BANKRUPTCY PREDICTION MODELS: PRELIMINARY THOUGHTS ON THE DETERMINATION OF PARAMETERS FOR THE EVALUATION OF EFFECTIVENESS AND EFFICIENCY

Greta Cestari, PhD
Department of Economics and Management, University of Ferrara, Italy

Gianluca Risaliti, Prof
Department of Accounting, Management and Economics, University of Naples Parthenope, Italy

Mariarita Pierotti, PhD
Department of Business Administration, Finance, Management & Law, University of Milano-Bicocca, Italy

Abstract
From initial developments (Beaver, 1966; Altman, 1968), the interest of experts, academics and others regarding models of bankruptcy prediction, has intensified, also in the light of the recent global economic-financial crisis. To date research in this field has mainly concentrated on the development of an instrument that has a higher level of reliability. Particular attention, has therefore been paid to the fine-tuning of the logical-technical system of models and proven efficiency, in terms of overall accuracy and proportional impact of errors made according to type. Nonetheless, one aspect that has not been explored in full in this field is that of accessibility of models that require the analysis of factors that affect usability of these instruments in an operational field. The analysis perspective is that of the user, therefore, over and above the diagnostic accuracy test, specific importance must be given to the implications which the adoption of the model can have on the user’s organisation and the costs that are involved with the use thereof. The diagnostic instrument can be effectively and profitably used only if these implications are sustainable for the specific user.
This article aims at identifying the parameters that can measure efficiency, in terms of diagnostic reliability and speed, and effectiveness, in terms of organisational and economic sustainability, of prediction instruments. This objective has been pursued with an in-depth analysis of existing literature in
matters of bankruptcy prediction models.

**Keywords**: Corporate crisis, bankruptcy prediction models, effectiveness, efficiency

**Introduction**

In recent years, research relating to the subject of bankruptcy prediction models has concentrated mainly on the development of instruments characterised by an ever-greater degree of diagnostic reliability. In particular, experts have concentrated on mathematical-statistical aspects, where improvement has allowed the fine-tuning of a logical-technical system and has allowed this instrument to be increasingly accurate and methodologically correct. This process of progressive perfectioning has led to a considerable improvement in terms of accuracy and precision in predictions. Nonetheless, this focus has set aside the aspects linked to the actual usability of the model on the part of potential users. By “analysis of usability” we mean the examination of factors that affect the actual possibility, for the various operators in the market, to profitably use the prediction models in their own assessment of the state of health of the companies (on the various uses of prediction models please see: Altman and Sametz, 1977; Altman and Hotchkiss, 2010). This must be mentioned if we wish to extend the results of research in bankruptcy prediction developed in an academic field, also to an operational context.

Having said this, we think that besides the traditional research aimed at developing the most accurate instrument, it would be interesting to also draw up studies that can examine and exploit existing instruments from the viewpoint of the potential user. The latter, faced with a wide range of models, must choose the one that can best fulfil his own specific requirements, in terms of quality-quantity characteristics and related costs. If it is true that now there are a large number of prediction instruments available, even rather accurate ones, it is equally true that some of these are extremely complicated to use and are rather expensive as regards acquisition/development and subsequent use. It is clear that these implications, if not sustainable for the user, can be detrimental to the use of said instruments in an operational sense. These aspects are equally important, as the sustainability of the model, both in terms of organisation and finance, constitutes a fundamental aspect in the choices made by the user.

This having been said, we think that over and above diagnostic efficiency (understood as being predictional accuracy), usually examined to assess how “good” the prediction model can be, we should also carefully analyse and assess the effectiveness of the instrument (in terms of
economic-organisational sustainability). In the light of this new analysis perspective, there is the problem of establishing factors which the potential user can examine and compare in order to be able to assess the existing models and choose the one most suited to his own operation.

This research falls within this aspect. This study aims at formulating some initial thoughts on diagnostic accuracy and the economic-organisational sustainability of prediction instruments. Specifically, this research aims at identifying the parameters aimed at assessing the degree of effectiveness and efficiency of insolvency prediction models, developed in literature or professional practice, through the perspective of the end user. Precisely, this research aims at answering two main questions:

1. Which parameters must be examined in order to assess the effectiveness of insolvency prediction models?
2. Which parameters must be examined to assess the efficiency of insolvency prediction models?

The research was developed along the lines of a methodological approach which, in keeping with traditional Italian research, is defined as being deductive (Ferraris Franceschi, 1978, 1998): the conclusions have been drawn following an in-depth critical analysis of the existing literature on the matter.

The paper is divided into five sections. The paragraph below gives a brief theoretical outline. The third paragraph aims at determining the parameters of effectiveness, in terms of both diagnostic reliability (sub-paragraph 3.1) and speed in deciding the state of health of companies (sub-paragraph 3.2). The fourth paragraph explains the parameters of efficiency: the criteria of organisational sustainability are discussed in sub-paragraph 4.1, while those regarding economic sustainability are discussed in sub-paragraph 4.2. The fifth paragraph puts forward the application to a sample of known models of efficiency and effectiveness parameters. The last paragraph contains some conclusions that have been drawn during the research and future developments of the research.

Theoretical framework

To date, studies undertaken in the field of bankruptcy prediction models, both on a national and international level, appear rich and particularly diversified: in fact, in this area of research the contributions offered by experts and economic operators (from the banking sector and non) are numerous and heterogeneous (Scott, 1981; Altman and Avery and Eisenbeis and Sinkey, 1981; Zavgren, 1983; Altman, 1984; Jones, 1987; Altman and Narayanan, 1997; Poddighe and Madonna, 2006; Bellovary and Giacomino and Akers, 2007).
In the 80’s and 90’s, these studies were more frequent following the mass diffusion, on the one hand, of information system tools and technologies, and, on the other, the discovery and the consequent introduction of new mathematical and statistical methodologies (Zhang et al., 1999; Aziz and Dar, 2006). This tendency has had the aim of elaborating an increasingly reliable model for the forecasting of corporate critical statuses. For this reason, the trend over the last decades has attracted the attention of experts to the methodological features adopted in the elaboration process of the forecasting model. In particular, these tools have been the subject of a gradual process of refined improvement, especially from a mathematical and statistical point of view: over the years, in fact, corrective interventions have been introduced with the simple aim of improving the logical-technical structure (Lim and Yun, 2012).

To this end, some studies have focused on the correct composition of the elaboration sample, both in a qualitative (Taffler, 1982; Gilbert et al., 1990) and a quantitative (Altman et al., 1981; Zmijewski, 1984) perspective. Other observations have been carried out in relation to the number (Altman 1988, Teodori 1989) and the nature (Edmister, 1972; Tennyson et al, 1980; Peel and Peel and Pope, 1985, D’Aveni, 1989) of the indicators able to feed the model, as well as to the relative importance (Eisenbeis, 1979) and the correlation degree (Altman, 1988) of the independent variables.

Moreover, since the first elaborations, most of the literature has focused on the detection of the statistical methodology that is most appropriate for the formulation of increasingly reliable diagnosis models. At first, the studies focused on the traditional techniques of the univariate discriminant analysis (Beaver, 1966; Ruozzi, 1974) and the multiple discriminant analysis (Altman, 1968; Alberici, 1975; Altman, Haldeman, and Narayanan, 1977), bayesian analysis (Forestieri, 1986), the principal components analysis (Cascioli and Provasoli, 1986), and the logit analysis (Martin, 1977; West, 1985; Platt and Platt, 1991). More recently, instead, the attention has concentrated on more innovative methodologies of the Recursive partitioned decision tree (Frydman and Altman and Kao, 1985; Pompe and Feelders, 1997), neural networks (Altman and Marco and Varetto, 1994; Yang et al., 1999) and genetic algorithms (Kingdon and Feldman, 1995; Varetto, 1998; Shin and Lee, 2002).

These processes of methodological investigation and refinement have been carried out to achieve the elaboration of particularly sophisticated instruments, methodologically speaking, that are effective in predicting the degree of vitality of the companies. Nevertheless, to this day no research has been carried out with the aim to verify the sustainability of this instrumentation by potential users. In this context, in fact, experts’ interest has focused on the evaluation of the costs of errors made during the
application and the potential employment of this information in the architecture model (Altman, 1984; Johnson et al., 1993; Chen et al., 1997; Gaber, 1986). This burden represents an undoubted aspect of notability in the estimation of the cost of a diagnostic model but does not allow for an exhaustive discussion of the problem of evaluation of the effects produced by the application of this instrumentation.

In addition to the economic impact, of which the cost of error represents a relevant, but not the only element, the organizational consequences have to be analyzed. These effects concern managerial complexity, system bureaucratization, and logistical superstructure due to the adoption of the model (Madonna and Cestari, 2012). These aspects must not be underestimated because they could affect, even in a decisive way, the effective usability of the forecasting models. Therefore, the operators that intend to avail themselves of the information obtained in the application of this instrumentation must not limit themselves to examining and appreciating its diagnostic performance. In fact, this forecasting ability should not be considered as a pre-eminent peculiarity but it should be reasonably devised and pondered through considerations inherent in the effective usability of this instrument.

In choosing the potentially more suitable model in the features of one’s own operational reality, therefore, the analysts have to carry out an evaluation process that compares the forecasting effectiveness and the economic and organizational sustainability degree of the instruments offered in literature and the operative practice. That being said, this study intends to offer an initial contribution to the existing literature by formulating some observations on the economic and organizational sustainability of the forecasting models to predict bankruptcy. Furthermore, research proposes to define, with a greater degree of certainty, the parameters able to endorse the diagnostic effectiveness of this instrumentation.

**The identification of effectiveness parameters**

It is important to clarify the effectiveness criteria as this allows for the evaluation of only the technical aspects of the model, in other words the diagnostic performance demonstrated in the classification of companies under examination as being either anomalous or virtuous.

The technical performances (effectiveness parameters) of a diagnostic instrument refer to:

1. the accuracy/reliability, i.e. the capacity to correctly predict the operative status of the firm under investigation;
2. the timeliness, or rather the speed with which the correct information on the state of health of the studied company is available. As a matter of fact, it is clear that an instrument results as more effective the more precise the evaluation obtained of the real operative status of the firm.
is, and the quicker this correct information is available to the analyst. In particular, timeliness is a decisive factor in this class of instruments. In fact, the forecasting models must not be limited to providing a punctual diagnosis but they must attempt to formulate a founded hypothesis on the state of health that the firms will have in the mid-long term. Therefore, this information must be acquired as soon as possible: to be notified of the company crisis well in advance allows the internal or external analyst to have sufficient margin of time to make more appropriate and more efficient decisions (Giannessi, 1979).

In order to determine the parameters capable of measuring the effectiveness of predictive models we have relied mainly on the contributions of Aziz and Dar (2006). In their work, experts have compared the reliability of 89 diagnostic instruments elaborated adopting different theories, sampling approaches and methodologies. In comparing the reliability of models, experts have examined three fundamental parameters: 1) the percentage of correct classifications for the year immediately prior to the critical event being investigated by each predictive instrument; 2) errors of the first type (classification of a company in crisis as being healthy); 3) errors of the second type (classification of a healthy company as being in crisis). These concern criteria usually taken into consideration at the time when the models are drawn up in order to assess specific performance.

Nonetheless, we think that the aforementioned parameters, while being important, are not enough to give a complete level of accuracy to the predictive instruments. There are two main reasons that bring us to this conclusion. Firstly, Aziz e Dar only compared the predictive performances stated by the authors without any verification of the percentages of success in applying the instruments onto new samples of analysis. It is a well known fact that when the studies on predictions of companies in crisis are repeated on samples other than those under study, the results obtained are less satisfactory. (Plat and Plat, 1990). Therefore, in addition to these parameters we have deemed it important to consider a few parameters that aim at stating an opinion on the potential generalizability of the models (Grice and Ingram, 2001). In short, this term indicates the capability of predictive instruments to show they are reliable also in relation to time and space, other than those under assessment. Furthermore, the contribution by Aziz and Dar does not take into account parameters of time, i.e. the times within which the information on the state of health of the companies are made available. The authors in fact, only state conclusions of a general nature, stating that the examination of accuracy one year prior to the critical event investigated (failure, bankruptcy, insolvency, distress, etc.) is not sufficient to state a judgement of the predictive ability of the models. Indeed, it is understood
that the further one moves away from the pathological event, the more the reliability of the instrument diminishes. This having been said, in order to appreciate the effectiveness of the prediction instruments, over and above the reliability criteria, we have also taken into account parameters that state timeliness in the diagnosis of critical states. Figure 1 shows a diagram of the identified effectiveness criteria, which will be dealt with in the sub-paragraphs that follow.

**Fig. 1 – The effectiveness parameters**

![Diagram of effectiveness parameters](image)

**Source:** authors’ own elaboration

**The reliability parameters of models**

According to what has been stated above, we think that the reliability of the predictive models must be examined in relation to five fundamental criteria:

1) the overall percentage of correct classifications;
2) the percentage of first type errors (when the model classifies as healthy a critical firm);
3) the percentage of second type errors (when the model classifies as critical a healthy firm);
4) the possible application of the model to a control sample;
5) the possible application of the model to subsequent samples compared to the elaboration one.

The first criteria (the overall percentage of correct classifications) estimates the ability of the model to correctly classify the sample under examination considered as a whole. Therefore, the observation of this
percentage allows us to appreciate the overall reliability of the diagnostic model. Moreover, the two parameters that follow (the percentage of the first type errors and the second type errors) highlight the instrument potential in correctly classifying the classes of, respectively, the “anomalous firms” and the “healthy firms”. So these measures, unlike the one above, allow us to estimate the specific reliability of the model as the assessment is circumscribed to a subgroup of the sample under examination.

As easy as it may be to understand, this is an indirect evaluation because it follows from the examination of errors made in classifying the two classes of companies. Consequently, the analysis of first type errors allows us to appreciate the performance of the model in diagnosing the status of the anomalous companies. The analysis of second type errors permits us to draw analogous but diametrically opposed considerations: in fact, in this case the potential of the instrument in qualifying the physiological companies, can be deducted.

For an accurate evaluation of forecasting models, a combined appreciation of both overall and specific effectiveness parameters had become unavoidable. Indeed, only the total percentage of correct classifications, while being a relevant criterion, could bring to the formulation of approximate and, sometimes even misleading judgments. In particular, this occurs when the specific analysis highlights, for one of the company groups, a percentage of correct classifications close to zero. In this case it is easy to understand that the effectiveness of the model would be neutralized because the discriminant activity between the virtuous and the distressed companies regresses. As a matter of fact, the model could similarly qualify almost all the firms.

The fourth and the fifth evaluation criteria (model application to a holdout sample and to subsequent samples compared to the elaboration one) intend to verify the subsistence of validations of the degree of accuracy of the model on different samples relating to the estimation one. It is known that the predictive capability of an instrument is strengthened and confirmed by the application of diagnostic models on control samples (Jones, 1987). Nonetheless, as mentioned by Aziz and Dar (2006), there are numerous studies that lack this step to check the degree of reliability (only 46% of models investigated by experts, for example, used a holdout sample to verify diagnostic capability). It occurs frequently that the diagnostic reliability is observed by applying the instrument to the same sample studied during the elaboration process. It is clear that the verification of the tool potential through the mere re-application to the group of companies that was observed in order for the model to be formulated allows us to obtain undoubtedly positive results, which are generally overestimated. Therefore, in this case the accuracy could only be ascribed to the capability of the model to represent in detail the
managerial features of a few sampled companies. With considerable certainty, the same instrument would be able to guarantee a similar diagnostic performance if extrapolated and applied outside the estimation context. In fact, the features of firms could be affected, even deeply, by the contingencies characterized by the development of the surrounding economic environment. Consequently the spatial and temporal stability of the models – the so-called “robustness” test (Haber, 2005) – could depend on factors such as: the structural changes in the economic cycle, the variations of the inflation rate and the conditions of monetary permissiveness. Despite the congenital limitations in the composition of control samples (Grice and Ingram, 2001), in evaluating the real forecasting abilities of the model it is necessary to extend the verification of specific performances even to control samples, that would be derived from the same estimation sample or an ex novo formation sample, or to groups of firms referring to different historic periods compared to the estimation one.

**The timeliness parameters of models**

With regards to the appreciation of the characteristic of timeliness, the examination of two parameters has been considered interesting:

1) the critical event investigated;
2) the number of reliable years of the model.

It is the definition of failure that requires the appreciation of the first of the two evaluation criteria (investigated critical event). Karels and Prakash (1987) state that, within the field of predictive models, there is no uniform definition of this phenomenon. Many experts interpret failure as synonymous with bankruptcy or liquidation. Others extend the phenomenon also to the concept of financial distress or insolvency. In any event, failure is not identifiable in a specific episode, where the occurrence thereof can qualify the firm as critical, but in a process of progressive worsening of the business performance. This regression is marked by the succession of several sequential events characterized by an increasing degree of seriousness (Argenti, 1976; Sharma and Mahajan, 1980). It follows therefore that there is a need to immediately recognize the critical state, even before the disease becomes irreversible.

While recognizing the dynamic nature of the crisis, for the elaboration of the model, it becomes essential that the degenerative process as a whole be traced back to a specific event that unmistakably identifies the condition of corporate distress. This is clearly a simplification from which, however, practical advantages can be obtained. Recognizing the pathological condition in a specific critical event allows us to definitely qualify the firms as anomalous or healthy simply by confirming, in each one, the existence or absence of the investigated fact. In this way, it is believed that any recognizable event of a pathological phenomenon can be properly diagnosed.
Nevertheless, in practice, events rendered formal by a legal judgment relating to the final stages of involutional dynamics, are preferred. With these events the differences between physiological and distressed firms are definitely accentuated and easy to appreciate. For this reason, the authors usually elaborate models to predict corporate bankruptcy, an event that usually precedes a company break up. There are, however, numerous models that have been elaborated with the aim of predicting the corporate disease events prior to a legal judgment. Therefore, the survey of econometric models relating to the investigated event, is profoundly heterogeneous. In judging the timeliness of this instrumentation, this aspect must not be overlooked: it is clear that a model that intends to recognize in advance the early moments of involutional dynamics must be considered more well-timed than a tool that forecasts its last steps. In fact, in the first case the internal or external users would have a greater amount of time to make more appropriate decisions on “corporate investments”.

The second parameter (number of reliable years of the model), instead, allows us to verify and evaluate the number of years before the investigated event, in which the model proves to be reliable in classifying the operative status of the firms. Numerous studies have highlighted the predictive accuracy of the model one year before bankruptcy. Nonetheless, some instruments have been shown to correctly predict bankruptcy significantly in advance. For example, the Deakin model (1972) has been shown to be 96% reliable already two years prior to bankruptcy. The Dwyer model (1992) has been able to diagnose the critical state of 97% of sampled companies three years prior to the critical event investigated. And further, the Hennawy and Morris instrument (1983) is 100% reliable already five years prior to bankruptcy. It is clear that the models able to produce correct information on the state of health of companies in good time, must be evaluated positively. This statement is valid should the instrument be used by external subjects (banks; courts of law; auditors; public administrations; etc.) or internal subjects (managers). The former in fact, will have the time to take the decisions they deem to be more appropriate in relation to the company under investigation. The latter instead will be able to assess, once the reasons for the critical state have been identified, the existence of conditions to initiate a reorganisation process.

The identification of efficiency parameters

The efficiency criteria allow for the estimation of sustainability of the model from the point of view of the economic operators. The assessment must be formulated by observing two different aspects:

1) the organizational sustainability that involves the implications for the organizational structure of an economic operator following the choice of an econometric model;
2) the economic sustainability that measures the cost of the adoption and/or the employment of the diagnostic instrument.

The observation of these parameters in relation to the instruments, characterized by a satisfactory degree of reliability, allows us to appreciate their operational usability. It is important to clarify that the sustainability of a model must not be estimated in absolute terms: it will, in fact, depend on the specific features of the user. It is clear that more onerous models, from an organizational and economic point of view, would be employed only by the companies that have an adequate business.

The search for efficiency criteria has been the largest phase of the research. There have been very few studies that have compared models outside the, by now traditional, effectiveness parameters (overall reliability; errors of the first type; errors of the second type). In regards thereto Lace and Koleda (2008) presented a study which compares fourteen diagnostic models taking into consideration four classes of parameters: 1) credibility of information; 2) factor completeness; 3) complexity of calculations; 4) effectiveness of results.

Within these categories there are criteria which aim at expressing judgements on the correctness of the structure of the model and the variables contained therein as well as the proven predictive accuracy. Nonetheless, albeit marginally, the authors also consider parameters which highlight the organisational sustainability of the models. Examples of this are: accessibility of the information required; mathematical expertise of the analyst; expertise in business process; expertise in accounting; labour intensity of calculations.

Research by Lace e Koleda, however, pinpoints two fundamental limitations: 1) there is no systematic and autonomous examination of the efficiency parameters in regard to efficiency criteria and 2) the economic sustainability of models is not taken into account. This latter aspect is essential in the assessment of usability by potential users.

The approach taken to determine efficiency criteria attempts to fill these blanks and aims at highlighting, in an orderly and systematic manner, the parameters that can measure, on the one hand, the organisational sustainability of diagnostic instruments, and on the other, the economic sustainability. Figure 2 shows a diagram of identified efficiency parameters that will be investigated in the continuation of negotiations.
The organizational sustainability of models

With regard to organizational sustainability, in keeping with the study by Lace and Koleda (2008), it is interesting to focus on three criteria:

1) the need and degree of complexity of recording systems for determining the parameters to feed the model;
2) the need for a specific qualified figure in the organizational structure;
3) the degree of interpretative complexity of the results of a diagnostic model.

The first parameter of comparison (need and complexity of recording systems) focuses on analyzing the effects of the introduction of the model in the corporate instrumentation in terms of intricacy of the information systems to support. This is mainly a result of the more or less detailed outline of the logical and technical framework of forecasting models. The complexity of this could depend on several factors. Among the main elements, the most important are: the number of variables provided, the degree of sophistication of the methodology adopted, the detail of the data used and the nature of the information to be acquired (quantitative, qualitative, accounting, non-accounting data).

It is clear that the need for data not provided in the ordinary corporate information systems requires the activation of a process of acquiring and/or
processing from scratch, of this information. This process often requires the adoption of new and more sophisticated recording systems. Moreover, the use of a specific predictive model may also involve the introduction of an appropriate figure in the organizational structure of the user (second parameter of comparison). The mind, specific professional qualities, intuition and the experience of the analyst may considerably affect the results of the assessment of the conditions of corporate vitality. For this reason, the use of some models requires that economic operators have a solid knowledge of business economics, mathematics and statistics and even engineering and computer studies. The word “use” has to be understood in its broad sense, or rather not only as a mere application of the instrument but also, and especially, as management, updating, remodelling and maintenance over time. Indeed, some models are characterized by a static logical-technical framework, for which further interventions of adaption and modernization are not needed. Nevertheless, others have a very dynamic structure and, therefore, require more radical correctives or less, in order to maintain the same degree of predictive reliability.

Finally, our attention must focus on the degree of complexity following the interpretation of the results obtained from the diagnostic instrument (third parameter for comparison). As everybody knows, the forecasting models formally define the parameters against which the information obtained may be assessed. Therefore, the application of this tool allows us to appreciate the status of the company investigated by minimizing (if not even eliminating) the subjectivity of the analyst’s evaluation. This determinism may involve a tightening up of the interpretative process, so much as to make it uncomfortable during its usage. Obviously, this depends on the type of model adopted and, in particular, on the methodology used for processing the data entered into the analysis process.

**The economic sustainability of the models**

The economic sustainability of diagnostic models – in the writer’s opinion – can be assessed by examining two classes of costs:

- **direct costs**, i.e. the economic sacrifices made during the acquisition and introduction of the diagnostic model in the users’ analysis tools;
- **indirect costs**, i.e. the advantages resulting from future employment of the model.

Different evaluation parameters can be identified for each of these classes. More precisely, in the direct charges it is interesting to observe the costs of:

1) elaboration or acquisition;
2) maintenance;
3) support instrumentation;
4) management.
The first evaluation criteria examines the economic sacrifice made by the user to internally elaborate or to externally acquire the econometric model. These costs refer to the complexity implied by the instrument: the high degree of articulation of the model and the employment of highly sophisticated technologies involve greater monetary investment than simpler supports, elaborated through traditional methods.

Once acquired, the instrument may require the incurring of costs for maintaining, updating, reshaping the logical-technical framework (second parameter of assessment). These expenses may be minimal or nonexistent for the models characterized by a “fixed or static structure”. These instruments do not require sudden and frequent modifications to be adapted to environmental and time changes. Several observations can be drawn for the most dynamic instruments, or rather for models that: a) propose to update its structure in considering the changes in time and space that have occurred before the application; b) need to have a continuous influx of new information in order to maintain a satisfactory degree of diagnostic reliability.

The introduction of a predictive model usually requires incurring costs for the acquisition of software and hardware that may support its functioning (third parameter of assessment). Once again, the complexity of the model influences its costs: complex models require more sophisticated and, therefore more expensive instrumentation; meanwhile opposing considerations can be formulated in the case of more traditional tools.

Finally, the costs ascribed to the management of the models have to be examined (fourth parameter of assessment). This expense mainly follows from the organizational implications deriving from the introduction of a forecasting instrument. In more detail, the evaluation parameters examine the increase of labour costs due to the introduction of a new figure in the business management of the firm or for the periodic training of the analyst already in the company.

Over and above the indirect costs resulting from the use of forecasting instruments, the assessment should consider the following parameters:

5) costs for the retrieval of information required by the model;
6) costs associated with the stabilisation of the organizational structure;
7) costs associated with errors made by the model.

As explained previously, the complexity of the elaboration models adopted may have implications in terms of the reduction of organizational flexibility. This decrease in flexibility is due to the adaptation of information systems and/or the implementation of the organization chart of the user. On the one hand, this generates an increased rigidity of the analysis procedures of the conditions of vitality of the firms. On the other, it generates important
economic repercussions (fifth and sixth parameter of assessment). If one thinks of the investments required for the research, one thinks of the selection, processing and verification of information, in order to feed the forecasting model or of the costs resulting from the possible embitterment of the decisional procedures of the user. This increase in complexity may ensue from the need, derived from corporate leadership, to subordinate decisions to the application of the predictive instrument. The need arises from the purpose of verifying, by means of results obtained from the model, the effects of these strategic choices on business performance.

Finally, the costs correlated to the type of errors made by the model must not be underestimated (seventh parameter of assessment). It is easy to understand that to classify a distressed company as healthy (first type error or false negative) generates different costs compared to a situation in which the model qualifies a firm as critical (second type error of false positive). In short, a false negative is more expensive than a false positive. Even when this is known, it is not possible to have a precise estimation of the size of the gap of the costs of the error type made. This falls within the specific tasks of the analyst. Therefore, for credit institutes the costs of first type error are a result of losses generated from the non (or partial) return of the loan granted to the company that has become insolvent. On the other hand, the losses implied by a first type error must be evaluated in terms of opportunity cost, therefore a comparison of the losses from the relinquishment of proceeds on a loan that has not been granted and the profit obtained by using alternative forms of investment.

For the firm using the models with an internal monitoring purpose, costs for the second type error can be related to unnecessary costs coming from the activation of unnecessary recovery processes. Vice versa, the economical consequences of second type errors can sometimes be critical: for instance in the case of amplification of annual losses and the consequential effects on the assets and financial points of view if the critical status is not recognized in good time and its causes are not definitively removed. The functional differences of the users and the concrete difficulties in the measurement have practically discouraged study in the area of evaluation of costs of classification errors. At present, the greatest contribution to this field of study has provided an estimate of the costs of first and second type errors obtained by examining the behaviour of 58 American institutes of credit in their ordinary work of granting of loans (Altman, 1977). The analysis found an average cost for second type error of 70% on the granted financing. The first type error, on the other hand, was estimated as being 2%. Therefore, in comparing these percentages it emerges that the cost of an incorrect classification of a critical company is 35 times higher than an error made for the healthy firms.
Effectiveness and efficiency of bankruptcy prediction models: an initial application

In this paragraph we will deal with an initial application of the aforementioned effectiveness and efficiency parameters on some of the main bankruptcy prediction models found in literature. We refer specifically to: Altman's Z Score (M1); Alberici's discriminant function (M2); Altman's quadratic Zeta (M3); the reduced decisional tree of Altman, Frydman and Kao (M4); Forestieri's bayesian model (M5); the neural network (15, 6, 1) of Altman, Varetto and Marco (M6); Altman and Sabato's model based on logit analysis (M7).

This analysis – which we will call a “pilot research” – will set forth initial thoughts and assessments on the sustainability and reliability of prediction models for potential users. While the study has been conducted on a limited number of instruments, we think that this application nonetheless highlights a few interesting points that could be further developed and proven in subsequent research. The results obtained are summarised below.

Concerning the assessment of effectiveness parameters, the comparison was done taking into account the data obtained from an examination of contributions, where the experts outlined the main specifications and the results drawn from the respective diagnostic models. This therefore concerned the comparison of the “declared effectiveness” by the same authors in the articles on the presentation of the models. Figure 3 summarises the results of said comparison.

<table>
<thead>
<tr>
<th>Model application to original sample</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall correct classifications</td>
<td>95%</td>
<td>85,7%</td>
<td>92,85%</td>
<td>91,5%</td>
<td>93,7%</td>
<td>87,3%</td>
<td>87,22%</td>
</tr>
<tr>
<td>First type error</td>
<td>6%</td>
<td>14,3%</td>
<td>5,7%</td>
<td>8,62%</td>
<td>25,9%</td>
<td>9,50%</td>
<td>11,76%</td>
</tr>
<tr>
<td>Second type errors</td>
<td>3%</td>
<td>14,3%</td>
<td>8,6%</td>
<td>8,45%</td>
<td>3,6%</td>
<td>15,90%</td>
<td>27,92%</td>
</tr>
<tr>
<td>Model application to a control sample</td>
<td>Yes</td>
<td>Not</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not</td>
</tr>
<tr>
<td>Model application to subsequent samples</td>
<td>Yes</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Investigated critical event</th>
<th>bankruptcy</th>
<th>bankruptcy</th>
<th>bankruptcy</th>
<th>bankruptcy</th>
<th>Opening credit recovery procedure</th>
<th>insolvency</th>
<th>bankruptcy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reliable years</td>
<td>2 years.</td>
<td>5 years</td>
<td>5 years</td>
<td>1 year</td>
<td>2 years</td>
<td>3 years</td>
<td>1 year</td>
</tr>
</tbody>
</table>
We focused on the analysis of parameters that could show diagnostic reliability and the timeliness of bankruptcy models.

Concerning prediction reliability, the first parameters we examined were the overall correct classifications, i.e. the overall percentage of correct predictions. The table shows how all the models gave more or less positive results. Indeed, four models out of seven (M1; M3; M4; M5) had correct assessment percentages that exceeded 90%. The other three models examined showed a degree of overall reliability that varied from 85.7% (M2) to 87.33% (M6).

As regards errors committed, the models M1, M3, M4 and M6 show a lesser percentage than the first, which level out at less than 10%. The highest percentage of this type of error was given by M5 which showed that, effectively, this instrument classifies one company in four as being healthy, while the same company is actually in distress. For second type errors however, the models M1, M3, M4 and M5 are more effective as the classification errors committed were less than 10%. On the other hand, in view of this parameter, it is M7 that gives the worst model as it incorrectly classifies 27.92% of healthy companies as being critical. Naturally observations on the importance (and consequently the assessment) of the different types of errors made must be modified in relation to the accompanying costs these entail. As often mentioned previously, first type errors are undoubtedly more costly than second type errors. Nonetheless these costs cannot be calculated in a single and absolute manner but rather they depend on: 1) the specific nature and sensitivity of the user (namely banks, rating companies, public administrations, managers, etc.); 2) the specific context of use and 3) the specific purposes for using the model. This having been said, it is important for the user to have available the fragmentary and specific data (percentage of errors committed, in types) in order for it to be possible to distinguish and assess the different types of errors and relative consequences.

The fourth and fifth reliability parameters analysed (application to control sample and to sample groups subsequent to the one used to elaborate the data) are to verify if the prediction models used are able to guarantee a suitable flexibility of use (versatility). Indeed, models are often conceived in such a way as to maximise prediction reliability with reference to a specific original sample, therefore it is evident that they are particularly accurate as regards that specific area. Nonetheless it is not unusual that these lose prediction effectiveness when used outside their original context. For this reason it is useful to test prediction instruments (and therefore have proof of the degree of reliability) on samples other than the original one from which the data are drawn. In this regard we must point out that only Altman’s Z Score (M1) responds positively to both parameters, while there are some
(M2 e M7) that have not been tested outside their area of elaboration (neither with control samples nor with samples subsequent to the ones from which data were drawn).

The assessment of the timeliness of the prediction, as mentioned previously, depends mainly on two parameters. Firstly, we must consider the critical event under investigation which can be distributed along the entire progress of company crisis: clearly the earlier the symptom we are trying to predict is manifested, the sooner we can have the diagnosis. Secondly we must assess the level of reliability of the model with the modification of the time variable; i.e. the ability of the model to offer correct predictions even long in advance compared to the critical event under investigation. In this regard the comparative analysis of models allows us to observe that there are some (M2 and M3) which, while showing great diagnostic prematurity, do so in reference to a symptom that is late in manifesting (and therefore serious) in the process of the progressive degeneration of the crisis. There are then some other models (M5 and M6) that show they are able to predict an early symptom with a significant margin of forewarning. The others show varying levels of contradiction regarding this element of assessment.

As regards effectiveness parameters, we think that it is appropriate to mention that, to date, there are no documented attempts aimed at “quantifying” organisational and economic sustainability of the predictability models drawn up thus far. Therefore our contribution will not be exhaustive on this matter but it aims at being an incentive for further study. In other words this is an attempt at measuring the parameters not yet emphasised by experts in the sector. As illustrated in previous paragraphs the parameters for organisational and economic sustainability are mainly quality elements that do not easily fit in with numerical quantifications. For this reason the effectiveness measurement phase has led us to call on a panel of users of the model. We submitted a questionnaire to the panel and, on the basis of the initial results that emerged, we began assigning the assessment to the selected parameters. This clearly concerns a research in the initial stages, but which has already produced some results. The results of this initial analysis are stated in figure 4.

---

129 This concerns an assessment begun in January 2010 which provided for the submission of a questionnaire to a sample group of eighty users (including banking houses, trade associations and public administrations) in the area of Emilia Romagna and Lombardy (in northern Italy).
As stated previously, we have decided to focus on the effectiveness of models by referring to two sets of parameters: those relating to organisational sustainability and those relating to economic sustainability, in reference to direct and indirect costs connected to the use of the models. Initially we will observe the criteria examined and assessed in reference to the first class (organisational sustainability).

In reference to the first parameter (necessity and complexity of appropriate IT systems), the models drawn up use the technique of multivariate discriminant analysis (linear or quadratic, i.e. M1, M2 and M3) shown to be far less costly, while subsequent ones (from M4 to M7) are characterised by far more demanding requirements, namely those using a neural type of technique (M6).

The need to make use of a suitably qualified professional figure is the second parameter under consideration. In this sense, one can see, once again, that models based on a multivariate discriminant analysis are characterised by less demanding requirements. On the other hand other models -

---

**Figure 4 – The efficiency comparison**

<table>
<thead>
<tr>
<th>Need and complexity degree of recording systems</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Middle</td>
<td>Middle</td>
<td>High</td>
<td>Middle</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Need for a specific qualified figure in the organizational structure</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not in high Middle High High High</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>Middle</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpretative complexity degree of the results</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not in high High High High</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

**ECONOMIC SUSTAINABILITY PARAMETERS**

<table>
<thead>
<tr>
<th>DIRECT COSTS</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elaboration/acquisition costs</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>Middle</td>
</tr>
<tr>
<td>Maintenance costs</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>Not</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Support instrumentation costs</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Management costs</td>
<td>Not</td>
<td>Not</td>
<td>Not</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INDIRECT COSTS</th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>M7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costs for the information retrieval required by the model</td>
<td>Low</td>
<td>Low</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Cost associated to the errors made by the model</td>
<td>4,26</td>
<td>10,30</td>
<td>4,16</td>
<td>6,203</td>
<td>18,20</td>
<td>6,97</td>
<td>8,7904</td>
</tr>
</tbody>
</table>
essentially those adopting a neural-type technique – require complex systems of *ad hoc* measurement and appropriate professional figures.

Lastly we assessed the degree of complexity in the analysis of results offered by the prediction model. In this case, as for the previous one, much depends on the specific properties of the model used: in some cases the strong determinism of the technical or engineering techniques at the basis of the models implies the absence of flexibility in the interpretative process, with resulting problems in use which are easy to imagine. In this regard nonetheless we must acknowledge that the majority of models do not present particular problems, if one excludes those based on neural-type techniques or logit analysis (M6 and M7).

As regards economic sustainability, direct costs were examined first. The first of these parameters (realisation/acquisition costs) is closely connected to the complexity of the specific instrument: very refined models, that manage a considerable amount of data, according to complex and "customised" algorithms with procedures that concentrate on various steps normally require more important monetary investments than more simple instruments that are based on traditional methods. Similarly to what has just been stated in reference to the complex interpretation of results, it is necessary to consider that the majority of analysed models does not offer any particular problems in this respect, with the exception, in this case also, of those based on neural type techniques or on logit analysis (M6 and M7) which instead require rather more important investments. Maintenance costs are instead connected to the necessary activities of revision, periodic calibration, updating, re-modulation of the logic-technical systems. These costs seem to be negligible when the structure of the model is essentially static (namely models that are based on multivariant discriminant analysis – up to M3 – and the Forestieri model – M5) while becoming extremely important for more dynamic diagnostic instruments which, in order for them to function correctly, require on-going and important re-elaborations, especially the one using a neural-type technique (M6). It is also important to take into consideration costs relating to support instruments (namely hardware and software). While these are not normally costly investments, it is also necessary to point out that the models that are based on multivariant discriminant analysis have more limited management costs. The same can also be said in reference to the last parameter taken into consideration, i.e. costs attributable to the management of models, mainly referring to acquisition costs and/or the training of staff necessary for the management of prediction instruments: models that are based on the multivariant discriminant analysis are characterised by being less costly to use.

Lastly we examined indirect costs resulting from the use of prediction models. The need to acquire information necessary for the operation of the
instrument (first parameter) can, in some cases, lead to the adoption of detection systems that can be complex and complicated, taking up time and attention at various levels of organisation. Models that are based on the multivariant discriminant analysis, being operated with data already available, do not usually cause problems which unfortunately characterise those models that are based on neural-type techniques or logit analysis (M6 and M7).

In the second parameter analysed (rigidity of the organisational structure) we can see that the most complicated and sophisticated models usually determine a cumbersome organisation, at times rather serious. These concern not only the data collection and processing phases but also the consequent fluidity of the decision-making process, attributable to the slowing down necessary to allow for the analysis of results and more simply, the complexity of processing management decisions, to assess a growing number of variables with due care, these not always being clear and easy to interpret. In this regard it is necessary to state that to date the survey has not yielded satisfactory results: as regards the estimate of costs relating to cumbersome organisational structures, we have not yet, in this phase of work, identified a mechanism that would enable us to achieve a reliable measurement. We have therefore preferred not to expose ourselves by putting forward unsatisfactory predictions, with the intent of better defining this aspect in future developments on this matter.

Lastly, we have considered costs connected to the fallacy of various models (the last parameter of assessment). The parameter in question is obtained as a function of the probability of error of each individual model and the costs attributable to the specific category of error. The table clearly shows that only Altman’s model Z score is associated to tolerable levels of cost error.

In summary, in reference to prediction effectiveness, we must point out that the models under examination show (exception made for M7 and M2) a satisfactory level of diagnostic reliability accompanied by a limited percentage of first type error. What is surprising is the absence, in two out of the seven models under examination, of application in control samples and, in six models out of seven, of verification on sample companies at a time later than the assessment. Reservations arise with reference to the stability of the rule of classification in time (verified with the application on samples at a

---

130 Concerning the assessment of costs connected to classification errors we have drawn on what emerged from research conducted by Altman in 1977 (see previous paragraph). This study was conducted on a sample of 58 regional and provincial banks and showed an average cost for first type errors of 70% of the premium granted and for second type errors around 2%. The overall cost of errors for the model, therefore, was established on the basis of: \( x_1 \times 70\% + x_2 \times 2\% \). Where the variable “\( x_1 \)” is the percentage of first type errors committed by the model; and the variable “\( x_2 \)” is the percentage of second type errors committed by the model.
later date) and space (assessed with the application on samples relating to the same time frame as the one in the elaboration of the model) in the absence of verification on control groups. In the absence of such stability, the models would not be effective in decontextualised applications compared with the area of assessment and thus be unusable in operation.

In reference to effectiveness instead, it appears that the models elaborated with what we can call «complex» techniques (neural networks and genetic algorithms) are difficult to sustain from an organisational (due to implications concerning the IT systems necessary to support the use of the instrument and on the organisational structure of the user) and economic (due to high direct and indirect costs that accompany the same) points of view.

The decision regarding sustainability is clearly conditioned by the size and specific nature of the user: it is clear that it would be difficult for small companies – such as those characterising the Italian market – to economically and organisationally sustain models elaborated using particularly complex techniques.

**Conclusions and future developments of the research**

In this paper we have tried to develop some initial thoughts on the assessment/comparison parameters of prediction models for company crises, elaborated until now with a view to the end user. Once the assessment criteria have been envisaged, the user can knowingly select and adopt the instrument that is more appropriate to his own operative needs, in terms of technical and economical characteristics. In particular, we have deemed it appropriate to consider the performance of predictive instruments (effectiveness) but also, and mainly, the economic and organisational sustainability (efficiency) of the same. The efficiency analysis – in our opinion – represents a fundamental view of the possible uses of the diagnostic instruments within a company/profession. Indeed, models that are particularly effective but excessively costly and cumbersome for the organisational apparatus of the user would still be impossible to use. The effectiveness and efficiency parameters, therefore, must be examined within a systematic viewpoint. Naturally, the assessment of these criteria is characterised by a certain degree of subjectiveness as it depends on the specific characteristics and needs of the user.

As we have stated, the study wishes to offer some considerations, within the limits posed by an initial approach, regarding this particular view of model analysis. The proposed application, as stated in the previous paragraph, has pinpointed some relevant aspects of the “declared” effectiveness and efficiency of models as perceived by the users.

Concerning this latter aspect, we have seen that instruments prepared with more up to date techniques (neural networks and genetic algorithms) are
difficult to sustain by small-medium Italian users. The complexity that accompanies the use of said instruments is particularly demanding on the user’s organization as well as having considerable direct and indirect costs. What emerges – at least as far as a full application is concerned – is a type of trade-off between what has been researched in literature on matters of prediction models and what, instead, is required in a day-to-day operation. Even though in recent years researchers have moved towards the elaboration of instruments using increasingly sophisticated techniques, (largely based on engineering studies), what operators are asking for is the availability of reliable models that are sustainable at the same time. Only by respecting these conditions will models be used profitably on the field.

Naturally the proposed application has congenital limitations (in terms of size of the analysis sample) on the “pilot research”. It would nonetheless be interesting to understand if there is a real relationship between efficiency and effectiveness of prediction models. For this reason, in the future, we aim at extending the comparison mentioned in this paper with reference to a sample of diagnostic models (different in their elaboration approach: number and nature of variables examined and the quality-quantity ratio of the assessment samples) that is wider and statistically more important. This further development in the research would allow us to further validate or refute what is shown in the proposed application.

Acknowledgments

With grateful thanks to Prof. Salvatore Madonna (University of Ferrara – Italy).

References:


Haber J.R. Assessing bankruptcy prediction models are evaluated. *Journal of business & economics research*, vol. 3, no. 1, 2005.


Platt H.D., Platt M.B. Development of a class of stable predictive variables:


