Chapter 4
Credit Scoring

Abstract The recent financial crisis has highlighted the importance of credit risk assessment for financial institutions, firms, and supervisors. Credit scoring systems are important tools for credit risk evaluation and monitoring. This chapter describes the process for building and testing credit scoring models and illustrates how multicriteria techniques based on disaggregation analysis can be used in this area. Empirical results are also presented, derived from an application to a large sample of Greek firms.

Keywords Credit risk · Preference disaggregation · Multicriteria classification

4.1 Credit Scoring Systems

Credit risk modeling plays a crucial role in financial risk management, in areas such as banking, corporate finance, and investments. Credit risk management has evolved rapidly over the past decades, but the global credit crisis of 2007–2008 highlighted that there is still much to be done at multiple levels. Altman and Saunders [3] list five main factors that have contributed to the increasing importance of credit risk management:

1. the worldwide increase in the number of bankruptcies,
2. the trend towards disintermediation by the highest quality and largest borrowers,
3. the increased competition among credit institutions,
4. the declining value of real assets and collateral in many markets, and
5. the growth of new financial instruments with inherent default risk exposure, such as credit derivatives.

Credit risk refers to the probability that an obligor will not be able to meet scheduled debt obligations (i.e., default). Early credit risk management was primarily based on empirical evaluation systems of the creditworthiness of a client. CAMEL has been

M. Doumpos and C. Zopounidis, *Multicriteria Analysis in Finance*,
SpringerBriefs in Operations Research, DOI: 10.1007/978-3-319-05864-1_4,

the most widely used system in this context, which is based on the empirical combination of several factors related to capital, assets, management, earnings, and liquidity. It was soon realized however, that such empirical systems cannot provide a solid and objective basis for credit risk management. This led to an outgrowth of studies from academics and practitioners on the development of new credit risk assessment systems. These efforts were also motivated by the changing regulatory framework that now requires banks to implement specific methodologies for managing and monitoring their credit portfolios [18].

The existing practices are based on sophisticated analytic modeling techniques, which are used to develop a complete framework for measuring and monitoring credit risk. Credit scoring systems are in the core of this framework and are widely used to assess the creditworthiness of firms and individuals, estimate the probabilities of default, and classify the obligors into risk groups.

The aim of credit scoring models is to assess the probability of default for an obligor and differentiate individual credits by the risk they pose. This allows creditors to monitor changes and trends in risk levels, thus promoting safety and soundness in the credit granting process. Credit scoring models are also used for credit approval and underwriting, loan pricing, relationship management and credit administration, allowance for loan and lease losses and capital adequacy, credit portfolio management and reporting [48].

Generally, a credit scoring model can be considered as a mapping function $F(\mathbf{x}; \boldsymbol{\alpha})$, defined by a vector of modeling parameters $\boldsymbol{\alpha}$, such that $F(\mathbf{x}; \boldsymbol{\alpha})\colon \mathbb{R}^K \to G$. The credit scoring model provides estimates for the probability of default for an obligor described by a vector $\mathbf{x} \in \mathbb{R}^K$ of K attributes and maps the result to a set G of risk categories.

The attribute vector $\mathbf{x}$ represents all the relevant information that describes the obligor, including financial and non-financial data. For instance, for corporate loans, financial ratios, measuring the company's profitability, liquidity, leverage, etc., are usually considered to be important quantitative attributes. Non-financial criteria are related to the company's activities, its market position, management quality, growth perspectives, credit history, the trends in its business sector, etc. Empirical evidence has shown that such non-financial attributes significantly improve the estimates of credit scoring and default prediction models [107]. Furthermore, market data and estimates from the Black-Scholes-Merton model have also been shown to be strong predictors of credit risk [68, 250].

Credit risk assessments can be obtained either through models developed internally by financial institutions [244] or are provided externally by credit rating agencies (CRAs). The latter, provide credit ratings for firms in a multi-grade risk scale. Despite the criticisms on their scope and accuracy [93, 190, 241], they are widely used by investors, financial institutions, and regulators, and they have been extensively studied in academic research [130]. However, external ratings, even if considered to be reliable, they do not have a global coverage as they are available only for large corporations, they are not always provided in a timely manner, and they do not differentiate between companies in the same rating class. On the other hand, credit scoring models provide a unique credit score to each rated borrower and they are applicable

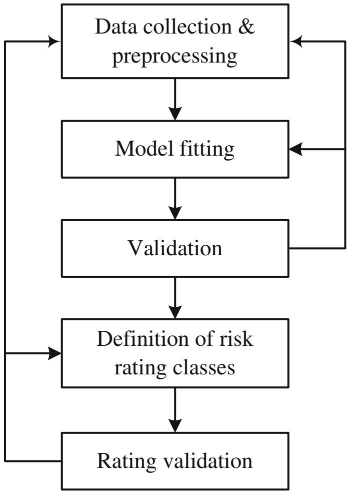

Fig. 4.1 The process for developing credit rating models

to all borrowers (including corporate loans and consumer credit), thus providing a full coverage of a loan portfolio.

4.2 Construction and Validation Process

The credit scoring modeling process can be described through the five steps illustrated in Fig. 4.1.

The process begins with the collection of appropriate data involving obligors with known creditworthiness status. In a typical setting, data for defaulted and non-default cases are collected. These data can be obtained from the historical data base of a credit institution or from external sources. At this stage, some preprocessing of the data is necessary in order to transform them into meaningful attributes, to eliminate outliers, and to select the appropriate set of attributes for the analysis. These steps lead to the final data $\{\mathbf{x}_i, y_i\}_{i=1}^m$, where $\mathbf{x}_i$ is the input attribute vector for obligor i, y_i in the known status of the obligor, and m in the number of observations in the data set. These data, which are used for model development, are usually referred to as *training data*.

The second stage involves the model fitting process, which refers to the identification of the model's parameters that best describe the training data. For instance, assume the following linear model:

$$F(\mathbf{x}) = \alpha_0 + \mathbf{x}\boldsymbol{\alpha}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^K$ is the vector with the coefficients of the selected attributes and α_0 is a constant term. In this case, model fitting is involved with finding the optimal parameters $\boldsymbol{\alpha}$ and α_0 on the basis of the information provided by the training data.

This can be expressed as an optimization problem of the following general form:

$$\min_{\boldsymbol{\alpha} \in \mathscr{A}} L(\boldsymbol{\alpha}, \mathbf{X}) \tag{4.1}$$

where $\mathscr{A}$ is a set of constraints that define the feasible (acceptable) values for the parameters of the model, $\mathbf{X}$ is the training data set and L is a loss function measuring the differences between the model's output and the given classification of the training observations.

On the algorithmic side, several statistical, data mining, and operations research techniques are used to implement the model fitting process. The most widely used methods include logistic regression and probit models, but non-parametric techniques have also gained much interest among researchers and practitioners. Some examples include, neural networks, rule-induction algorithms, support vector machines, fuzzy models, ensembles, and hybrid systems (e.g., neuro-fuzzy models). Comprehensive reviews and discussion of popular methods can be used in Abdou and Pointon [1], Crook et al. [54], and Papageorgiou et al. [193].

The result of the model optimization process are validated using another sample of obligors with known status. This is referred to as the *validation sample*. Typically it consists of cases different than the ones of the training sample and for a future time period. The optimal model is applied to these new observations and its predictive ability is measured, usually using statistical measures (for an overview see [228]). The economic aspects of the model's predictive results are also important [30, 92, 147, 188].

The validation of the scoring model is followed by mapping the model's outputs (credit scores) to risk rating classes consisting of borrowers with similar levels of creditworthiness [157]. The defined rating needs also to be validated in terms of its stability over time, the distribution of the obligors in the rating groups, and the consistency between the estimated probabilities of default in each group and the empirical ones which are taken from the population of rated obligors.

4.3 Multicriteria Aspects of Credit Scoring

From the methodological point of view, credit scoring for business and consumer loans is a statistical pattern classification problem, as the decision models are constructed on the basis of historical default data.[1] Nevertheless, some features that analysts often require scoring models to have [147], make MCDA techniques appealing in this context. In particular:

- *Credit scoring models are usually required to be monotone with respect to the inputs*. From an economic and business perspective, the monotonicity assumption

[1] In other specialized credit granting contexts (e.g., project finance), the risk assessment process is mostly based on empirical quantitative and qualitative models [181] (Chaps. 8, 10), which fit well the context of MCDA.

implies that as the input information for a given applicant improves, the estimated probability of default should decrease. Assuming that all attributes are in a maximization form, the monotonicity assumption can be formally expressed as follows:

$$\Pr(D|\mathbf{x}_i) \leq \Pr(D|\mathbf{x}_j), \quad \forall \mathbf{x}_i \succ \mathbf{x}_j \tag{4.2}$$

where $\Pr(D|\mathbf{x}_i)$ is the estimated probability of default for credit applicant i and $\succ$ represents the dominance relationship, defined as follows: $\mathbf{x}_i \succ \mathbf{x}_j \Leftrightarrow \mathbf{x}_i \geq \mathbf{x}_j$ and $x_{ik} > x_{jk}$, for at least one attribute k.
Models that violate monotonicity in an arbitrary manner may fail to be accepted, simply because they lack economic sense, thus providing counterintuitive results from an economic perspective. Furthermore, empirical results have shown that introducing monotonicity in credit scoring models actually improves their predictive performance and robustness, through the elimination of the over-fitting effect [72].

- *Credit scoring models should be transparent and comprehensible*. The predictive accuracy of credit scoring models is not the sole decisive factor for their success in practice. In addition to being accurate, the modes should also be easy to understand by analysts, end users, and regulators. A comprehensible model enables its user to understand its underlying logic and provide justifications on its recommendations [170, 172], instead of simply being used as a black-box analytic recommendation tool.
- *Risk grades are ordinal*. This is often ignored by many popular statistical and computational intelligence techniques used for model building, which often assume that the classes are nominal (i.e., in no particular order).

Multicriteria decision models fit well these requirements: (a) they are by definition ordinal, (b) they provide evaluation results that are monotone with respect to the evaluation criteria, and (c) they promote transparency, enabling the credit analyst to calibrate them on the basis of his/her expert domain knowledge, and allow for justification of the obtained results. Among others, MCDA methods have been used in the area of credit scoring (and the relevant field of bankruptcy prediction) in different ways:

1. As tools for building accurate and transparent credit scoring systems, customized to the needs of particular financial institutions [51, 99]. This is particularly important for special types of credit (e.g., project finance) for which historical data may be lacking. In such cases, MCDA methods can greatly enhance peer expert judgment scoring systems, facilitating the structuring of the credit granting evaluation process and providing formal procedures for aggregating multiple credit evaluation criteria.
2. In combination with other modeling and learning techniques, including rough sets, fuzzy models, case-based reasoning, and neural networks [39, 120, 252, 262]. Such computational intelligence techniques provide strong data analysis capabilities. MCDA on the other hand, provides axiomatic decision models of different

forms. The combination of these paradigms [64] provides a new set of powerful hybrid systems for credit scoring.

3. As optimization approaches for model fitting under multiple performance measures [113, 154, 186]. The performance of a credit scoring model has different aspects, including statistical (e.g., different measures of predictive accuracy) and economic (profit/costs derived from actions taken on the basis of the results of a credit scoring model). Multiobjective optimization techniques enable the consideration such multiple performance measures when building a credit scoring model.
4. As alternatives to popular statistical and machine learning approaches providing more accurate rating results [69, 74, 121]. The results from several studies show that credit scoring models constructed using MCDA preference disaggregation techniques provide robust and accurate results, and often actually outperform other popular approaches. Thus, they could be considered as potential candidates for constructing credit scoring and rating models.

The next section illustrates the application of a multicriteria methodology for developing a credit scoring model and its comparison to popular statistical and non-parametric techniques.

4.4 Using Preference Disaggregation Analysis for the Construction of a Credit Scoring Model

MCDA provides a variety of approaches for credit risk modeling and the construction of credit scoring systems, including outranking techniques [74, 121], rule-based models [39, 61, 252, 262], and value models [51, 69, 63].

To facilitate the presentation we shall focus on additive value models in the framework of the UTADIS method [70, 265]. Additive models are popular approaches for credit risk modeling, as they are intuitive scoring systems, that are simple to understand and implement, as they are compatible with the scorecard structure of credit rating systems used in practice [219]. For instance, Krahnen and Weber [147] conducted a survey among major German banks and found that all of them used credit scoring models expressed in the form of an additive value function:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} w_k v_k(x_{ik}) \tag{4.3}$$

where the global value $V(\mathbf{x}_i)$ is an estimate of the overall creditworthiness and default risk of obligor i.

In this model, the overall assessment is a weighted average of partial scores $v_1(x_{i1}), \ldots, v_K(x_{iK})$ defined over a set of K credit risk assessment criteria. Without loss of generality, we shall assume that the weighting trade-off constants are non-negative and normalized such that $w_1 + w_2 + \cdots + w_K = 1$. On the other hand, the marginal value functions $v_1(\cdot), \ldots, v_K(\cdot)$, which define the partial scores, are scaled such that $v_k(x_{k*}) = 0$ and $v_k(x_k^*) = 1$, where x_{k*} and x_k^* are the most and least risky level of risk attribute k, respectively. For simplicity, henceforth it will be assumed that all risk assessment criteria are expressed in maximization form (thus implying that all marginal value functions are non-decreasing).

The construction of the credit scoring model (4.3) can be simplified by setting $u_k(x_k) = w_k v_k(x_k)$, which leads to a rescaled set of marginal value functions $u_1, \ldots, u_K$ normalized in $[0, w_k]$. With this transformation, the evaluation model (4.3) can be re-written in the following equivalent form:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} u_k(x_{ik}) \tag{4.4}$$

This decision model can be linear or nonlinear depending on the form of the marginal value functions. The marginal value functions can be either pre-specified by the decision maker or inferred directly from the data using a preference disaggregation approach. In the context of credit scoring the latter approach is the preferred one, particularly when there are historical data available for constructing the model. Under this scheme, a convenient and flexible way to take into consideration a wide class of monotone marginal value functions, is to assume that they are piecewise linear. In that regard, the range of each risk criterion k is split into $s_k + 1$ subintervals defined by s_k break-points $\beta_0^k < \beta_1^k < \cdots < \beta_{s_k+1}^k$, between the least and the most preferred levels of the criterion (denoted by β_0^k and $\beta_{s_k+1}^k$, respectively), as illustrated in Fig. 4.2. Thus, the marginal value of any alternative i on criterion k can be expressed as:

$$u_k(x_{ik}) = \sum_{r=1}^{s_k} p_{ik}^r d_{kr} \tag{4.5}$$

where $d_{kr} = u_k(\beta_r^k) - u_k(\beta_{r-1}^k) \geq 0$ is the difference between the marginal values at two consecutive levels of criterion k and

$$p_{ik}^r = \begin{cases} 0 & \text{if } x_{ik} < \beta_{r-1}^k \\ \frac{x_{ik} - \beta_{r-1}^k}{\beta_r^k - \beta_{r-1}^k} & \text{if } x_{ik} \in [\beta_{r-1}^k, \beta_r^k] \\ 1 & \text{if } x_{ik} > \beta_r^k \end{cases} \tag{4.6}$$

With the above piecewise linear modeling of the marginal value functions, the scoring model (4.4) can be expressed as a linear function of the step differences in the marginal values between consecutive break-points in the criteria's scale:

$$V(\mathbf{x}_i) = \sum_{k=1}^{K} \mathbf{p}_{ik}^\top \mathbf{d}_k \tag{4.7}$$

where $\mathbf{p}_{ik} = (p_{ik}^1, p_{ik}^2, \ldots, p_{ik}^{s_k})$ and $\mathbf{d}_k = (d_{k1}, d_{k2}, \ldots, d_{ks_k})$.

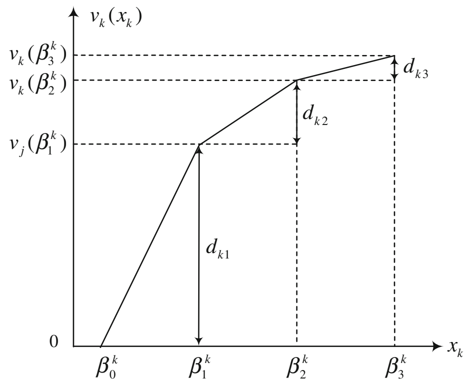

Fig. 4.2 Piecewise linear modeling of a marginal value function

The parameters of model (4.7) can be estimated in the context of the MCDA disaggregation paradigm [128] with non-parametric linear programming formulations, using data for obligors classified into predefined risk classes. Such data can be collected from historical data bases of financial institutions. Usually the data consist of defaulted and non-defaulted obligors, but multi-grading schemes are also possible, such as the credit ratings issued by credit rating agencies [69].

In a general setting, let us assume that reference (training) data for $M_1, M_2, \ldots, M_N$ obligors are available from N risk classes $C_1, \ldots, C_N$, defined such that C_1 is the low risk category and C_N the higher risk one. The decisions based on a credit scoring model $V(\mathbf{x})$ are made on the basis of the following classification rule:

$$\text{Obligor } i \text{ belongs in risk category } \ell \Longleftrightarrow t_\ell < V(\mathbf{x}_i) < t_{\ell-1} \tag{4.8}$$

where $1 > t_1 > t_2 > \cdots > t_{N-1} > 0$ are score thresholds that distinguish the risk classes. The scoring model and thresholds that best fit the above rule, according to the available training data for M obligors can be estimated through the solution of the following linear programming problem [71]:

$$\begin{aligned}
&\min \quad \sum_{\ell=1}^{N} \frac{1}{M_\ell} \sum_{\mathbf{x}_i \in C_\ell} (\varepsilon_i^+ + \varepsilon_i^-) + \lambda \sum_{k=1}^{K} \mathbf{1}^\top \mathbf{d}_k \\
&\text{s.t.} \quad V(\mathbf{x}_i) = \sum_{k=1}^{K} \mathbf{p}_{ik}^\top \mathbf{d}_k && i = 1, 2, \ldots, M \\
&V(\mathbf{x}_i) - t_n + \varepsilon_i^+ \geq 1, && \forall\, \mathbf{x}_i \in C_\ell,\ \ell = 1, \ldots, N-1 \\
&V(\mathbf{x}_i) - t_{n-1} - \varepsilon_i^- \leq -1, && \forall\, \mathbf{x}_i \in C_\ell,\ \ell = 2, \ldots, N \\
&t_{\ell-1} - t_\ell \geq 0, && \ell = 2, \ldots, N-1 \\
&\mathbf{d}_k, t_\ell, \varepsilon_i^+, \varepsilon_i^- \geq 0 && \forall\, i, k, \ell
\end{aligned} \tag{4.9}$$

where **1** is a vector of ones. The first set of constraints defines the credit scores for the training cases according to the additive model (4.7). The second set of constraints defines the violations (ε^+) of the lower bound of each risk class (this applies only to obligors belonging to classes $C_1, \ldots, C_{N-1}$), whereas the third set of constraints defines the violations (ε^-) of the upper bound of each risk category (this applies only to the obligors belonging to classes $C_2, \ldots, C_N$). The last constraint is used to ensure that the thresholds are monotonically non-increasing.

The objective function combines two terms. The first involves the minimization of model's fitting error. This is defined as the weighted sum of the errors for cases belonging into different classes, where the weights are defined in terms of the number of sample observations in each class. In this way, it is possible to handle reference sets with considerable imbalanced class sizes, which are very common in credit scoring (e.g., the number of obligors is default is much lower than the non-defaulted obligors). The second term in the objective function is a regularization term in accordance with Tikhonov's regularization principle [242]. The parameter $\lambda > 0$ defines the trade-off between the minimization of the fitting error and the complexity of the model, which can be set by trial-and-error or with statistical resampling techniques such as such as cross-validation [233] and the bootstrap [79].

Denoting by $\mathbf{d}_k^*$ ($k = 1, \ldots, K$) the optimal parameters of the model resulting from the solution of the above linear program, the constructed additive value function is scaled between zero and $\theta = \sum_{k=1}^{K} \mathbf{1}^\top \mathbf{d}_k$. Rescaling the model in [0, 1] can be easily done simply by dividing the optimal solution by θ.

The use of linear programming for model fitting enables the handing of big data sets. This is particularly important for credit scoring, as the available data become larger, particularly after the introduction of the Basel II regulatory framework. Furthermore, a linear programming model enables the risk analyst to incorporate special domain knowledge, which can be very useful for calibrating the model with expert judgment, in order to capture aspects of the problem not adequately covered by the data. Finally, post-optimality techniques can be employed to analyze the robustness of the results and the obtained model [75].

This modeling framework is also applicable with other types of decision models for credit scoring and alternative optimization techniques for model fitting. For instance, Doumpos [63] presented an evolutionary algorithm for constructing a non-monotone value function, whereas Doumpos and Zopounidis [74] used a similar algorithm for an outranking model. Bugera et al. [37] introduced goal programming models for developing a credit scoring model in the form of a quadratic value function, whereas Doumpos et al. [65] used the MHDIS method [266], which is based on multiple additive value models. Other optimization formulations (linear and non-linear) for fitting multicriteria credit scoring models have been also been proposed in several studies [99, 113, 197, 263].

Table 4.1 Number of sample observations in each year and category

Years	Non-defaulted	Defaulted	Total
2007	2,748	252	2,800
2008	2,846	253	2,899
2009	2,731	299	2,830
2010	2,143	244	2,187
Total	10,468	248	10,716

4.5 An Application

4.5.1 Data

To illustrate the usefulness and performance of MCDA methods in credit scoring, a large sample of Greek firms from the commercial sector (wholesale and retail trade) is used. The sample is taken from the database of ICAP S.A., which is a leading business information and consulting firm in Greece. The data span the period 2007–2010. In each year throughout that period, the firms in the database were classified either in the default or in the non-default group. The default group consists of firms that declared bankruptcy as well as firms with other default events such as protested bills, uncovered cheques, payment orders. Table 4.1 presents the number of observations from the two groups for each year in the sample.

The firms in the sample are described over seven financial ratios (Table 4.2), which cover three main aspects of corporate performance in accordance with the framework of Courtis [53]:

- *Profitability*: Profitability ratios assess the ability of a firm to generate earnings. The profitability ratios considered in this study include the gross profit margin (gross profit/sales) and return on assets (earnings before interest and taxes/total assets). The gross profit margin ratio is used to assess the sales profitability of the firms, after controlling for the cost of sales, whereas the return on assets ratio provides an overall evaluation of the operating profitability of the firms, taking into consideration all types of operating expenses.
- *Solvency and liquidity*: Solvency assesses the dependency of the firms on debt financing and their overall level of leverage. Liquidity, on the other hand, determines a company's ability to pay off its short-term debt obligations. In this study, total liabilities/total assets is used to assess the firms' solvency, whereas the liquidity of the firm is considered through the current ratio (current assets/current liabilities).
- *Managerial performance*: Managerial performance ratios focus on the efficiency of a firm's policies towards its creditors and clients as well as its financial efficiency. The former dimension is taken into consideration through two ratios, namely the receivables turnover ratio (accounts receivables $\times$ 365/sales) and the sales to current liabilities ratio. On the other hand, financial efficiency is analyzed through the interest expenses ratio (interest expenses/sales).

Table 4.2 Financial ratios for credit risk assessment

Financial ratios	Abbreviation	Relationship to credit risk
Profitability		
Gross profit/sales	GP/S	–
Earnings before interest and taxes/total assets	EBIT/TA	–
Solvency and liquidity		
Total liabilities/total assets	TL/TA	+
Current assets/current liabilities	CA/CL	–
Managerial performance		
Accounts receivables $\times$ 365/sales	AR/S	+
Sales/current liabilities	S/CL	–
Interest expenses/sales	IE/S	+

The risk of default increases with ratios that are positively related to credit risk

Table 4.3 Averages and standard deviations (in parentheses) of the financial ratios for each group of firms

	Non-default	Default
GP/S	0.30 (0.20)	0.23 (0.20)
EBIT/TA	0.04 (0.12)	−0.04 (0.14)
TL/TA	0.72 (0.27)	0.88 (0.25)
CA/CL	1.67 (1.55)	1.22 (1.07)
AR/S	237.31 (247.79)	342.55 (371.90)
S/CL	2.57 (2.96)	1.51 (2.41)
IE/S	0.03 (0.04)	0.07 (0.08)

The selection of the financial ratios was based on the combination of three main factors: (a) the research literature and the current best practices in the area of credit scoring by international organizations, (b) the judgment of credit scoring analysts with significant expertise on the characteristics of Greek firms, and (c) the discriminating power of the ratios.

Table 4.3 provides some basic statistics for the selected financial ratios for each group of firms. As expected, firms in default have lower profitability, higher debt burden and lower liquidity, higher interest expenses, and are less efficient in terms of the credit they provide to their clients (accounts receivable turnover) and the management of their short-term liabilities (S/CL ratio). The differences between the two groups are all found to be significant at the 1 % level through the Mann-Whitney non-parametric test.

4.5.2 Results

In order to be able to assess the predictive performance of a credit scoring model, a holdout sample is required, ideally involving data for different obligors and time period compared to the data set used to construct the scoring model [228]. In that

Table 4.4 Contribution of the financial rations in the model

	UTADIS	LR
GP/S	0.009	−0.684(0.068)
EBIT/TA	0.184	−3.645(0.195)
TL/TA	0.180	1.791 (0.207)
CA/CL	0.199	−0.332(0.158)
AR/S	0.104	0.001 (0.152)
S/CL	0.131	0.047 (0.069)
IE/S	0.194	8.113 (0.151)
Constant		−1.497

regard, the sample described in the previous section is split in two parts. The first covers the period 2007–2008 and it is used as the training sample, whereas the data for the period 2009–2010 are employed to test the performance of the credit scoring models (i.e., holdout sample).

Following this procedure, Table 4.4 reports the trade-offs of the financial ratios in the multicriteria additive model, as estimated through the solution of the linear program (4.9) using the 2007–2008 data. For comparative purposes the coefficients of the ratios in a logistic regression (LR) model are also reported, together with their relative contribution in the model (in parentheses). LR is the most popular statistical approach for constructing credit scoring models and it is widely used in this field by both researchers and practitioners. In the context of LR, the relative importance of the ratios can be assessed through the following index:

$$w_k = \frac{|\alpha_k|\sigma_k}{\sum_{k=1}^{K} |\alpha_k|\sigma_k}$$

where α_k is the regression coefficient of ratio k and σ_k is the standard deviation of the ratio. Measured in this way, w_k represents the relative influence of ratio k on the LR result in terms of the absolute impact of a standard deviation change in the ratio as a proportion of the total absolute change in the dependent variable, given a standard deviation change in all ratios [2].

In the multicriteria model, liquidity (CA/CL), interest expenses (IE/S), return on assets (EBIT/TA), and solvency (TL/TA) have the highest trade-offs and consequently they are important factors in the credit scoring process. The same variables also contribute significantly in the LR model (all coefficients are significant at the 1 % level). However, it is worth noting that the coefficient of the sales/short-term liabilities ratio has an incorrect sign in the LR model, as its positive regression coefficient indicates that the probability of default increases with this ratio, which does not comply with the economic interpretation of this ratio.

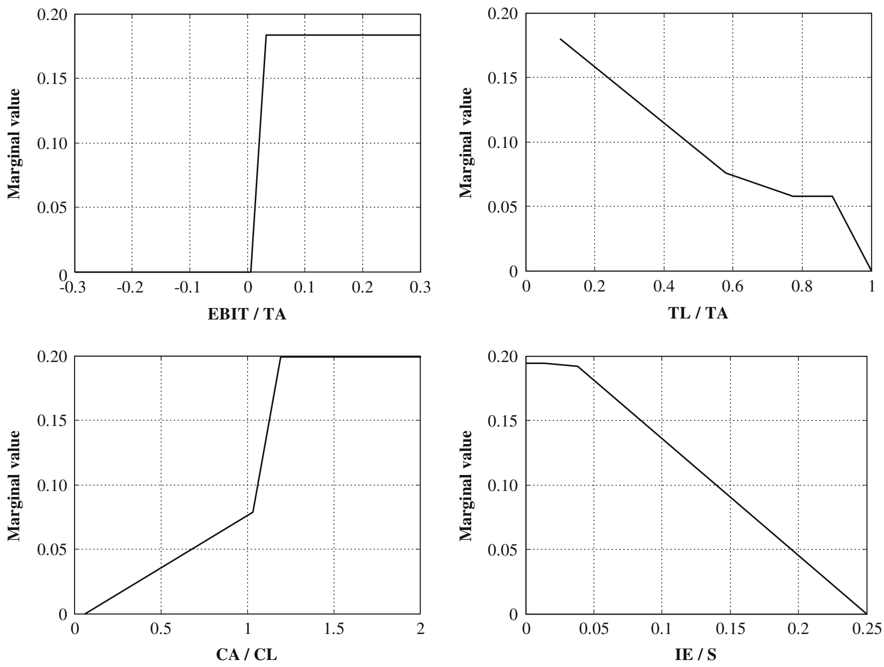

Fig. 4.3 Marginal value functions

Figure 4.3 illustrates the marginal value functions of the four most significant ratios in the multicriteria model. These functions provide further insights into how the overall credit score of the firms is affected by their performance on these ratios. For instance, EBIT/TA has a type of a step function, with the marginal value (partial credit score) increasing for positive values of the ratio. Thus, the likelihood of default is significantly lower for firms with positive return on assets. A similar behavior is also observed for liquidity; the credit score increases (improves) linearly when CA/CL is below one, but improves significantly for higher values. On the other hand, the marginal value functions for the solvency and the interest expenses ratios have a nearly linear form. As far as the TL/TA ratio is concerned, the partial credit score decreases almost linearly for firms with TL/TA < 0.9, but it is significantly lower for firms facing a higher debt burden. On the other hand, the marginal value for the IE/S ratio remains at high levels for firms with IE/S lower than 5 % and decreases linearly for firms with higher interest expenses. This kind information derived from the marginal value functions of the credit assessment criteria can be of great help for credit analysts, as it enables them to have a better understanding of the credit scoring model.

Except for analyzing the structure of a credit scoring model and the role of the credit assessment criteria, the relationship between the probability of default and the credit scores of the model as well as the model's predictive performance, are also

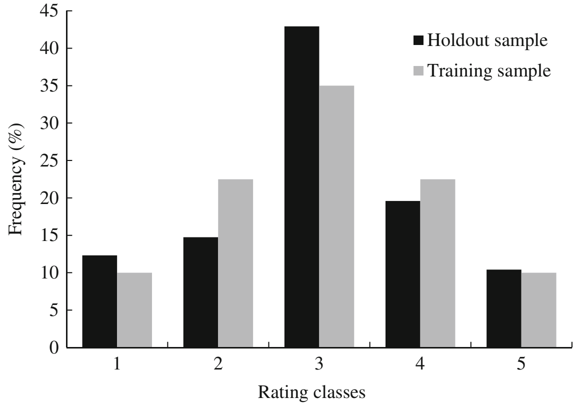

Fig. 4.4 Distribution of sample observations in the credit rating classes

critical issues for the implementation of the model in practice. In this application, these issues are analyzed by applying the constructed multicriteria model to the 2009–2010 holdout data. In order to examine the relationship between the probability of default and the credit scores of the model, the latter are mapped to a five-point credit rating scale. The rating scale is defined on the basis of the global values (credit scores) of the observations in the training sample, as follows:

- class 1: very low risk firms with $V(\mathbf{x}_i) \geq 0.899$,
- class 2: low risk firms with $0.82 \leq V(\mathbf{x}_i) < 0.899$,
- class 3: medium risk firms with $0.617 \leq V(\mathbf{x}_i) < 0.82$,
- class 4: high risk firms with $0.513 \leq V(\mathbf{x}_i) < 0.617$,
- class 5: very high risk firms with $V(\mathbf{x}_i) < 0.513$.

The thresholds are set such that the firms are approximately normally distributed in the five rating classes, according to the available data for the calibration of the model (i.e., the training data). In that regard, the top 10 % of the training cases are assigned to class 1 (i.e., the threshold 0.899 is the 90 % percentile of the global values of the training observations). The next 22.5 % of the training cases are assigned to class 2 (i.e., the threshold 0.82 is the 67.5 % percentile of the global values of the training observations). Following the same approach, the threshold 0.617 that distinguishes medium risk firms from high risk ones, corresponds to the 32.5 % percentile of the scores in the training sample (i.e., the medium risk category consists of 35 % of the cases), whereas the threshold 0.513 that distinguishes high risk firms from very high risk ones corresponds to the 10 % percentile of the scores in the training sample (the high risk group includes 22.5 % of the training observations and the very high risk class includes the bottom 10 %). The score thresholds specified in this way are then used to rate the firms in the holdout sample. As shown in Fig. 4.4, the distribution of the sample observations in the five rating classes exhibits good stability when comparing the results for the training and holdout samples.

On the basis of this five-point rating, the probability of default in each rating class can be estimated as the number of cases in default to the total number of cases in each

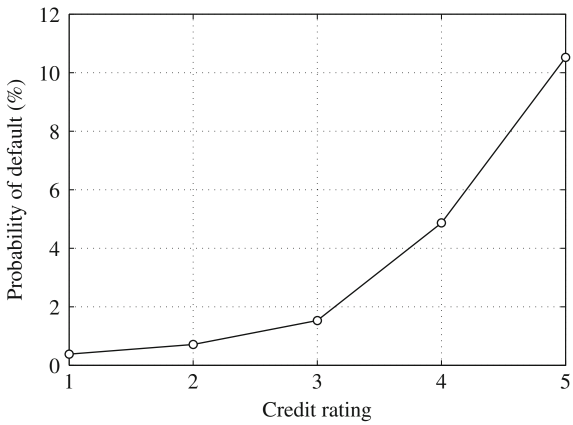

Fig. 4.5 Empirical probability of default for each rating class (holdout sample)

category. Figure 4.5 illustrates the results for the holdout sample. It is evident that the risk of default increases exponentially when moving from low risk grades to high risk ones (by a factor of about two). This is an appealing feature for a credit scoring model, as it indicates that the model provides a clear differentiation of the obligors in terms of their risk level, and its results are in accordance with the empirical default frequency in the data.

To further analyze the predictive ability of the multicriteria model, different performance measures are employed:

- *Accuracy rates*: On the basis of the credit scores estimated through a model and a cut-off point, obligors are classified in the pre-defined risk categories (default and non-default). Then, different accuracy measures can be defined. In this application we use two main accuracy criteria:
  - *Overall classification accuracy* (*OCA*): the ratio between the model's correct classifications to the total number of obligors evaluated. Similar calculations can be made (separately) for each risk category. Thus, the accuracy rate α_{ND} for the non-default group is defined as the percentage of non-defaulted obligors classified correctly by the model. The accuracy rate α_D for the default group is defined in the same way (i.e., the percentage of defaulted obligors classified by the model in the default category).
  - *Average classification accuracy* (*ACA*): the average of α_{ND} and α_D. This averaging can be justified for the most common setting where a credit scoring model is constructed with data for defaulted and non-defaulted obligors, taking into account the expected misclassification cost. In particular, denoting by p_D the a-priori probability of default, the expected misclassification cost of a credit scoring model is:

$$E(C) = p_D C_D \alpha_D + (1 - p_D) C_{ND} D a_{ND},$$

where C_D is the cost of misclassifying an obligor in default and C_{ND} is the cost for a non-defaulted obligor. The former is associated with losses due to default, whereas the latter is related to the opportunity cost derived by rejecting credit to a creditworthy client. Obviously C_D is much higher than C_{ND}, but on the other hand, the a-priory probability of default (p_D) is generally low (e.g., typically around 5 %). Thus, it is reasonable to assume that $p_D C_D \approx (1 - p_D) C_{ND} = P$, in which case:

$$E(C) \approx 2P \frac{\alpha_{ND} + \alpha_D}{2} = 2P \times ACA$$

Thus, ACA is (in general) a reasonable proxy for the expected cost that arises from using a credit scoring model.

- *Area under the receiver operating characteristic curve*: (*AUROC*) The AUROC provides an overall evaluation of the generalizing performance of a classification model without imposing any assumptions on the misclassification costs or the prior probabilities [87] and it is commonly used to assess the discriminating power of credit rating models [30, 82, 222]. Formally, the AUROC represents the probability that a non-defaulted obligor will receive a higher credit score compared to one in default. Thus, it can be calculated as follows:

$$\text{AUROC} = \frac{1}{M_D M_{ND}} \sum_{i \in ND} \sum_{j \in D} I(\mathbf{x}_i, \mathbf{x}_j)$$

where M_D, M_{ND} denote the number of observations in default and non-default respectively and $I(\mathbf{x}_i, \mathbf{x}_j)$ is defined for a credit scoring model $V(\mathbf{x})$ as follows:

$$I(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1 & \text{if } V(\mathbf{x}_i) > V(\mathbf{x}_j) \\ 0.5 & \text{if } V(\mathbf{x}_i) = V(\mathbf{x}_j) \\ 0 & \text{if } V(\mathbf{x}_i) < V(\mathbf{x}_j) \end{cases}$$

- *Kolmogorov-Smirnov distance*: (*KS*) The Kolmogorov-Smirnov distance is the maximum absolute difference between the cumulative distribution functions of the credit scores of the obligors belonging into different groups. The highest is this difference the more powerful is a credit scoring model in discriminating the risk classes.

Table 4.5 presents detailed results on the predictive ability of the multicriteria credit scoring model according to the above performance measures. For comparison purposes, the results of LR are reported as well as those of a support vector machine (SVM) model developed with a radial basis function kernel (SVM-RBF) using the LIBSVM library in MATLAB R2013 [42]. SVMs have become an increasingly popular statistical learning methodology for developing classification and regression models [249] with many successful applications in financial decision-making

Table 4.5 Comparative results for the predictive performance of the models

Measures	Methods	2009	2010	2009–2010
α_{ND}	UTADIS	0.720	0.723	0.721
	LR	0.676	0.690	0.682
	SVM-RBF	**0.735**	**0.745**	**0.739**
α_D	UTADIS	**0.707**	**0.818**	**0.741**
	LR	0.697	0.795	0.727
	SVM-RBF	0.687	0.750	0.706
ACA	UTADIS	**0.713**	**0.771**	**0.731**
	LR	0.687	0.743	0.705
	SVM-RBF	0.711	0.748	0.723
OCA	UTADIS	0.719	0.725	0.722
	LR	0.677	0.692	0.683
	SVM-RBF	**0.733**	**0.745**	**0.738**
KS	UTADIS	**0.442**	**0.567**	**0.479**
	LR	0.393	0.515	0.429
	SVM-RBF	0.435	0.512	0.453
AUROC	UTADIS	**0.769**	**0.826**	**0.786**
	LR	0.756	0.815	0.775
	SVM-RBF	0.767	0.825	0.785

The best result for each performance measure is marked in bold

problems, including credit scoring [23, 123, 171, 234]. The use of the RBF kernel enables the development of nonlinear classification models, as opposed to the linear modeling setting of LR and the additive nature of the MCDA approach used in this analysis.

The results of Table 4.5 indicate that the multicriteria credit scoring model consistently outperforms LR on all performance measures and time periods, while being quite competitive to the SVM-RBF nonlinear model. In particular, the UTADIS model outperforms LR and SVM-RBF in identifying firms in default. Throughout the two-years period 2009–2010, the accuracy of the multicriteria model for the firms in default is 74.1 % versus 72.7 and 70.6 % for LR and SVM-RBF. On the other hand, the SVM-RBF model performs better for the non-default group, whereas LR performs poorly compared to the other methods. Overall, the UTADIS model achieves the best balance between the accuracy rates for the two risk groups. As a result, it outperforms the other models in terms of ACA (73.1 % overall versus 72.3 % for the SVM-RBF models, and 70.5 % for the LR model). The good performance of the SVM-RBF model for the non-default group (which is the largest one; cf. Table 4.1) leads to its high OCA. The UTADIS model follows second in terms of its OCA. Finally, as far as the two performance measures that do not involve accuracy rates are concerned, the UTADIS model performs consistently better than LR and SVM-RBF. The differences are higher for the KS distance, whereas in terms of AUROC the multicriteria model and SVM-RBF perform almost equally well.