  0160-5682/16

www.palgrave-journals.com/jors/

Exploring Management Capability in SMEs using transactional data

Yigui Ma[1,2*], Jake Ansell[2] and Galina Andreeva[2]

[1]Credit Reference Centre, People's Bank of China, Beijing, China; and [2]University of Edinburgh, Edinburgh, UK

Small- and medium-sized enterprises (SMEs) have become very important in most world economies. Governments have developed policies to support them and within the United Kingdom the government has encouraged lending to SMEs. Traditional relationship banking is based on the confidence a banker may have in the quality of SME's management. Yet with a shift towards transactional quantitative risk assessment, there is a concern that Management Capability, which is critical to the success of a SME, is not necessarily captured by risk modelling. This paper reports findings from work on determining Management Capability from quantitative transactional measures. The study has used principal component analysis and partial least squares regression to elicit manifestations of Management Capability. The results indicate some success in determining measures for Management Capability.

Journal of the Operational Research Society (2016) **67**(1), 1–8. doi:10.1057/jors.2014.90

Published online 12 November 2014

Keywords: 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). There is also the argument that some SME are innovatory and so may subsequently develop into the major companies of the future, and hence can become important for the prosperity of an economy (de la Torre *et al*, 2010). The problem for the banks is related to some difficulties in risk assessment specific to SME sector. For example, often it is hard for a bank to discern the relevant information from SMEs. The previous research has highlighted SMEs informational opacity as a major problem: a lot of smaller companies are not listed, the financial statements may not be audited or complete (Berger and Frame, 2007). It therefore becomes more critical to use the available information to enhance the risk modelling of SMEs.

In the past, credit managers have attempted to overcome this issue via relationship lending, and aspect of that is the confidence a banker may have in the SME's management. This may be even more important when there is a downturn in the economy. Understanding the business customer becomes a major issue. One aspect of this is being able to assess the Management Capability of the enterprise owner(s). In this research, the term 'Management Capability' is borrowed from Organizational Theory and Information Management. Building on the concept from Chen and Wu (2011) who define a capability 'as a concept encompassing the possession of skills/ knowledge for the effective execution of specific activities', the term 'Management Capability' in this paper describes the ability of the owner/management team within the enterprise to control the running of the business to achieve good overall performance of the business. It is a complex phenomenon composed of a number of attributes. In the context of this paper the interest focuses on those attributes that may facilitate lending and can be derived from transactional data on SMEs obtained from a bank's internal records.

It is believed 'that managerial resources or capabilities are key contributors to the entire bundle of firm resources' (Thomson and Heron, 2005), and therefore significant contributors to the success of a SME and its access to finance. The traditional development of a face-to-face relationship between the bank and especially micro-SME customers is increasingly under challenge due to the costs to the bank that it represents. As a consequence many banks in the United Kingdom are trying to identify ways to understand Management Capability in the absence of such a close relationship.

Access to finance is conditional on the firm proving its credit worthiness to the lender. This process which is also referred to as credit risk assessment is mainly based on 'hard information', which includes financial accounts and payment performance (Grunert *et al*, 2005; Grunert and Norden, 2012). Yet the importance of 'soft information', which includes 'borrower's management skills, the product-market position and his strategy' (Grunert and Norden, 2012), is increasingly emphasised.

**Correspondence: Yigui Ma, Business School, University of Edinburgh, Room 3.02, 29 Buccleuch Place, Edinburgh EH8 9JS, UK; Credit Reference Centre and Financial Research Institute, People's Bank of China, No. 32, Chengfang Street, Xicheng District, Beijing 100800, China.*
E-mail: mayiguicn@gmail.com

Furthermore, the authors quoted above observe that 'the notion of soft information is not well defined in the literature' (*ibid*.).

While there is no universally agreed definition of the soft information, most of the authors consider some aspects of Management Capability as being part of it. Examples of such aspects are long-term planning (Gaskill *et al*, 1993; Lussier, 1995; Perry, 2001; Maes *et al*, 2005; Carter and Van Auken, 2006; Lussier and Halabi, 2010), financial/accounting knowledge (Wichmann, 1983; Gaskill *et al*, 1993; Lussier, 1995; Lussier and Halabi, 2010), use of information systems or of internet (Maes *et al*, 2005; Carter and Van Auken, 2006), openness to advice and second opinion (Gaskill *et al*, 1993; Lussier, 1995; Maes *et al*, 2005; Lussier and Halabi, 2010). For SMEs banks would regard the cost of collection of the information as too high for the size of the loans being considered since it would impose costs that would have to be borne by either the SME or the bank. For larger SMEs, where there is a greater interaction, bank managers may be asked to subjectively assess the quality of the management team. This will be based on the views the bank manager will gain from visits and so it may be hard to ensure consistent measurement of the quality of the team. Moreover, the banks would wish to explore alternative measures based on data they can directly obtain that can be verified and compared across SMEs. In discussion with the bank (the anonymous data provider) the interest has been expressed in measuring 'Management Capability' based on verifiable data they had immediate access to, which implies use of bank account transactions. This paper will therefore focus on exploring methods to elicit Management Capability using companies' quantitative transactional characteristics (hard information). This may allow banks to assess SMEs' management quality, and therefore improve assessment accuracy of SMEs credit risk. It will also contribute to the SMEs credit risk modelling by exploring the relationship between hard quantitative measures and underlying soft concepts. As only anonymised records for the SMEs are available, it is not possible to determine the relationships directly between the derived quantitative measures and soft characteristics within specific SMEs.

This paper is organised as follows. The next section summarises previous research on Management Capability and its relation to the firm's performance. The third section is methodology, including principal component analysis (PCA) and partial least squares (PLS) regression. The following section describes a data sample obtained from a UK bank. Then the results of PCA and the results of PLS regression are presented. The subsequent section compares the predictive power of principal components, PLS regression and often-used logistic regression model. The final section concludes.

Management Capability and small business performance

Management Capability is believed to be an important aspect of small business success, as shown in the studies below that attempted to link SMEs performance to various aspects of management. This section reviews only a selection of studies that are most relevant within the context of this paper (see Ma and Lin, 2010 for a more detailed literature review on SME performance modelling with summary tables).

A conceptual model of small firm performance was proposed by Keats and Bracker (1988) based on strategic, entrepreneurship and organisational theory. The paper suggested that Entrepreneurship Intensity (an owner/manager characteristic) would impact on Behavioural Strategic Sophistication through Task Motivation and Percieved Strength of Environmental Influence. Next, owner's behaviour would impact on firm's performance mediated by Cognitive Strategic Sophistication and Task Environment. This theoretical model established a basis for the explanation of how success or failure of SME relate to owner characteristics, behaviour and contextual factors. Later empirical study by Maes *et al* (2005) confirmed that owner–manager characteristics (education, financial experience, etc) and company characteristics (company size and age) had no direct significant impact on financial results, but were indirectly connected to the performance via the choice of management practices.

As for the aspects of manager/owner characteristics, a number of studies investigated them and their relation to success/failure in different industries and counties. An early empirical paper exploring the relationship between owner–manager characteristics and business failure was by Larson and Clute (1979). They extracted common characteristics from the sample of 359 US small firms with financial difficulties and grouped them into three main categories. These were personal characteristics of the owner (limited education, lack of insight, inflexibility, etc), managerial shortcomings (inadequate knowledge of pricing, target market, lack of advertising and promotional strategy, failure to generate effective plans, etc) and financial deficiencies (incomplete accounting background, poor cash flow analysis, etc).

Haswell and Holmes (1989) reviewed the research on small business failure and their causes in Australia. They concluded that managerial inadequacy, which is composed of management inexperience and incompetence, is often a major contributor to SME failure. Inadequate management is related to several aspects: inadequate or poor accounting records, deficiency in accounting knowledge, limited access to the information required to assist business decision and lack of good managerial advice.

Gaskill *et al* (1993), in an empirical study on apparel and accessory industrial sector, found the following factors playing an important role in explaining small enterprise failure: managerial and planning functions, working capital management, competitive environment, and unsustained growth and expansion. Lussier (1995) suggested a 15-variable business success model, which included such characteristics of managers as education, industry and management experience, marketing skills. Business functions considered important were record keeping and financial control, planning, staffing, the use of professional advice. The original study looked at the

United States, but later confirmed the model in Croatia (Lussier and Pfeifer, 2001) and Chile (Lussier and Halabi, 2010). In a comparative study of the key factors influencing SMEs' failure between the United Kingdom and Nigeria, it was found that internal factors such as 'management inability or ineffciency' were the most significant factors for UK SMEs, while external factors such as economic condition and infrastructure played more important role for Nigerian SMEs (Ihua, 2009). In terms of specific managment functions, planning marketing and accounting were percieved important.

Lack of knowledge was determined to be one of the crucial factors leading to bankruptcy of US small firms by Carter and Van Auken (2006). The latent factor Knowledge comprised dimensions of general managerial skills, accounting, marketing and planning functions, which is consistent with the studies previously described.

An importance of planning function was further emphasised by Perry (2001) in investigation of the US small businesses. Later study by van Gelder *et al* (2007) confirmed that a detailed and long-term plan distinguished operational firms from bankrupt ones in the Fiji Islands. Other significant factors included specific and more difficult goals and higher degree of human capital, which consisted of education and experience indices.

The concept of human capital was reiterated by Isachenkova and Weeks (2009), who used the term Managerial Capital and investigated its connection to SMEs failure in the United Kingdom. The authors followed Becker's (1964) definition of human capital as 'knowledge and skills obtained through formal education and professional training, and accumulated through work experience' (Isachenkova and Weeks, 2009) and defined three elements within managerial (human) capital: firm-specific (measured by years with the firm), professional-specific (measured by education) and generic components (measured by age). Additional information such as previous instances of unemployement and plans about firm's future growth were also included. Their results demonstrate that managers with higher managerial capital and more ambitious plans for growth provide better survival propects for their firms.

Cohen and Kaimenakis (2007) explored a related concept of Intellectual Capital. Using a sample of knowledge intensive SMEs from Greece the authors found 'Hard' Intellectual Assets Factor (including capabilities and skills) to be significant in explaining financial performance.

Although there are differences in terminology, one can conclude that owner/manager characterisics are important for business success. Table 1 summarises the aspects/dimensions of Management Capability that have been found to be significant in the studies described above. Intellectual Capital/Competence/Knowledge is the most recurring theme, with accounting and planning being most mentioned among specific management areas.

Yet much of the previous research described above mainly focuses on the qualitative aspects or owner–manager characteristics and performance measures derived from financial accounts, which are usually submitted on a yearly

Table 1 Summary of Management Capability aspects found important in previous research

	Larson and Clute (1979)	*Keats and Bracker (1988)*	*Haswel and Holmes (1989)*	*Gaskill et al (1993)*	*Lussier (1995), Lussier and Halabi (2010)*	*Perry (2001)*	*Maes et al (2005)*	*Carter and Van Auken (2006)*	*Cohen and Kaimenakis (2007)*	*Van Gelder et al (2007)*	*Ihua (2009)*	*Isachenkova and Weeks (2009)*
Intellectual/Managerial/Human capital									X			X
Education	X				X		X					X
Business experience/Competence			X	X	X		X	X		X	X	X
Entrepreneurship		X										
Strategic sophistication		X										
Financial/Accounting knowledge	X		X	X	X		X	X			X	
Marketing knowledge	X				X			X			X	
Planning	X			X	X	X	X	X		X	X	
Innovative insight/Flexibility	X											
Leadership/Team management	X						X					
Ambitious goals										X		X
Well-grounded decision making/Efficient use of information	X		X				X	X				
Use of professional advice	X		X		X							
Use of internet								X				

basis. There is no exploration of the relationship between qualitative Management Capability and quantitative information arising from credit transactions. The aim of the current paper is to fill in this gap and to investigate whether it is possible to derive measures of Management Capability from the information the bank has already available, and thus to explore a relationship between Management Capability and credit performance.

Modelling methodology

The most common methodology used in the previous research is factor or PCA, the representative papers being Gaskill *et al* (1993) and Carter and Van Auken (2006). This paper continues this line of investigation by PCA to a new type of data, and in additon employs PLS in order to improve the accuracy of predicting the credit performance. The latter technique has not been considered in previous research on SMEs credit performance, although it is widely used in other areas, including modelling capability of IT management personnel (Chen and Wu, 2011).

The information from the bank, which will be described in more detail later, consists of two elements: a set of transactional characteristics (predictor variables *X*) and subsequent credit performance (outcome variables *Y*). Both PCA and PLS aim at constructing new variables from the original set of variables with specific properties. In both cases, it is desired that the newly derived variables will be interpretable within the research context and shed further light on the subject of study. In the case of the current study, it is hoped that further insight into Management Capability will be derived.

The objective of PCA is to produce a limited set of new variables accounting for most of the variation within the old set of variables. These new variables will be a linear combination of the old variables that are orthogonal to each other and so remove the issues of multicollinearity when using the new variables in further analysis. Only the predictor *X* variables are considered. PLS was originally developed by Wold (1966). PLS could be said to follow a similar pattern to PCA, but the difference is that instead of simply using the predictor *X* variables it takes into account also the outcome *Y* variables. Hence PCA can be regarded as unsupervised dimension reduction since the outcome *Y* variables do not influence the reduction, whereas PLS is supervised.

PCA is a technique used generally across a wide range of domains, often used within Social Sciences and Business under the name of factor analysis. Frequently it is used to provide components that will be interpreted within the research context. PLS is frequently used in chemical engineering and related areas (Geladi and Kowalski, 1986; Kleinbaum *et al*, 1998; Mevik and Wehrens, 2007). An introduction to the technique and statistical details can be found in Geladi and Kowalski (1986) and Tobias (2003). Maitra and Yan (2008) provide a comparison of PCA and PLS. In the following two sections the main elements of the two approaches are discussed.

Principal component analysis

PCA rotates the original axes of the data to produce a set of axes that transforms the original correlated *X* variables into a set of new uncorrelated variables, the principal components (PC). It is often used as an exploratory technique to identify structure within the original *X* variables. Each of the principal components comprise linear combinations of the original variables, $\mathrm{PC}_i = W_{i1}X_1 + W_{i2}X_2 + \cdots + W_{ip}X_p$. PCA achieves the rotation by finding the eigenvalues and eigenvectors of the correlation matrix, though sometimes the variance–covariance matrix may be used. The eigenvalues reflect the shares of the total variation and eigenvectors provide the weights for the linear combinations of the variables, often referred to as scores. The principal components are usually ordered so that the first principal component contains the highest amount of variation and the second principal component contains the next highest amount of variation perpendicular to the first, and so on for all principal components. Often only the first few principal components account for the majority of the variance contained in the original set of variables (Stevens, 1992; Alexander, 2001). This enables the reduction in the dimensionality and it is hoped that these reduced set of new variables can be appropriately interpreted in the context of the research.

PLS regression

The difference from PCA of PLS is that allowance is made for the set of outcome *Y* variables, and so unlike PCA, PLS is supervised extraction of factors. Following Maitra and Yan (2008), PLS will find linear compositions of *X* and *Y* such that

$$M = TP' + E$$

$$N = UQ' + F$$

where M is a $n\times p$ matrix of the X variables and N is a $n\times q$ matrix of the Y variables

$$T = X\text{-scores}, \quad U = Y\text{-scores}$$

$$P = X\text{-loadings}, \quad Q = Y\text{-loadings}$$

$$E = X\text{-residuals}, \quad F = Y\text{-residuals}$$

The algorithm is an iterative procedure at each step producing *T* and *U* so that the covariance between them is maximised. The algorithm can also produce the estimates of B and B_0 in the regression model $\hat{Y} = XB + B_0$, referred to as PLS regression. In the current study, *Y* will be assumed to be a single outcome variable. The estimates of B and B_0 will be used to produce predicted *Y*-score for comparison with the results obtained from the logistic regression applied to original non-transformed variables and principal components obtained from PCA.

Data sample

The data used in this study is from the SME portfolio of a leading UK bank. Full details of this data set and complete description of the variables used cannot be disclosed due to the conditions of Confidentiality Agreement with the data provider. 35 000 observations on SMEs have been randomly selected. These have been split randomly into 2/3 for a development sample and 1/3 for a hold out or test sample.

The predictor variables consist of 70 numeric variables and 5 categorical 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 3 months, the past 6 months, and the past 12 months. A large number of characteristics have high correlations. All predictors (including the categorical ones) were coded using WOE (weights of evidence transformation) as described in Lin *et al* (2012).

The predictors are recorded at or before October 2007. The outcome variable Y is measured at the end of 12 months of observation, in October 2008. This is regarded as the norm for the development of a scoring system (see Thomas *et al*, 2002). Two dependent variables are used. One is GBI (Good/Bad/Indeterminate) flag, which has three values of ‘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 outcome variable is used to assess the prediction power of PLS regression and logistic regression based on extracted principal components and original variables. It is a standard practice in credit scoring to reduce the three outcome categories outlined above to a binary indicator (Thomas *et al*, 2002). There is no definitive answer how one should deal with ‘indeterminate’ category, yet the industry most often combines ‘indeterminate’ with ‘good’ (Hand and Henley, 1997; Tebboth and Gadi, 2009). However, Ma (2011) through the application of cluster analysis to the same data set as used in this paper has found that most ‘indeterminate’ observations show similar pattern with ‘bad’. Therefore, in the analysis presented in this paper ‘indeterminate’ are combined with ‘bad’ as one group, whereas ‘good’ were left separate when setting binary outcomes in this study.

Results of PCA

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 the levelling of the change in variation occurred starting from fifth component (see scree plot in Figure 1). On the basis of the plot, one could argue for either using the first four or first five components. Yet when fitting a logistic regression on the first five components to predict performance (as described later in ‘A comparative analysis of predictive power across the models’), the first four components were significant and the fifth component was not, testing at 5% level of significance (Table 2). Therefore, only the first four components are considered. The first four components account for almost half (49%) of the total variation in the original variables (Table 3). This table presents five components for information and for comparison to a finally adopted four-component solution.

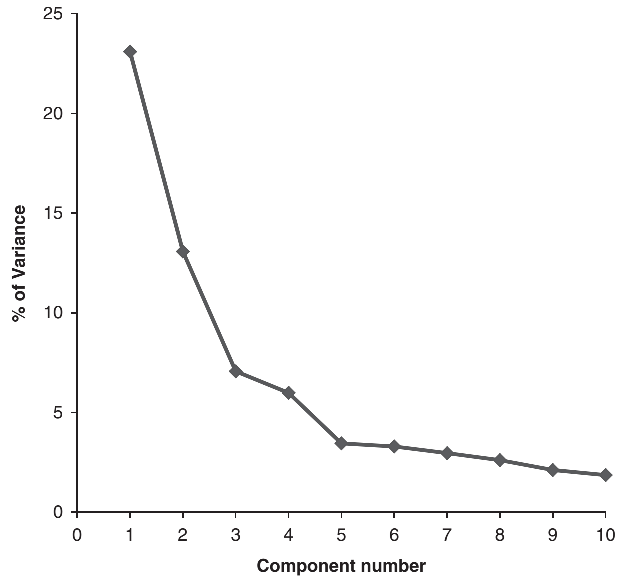

Figure 1 Scree plot for principal components (training sample).

Table 2 Logistic regression estimates of five principal components when predicting credit performance

Parameter	*DF*	*Estimate*	*Standard error*	*Wald* χ^2	*Probability* $>\chi^2$
Intercept	1	2.6091	0.0341	5838.7597	<0.0001
PC1	1	1.2752	0.0303	1770.3210	<0.0001
PC2	1	0.0785	0.0259	9.2054	0.0024
PC3	1	0.7426	0.0251	874.1317	<0.0001
PC4	1	0.1329	0.0242	30.1109	<0.0001
PC5	1	−0.0125	0.0240	0.2700	0.6033

Table 3 Percentage variation accounted by principal components (training sample)

Component	*Initial eigenvalues*		
	Total	*Percentage of variance*	*Cumulative percent*
1	17.320	23.094	23.094
2	9.804	13.072	36.166
3	5.297	7.062	43.228
4	4.488	5.984	49.213
5	2.587	3.450	52.662

The meaning of each principal component is assessed by examining the loadings of contributing variables. The variables with loadings above 0.5 (Comrey and Lee, 1992) are used in this study to interpret the meaning of each component. First component (PC1) is perceived to be a measure of size with many variables contributing. The other components have fewer major variables, and yield some insight into Management Capability. After exploring the major contributing variables (with loadings above 0.5) to the principal components, it can be deduced.

Second component (PC2) relates to the credit turnover and debit turnover, for example, average monthly credit turnover of last 3 months, ratio of total debit to total credit of last 3 months. This reflects transactional accounting information. Tight accounting control will result in efficient and prudent use of credit. Some external factors (such as the 'Credit Crunch') may in themselves have an impact on turnover. However, good management (and accounting) control will identify these trends early and enable the business manager to mitigate the issue by appropriate actions. This factor can be attributed to Management Capability.

Third component (PC3) consists predominantly of variables describing delinquency in payment, for example, average number of days in excess of last 3 months, the worst consecutive Days Past Due during this month. Delinquency in payment is considered to be connected to Management Capability, as good control of the account will not yield days in excess in payment.

Fourth component (PC4) comprises the age of the account and its balance, such as age of the account, time associated with the bank, and trend in the balance.

Results of PLS regression

Table 4 reports percentage of variance explained by PLS latent factors, the first five latent factors explain 48.6% of variance in the *X* variables and 33.4% of variance in the *Y* variable. Increasing the number of latent factors to 6, the variation in *X* variables rises to 52.8%, but variance of *Y* stays almost the same at 33.5%. Obviously the greater the variation within *X* variables accounted for by the latent factors the better it will reflect these variables. Similarly, the higher the amount of variance accounted for in the *Y* variable the more powerful the likely predictive power of it into the future. Yet it would seem from parsimonious point of view one should use the lower number of latent factors. For the purpose of comparison to PCA it would seem logical to maintain approximately the same percentage of *X* variance explained by retained components, that is, 49%. Therefore, five PLS latent factors are retained.

Table 4 Percentage variation accounted by partial least squares factors (training sample)

Latent factor	*X variance*	*Cumulative X variance*	*Y variance*	*Cumulative Y variance (R^2)*
1	21.8	21.8	25.0	25.0
2	7.7	29.5	6.9	31.9
3	3.3	32.8	1.2	33.1
4	11.6	44.4	0.1	33.2
5	4.2	48.6	0.2	33.4
6	4.2	52.8	0.1	33.5

The meaning of each latent factor is understood by exploring the significant weights of predictors contributing to latent factors. This study chooses the variables that are higher than 0.4 as contributing variables for a latent factor.

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 summary of past performance.

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

Fifth factor (LF5) is delinquency in payment, similar explanation as PC3 in PCA.

The extracted latent factors from PLS are close to the extracted principal components from PCA. Factors 4 (LF4) and 5 (LF5) from PLS are similar to second factor (PC2) and third factor (PC3) from PCA. It can be concluded that they both reflect Management Capability of the company. In addition, PLS gives more weight to soft features of the company. The reason could be that PLS captures response information. It confirms the importance of soft features in prediction of credit risk of businesses. In the next part, the quality of prediction using latent factors based on PLS is compared with PCA and stepwise logistic regression.

A comparative analysis of prediction power across models

Logistic regression is regarded as the most widely used model in credit scoring (Thomas *et al*, 2002). The predictive power of PLS regression is compared with logistic regression based on principal components and to logistic regression based on original predictors. The latter model employs forward stepwise selection mechanism to identify statistically significant predictors (5% significance level) in predicting a binary target variable, which is G/BI (Good *versus* Bad + Indeterminate) flag. The first four principal components from PCA are used as inputs into a logistic regression to predict G/BI flag. In the case of PLS it is possible to predict directly the G/BI flag. Table 5 shows the pseudo R^2 and number of variables fitted for the training sample, and measures of predictive accuracy for the test sample the specificity, the sensitivity, the area under the receiver operating curve (AUROC) and *H*-measure.

Pseudo R^2 is used to test the model fit. Logistic regression on 29 variables shows the best model fit among the three models.

Table 5 Comparison of predictive power across three models (test sample).

	PCA regression	*PLS regression*	*Logistic regression*
Variables/Factors fitted	4	5	29
Pseudo R^2	0.40	0.33	0.456
Sensitivity	0.533	0.565	0.573
Specificity	0.932	0.937	0.938
AUROC	0.860	0.883	0.886
H-measure	0.418	0.468	0.476

To compare the predictive power of the three models, Sensitivity, Specificity, AUROC and *H*-measure are used. AUROC provides a measure of discrimination, which is the probability that a good subject will score better than a bad subject for an entire range of possible cut-off points (Thomas *et al,* 2002). According to Hand (2009a), the AUROCs are not comparable across models in different frameworks and *H*-measure is proposed, which considers the loss associated with misclassification of 'default' and 'non-default'. Hand argued that AUROC is equivalent to measuring the performance of classification rules using metrics that depend on the rules being measured, while *H*-measure uses a universal standard cost ratio distribution, which avoids the relative misclassification costs that differ from classifier to classifier. Software R is used to compute *H*-measure with misclassification costs set to be equal to observed ratio of Goods to Bads and Indeterminates in the test sample (Hand, 2009b).

Sensitivity (the proportion of positives—or Bads in this case —that are predicted to be positive) and Specificity (the proportion of negatives—or Goods—that are predicted to be negative) both require a cut-off, which was set to observed numbers of Goods and Bads in the test sample (Anderson, 2007). From Table 5 all measures indicate that PLS performs better than PCA, while logistic regression seems to perform best. Yet PLS is only marginally short of logistic regression. The PLS is more efficient than the PCA due to the supervised nature of its algorithm. Although logistic regression is best in prediction, some care has to be taken because of the multicollinearity between the predictor variables. PLS provides good prediction and at the same time avoids the problem of multicollinearity.

In summary, logistic regression is best at prediction compared with PCA and PLS. PLS can achieve predictive accuracy close to that of logistic regression and avoiding the problem of multicollineairty. In addition, PCA and PLS provide insights into the interrelationship among predictors.

Discussion and Conclusions

Management Capability is acknowledged to be important for the success or failure of SMEs. Previous research has focused primarily on investigation of dimensions of this concept associated with bankruptcy or financial performance based on financial accounts. The primary aim of this paper is to derive Management Capability from credit transactional characteristics, using PCA and PLS regression. It is the first reported study to investigate the concept of Management Capability on transactional data and its association with credit performance. This study will benefit banks by helping them to get insights into SMEs management quality in the absence of high-cost relationship banking.

There is an overlap between the latent dimensions derived from PCA and PLS with LF1 related to PC4 (time), LF4 to PC2 (financial management) and LF5 to PC3 (credit delinquency). The findings from both techniques are also consistent in terms of interpretation. Financial management measure (credit turnover and debit turnover) and the credit delinquency measure (number of days in excess of the account) could be considered as reflecting Management Capability. Good management can identify trends at a very early stage and takes action to mitigate the issue.

One can relate these dimensions to previous research: financial/accounting knowledge was considered important in the majority of studies in Table 1. This dimension is also related to concept of planning. Temporal dimension reflects that mature SMEs perform better than newly formed SMEs, and this is again in line with previous research on importance of experience (Table 1) and also of age of the owner/ manager (Isachenkova and Weeks, 2009) and of the company (Maes *et al*, 2005). Obviously the credit delinquency measure is specific to the context of this study and have not been picked up in previous research, although an argument can be made that it is a component of financial/accounting knowledge.

An additional latent factor is extracted from PLS, which predominately consists of soft features of the company such as 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. Therefore it is suggested that soft information should be included in predicting SMEs credit risk.

The performance of PCA and PLS are compared with widely used logistic regression using several measures of predictive accuracy. Across the three models, logistic regression has the best prediction power, but the difference between logistic regression and PLS is marginal. PLS has advantage in achieving high predictive ability while coping with multicollinearity among predictors.

This study provided an initial exploration of the latent concept of Management Capability and its dimensions in the context of credit management. Further research can consider additional sources of information, for example, survey data, and investigate a variety of behavioural and environmental aspects following Keats and Bracker (1988) and Maes *et al* (2005).

Acknowledgements—The authors are grateful to the data provider that chose to remain anonymous and to the provider's contact persons for their outstanding help and support throughout the project.

References

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

Anderson R (2007). *The Credit Scoring Toolkit*. Oxford University Press: Oxford, UK.

Becker G (1964). *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. Columbia University Press: New York.

Berger AC and Frame WS (2007). Small business credit scoring and credit availability. *Journal of Small Business Management* **45**(1): 5–22.

Carter R and Van Auken H (2006). Small firm bankruptcy. *Journal of Small Business Management* **44**(4): 493–512.

Chen YC and Wu JH (2011). IT management capability and its impact on the performance of a CIO. *Information and Management* **48**(4–5): 145–156.

Comrey AL and Lee HB (1992). *A First Course in Factor Analysis*. Lawrence Erlbaum: Hillsdale, NJ.

Cohen S and Kaimenakis N (2007). Intellectual capital and corporate performance in knowledge-intensive SMEs. *The Learning Organization* **14**(3): 241–262.

de la Torre A, Martínez Pería MS and Schmukler SL (2010). Bank involvement with SMEs: Beyond relationship lending. *Journal of Banking & Finance* **34**(9): 2280–2293.

Hand DJ (2009a). Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning* **77**(1): 103–123.

Hand DJ (2009b). R code for H-measure, http://www2.imperial.ac.uk/~djhand/R%20CODE%20for%20H%20measure%20of%20classification%20performance.txt, accessed 16 December 2011.

Hand DJ and Henley WE (1997). Statistical classification methods in consumer credit scoring: A review. *Journal of the Royal Statistical Society Series A* **160**(3): 523–541.

Gaskill LR, Van Auken HE and Manning RA (1993). A factor analytic 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**(C): 1–17.

Grunert J and Norden L (2012). Bargaining power and information in SME lending. *Small Business Economics* **39**(2): 401–417.

Grunert J, Norden L and Weber M (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance* **29**(2): 509–531.

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

Ihua UB (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, UK.

Keats BW and Bracker JS (1988). Toward a theory of small firm performance: A conceptual model. *American Journal of Small Business* **12**(4): 41–58.

Kleinbaum DG, Kupper LL, Nizam A and Muller KE (1998). *Applied Regression Analysis and Multivariate Methods*. 3rd edn, PWS-Kent Publishing Company: Boston, MA.

Larson CM and Clute RC (1979). The failure syndrome. *American Journal of Small Business* **4**(2): 35–43.

Lin SM, Ansell J and Andreeva G (2012). Predicting default of a small business using different definitions of financial distress. *Journal of the Operational Research Society* **63**(4): 539–548.

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

Lussier RN and Halabi CE (2010). A three-country comparison of the business success versus failure prediction model. *Journal of Small Business Management* **48**(3): 360–377.

Lussier RN and Pfeifer S (2001). A cross-national prediction model for business success. *Journal of Small Business Management* **39**(3): 228–239.

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

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 BH and Wehrens R (2007). The PLS package: Principal components and partial least squares regression in R. *Journal of Statistical Software* **18**(2): 1–24.

Perry SC (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.

Rencher AC (2002). *Methods of Multivariate Analysis*. John Wiley & Sons: USA.

Stevens J (1992). *Applied Multivariate Statistics for the Social Sciences*. Erlbaum: New York.

Tebboth J and Gadi M (2009). Assessing the use of indeterminates for scorecard model development. Presented at *Credit Scoring & Credit Control XI Conference*, University of Edinburgh: Edinburgh, UK.

Thomas LC, Eldelman DB and Crook JN (2002). *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics: Philadelphia, PA.

Thomson M and Heron P (2005). Management capability and high performance work organization. *The International Journal of Human Resource Management* **16**(6): 1029–1048.

Tobias RD (2003). An introduction to partial least squares regression, http://www.ats.ucla.edu/stat/sas/library/pls.pdf, accessed 26 October 2011.

van Gelder JL, de Vries RE, Frese M and Goutbeek JP (2007). Differences in psychological strategies of failed and operational business owners in the Fiji Islands. *Journal of Small Business Management* **45**(3): 388–400.

Wichmann H (1983). Accounting and marketing—Key small business problems. *American Journal of Small Business* **7**(4): 19–26.

Wold H (1966). Estimation of principal components and related models by iterative least squares. In: Krishnaiaah PR (ed). *Multivariate Analysis*. Academic Press: New York, pp 391–420.

Received 18 January 2012;
accepted 4 September 2014 after three revisions