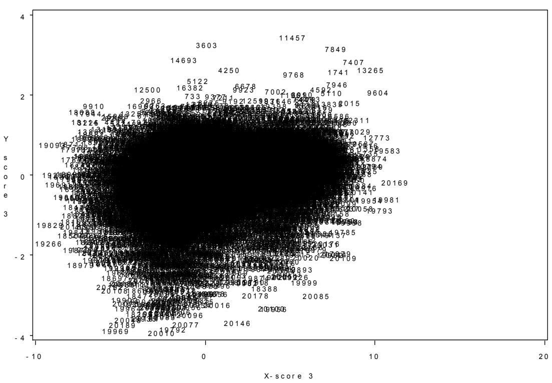
Figure 2: Plots of X-score and Y-score for first four latent factors of partial least squares regression



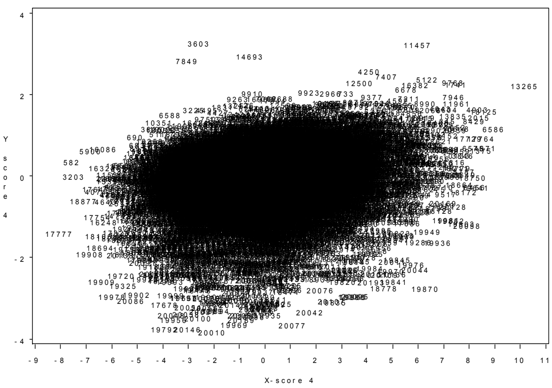
First latent factor



Second latent factor



Third latent factor



Fourth latent factor

Table 1: Percentage variation accounted by partial least squares factors

Number of extracted factors	Model effect		Dependent variables	
	Current	Total	Current	Total
1	21.8484	21.8484	45.9352	45.9352
2	7.7047	29.5532	9.7530	55.6882
3	10.7821	40.3352	1.5677	57.2558
4	5.5411	45.8763	1.5020	58.7578
5	2.9750	48.8513	0.4496	59.2074
6	4.5895	53.4408	0.0953	59.3027

Table 2: Comparison of predictive power across three models

	PCA Regression	**PLS Regression**	**Logistic Regression**
Adjusted R-square	0.42	0.45	0.35
AUROC	0.86	0.88	0.89

References:

Alexander C (2001). *Market Model: A Guide to Financial Data Analysis*. John Wiley&Sons Ltd: UK.

Carrascal L M, Galvan I, and Gordo O (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos.* 118: 681-690.

Carter R., and Auken H V (2006). Small firm bankrupcy. *Journal of Small Business Management.* 44(4): 493 - 512.

Gaskill L R, Van Auken H E, and Manning R A (1993). A factor analystic study of the perceived causes of small business failure. *Journal of Small Business Management.* 5(4): 18-31.

Geladi P and Kowalski B (1986). Partial least squares regression: a tutorial. *Analytica Chimica Acta*. 185: 1-17.

Haswell S and Holmes S (1989). Estimating the small business failure rate: a reappraisal. *Journal of Small Business Management.* 27(3): 68-74.

Hulland J (1999). Use of Partial Least Sqaures (PLS) in strategic management research: a review of four recent studies. *Strategic Management Journal.* 20: 195-204.

Ihua U B (2009). SMEs key failure-factors: a comparison between the United Kingdom and Nigeria. *Journal of Social Science.* 18(3): 199-207.

Isachenkova N and Weeks M (2009). *Acquisition, Insolvency and Managers in UK Small Companies.* University of Cambridge: Cambridge, U.K..

Keats B W and Bracker J S (1988). Toward a theory of small firm performance: a conceptual model. *American Journal of Small Business.* 12(4): 41-58.

Kleinbaum DG. et al.(1998). *Applied Regression Analysis and Multivariate Methods*, 3rd Edition, Pacific Grove, Ca.:Brooks/Cole Publishing Company.

Larson C M and Clute R C (1979). The failure syndrome. *American Journal of Small Business.* 4(2): 35-43.

Lussier R N (1995). A nonfinancial business success versus failure prediction model for young firms. *Journal of Small Business Management.* 33(1): 8-20.

Ma Y (2011). *Modelling Credit Risk of Small and Medium Sized Enterprises using Transactional, Accounting and Market Variables.* University of Edinburgh, Edinburgh.

Ma Y and Lin S (2010). 'Credit crunch' and Small- and Medium-sized Enterprises: Aspects affecting survival. *Journal of Financial Services Marketing.* 14(4), 290-300.

Maes J, Sels L, and Roodhooft F (2005). Modelling the link between management practices and financial performance: evidence from small construction companies. *Small Business Economics.* 25(1): 17-34.

Maitra S and Yan J (2008). Principle Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression. http://www.casact.org/pubs/dpp/dpp08/08dpp76.pdf, accessed 26 October 2011.

Mevik B H and Wehrens R (2007). The pls package: principal components and partial least squares regression in R. *Journal of Statistical Software.* 18(1): 1-24.

Perry S C (2001). The relationship between written business plans and the failure of small businesses in the U.S. *Journal of Small Business Management.* 39(3): 201-208.

Raubenheimer J E (2004). An item selection procedure to maximize scale reliability

and validity. *South African Journal of Industrial Psychology.* 30(4): 59-64.
Rencher A C (2002). *Methods of multivariate analysis*. John Wiley&Sons, Inc: US.
Steven J (1992). Applied Multivariate Statistics for the Social Sciences. Erlbaum, New York.
Thomas L C, Eldelman D B, and Crook J N (2002). *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics: US.
Tobias R D (2003). An Introduction to Partial Least Squares Regression. http://www.ats.ucla.edu/stat/sas/library/pls.pdf, accessed 26 October 2011.
Wichmann H (1983). Accounting and marketing-key small business problems. *American Journal of Small Business.* 7(4): 19-26.
Wold H (1994). PLS for multivariate linear modeling. In H. V. D. Waterbeend (Ed.). *QSAR: Chemometric Methods in Molecular Design: Methods and Principles in medicinal chemstry*. Verlag-Chemie: Weinheim, Germany.
Wold H (1975). Soft modelling by latent variables; the nonlinear iterative partial least squares approach.In:Gani,J.(ed.) Perspectives in probability and statistics. Papers in honour of M.S.Barlett. Academic Press, pp.117-142.