Review of Economics & Finance
Submitted on 18/03/2014
Article ID: 1923-7529-2014-03-16-17 Francesco Campanella

Assess the Rating of SMEs by using Classification And Regression Trees (CART) with Qualitative Variables

Francesco Campanella
Assistant Professor of Corporate Finance, Department of Economics, Second University of Naples
Corso Gran Priorato di Malta, 1, Capua, Caserta, 81043, ITALY
Tel: +39-082-327-4353, E-mail: francesco.campanella@unina2.it
Homepage: http://www.economia.unina2.it/suneco/site/it/docenti/details?username=francamp

Abstract: Italian banks have never used the credit rating system to grant funds to SMEs until the introduction of Basel II accord. Credit Rating Systems use financial ratios that are often not adapted to SMEs' assessment. In fact, small and medium-size enterprises are characterized by a high level of intangible assets. Some researchers focus their attention on the evaluation of qualitative variables of SMEs (management; corporate governance; SMEs-territory relationship), but no research is able to integrate these SMEs' qualitative variables into a single scoring model, or to sufficiently consider the characteristics of SMEs-financial markets relationship. This paper proposes a specific credit scoring model to SMEs' assessment which includes all these variables combining two methods: Altman's 'EM-Score' and CART (*Classification and Regression Tree*). This model is performed on a sample of 6,534 Italian manufacturing firms getting a high level of reliability.

JEL Classifications: G01, G21, G24

Keywords: Credit scoring model, Hierarchies of qualitative variables, Classification and Regression Trees, Rating, Relationship between SMEs and financial markets

1. Introduction

This research focuses on the problem of evaluating the actual credit standings of Italian small and medium-size enterprises (SMEs) characterised by intangible assets whose role is not correctly appraised by conventional credit scoring systems.

Over these past few years, this problem has been eagerly debated especially in connection with the recent economic crisis. Indeed, besides exacerbating economic and financial imbalances, the present recessionary trend has seriously undermined the credibility of some rating agencies which caused serious damage to investors for failing to provide correct estimates of the actual risk profiles of financial and industrial corporations.

To counter this adverse trend, the banking system has resolved to adopt more stringent credit risk monitoring procedures and credit initiation processes.

This is why credit scoring systems are ever more often used to enable the banking system to assess the actual creditworthiness levels of corporate credit applicants.

Regrettably, most of the scoring models in use for credit risk assessment purposes fail to generate reliable estimates of the actual credit standings of the applicants since the banking system has not been able to integrate them with a number of qualitative variables assumed to play a key role in determining the success or failure of a business firm. This is one of the reasons why numerous Italian firms are finding it harder and harder to gain access to credit.

Considering the empirical nature of this research, the author has divided the report into five sections after the introduction. The first section offers a review of the literature discussing the obstacles standing in the way of the integration of qualitative variables into credit scoring systems; the second defines the purpose and the hypotheses of the research; the third describes the research sample and the analytical procedure; the fourth analyses the research findings and reports the discussion and the fifth the conclusions.

2. Literature Review

Italian banks have never used the credit rating system to grant funds to SMEs until the introduction of Basel II accord. In Italy, Basel II introduced for the first time the use of credit scoring models in the banking system. This was a cultural turning point both for banks and for Italian firms, especially for SMEs. In fact, in Italy the credit was granted primarily on the basis of collateral and guarantees offered by SMEs. In addition, the relationship between banks and firms was characterized by a high level of information asymmetry.

The Credit Rating System is regulated in a generic way by the Basel document. Essentially, each bank can use its own rating system different from that used by other banks (Basel Committee on Banking Supervision, 2005, Gupton et al., 2000, Lemon et al., 2003). In this regard, many banks have referred to the credit scoring models proposed in scientific literature for the implementation of the Credit Rating System.

Despite the fundamental structure underlying the scoring models have already been developed in the thirties by authors such as Fisher (1938) and Durand (1941), the decisive impulse to the development and dissemination of these models has occurred by Altman (1968). Altman, using a set of economic and financial indicators, created a predictive model of companies default known as "Z-score" based on multivariate discriminant analysis (Durand, 1941; Fisher, 1938).

However, in attempting to obtain a more reliable estimate by scoring models, many researchers in more recent studies have used different statistical methods (genetic algorithms; classification trees; Multivariate Adaptive Regression Splines; Case-Based Forecasting; applicative models of Rough Set Theory; neural networks).

In general, these models have shown a higher level of precision than the best-known Multiple Discriminant Analysis and logistic regression (Crouhy et al., 2000), although according to some authors, the logistic regression has higher forecast accuracy in the long run (Haswell et al., 1989).

In spite of the increased accuracy in the prediction of default, all these models use financial ratios in accounting that are often not adapted to SMEs' evaluation. In fact, the disclosure of financial statements produced by SMEs' evaluation is less structured, detailed and reliable than that of larger firms (Ciampi, 1994; Ciampi and Gordini, 2013).

In addition, SMEs have a low level of tangible assets to ensure the repayment of debt to the banks and a high level of intangible assets (trademarks, patents and skills). This has generated the need to create scoring models specific to SMEs' evaluation which include qualitative variables.

Numerous scholars have tried to insert qualitative variables within the scoring models for SMEs. In particular, some researchers have focused their attention on qualitative variables relating to the characteristics of SMEs' management (intellectual capital, management skills, family relationship, risk appetite) (Cooper et al., 1991; Gabbi et al., 2006; Kwan, 1996; Lopez and Saidenberg, 2000). Instead, other scholars have focused their attention on SMEs' corporate governance characteristics (size and composition of the Board of Director) (Chaganti et al., 1985; Daily and Dalton, 1994).

In the end, some scholars have improved the previous empirical researchers introducing qualitative variables concerning SMEs' territory characteristics and SMEs-territory relationship (socio-economic characteristics of SMEs' territory).

In this regard, Ciampi and Gordini (2013) show that the variables related to the area of establishment and their SMEs-territory relationship improve the accuracy of predictive models. Furthermore, the predictive potential of these variables in the scoring models is particularly relevant for SMEs. However, all the empirical predictive models ignore the scoring of the qualitative variables relating to the characteristics of the relationship between financial markets and SMEs.

On that latter point, no research is able to integrate a single scoring model with all SMEs' qualitative variables related to:

1. characteristics of management;
2. characteristics of corporate governance;
3. characteristics of SMEs-territory relationship;
4. characteristics of SMEs- financial markets relationship.

Regarding the latter point, it is known that SMEs have relationships primarily with banks, so it is necessary to identify the characteristics of the credit relationship that actually can affect the rating calculated using economic and financial variables.

One major side effect of this issue is a steadily widening knowledge gap between SMEs and financial markets, i.e. a situation where asymmetrical information problems tend to generate adverse credit selection processes (Allee,2000; Lussier, 1995; Olshen et al., 1984).

Consequently, there are reasons for arguing that the increasing dematerialisation of corporate strategic assets may prevent the banking system from obtaining correct information regarding credit applicants and from making reliable estimates of the growth potentials of both large-size companies and SMEs.

With regard to the aforementioned assumptions, there are two major reasons why the banking system should make a concerted effort towards appraising the real prospects of SMEs to gain and sustain a competitive advantage. Firstly, despite comparatively small stocks of tangible capital assets, many of the existing SMEs may head towards a major performance thanks to firm-specific intangibles. Secondly, most of these firms have difficulty providing banks with adequate information on the qualitative variables that are the primary determinants of their success (Kavoussi, 1984).

With reference to the first of these points, numerous research studies on the assumed performance-boosting potential of qualitative variables have emphasised that family control may with equal probability be a key for the strength of a firm or the true cause of its weakness.

On the one hand, control by a single owner-founder may ensure prompt response to changing market requirements; on the other, it may become an obstacle to the access of the firm to the financial market or cause frictions upon the succession of a new owner to firm control (Rosa et al., 2003). Based on the higher success rate of family businesses in north-eastern Italy than in the South, some authors have argued that this kind of firm is heavily affected by exogenous variables of a geographical nature.

Other studies have found that the less formal governance patterns of these firms make both for brisker exchanges of information in the production chain and for a greater innovation focus than is usually observed in large-size businesses. With respect to this point, however, they have also emphasised that due to R & D funding difficulties this trend is more marked in process than product innovation (Cumming et al., 2010).

Concerning funding difficulties, some researchers have highlighted the fact that successful SMEs are often organised into districts or networks through which they obtain access to the financial resources required for their product development projects (Resti and Sironi, 2008). In their opinion, this is the reason why many of them rank higher in Pavitt's innovation-focused taxonomy[1] and have successfully developed the qualifications of their in-house intellectual capital. Indeed, it is clear that educated and well-training human resources are the prerequisite for a firm to take over and interiorise the tacit knowledge developed by other district or network firms.

Lastly, many SMEs have successfully combined flexibility with economies of scale and are using this combination as a springboard for a major performance. Specific factors accounting for the greater growth potential of district or network business firms include a marked ability to export products to world markets and a lesser need to delocalise production for cost-cutting purposes (Archibugi and Michie, 1997; Jorion, 2003).

With regard to the aforementioned assumptions, on the hypothesis that a tightly structured legal organisation form such as a group of firms is better able to meet the challenge of discouraging and preventing opportunity actions, some academics maintain that groups of SMEs are more efficient than firms organised into districts or networks.

Starting out from the observation that district, network and group businesses rank higher than stand-alone firms in business survival indexes, some researchers have argued that firms carrying on business in collaboration with others: 1) have a better survival potential than stand-alone firms, and 2) tend to outperform stand-alone businesses on financial markets.

Besides facing the challenge of enhancing credit scoring systems through the inclusion of performance-boosting qualitative variables, the banking system comes up against the problem of effectively handling relationships with small-size firms characterised by considerable information opacity. This is why several banking supervisory authorities have stressed the need to shape bank-firm contacts in line with the relationship lending model (Razi and Athappilly, 2005).

As this model envisages frequent intensive contacts between the bank and its corporate customers, its main strong points are long-term personal acquaintance and the resulting climate of mutual trust. The medium and long-term lender-borrower relationships typifying this model enable banks to obtain detailed information on the way a firm is being managed and, hence, to conduct their credit scoring operations in conditions of reduced asymmetries of information (Schwartz et al., 2014). Occasionally, an entrepreneur who has managed to establish a relationship of this kind is likely to abstain from multiple-bank borrowing and to obtain credit on more favourable terms (Bumacov and Ashta, 2011).

Other academics have highlighted a positive role of innovative finance when structured products and/or derivatives are used for good purposes such as financial hedging, rather than for speculative investment. In such situations, they remark, these transactions may help reduce financial risks and pave the way for easier access to credit.

Others still have pointed to a 'business angel' and venture capital as major signs of a firm's viability and creditworthiness (Ho et al., 1992). Specifically, some research studies have found that

[1] Pavitt's taxonomy identifies four categories of firms characterised by increasing levels of innovation focus: 1) supplier dominated, 2) scale intensive, 3) specialised suppliers and 4) science based business enterprises.

firms using risk capital investors find it easier to obtain resources and create the assumptions for their future growth.

Stock-Exchange listing appears to produce the same effect, since banks are in a position to rely on the control that is exercised on listed companies by the financial market.

Economic-financial data disclosure in SME-bank relationships is an additional point on which researchers have focused attention (Campanella et al., 2013). Their findings reveal that customer-bank relations are greatly enhanced by the routine submission of qualitative and quantitative reports on the results of a firm's operations and consequently confirm the crucial role of information in lender-borrower relationships.

3. Object and Hypotheses

The aim of this research is to improve the existing literature in two ways: 1) identifying the qualitative variables of the relationship SMEs-financial markets that affect the rating which is estimated using financial data; 2) identifying a "hierarchies of qualitative variables" between the qualitative aspects of management, governance, SMEs-territory relationship, the intrinsic characteristics of SMEs and the characteristics of SMEs-financial markets relationship.

Based on these objectives, the author formulated the following hypotheses:

- **H1.** The qualitative variables concerning the relationship between financial markets and SMEs affect the rating calculated by economic and financial ratios;
- **H2.** The qualitative variables concerning the relationship between financial markets and SMEs play an active role in the hierarchies of qualitative variables that influence the rating measured by economic and financial ratios.

Regarding the hypothesis H2, the need to establish hierarchies of qualitative variables required the use of the CART approach (Classification And Regression Trees), which allows to classify the variables more clearly than the most well-known logistic regression (Leea, 2006; Shmueli and Mani, 2013). Overall, the aim of this paper is to highlight what are the main qualitative variables to consider in order to better estimate the borrowers.

4. Materials and Methods

The sample adopted for this empirical research includes 6,534 Italian SMEs operating in 17 industries (Table 1) during 2011-2012. It can be described as homogeneous because firms from each sector range between a minimum of 4.6% and a maximum of 9.9% of the total.

The 'rating' variable was determined by processing the 2011 financial statements of the individual sample firms through the EM-Score model.

The EM-Score model is an improved version of Altman's Z-Score discriminant analysis model (Brida et al., 2010), from which it differs because score intervals are not described as '*safety zone*', '*distress zone'* or '*grey zone*', but directly associated with a *Standard & Poor's* rating value (Table 2). The score is generated by the following equation (1):

$$\text{EM-Score} = 3.25 + 6.56(X_1) + 3.26(X_2) + 6.72(X_3) + 1.05(X_4) \qquad (1)$$

where X_1 = working capital / total assets;
X_2 = retained earnings / total assets;
X_3 = EBIT / total assets;
X_4 = book value of equity / book value of total long-term liabilities.

The values before the variables, named 'predictors', are the results of the OLS regression analysis conducted by Altman on a sample of companies before constructing his model. The table of rating equivalents has been taken over from Altman, Hartzell and Peck (Table 2) (Andone and Sireteanu, 2009; Altman et al., 1998).

Table 1. Structure of the sample

	Manufacturing industries	***N***	**%**
1	Food products processing	650	9.95%
2	Beverages	496	7.59%
3	Textiles	365	5.59%
4	Apparel and leather and textile products	346	5.30%
5	Basketwork and wickerwork	380	5.82%
6	Wood and cork product (except furniture); straw materials	402	6.15%
7	Paper and paper product	339	5.19%
8	Recorded media printing and duplication	307	4.70%
9	Chemical products	341	5.22%
10	Medicinal chemicals and pharmaceutical preparation	324	4.96%
11	Rubber and plastic articles	479	7.33%
12	Non metal ore products	325	4.97%
13	Metal Working	302	4.62%
14	Metal products (Excluding machinery and equipment)	353	5.40%
15	Computers, electronic equipment and optical instruments	302	4.62%
16	Electrical equipment and non-electrical household appliances	356	5.45%
17	Other manufacturing industries	467	7.15%
	TOTAL	**6,534**	**100%**

Accordingly, the Altman, Hartzell and Peck model adopted for this study offers the advantage of a more detailed analysis of risk (since firms are not simply assigned to one of the above-mentioned three 'zones') and a combination with one of the best-known risk classification scales developed by Standard & Poor's (Caouette et al., 1998).

For a better understanding or our research findings, let us mention that *S&P* ratings fall into two main categories:

a) *investment grade* (from AAA to BBB-)[2]. As companies with a rating score in this interval have traditionally been able to meet their obligations, even risk-adverse investors may safely consider investing in them;

b) *below investment grade* businesses (from BB+ to D)[3]. As companies with scores in this interval are at high risk of default, risk-adverse investors are advised to abstain from investing in them.

This wide subdivision helps distinguish between companies in which investment is recommended or not recommended.

[2] This category includes the investment classes ranging from *Highest Grade* (companies with a very low probability of default value) to *Medium Grade* (companies with a medium-high probability of default value).

[3] The investment classes in this category range from *Speculative Grade* (companies with a medium-high PD value) down to *Default Grade* (companies with a proven inability to honour their obligations).

Table 2. S&P Rating Equivalents

	Rating	Scores
1	AAA	>8.15
2	AA+	7.60
3	AA	7.30
4	AA-	7.00
5	A+	6.85
6	A	6.65
7	A-	6.40
8	BBB+	6.25
9	BBB	5.85
10	BBB-	5.65
11	BB+	5.25
12	BB	4.95
13	BB-	4.75
14	B+	4.50
15	B	4.15
16	B-	3.75
17	CCC+	3.20
18	CCC	2.50
19	CCC-	1.75
20	D	0

For the purposes of our research, the inclusion of the sample companies in the 'investment grade' or 'below investment grade' group highlights the level of risk assumed to be associated with each of them. Altman's EM-Score model classifies firms by risk class exclusively by reference to quantitative accounting variables.

To establish if and to what extent qualitative variables may shed further light on the ratings obtained, the author has used the classification analysis method.

The need to define a "hierarchies of qualitative variables" has imposed the use of the Classification trees that classify more clearly the independent variables with respect to the best-known logistic regression.

The classification rule for a research sample has to be laid down before the classification tree is constructed (Brida et al., 2009). The classification rule for this research is the distinction between "*investment grade*" and "*below investment grade*" businesses made via Altman's EM-Score model. This classification rule is reflected in the Y variable of the classification tree.

Source of Table 2: *Altman E. (2000). Predicting financial distress of companies: revisiting the Z-score and ZETA models, Working paper Stern School of Business, New York University, 9-12.*

At this point, a recursive partition technique was applied to assign each statistical unit to one of the a priori classes defined by Y. The sample units were repeatedly split into groups whose composition will be ever more homogeneous with the dependent variable Y. The splitting procedure was conducted by reference to the explanatory variables $X=(X_1,X_2,\ldots, X_s, \ldots, X_p)$.

In line with the research aim, the author had to define those qualitative variables (X) which in previous research studies had been shown to have a bearing on the ratings assigned to firms.

Accordingly, he defined 17 qualitative X variables divided in different areas as suggested by the literature review (see Annex 1).

The values of the variables were determined by circulating a closed-response questionnaire via the CATI (computer-assisted telephone interviewing) system. Thanks to this computerised system, the response rate was over 70%.

A closed-response questionnaire suggests two possible answers for each question. For this reason, the author exclusively used dichotomous questions for his research (0;1) (Table 3).

Based on the predetermined category reflected in the dependent variable Y (basis) and the explanatory variables $X=(X_1, X_2,\ldots,X_s,\ldots,X_p)$, the set of 6,535 firms constituting our sample was gradually split into smaller and smaller partitions characterised by increasing internal homogeneity with respect to the dependent variable Y.

To minimize the number of terminal nodes, the tree was constructed in line with the following specifications:

- minimum number of cases in parent node: 100;
- minimum number of cases in child node: 50.

Impurity was measured by reference to the Gini index, i.e. by fixing the minimum change rate at 0.0001. Lastly, to ensure the construction of a possibly reliable tree, the partition was validated by means of a training sample including 60% of the firms in the aggregate sample. The remaining 40% was used to validate the model (test sample).

The overall efficiency of the proposed model was evaluated by the Receiver Operating Characteristic Curve (ROC Curve).

Table 3. Model summary

Specifications	Growing Method:	Classification and Regression Trees
	Dependent Variable:	Rating
	Independent Variables:	Area, District, Organization, Pavitt, Group, Network, Family business, Intellectual capital, Years, Export, Outsourcing, Commitment, Relationship lending, Innovative finance, Venture capital, Quotation, Information opacity
	Validation:	Split Sample
	Maximum Tree Depth:	5
	Minimum Cases in Parent Node:	100
	Minimum Cases in Child Node:	50
Results	Independent Variables Included:	Commitment, Innovative finance, Group, District, Export, Relationship lending, Venture capital, Pavitt, Information opacity
	Number of Nodes:	11
	Number of Terminal Nodes:	6
	Depth:	3

5. Results and Discussion

Before the discussion of the research findings, it is convenient to provide an overview of the CART model used for the research sample.

Table 3 reports both the model specifications and the analysis results. The ‘Specifications’ section offers information on the tree model construction criteria, including the analysis variables, whereas the values reported in the ‘Results‘ section reflect aggregate node number, number of terminal nodes and tree depth.

Initially, the author identified seventeen independent variables. Only nine of these were included in the model since the potential contribution of the remaining eight was assumed to be fairly negligible. Figure 1 shows the tree diagram of the training sample at the end of the pruning process. Figure 2 shows the test sample. The optimal sub-tree was found to include five variables.

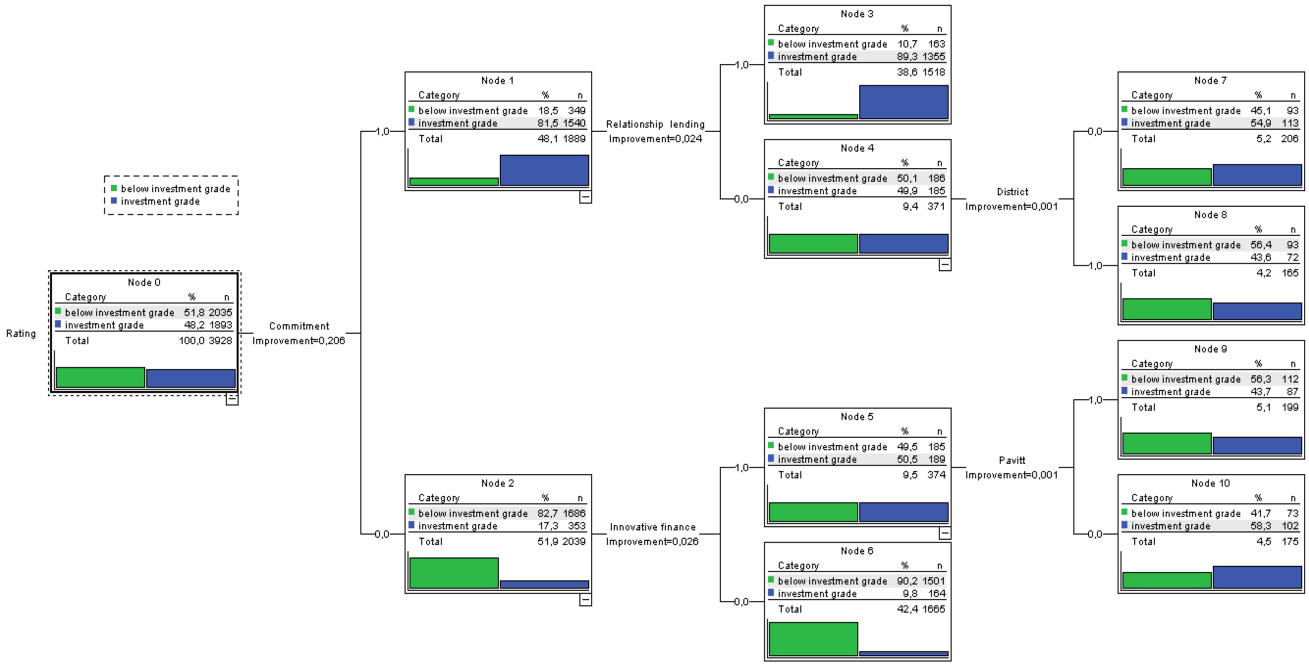


Figure 1. Training sample

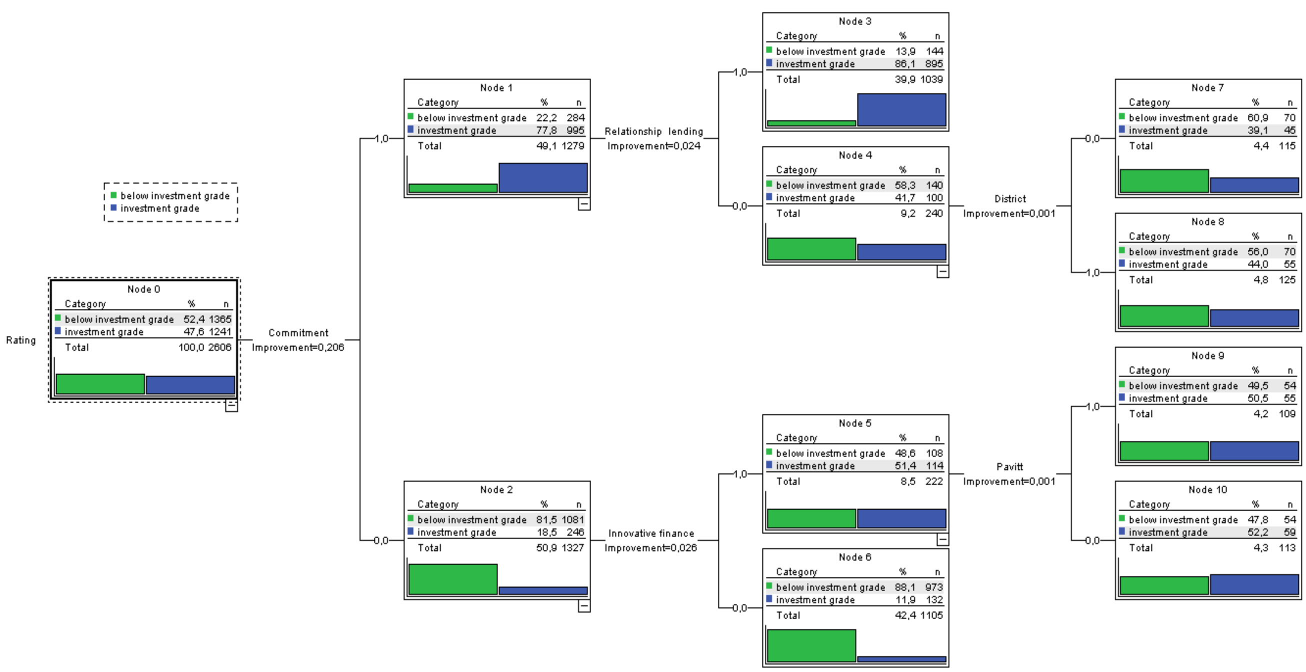


Figure 2. Test sample

Analysing the CART classification sub-tree of the training sample, it is worth noting that:

a) the ‘commitment’ variable splits the data into two nodes, node 1 and node 2. The data in node 1 show that 81.5% of the firms entertaining credit relations with only one bank fall within the ‘investment grade’ category; node 2 is evidence that 82.7% of the companies holding multiple-bank credits fall within the ‘below investment grade’ category;
b) the subsequent best classification variable for firms entertaining a single-bank credit relationship (node 1) is relationship lending (node 3). In node 3, 89.3% of the firms entertaining a medium-long credit relationship with the bank are categorised as 'investment grade'. This means that the procedure is an effective way to classify firms with high rating scores. Node 3 is a terminal node;
c) the subsequent best classification variable for firms entertaining just one lending relationship (node 2) is innovative finance (node 6). In node 6, 90.2% of the firms which do not use any structured credit products and/or derivatives for financial hedging purposes fall within the 'below investment grade' category”, which shows that the procedure is an effective way to classify firms in lower rating classes. Node 6 is a terminal node;
d) no efficient subset classifications were found in nodes 4, 5, 7, 8, 9 and 10.

Parallels between the findings for the training sample and those of the test sample (see figure 1 and figure 2 above) confirm the appropriateness of the model.

The Risk and Classification (Tables 4 and 5) show that the correct classification rate obtained for the training sample is 83.4% and that the classification error risk rate of the model (see Tab. 5) stands as low as 16.6%, with a 0.006 standard error value. As a result, the model can be classed as reliable.

Table 4. Risk

Sample	Estimate	Std. Error
Training	.166	.006
Test	.196	.008

Table 5. Classification (Growing Method: CRT; Dependent Variable: Rating)

Sample	Observed	Predicted		
		Below investment grade	Investment grade	Percent Correct
Training	Below investment grade	1,706	329	83.8%
	Investment grade	323	1570	82.9%
	Overall Percentage	51.7%	48.3%	83.4%
Test	Below investment grade	1,097	268	80.4%
	Investment grade	242	999	80.5%
	Overall Percentage	51.4%	48.6%	80.4%

Table 6 summarizes the gain for nodes showing the number of nodes, the number of cases, the average profit and ROI (Return On Investment). The best performance is the node 3, the worst one is the node 6.

Table 7 is referred to the target variable “below investment grade” and it includes the gain in percent, the response rate and the percentage index (lift) per node.

The figures 3 and 4 illustrate the node performance respectively compared to the gain and the index. In the figure 3, the gain chart quickly increases towards the 100% than fall on the diagonal. This graph indicates that the model is quite reliable. In fact, a model that does not provide information follows the baseline of the diagonal.

Table 6. Gain Summary for Nodes
(Growing Method: CRT; Dependent Variable: Rating)

Sample	Node	N	Percent	Profit	ROI
Training	3	1,518	38.6%	4.248	4.9%
	10	175	4.5%	2.080	3.4%
	7	206	5.2%	1.840	3.2%
	9	199	5.1%	1.060	2.2%
	8	165	4.2%	1.055	2.2%
	6	1,665	42.4%	-1.311	-6.5%
Test	3	1,039	39.9%	4.030	4.8%
	10	113	4.3%	1.655	3.0%
	7	115	4,4%	0.739	1.7%
	9	109	4.2%	1.532	2.8%
	8	125	4.8%	1.080	2.2%
	6	1,105	42.4%	-1.164	-5.3%

In the figure 4 the cumulative indexes plots tend to start above 100% and gradually decrease until they reach 100%. This graph shows that the model is reliable. Indeed, in a reliable model, the value of the index starts well above 100%, it remains stable as you move and then rapidly drops to 100%. For a model that does not provide information, the line will overlap 100% in the whole chart.

Table 7. Target category: below investment grade – Gain for Nodes

Sample	Node	Node		Gain		Response	Index
		N	Percent	N	Percent		
Training	6	1,665	42.4%	1,501	73.8%	90.2%	174.0%
	8	165	4.2%	93	4.6%	56.4%	108.8%
	9	199	5.1%	112	5.5%	56.3%	108.6%
	7	206	5.2%	93	4.6%	45.1%	87.1%
	10	175	4.5%	73	3.6%	41.7%	80.5%
	3	1,518	38.6%	163	8.0%	10.7%	20.7%
Test	6	1,105	42.4%	973	71.3%	88.1%	168.1%
	8	125	4.8%	70	5.1%	56.0%	106.9%
	9	109	4.2%	54	4.0%	49.5%	94.6%
	7	115	4.4%	70	5.1%	60.9%	116.2%
	10	113	4.3%	54	4.0%	47.8%	91.2%
	3	1,039	39.9%	144	10.5%	13.9%	26.5%

Note: Growing Method: CRT; Dependent Variable: Rating

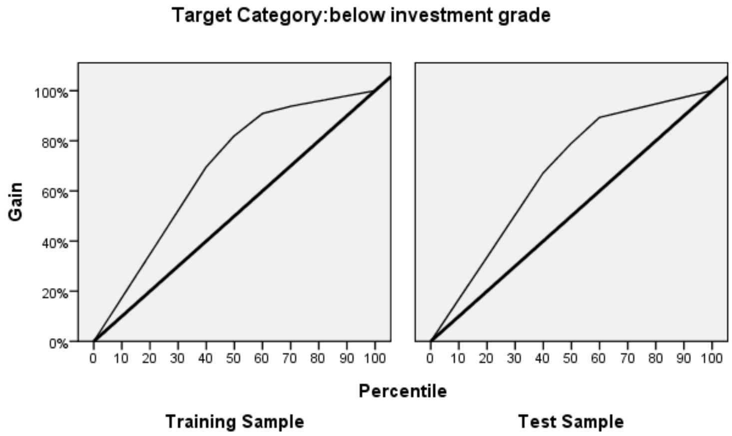


Figure 3. Target category: below investment grade – Node performance: gain

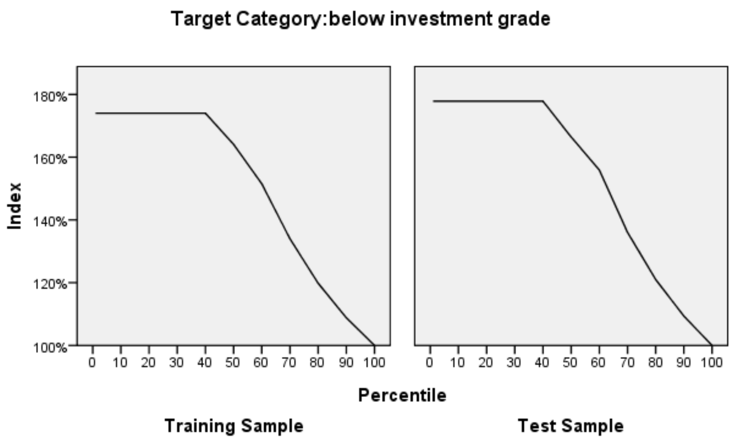


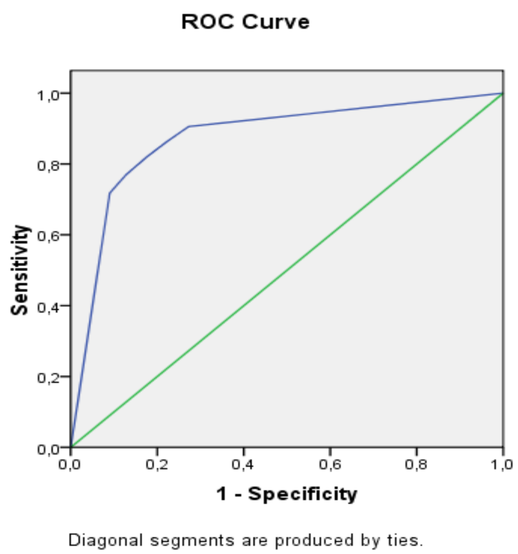
Figure 4. Target category: below investment grade – Node performance: index

The overall efficiency of the optimal tree was evaluated by the Receiver Operating Characteristic Curve using predicted probability of the model (Table 8, Table 9 and Figure 5).

Table 8. Case Processing Summary

Rating	Valid *N* (listwise)
Positive [a.]	3,134
Negative	3,400

Notes: Larger values of the test result variable(s) indicate stronger evidence for a positive actual state; (a.) The positive actual state is 1.



The Area Under the Curve (AUC) in figure 5 is equal to 0.875. The best cut off is at 0.4929 level (Youden's index = 0.644) (Table 9 and Table 10).

Figure 5. Receiver Operating Characteristic Curve

Table 9. Area Under the Curve (AUC)

Area	Std. Error [a.]	Asymptotic Sig. [b.]	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.875	.005	.000	.866	.884

Notes: Predicted Probability for Rating=1 has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased; (a.) Under the nonparametric assumption; (b.) Null hypothesis: true area = 0.5

Table 10. Coordinates of the curve

Positive if Greater Than or Equal To [a.]	Sensitivity	1 – Specificity
.0000	1.000	1.000
.2674	.906	.272
.4368	.865	.224
.4929	.820	.176
.5657	.769	.128
.7377	.718	.090
1.0000	.000	.000

Notes: Predicted Probability for Rating=1 has at least one tie between the positive actual state group and the negative actual state group; (a.) The smallest cut-off value is the minimum observed test value minus 1, and the largest cut-off value is the maximum observed test value plus 1. All the other cut-off values are the averages of two consecutive ordered observed test values.

The analytical procedure has successfully identified qualitative variables concerning the SMEs-financial markets relationship which afford reliable estimates of the creditworthiness of Italian SMEs (Hypothesis H1 is confirmed).

The empirical analysis has highlighted that the best performing firms are typified by two qualitative attributes: 1) they entertain credit relationships with only one bank; 2) the credit relationship with this bank is a long-term one (in line with the relationship lending model. Instead, the worst performing firms are typified by the use of innovative finance.

In the “hierarchies of qualitative variables”, the variables regarding the SMEs-financial markets relationship are more important than the other variables, such as the strengths of SMEs-territory relationship (District) or the intrinsic qualities of SMEs (Pavitt) (Hypothesis H2 is confirmed). These findings are perfectly in keeping with recent recommendations by banking supervisory bodies.

Single-bank credit holding leads to relationship lending, a process which generates a climate of mutual trust between a bank and its corporate credit holders. Hence, this empirical analysis confirms the key role of mutual trust – which today seems to have become the exception – in the relations between banks and their customers.

Lastly, the analysis provides evidence that the least creditworthy firms typically entertain multiple-bank credit relationships and use derivatives.

This finding is consistent with the assumption that credit institutions mistrust multiple-bank borrowing as a clear sign that a firm unable to honour its debts has been obliged to contract fresh long-term loans with several banks and is at risk of default if this situation drags on for a long time.

As far as derivatives are concerned, they may seriously undermine a firm’s credit standing even if they are used for hedging, instead of speculation purposes. Today, this risk is further escalating as a result of the lasting financial crisis.

6. Conclusion

At this stage, within the overall debate on variables with a potential bearing on bank-firm relationships this study may take credit for demonstrating that some endogenous attributes of firms would enhance such relationships especially by creating a climate of mutual trust.

The mutual trust between banks and enterprises has previously had a key role in business development in some countries characterized by a large number of SMEs, like Italy. In fact, the Italian SMEs are characterized by a high level of intangible assets (trademarks, patents, entrepreneur’s skills). These intangible assets are evaluated with difficulties by banks.

In the past, the geographical proximity between banks and SMEs created a climate of mutual trust and it allowed the financing of companies that had few tangible assets to pledge as collateral of the debt. In the 2000s, mergers and acquisition in the banking sector have distanced from SMEs’ territories the decision-making centers of banks. This phenomenon has broken the climate of mutual trust. Moreover, the introduction of quantitative rating systems provided by Basel II on the one hand has decreased the information opacity between banks and enterprises, on the other hand it has substantially reduced the importance of SMEs-financial market relationship. These phenomena have caused the credit crunch in Italy.

Thus it is necessary more responsibilities and diligence of supervision and regulation bodies in the credit system (i.e. Basel Committee, Financial Stability Board) in order to give more prominence to entrepreneur’s qualitative variables in the evaluation of creditworthiness.

Obviously, a correct evaluation of credit standings of SMEs cannot rely solely on qualitative variables, but it is necessary they are integrated in quantitative credit scoring models.

Annex – The variables

A. Quality of management

1. Intellectual capital of management = an explanatory variable reflecting the average educational qualifications of the SMEs' management. This variable is set at 1 if graduates account for over 20% of the total management, at 0 if the relevant proportion falls short of 20%. The 20% level has been set because the number of graduates in Italy is among the lowest in Europe, and most of them work in the management of large enterprises;

B. Governance

2. Family business = an explanatory variable standing for the firm's governance system. The dichotomous variable is set at 1 if the firm is a family business, at 0 in the opposite case;

C. Quality of territory and strengths of SMEs-territory relationship

3. Area = a variable indicating the location where the firm is headquartered. This dichotomous variable is set at 1 if the registered office of the firm is in northern Italy, at 0 if it is in central or southern Italy;
4. District = an explanatory variable standing for membership in a district organisation. This dichotomous variable is set at 1 for district businesses, at 0 for stand-alone businesses;

D. Intrinsic characteristics of SMEs

5. Organisation = an explanatory variable standing for the organisational structure of the firm. This dichotomous variable is set at 1 if the firm adopts a formal organisation structure with clearly defined functional levels and at 0 if it has an informal organisation structure;
6. Network = an explanatory variable standing for inclusion in a network organisation. The 1 setting reflects a network business; the 0 setting designates a stand-alone business;
7. Group = an explanatory variable standing for the firm's membership in a group. This variable is set at 1 if the firm is a group business, at 0 if it is a stand-alone business;
8. Export = an explanatory variable reflecting the firm's or group's level of export focus. This variable is set at 1 if exports exceed 10% as a share of aggregate sales, at 0 if the relevant proportion is less than 10%;
9. Pavitt = the Pavitt taxonomy which classifies business firms by reference to the following homogeneous criteria: sources of technology, innovation modes, degree of appropriability of innovative technologies, existence of barriers to entry, firm size. Based on these factors, Pavitt identified four categories of firms typified by increasing levels of innovation focus: 1) supplier dominated, 2) scale intensive, 3) specialised suppliers and 4) science based business enterprises. The dichotomous variable associated with Pavitt's taxonomy is set at 1 if the firm is supplier dominated and at 0 if it is categorised as scale intensive, as a specialised supplier or as a science based business;
10. Outsourcing = an explanatory variable standing for product delocalisation. This variable is set at 1 if the firm's production processes have been delocalised to foreign countries, at 0 in the opposite case;
11. Years = a variable standing for the firm's life cycle. This dichotomous variable is set at 1 if the firm was established over 10 years ago, at 0 if its foundation dates back to an earlier point;

E. Characteristics of SMEs-financial markets relationship

12. Commitment = a variable explaining if the entrepreneur has or has not opted for multiple bank borrowing. This variable is set at 1 if the firm entertains credit relations with only one bank; at 0 if it is a multiple-bank credit holder;

13. Relationship lending = a variable highlighting the length, in years, of the relationship entertained with the main bank. This variable is set at 1 if the lending relationship has been lasting for over 10 years, at 0 in the opposite case;
14. Innovative finance = a variable explaining if the firm concerned uses structured products and/or derivatives for financial hedging purposes. This variable is set at 1 if it does, at 0 if it does not;
15. Venture capital = a variable indicating the existence of a venture capitalist. This dichotomous variable is set at 1 if a venture capitalist co-finances the firm, at 0 if the firm is not co-financed by any venture capitalists;
16. Quotation = a variable indicating if the company is listed on a stock exchange. This dichotomous variable is set at 1 if the firm is listed, at 0 if it is unlisted;
17. Information opacity = a variable indicating whether or not and how often the company submits information reports to banks. This dichotomous variable is set at 1 if the company submits to the main bank at least two reports a year, at 0 if it submits less than two reports per year.

Acknowledgments: Thanks to the anonymous referees for their valuable comments and suggestions.

References

[1] Allee, V. (2000). "Reconfiguring the value network", *Journal of Business Strategy*, 21(4):36-39.

[2] Altman, E., Hartzell, J., and Peck, M. (1998). *Emerging Market Capital Flows*, The New York University Salomon Centre Series on Financial Markets and Institutions, Springer US, New York, Chapter 5: 391-400.

[3] Andone, I., and Sireteanu, N.A. (2009). "A Combination of Two Classification Techniques for Businesses Bankruptcy Prediction", *Working paper*.

[4] Archibugi, D., and Michie J. (1997). *Technology, Globalisation and Economic Performance.* Cambridge University Press, Chapter 4: 83-95.

[5] Basel Committee on Banking Supervision (2005). Studies on the validation of internal rating system. Bank for International Settlements, BCBS Working Papers No 14.

[6] Brida, J.G., Gómez, D.M., and Risso, W.A. (2009). "Symbolic hierarchical analysis in currency markets: An application to contagion in currency crises", *Expert Systems with Applications*, 36(4):7721-7728.

[7] Brida, J.G., Pulina, M., Riaño, E.M.M., and Zapata-Aguirre, S. (2010). "A Classification and Regression Tree (CART) to Analyse Cruisers' Expenditure Pattern and Perception in a Port of Call", *Working paper.*

[8] Bumacov, V., and Ashta, A. (2011). "The conceptual framework of credit scoring from its origins to microfinance", Paper presented at the Second European Research Conference on Microfinance, Groningen, The Netherlands, June 2011.

[9] Campanella, F., Del Giudice, M., and Della Peruta, M.R. (2013). "The role of information in the credit relationship", *Journal of Innovation and Entrepreneurship*, 17(2):1-20.

[10] Caouette, J., Altman, E., and Narayana, P. (1998). *Managing Credit Risk: The Next Great Financial Challenge*, New York: John Wiley & Sons.

[11] Chaganti, R.S., Mahajan, V., and Sharma, S. (1985). "Corporate Board Size, Composition and Corporate Failures in Retailing Industry", *Journal of Management Studies*, 22(4): 400-416.

[12] Ciampi, F. (1994). Squilibri di assetto finanziario nelle P.M.I. Finanziamenti e contributi della Comunità Europea. *Banca Toscana*.

[13] Ciampi, F., and Gordini, N. (2013). Relazione impresa-territorio e modelli predittivi del default d'impresa. Primi risultati di una analisi statistica sulle piccole imprese italiane. *Sinergie, rivista di studi e ricerche.* 90:51-76.

[14] Cooper, A., Gascon, J., and Woo, C. (1991). "A Resource-Based Prediction of New Venture Survival and Growth", *Academy of Management Proceedings*, 1991(1):68-72.

[15] Crouhy, M., Galai, D., and Mark, R. (2000). "A comparative analysis of current credit risk models", *Journal of Banking and Finance*, 24(1):59-117.

[16] Cumming, D., Shmidt, D., and Walz, U. (2010). "Legality and venture capital governance around the world", *Journal of Business Venturing*, 25(1):54–72.

[17] Daily, C., and Dalton, R. (1994). "Bankruptcy and Corporate Governance: The Impact of Board Composition and Structure", *Academy of Management Journal*, 37(6):1603-1617.

[18] Durand D. (1941). *Risk Elements in Consumer Installment Financing*. NBER Books.

[19] Fisher R.A. (1938). "The statistical utilization of multiple measurements", *Annals of Eugenic*, 8(4):376-386.

[20] Gabbi, G., Matthias, M., and De Lerma, M. (2006). "CART analysis of qualitative variables to improve credit rating processes", *Society for Computational Economics*, 179: 1-15

[21] Gupton, G.M., Gates, D., and Carty, L.V. (2000). "Bank loan loss given default", Moody's Investors Service, NY, Global Credit Research.

[22] Haswell, S., Holmes, S. (1989). "Estimating the Small Business Failure Rate: A Reappraisal", *Journal of Small Business Management*, 27(3):68-74.

[23] Ho, K.L., Hsu, Y., and Yang, C. (1992). "Short term load forecasting using a multilayer neural network with an adaptive learning algorithm", *IEEE Transactions on Power Systems*, 7(1): 141-149.

[24] Jorion, P. (2003). *Financial Risk Manager Handbook*, Haboken: Wiley and Sons.

[25] Kavoussi, R.M. (1984). "Export expansion and economic growth: further empirical evidence", *Journal of Development Economics*, 14(1): 241-250.

[26] Kwan, S.H. (1996). "Firm-specific information and the correlation between individual stocks and bonds", *Journal of Financial Economics*, 40(1):63-80.

[27] Leea, T.S., Chiub, C.C., Chouc, Y.C., and Lud, C.J. (2006). "Mining the customer credit using classification and regression tree and multivariate adaptive regression splines", *Computational Statistics & Data Analysis*, 50(4):1113-1130.

[28] Lemon, S.C., Roy, J., Clark, M.A., Friedmann, P.D., and Rakowski, W. (2003). "Classification and regression tree analysis in public health: Methodological review and comparison with logistic regression", *Annals of Behavioral Medicine*, 26(3):172-181.

[29] Lopez, J.A., and Saidenberg, M.R. (2000). "Evaluating credit risk models", *Journal of Banking & Finance*, 24(1):151-165.

[30] Lussier, R. (1995). "A Non Financial Business Success versus Failure Prediction Model for Young Firms", *Journal of Small Business Management*, 33(1):8-20.

[31] Olshen, R., Breiman, L., Friedman, J.H., and Stone, C.J. (1984). *Classification and Regression Trees*. Wadsworth International Group.

[32] Razi, M.A., and Athappilly, K. (2005). "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models", *Expert Systems with Applications*, 29(1):65-74.

[33] Resti, A., and Sironi, A. (2008). *Rischio e valore nelle banche: misura, regolamentazione, gestione*. Egea, Milano.

[34] Rosa, R., Velayuthen, G., and Walter, T. (2003). "The share-market performance of Australian venture capital-backed and non-venture capital-backed IPOs", *Pacific-Basin Finance Journal*, 11(2):197-218.

[35] Schwartz, E.M., Bradlow, E., and Fader, P. (2014). "Model Selection Using Database Characteristics: Developing a Classification Tree for Longitudinal Incidence Data", *Marketing Science*, 33(2):188-205.

[36] Shmueli, G., and Mani, D. (2013). "Impact Assessment in Observational Studies: A Classification and Regression Tree Approach", *Working paper Indian School of Business (ISB).*