ELSEVIER

Expert Systems with Applications 30 (2006) 507–518

Expert Systems with Applications

www.elsevier.com/locate/eswa

Assessing a knowledge-based approach to commercial loan underwriting

Roger Kumra [a,1], Roger M. Stein [b,*], Ian Assersohn [a,2]

[a] *Moody's KMV, Pilgrims Court, Reigate RH2 9BL, UK*
[b] *Moody's Investors Services, 99 Church Street, NY 10007, USA*

Abstract

We discuss the challenges in developing decision support tools for commercial underwriting and discuss how several different approaches to the underwriting problem have been addressed. We then describe an expert system-based approach to credit underwriting that has been in commercial usage for over ten years in a variety of financial institutions. The expert system approach addresses many features of the underwriting process that alternative approaches do not. The system is characterized by a functional representation of knowledge and a graph-based inference mechanism. The inference mechanism is unique in its pragmatic approach to the implementation of probability theory. This approach offers flexibility for modeling various aspects of real world credit decisions not always treated by traditional approaches. We give examples of how this approach can be and is currently being applied to facilitating underwriting decisions in commercial lending contexts.

Keywords: Credit risk; Credit scoring; Expert systems; Functional systems

1. Introduction

Credit risk assessment is a key component in the process of commercial lending. A potential borrower's credit assessment determines whether the borrower will ultimately be granted credit, and if so at what cost in terms of underwriting fees and interest rates. Although many methods exist for evaluating credit risk, for a number of lenders, risk assessment in the underwriting process is fundamentally different from the evaluation of credits within a portfolio due to differing levels of analytic detail and opportunity costs.

This paper presents the details of a knowledge-based system (KBS) tool for aiding in loan underwriting decisions in the commercial credit domain. The system is unique in that it can complement statistical models of default by including both quantitative information and the non-quantitative, often subjective, information that typifies the underwriting process at many institutions. This tool is used by over 60 banks and non-banks ranging from leading edge global institutions to smaller regional banks[3].

We also provide a comparison of expert system-based underwriting tools and other approaches. One component of this comparison is a dimensioning of the credit-underwriting problem along business and technical components. The second part of this analysis is a discussion of how well various alternative techniques (expert systems, statistical default models, and scorecard-type systems) conform to the dimensions of the problem.

1.1. Underwriting vs. portfolio management

There are typically two associated but nonetheless distinct credit-related functions in most financial institutions. One area is responsible for originating or underwriting individual obligations, such as loans, with a credit component; and the other is responsible for monitoring and optimizing the organization's overall credit exposure in a portfolio sense.

* Corresponding author.
E-mail addresses: roger.kumra@mkmv.com (R. Kumra), roger.stein@moodys.com (R.M. Stein), ian.assersohn@mkmv.com (I. Assersohn).
[1] Tel.: +44 1737 229916; fax: +44 1737 229900.
[2] Tel.: +44 1737 229908; fax: +44 1737 229900.

0957-4174/$ - see front matter © 2005 Published by Elsevier Ltd.
doi:10.1016/j.eswa.2005.10.013

[3] It is worth noting that at the time of this writing, the tool we describe in this article is in the process of being revised to accommodate various technological developments that have arrived since the publication of the original article (Kumra, Stein, & Assersohn, 2000). Among the innovations is a revision of the knowledge representation language (Syntel) described in this paper to a newer representation framework that more readily permits thin client execution. However, this can be viewed as an implementation detail as the core ideas in terms of knowledge representation, uncertainty and knowledge processing are all being incorporated into the new implementation.

In a portfolio analysis context, what is often of most interest to an analyst is the credit riskiness of an asset relative to its comparative return and whether the obligation, on average increases or decreases the overall credit exposure and/or profitability of the portfolio. In particular, a portfolio risk manager is concerned with such things as the aggregate credit quality and diversity of an institution's portfolio and how an individual credit may contribute marginally to this risk profile along certain dimensions (e.g. geographic concentration, industry concentration, asset class concentration, etc.). This type of analysis usually requires quantitative information about default expectations, recovery expectations, correlations, net profitability, etc. In addition, a portfolio manager often requires the ability to rebalance the portfolio by either removing certain types of assets through sale or securitization, or by purchasing additional assets.

However, the fact that the analysis must be conducted across the entire portfolio (or a segment of the entire portfolio) and this may involve thousands or tens of thousands of obligors, makes it imperative that the analysis be executed quickly and with a minimum of manual effort. In addition, since most portfolio management techniques require various types of optimization or simulation, it is usually desirable to describe risks in terms of objective measures that can be recalculated conveniently.

On the other hand, at the time of underwriting[4], the credit analyst's objectives are often different. This juncture is often when the most detailed analysis of a potential borrower is done. It is also the time at which the borrower is most eager to provide information and to make concessions in the structuring and security provisions of a loan. In many cases, analysts wish to rely on information beyond the purely quantitative data available in financial markets and accounting statements. In addition, current regulatory proposals require that all relevant information (qualitative and quantitative) be incorporated into credit assessments. This would necessarily include judgments made by the credit analyst. In particular, the 'international convergence of capital measurement and capital standards' (Basel, 2004) indicates that

> Although mechanical rating procedures may sometimes avoid some of the idiosyncratic errors made by rating systems in which human judgement plays a large role, mechanical use of limited information also is a source of rating errors. Credit scoring models and other mechanical procedures are permissible as the primary or partial basis of rating assignments, and may play a role in the estimation of loss characteristics. Sufficient human judgement and human oversight is necessary to ensure that all relevant and material information, including that which is outside the scope of the model, is also taken into consideration, and that the model is used appropriately. (Paragraph 417)

[4] As well as, to a lesser degree, at the time of review.

For this reason, there is often a requirement for tools that support a rich and detailed analysis of a borrower both from a quantitative and subjective perspective.

This activity also has benefits in terms of supporting relationship banking and allowing an institution to gain more insight into its client base. It is one form of due diligence and many institutions use this point to help reinforce best practices. They see the underwriting function as the core of the firm's credit culture and each firm typically has its own unique approach to this part of the credit process. Underwriting activity also allows the institution to gain a specific understanding of a borrower's business, its management, and its market. These factors are often key drivers of the credit quality but can be difficult to measure objectively, particularly in markets where pricing information is not available or accounting data is scarce and unreliable. Including qualitative factors in a statistical model is difficult, since they rely on the underwriter's skill in assessing the business and must thus be considered subjective.

A model that supports the underwriting process should, ideally, structure its analysis in a way that is similar to the way an experienced underwriter would look at a risk and assist the underwriter in gaining an understanding of the business. This is important since potential borrowers usually expect the underwriter to understand their business. In addition, the underwriter must appreciate the business and the model's description of the business well enough to identify and incorporate factors that are not captured by the model. While statistical credit models can be extremely accurate tools for identifying potentially risky counterparties (c.f., Arora, Bohn, & Zhu, 2005; Stein, Roger, Ahmet Kocagil, Jeff Bohn, & Jalal Akhavein, 2003) in terms of default probability, they are not always sufficient to address the additional special needs at the time of underwriting within some organizations and market segments.

1.2. Credit culture and the commercial underwriting process

Credit risk assessment in its purest sense is cast as an attempt to quantify the risk of loss to the lender when making a particular lending decision (Arnold, 1989). This loss may be a loss of principal or interest or both.

Two components affect the risk of loss; the risk of default by the borrower and the severity of the loss in the event of default[5]. The former is typically a function of the borrower's financial strength and the general viability of the firm. The later is often a function of the transaction being underwritten, including the value of the security held against the loan, the seniority of the lender's lien with respect to other lenders, etc. It can be argued that even if the risk of default is high, as long as the loss in the event of default is sufficiently small, then the risk of loss to the lender should be slight. However, not all

[5] Banks usually adopt one of two approaches to quantifying these risks: either the analysis is broken into two separate components, the probability of default and the loss in the event of default, or the bank attempts to quantify the expected loss (typically the product of the two) in a single analytic quantity. See (Treacy & Carey, 1998) or (Basel, 2000) for discussions of banks' internal rating systems.

loans can be adequately secured and the value of any collateral held may go down. In addition it may be seen as the lender's responsibility to lend sensibly.

Traditionally, bankers have assessed credit risk by analyzing the historical account data of their customers and projecting these into the future based on quantitative analysis and expert judgment. This is usually achieved by analyzing various financial ratios, calculated from accounting data. These show the relationship of account variables to each other and are used to assess the state of a business. Leverage and profitability ratios are frequent favorites for this type of analysis, as are cashflow-based measures. These latter, in particular, help a lender determine the business' ability to generate cash from its core activity and allows the calculation and analysis of key indicators that help the lender to identify factors which have affected cash flow in the past and therefore may well affect cash flow in the future (Stimpson, 1991).

In addition to comparing each ratio against some absolute standard, the ratios are often assessed comparatively as well. Peer comparison and cross-sectional analysis is common since, it gives an analyst a more complete picture of the meaning of particular financial variables with respect to the borrower's overall industry. The strength of the industry and the prospective borrower's position within that industry are obviously important to the risk assessment. Time-series analysis is often used as well since, it allows an analyst to evaluate how each ratio changes in time and whether there are specific trends of interest for a borrower's financial status[6].

In many cases, analysts also wish to layer in non-financial information about management, industry direction, etc. For example, a good management team may keep a business solvent even in very adverse conditions. The information used in these analyses is often of a subjective nature and specific to the borrower's industry. This information is sometimes combined with an analysis of other non-financial data to provide better overall insight into the state of the borrower's business.

All of these factors contribute to traditional credit underwriting. But there is more to the underwriting process than the calculation of an expected default frequency, important though that is. Relationship managers must get to know the client, develop a 'feel' for the business, by making site visits, talking to staff and so on. He or she must also be aware of the specific credit culture within the organization and work towards establishing a commercial advantage, whether this is based on price, quality or breadth of service or some other factor.

A fully-rounded decision support tool must also support these aspects of the Relationship Manager's job, making it easier both for them to record and justify his or her recommendations and decisions, and for the organization to develop good management information tracking systems that ensure consistency in the application of its credit policies.

In order to develop a system to provide decision support for underwriting, each of these issues must be considered and incorporated into the analysis. One way to accomplish this is through the use of a carefully constructed knowledge-based system, but there are others as well. Defining the technical and business dimensions for such a system, selecting an appropriate method that balances both the business and technical constraints, and implementing such a system in a manner that banks and other institutions find valuable is the subject of this article.

Importantly, the issue of the validity of assertions about the need and value of combining qualitative judgments with quantitative assessments is an open one. Traditional model validation techniques (c.f., Stein, 2002) are difficult since, unlike the case of a pure quantitative model, human judgments are exceedingly difficult to recreate ex post.

It is hard, for example, for an analyst to recreate accurately what his or her view of, say, the quality of the management of a firm, would have been three years earlier, particularly if that firm has had adverse credit events in the intervening years. In addition, it is often the case that staff turnover within institutions results in inconstant interpretation and application of subjective components of the credit process. As a result, retrospective studies of the performance of hybrid (quantitative and qualitative) rating systems are rare and must typically rely on sparse data sets. While we do not address this specific issue in this article, we note in passing that to date, the evidence here appears mixed. However, as more data become available retrospective studies should become more prevalent and should provide additional insight on this open issue.

In the remainder of this paper, we provide a set of criteria for evaluating systems for supporting the underwriting process. We describe an expert system designed to facilitate the analysis of financial and subjective data for underwriting decisions that we feel addresses the significant dimensions of the underwriting problem. It provides a framework for analysis of financial data that mirrors the traditional credit analyst's approach as well as supporting qualitative analysis and enforcing an organization's chosen credit culture.

The paper proceeds as follows. In Section 2, we briefly discuss the general technical and business requirements for a system to support this process. In Section 3, we provide a description of a knowledge-based approach that we have developed to solving this problem. In Section 4, we compare the approach with one based on the use of pure statistical models. The final section is a discussion of the tool's ability to meet the requirements described earlier.

2. Dimensions of a system for underwriting decision support

Decision support systems can be evaluated along a wide variety of dimensions ranging from the accuracy of a system's ultimate assessments to the ease with which its output can be explained, or the flexibility with which a system can be modified or customized[7]. The nature of the task to which the

[6] See, for example (Foster, 1984) for a discussion of common financial statement analysis practices.

[7] Dhar & Stein, 1997) gives taxonomy of over a dozen dimensions along which such systems can be evaluated and we adopt that approach here.

system will be applied and the organization within which it will be used affect the relative importance of these dimensions in determining the suitability of a particular system.

Interestingly, although accuracy in default prediction is the primary goal of most internal rating systems, it is only one of several that some organizations consider when implementing underwriting tools. In fact, as we discussed in Section 1.2, although accuracy is typically the most important feature, other dimensions, such as the transparency of the model, its ability to support relationship management and the application of a consistent credit culture can also be essential.

To describe the dimensions of a credit underwriting system, we start with the requirements for the most obvious outputs of the system. Important outputs of such a system are: assessments of a prospective borrower's financial status, assessments of its industry and assessments of the status and quality of the firm and its management. As described in Section 1, this involves an analysis of both financial (numerical) and subjective data. Thus, if analysts believe this process to be a valid one, a decision support system based on this approach should be able to simulate judgments using both numerical and subjective data.

An underwriting system must also be able to function in realistic environments where data may be incomplete, uncertain, or noisy. It is sometimes the case, for example, that the information available to a lender on a potential borrower is inconsistent or conflicting. This may arise benignly, as a result of the complex nature of the business environment, or it may be due to potential data errors or inconsistency in information reporting. In either case, an automated decision support system for underwriting should be able to alert the lender if there are any problems or inconsistencies with the case in question. In addition, since some data may be uncertain, the system should also be able to represent and express varying levels of uncertainty.

Financial institutions are often plagued with data integration problems as well. Mergers in particular have made it difficult for many institutions to standardize data collection and archival. As a result, it is often the case that it is impossible or impractical to obtain a complete set of answers to all the questions that are relevant. To address this, the system should be able to deal with incomplete data.

In addition to integrating data from many physical data sources, it is important to be able to integrate data on credit factors that are conceptually disparate. Credit analysis, as we have described it here, involves weighing a variety of complex factors and interactions. Many of these are not directly comparable in their raw form. For example: is a profitability ratio of 28% offset by a management assessment of 'poor'? How about 'very poor'? A system for underwriting needs to be able to integrate information from various sources and in various forms. It needs to be able to support complex inference about subtle and often non-linear relationships among this information.

Furthermore, since it is not practical for an automated system to query every possible factor that may affect lending decisions, some institutions believe that the analyst should be able to override system generated) assessments when they feel there are extenuating circumstances that have not have been captured. What this also implies is a need for users to be able to understand with a fairly good level of transparency the underlying model of the credit process and how that model is represented in the system.

Importantly, it is often difficult for institutions to assess the impact of allowing such overrides. As we discuss in the introduction, it is not straightforward to validate qualitative components of rating systems. However, even small difference in model performance has been shown to result in substantial differences in profitability (Stein, 2005). In this context, institutions must struggle with the trade-offs implied by allowing analysts to override systems.

Since, each institution has its own policies and underwriting practices it is not uncommon for firms to need to customize a system to account for differences in business lines or changing economic or regulatory environments. This allows experts within an institution to impart their experience and foresight into the system. Accordingly, a decision support system needs to be flexible and customizable enough to fit in with the business practices and environment of its users.

However, the system also needs to be robust enough to ensure that the institution using it is able to follow best practices. In cases where the current organizational practices are inferior to industry best practices, it is useful to have a system that can be adjusted to integrate with the credit culture of the organization while also bringing about the necessary process changes required for implementing rigorous credit processes. While such organizational overhauls can be wrenching, customization can ease the transition.

Finally, the system must be able to interface not only with end-users, but other systems in use within the institution.

These dimensions are summarized in the Table 1 below. The second column shows that an ideal solution should possess at least the level of support for each attribute shown.

We feel that we have developed a system that addresses these dimensions well. In the remainder of this paper, we discuss this system and analyse it both with respect to the underwriting problem and with respect to other potential solutions.

3. A knowledge-based approach to modeling credit expertise

The system we describe in this paper, Moody's KMV Risk Advisor™ (MRA), is a knowledge-based system (KBS) for

Table 1
Requirements for an ideal decision making system for underwriting support

Attribute	Ideal solution
Accuracy	High
Explainability	High
Tolerance for noisy data	Moderate
Tolerance for sparse data	Moderate
Tolerance for complexity	High
Response speed	Moderate
Flexibility	High
Embedability	High

supporting the underwriting process. A KBS is a computer program that encapsulates expertise, elicited from experts in the form of business concepts and the relationships between them. The expertise within MRA was compiled over more than decade by bankers and credit experts from a variety of institutions[8]. The system is implemented in a proprietary expert systems language called Syntel™.

In this section, we provide some high level technical details on how the system represents experts' knowledge, and how the knowledge is used to infer credit quality in the underwriting process (readers less interested in these issues can safely skip to Section 3.3, which discusses how the system appears to analysts. However, many of the features discussed the next section have implications to our analysis of the appropriateness of our approach to credit underwriting).

3.1. Preliminaries: a general overview of the structure of knowledge representation

In this subsection, we provide a brief overview of some of the fundamental concepts we use for knowledge representation in MRA. In the current version of this system, these constructs are implemented in the Syntel™ language. In subsequent planned releases most of these concepts will be transitioned into an updated implementation.

The structure of the knowledge base that forms the core of MRA is similar in many respects to earlier work done by its primary architects in the 1980s. The representation can be viewed as a non-procedural dataflow language and has a rich history in the expert systems literature (Duda et al., 1978, 1979, 1987; Duda & Reboh, 1984; Hart, 1989; Reboh & Risch, 1987; Risch et al., 1988; Kumra, 1996). At its core, MRA uses a representation scheme tailored to designing decision networks based on user input. The basic component of the structure is called a node. Nodes represent business concepts and a particular knowledge base's network architecture describes how these concepts relate to each other.

3.2. Structure of the expert knowledge encoded in MRA

Since, the representation of knowledge in MRA is a functional one, the standard dynamic expert system model of a RECOGNIZE-ACT[9] cycle does not apply.

Knowledge representation, as it is used for MRA, is best understood as a one-directional network[10] of decision-components (nodes) that are arranged hierarchically from the most general overall assessment of risk (i.e. borrower rating) to the most specific (financial statement values such as interest expense, etc. and user inputs). In the current implementation of the MRA knowledge-base, the main network forms a tree in which the raw data enter at the bottom and assessments are produced to illuminate key business concepts through the tree, culminating in the pinnacle (root) node[11], the borrower rating.

Working backward from the root (borrower rating) node, the lower layers of the network represent various sub-assessments (e.g. financial assessment, subjective assessment, etc.) with a similar structure: an assessment being derived from combinations of lower nodes. This pattern repeats itself recursively until a node is reached that represents an input value.

As an example, consider Fig. 1. It shows the various components of the financial assessment component of the MRA knowledge base. Note how each sub-component feeds into a more abstract concept within the knowledge base. Similarly, each of the sub-components shown is derived from more basic concepts, eventually starting with raw data. For example, the debt service coverage assessment in Fig. 1 can be further broken down to its base components as shown in Fig. 2, below[12].

The Syntel™ language used to create MRA supports a wide range of functions that can be used to represent expert knowledge. These include logical, arithmetic, statistical and aggregation functions as well as rules. Many of these are used to assess the inputs to MRA. For example, the assessment of all non-debt service ratios in MRA consists of a procedure that requires the calculation of several factors including the ratio's historic trend, its volatility and an estimate of how the value of the ratio under analysis ranks within its industry group.

In addition, Syntel™ supports a powerful function called WEIGHT that is used throughout MRA. WEIGHT supports a concise representation of expert knowledge. WEIGHT takes one or more nodes as inputs and calculates an assessment of those inputs. For each input a weight rule is coded. The weight rule specifies a mapping from the input's value to a numeric vote that constitutes the input's contribution to the WEIGHT node's value. The contributions of the inputs are summed and then converted into a utility value. This value is then mapped into the WEIGHT node's type to produce an output value. The rule is thus a type of lookup-table parameterized with expert knowledge.

[8] A key component of the knowledge elicitation process is the ongoing participation of a select group of users who provide feedback and suggestions based on their institutions' experiences with the system and the local credit environments.

[9] This is an algorithm in which a proposition is identified as fitting a rule pattern (RECOGNIZE) and then used to generate a hypothesis (ACT) which then itself becomes a proposition in the next cycle.

[10] In mathematical terms the network has the property of being a directed acyclic graph (DAG).

[11] This is not to say that the structure of the knowledge base is a simple pyramid. In fact, to switch metaphors, there are a number of side branches of varying degrees of importance which do not lead to the root, but which illuminate for the user certain aspects of the analysis. For example, one section of the knowledge base is devoted to presenting the user with a number of so-called advisories. These are messages generated dynamically by the system which draw the user's attention to inconsistencies of input, and anomalous or potentially concerning conditions within the analysis. (e.g. if other non-current assets exceed 10% of total assets, an advisory fires warning the user that liquidity could be impaired and that there is a possibility that intangible assets may have been misclassified.) Other side-branches might exist to re-display sub-assessments in a different way, such as a mapping to a bank's own rating scale. In some implementations, the borrower rating itself acts as input to further trees, which culminate in ratings for individual transactions.

[12] A more detailed discussion of the weights and form of the MRA knowledge base can be found in (Moody's, 2004).

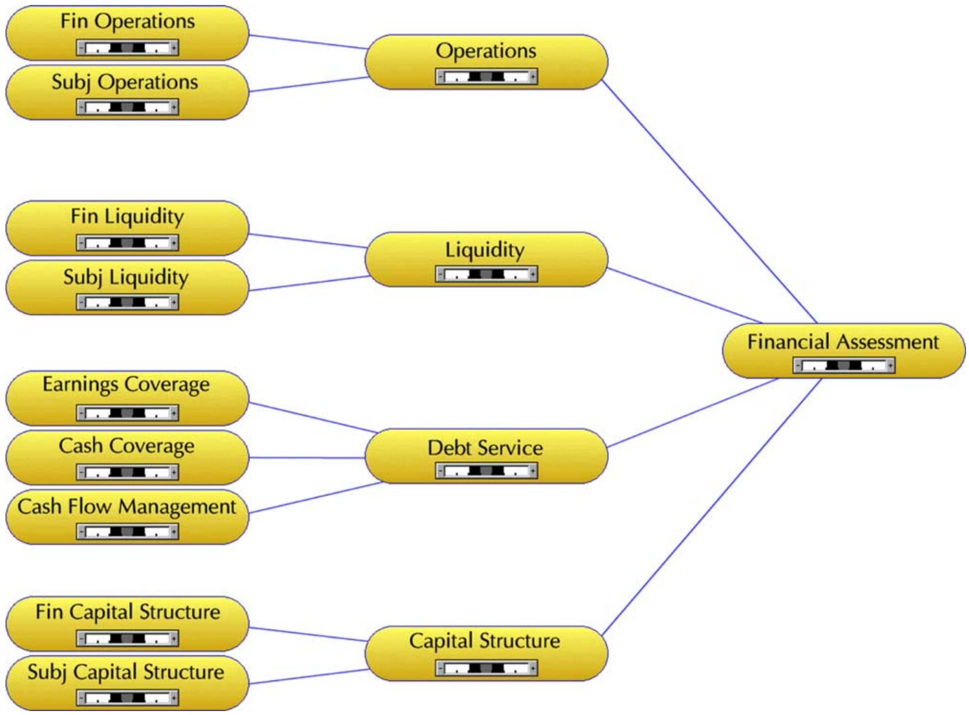

Fig. 1. The structure of the assessment network used to calculate the financial assessment. This figure shows how the various components of the financial assessment are combined. Sub-assessments for each component of the analysis are weighted and combined using expert rules. This representation makes the knowledge base more readily understandable to credit professionals and allows them to understand the dynamics of the model.

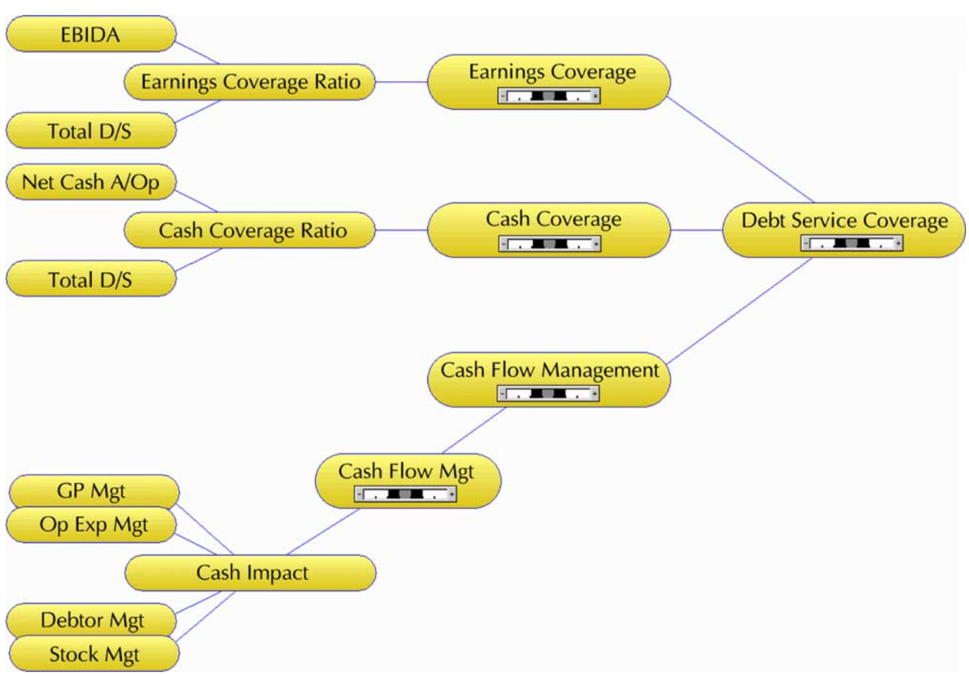

Fig. 2. The structure of the assessment network used to calculate the debt service coverage assessment. This figure shows a close up of the details of the components of the debt service sub-assessment pictured in. In some cases, the schematic terminates at an input such as EBIDA (upper left). In others, such as Stock Mgt., still lower levels exist below the one shown.

To illustrate the concept, consider a fictional quantitative input variable. If, as part of the management assessment, MRA had a question for the size of the senior management team, expert knowledge about the management team size could be encoded in a weight rule as shown in the Table 2 below.

Experts would first specify a series of fixed points with associated votes. If the rule feeds the management quality component, then the rule can produce positive and negative votes, which move the score for the Management attribute upward or downward. For example, a single manager would

Table 2
A weight rule

Fixed point	Vote
1	−20
5	40
8	0
15	−40

tend to reduce the score, while a small team of five would increase it, but a larger team of 15 would decrease it.

For values between the fixed points, the system interpolates votes linearly for intermediate values. For example, if the actual management team consisted of four individuals then the vote for this attribute would be 25[13].

From a practical perspective, this method is extremely useful since it means that industry experts can speak in terms they are comfortable with and these terms can be translated into Syntel™ and included in MRA.

In the next section, we give an example of the definition and application of a weight rule. These rules are used extensively in MRA and are especially useful for dealing with subjective inputs.

3.2.1. Handling non-numeric and subjective variables

In the previous section, we discussed conceptually how MRA aggregates quantitative information. Unfortunately, not all information that an analyst might wish to include in his analysis can be boiled down to simple ratios. What about things like management quality or industry competitiveness? How should these business concepts be addressed?

In order to reconcile the often incompatible scales (e.g. ordered classes, unordered-discrete, continuous, etc.) and orders of magnitude (e.g. dollars and ratios or dollars and 'GOOD MANAGEMENT') of different inputs, all assessments in MRA are internally represented as utilities (in the economics sense of measures of goodness). The utilities are mapped back and forth between discrete and continuous representations on arbitrary scales to facilitate understanding by the user.

For example, consider the case of the concept 'management quality'. Most lenders would agree that this is an important variable but would have difficulty describing it quantitatively. Rather they might discuss it in terms of the relative merits of different aspects of the management team and assign these a qualitative value based on their opinions. Fig. 3, below illustrates MRA's approach.

To transform a subjective concept into one amenable to aggregation with other quantitative data, the function WEIGHT, mentioned above, is used. At the lowest level the user is presented with a list of categories to describe each input. For example, if we take the management character branch of the tree, the user must perform two evaluations: commitment and integrity. For commitment the user must choose between:

[13] $(4-1)/(5-1)\times(40--20)-20=25$.

Table 3
A weight rule specifying the impact of commitment on the assessment of management character

User's answer	Votes
Very high	24
High	12
Average	0
Low	−12
Very low	−24

Table 4
A weight rule specifying the impact of integrity on the assessment of management character

User's answer	Votes
Outstanding	20
Acceptable	0
Questionable	−150

'very high', 'high', 'average', 'low' or 'very low' and for integrity the user has the options: 'questionable', 'acceptable' or 'outstanding'.

To evaluate these two inputs and form an assessment of management character MRA uses two weight rules in which the votes are chosen specifically to ensure the assessments match the experts' views. The first weight rule specifies the impact of commitment on management character. Table 3 below specifies this weight rule[14].

Fig. 4 illustrates the second weight rule, which specifies the impact of integrity.

Here, the interpretation is that negative values reduce the overall score for the management character and positive ones improve it. Note that the spacing of the votes for Integrity is unequal, thus illustrating one way in which non-linear relationships are represented in the system.

For example, moving from 'acceptable' to 'questionable' is significantly more costly than moving from 'outstanding' to 'acceptable', even though in both cases, the move was from one value to the next lower one (Table 4).

The answers to commitment and integrity are mapped into a utility value that forms the management character assessment. This is achieved by summing the associated votes, applying a sigmoidal function to this value and incorporating any uncertainty associated with the assessment.

The sigmoidal function serves two purposes: it ensures that the utility value is bounded and also incorporates the concept of diminishing marginal returns[15]. Uncertainty arises from two areas. There is an associated uncertainty with the

[14] Note that this weight rule is used as an example only and does not represent a view on the interpretation of this attribute.

[15] The system also adds a value of 50 to the raw utility value. The sigmoidal function transforms the range $(-\infty,\infty)$ into (0, 100) so that the center value of the assessment is 50. Note that in this example we are representing the assessment as a set. It is also possible to represent the assessment on a continuous scale.

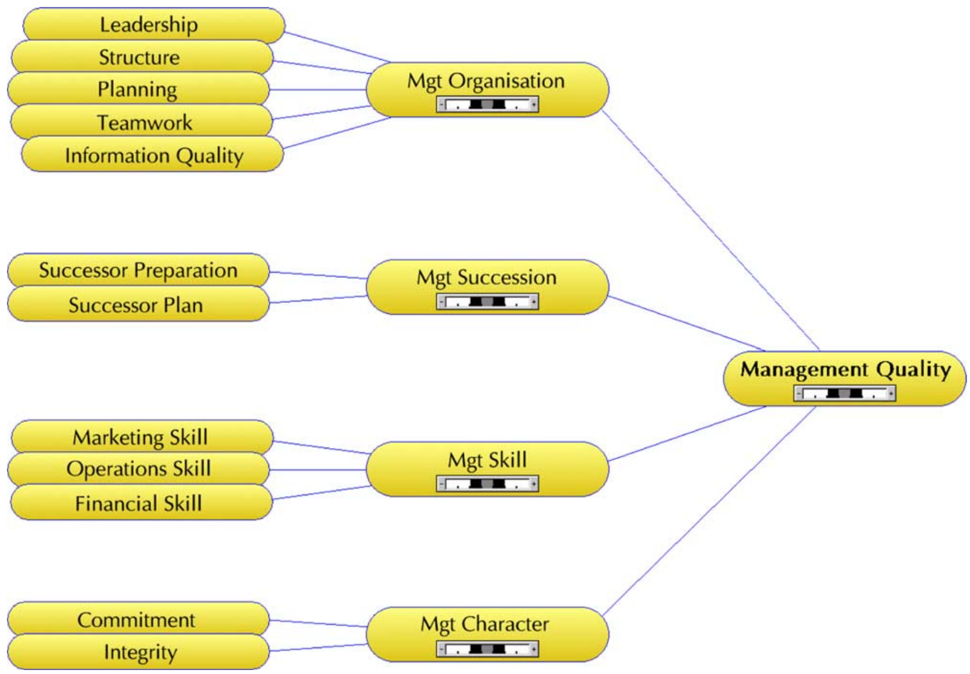

Fig. 3. The structure of the assessment network used to calculate the management quality assessment. MRA can incorporate both subjective and quantitative data. The assessment of management quality shows schematically how subjective elements are combined. Management quality concept is broken down into four sub-components: organization, succession, skill and character and these in turn are broken down further. MRA takes the approach that such complex values are very difficult to assess with a single input. Instead, the management quality concept is broken down into four sub-components: Organization, succession, skill and character and these in turn are broken down further. This simplifies the user's task as less needs to be considered when making an evaluation for each input and improves the consistency of such subjective assessments as management quality; for each borrower the user will consider the same factors and these will be evaluated in a consistent manner.

assessment; answers to two questions cannot be considered to determine the value of management unambiguously. In addition, if one of the questions is unanswered the associated uncertainties will increase[16]. The value of the assessment can then be mapped directly into a utility set as is illustrated within Table 5[17].

As mentioned, some assessments can have associated uncertainty. This is represented as a spread value and this means that an assessment can take values in more than one category. The system calculates a certainty value that the 'assessment' is within each category and ensures that the sum of the certainty values is one. For example the value of Management Character could be described using the values shown in Table 6.

Because of the mapping of all variable types to a common representation, the WEIGHT function can aggregate knowledge for inputs of different types. These can be both symbolic and numeric[18]. If any of the inputs are uncertain, WEIGHT will map that input's distribution across its weight rule to produce an appropriate contribution to the WEIGHT node's output. In this way the WEIGHT function can produce an assessment from a set of inputs that resembles compensatory reasoning. At the same time, it can compute the degree of uncertainty that is present. A tuning tool is provided which assists in the creation of appropriate weight rules[19].

3.3. Interacting with the analyst: the user interface

MRA uses the concept of a form as the basis for the user interface. A form may contain a number of input fields, output fields, buttons, text fields, buttons and boxes. The user navigates to any visible form at any time. This gives the user control over the system.

To display assessments graphically, MRA uses the concept of a meter. A meter shows graphically where a particular assessment falls with respect to the full range of possible values. In addition to the assessment, the meter shows the accompanying uncertainty associated with the assessment. An example of a meter is shown below in Fig. 4.

The black band shows the most likely value of the assessment based on the input data and the rules in the knowledge base. The grey bands show values that are possible but less certain, under the analysis. The spread of black

[16] Note also that any uncertainty associated with an input to an assessment will feed into that assessment. For example, the uncertainty associated with management commitment will increase the uncertainty associated with the management quality assessment.

[17] The apparent overlap in these figures is due to rounding.

[18] With the proviso that symbolic types are categorised and ordered.

[19] Note that to produce similar functionality using crisp IF-THEN rules would typically take many rules due to the continuous nature of the concepts being represented. This type of rule reduction is common between crisp and fuzzy domains.

Table 5
A utility set

Category	Utility value
Excellent	85.7–100
Very good	71.7–85.7
Good	57.1–71.7
Average	42.9–57.1
Below average	28.6–42.9
Poor	14.3–28.6
Unacceptable	0–14.3

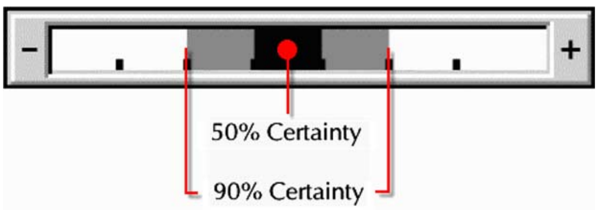

Fig. 4. A meter. Meters provide a graphical representation of the value of a particular assessment or variable in the knowledge base.

and grey bands gives an indication of the uncertainty in the assessment[20]; if the meter displayed a single black band and no grey ones the assessment would be more certain. If more than one band is colored black then the assessment value is fairly uncertain. By convention, the more the meter tends to the right the better the assessment. Hence, the meter shown in Fig. 4 indicates this to be an about average assessment.

MRA provides a feature called an alert that fires a message when a particular condition occurs. These are useful to warn users of situations that may require extra consideration, for example dangerous customer occupations or contraventions of normal practice.

Users can override system calculations and thereby influence the system's assessments further up the node-network. In addition users can 'footnote' specific fields. These are useful to document additional information relevant to a field, for example why a user has overridden it.

In the current implementation of MRA, the user may also examine the combined effect of all of the information. In the screen shot shown in Fig. 5, for example, the user can see the results of the expert system's analysis of the financial strength of a borrower, broken down into sub-components.

3.3.1. Providing explanations of underwriting recommendations

MRA provides a flexible reporting mechanism based on a technology similar to the mail merge function found in word processors. This enables business reports to be coded using a mixture of static text and dynamic node values.

The same mechanism is used to provide the in-built explanation facility that lets the user examine the structure of the analysis. A hierarchical set of reports, chained together using hyperlinks, lead the user through the analysis from the final assessment down to the input data, providing a natural language commentary on the meaning of and reasoning behind, each sub-assessment.

[20] The concept of 'certainty' here is the same as used in the expert system context (a measure of confidence based on the knowledge base), rather than 'confidence' in the statistical sense.

Table 6
An example of how certainty values are represented within MRA

Category	Certainty
Excellent	0.000199
Very good	0.340732
Good	0.621283
Average	0.037598
Below average	0.000069
Poor	0.000096
Unacceptable	0.000023

3.4. Parameterizing the knowledge base to reflect organizational credit culture: the knowledge elicitation and tuning process

Up until this point we have presented the mechanics of how knowledge gets defined, synthesized and abstracted in MRA. However, we have not yet discussed how the original knowledge in the model, or knowledge base, is created or how it is customized by organizations to reflect more accurately their own credit practices.

An expert system is only as good as the quality of the expertise it encodes. The expertise within the generic version of MRA is monitored and maintained by a committee of experts drawn from its user base. However, the system is generally also altered for each client organization to reflect the unique features of its approach to credit.

The process of customization is referred to as Tuning because it usually includes, although it is not necessarily limited to, fine-tuning of the weights feeding the assessments and sub-assessments. Because all weights are externalized into database tables, this process can be accomplished through a suite of tuning tools that assist the organization's experts to intelligently update the knowledge base.

Because raw weights are unintuitive to most users, the tools must present the user with a more abstract representation of the structure. Thus they can quickly set up default weights by specifying factors such as the relative importance of the input factors, and whether a positive answer should weight more or less heavily than a negative one. They can then fine-tune using the actual weights until the output meter gives appropriate results.

Commonly, in addition to changing weights, the client may wish to alter labels, add or remove subjective questions or change the list of values from which the user can select.

On occasions, however, more drastic changes need to be made to the knowledge base. In this case, a traditional knowledge elicitation exercise takes place involving interviews, storyboards, prototyping and the full panoply of knowledge engineering techniques.

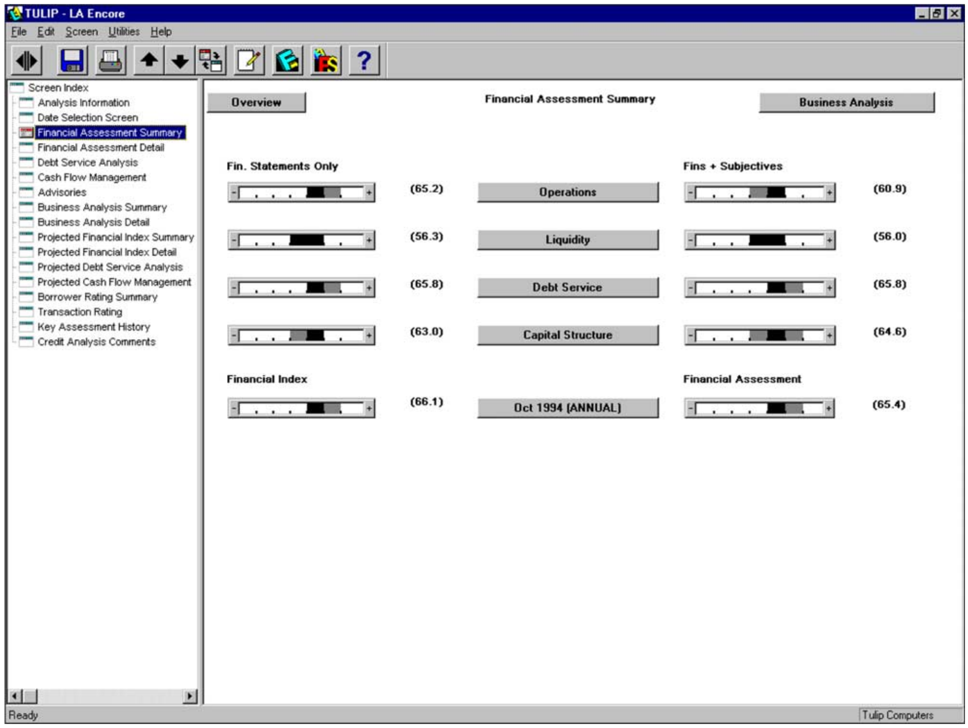

Fig. 5. Financial assessment summary screen. This screen is an example of the MRA interface. It shows the system's assessments for operations, liquidity, debt service and capital structure based purely on financial data and the same assessments based on both financial and subjective factors.

3.5. Summary of Moody's KMV risk advisor

In this section, we have described Moody's KMV risk advisor (MRA), an expert system developed in close consultation with industry lending experts over a period spanning almost a decade. The current version of the system was implemented in Syntel™, an expert systems language that allows the representation of complex and subtle business concepts in a straightforward manner. Part of MRA's power derives from its extreme flexibility which allows users to directly modify the underwriting models that are used by the system by augmenting the knowledge base with additional rules or by adjusting the weightings within the existing knowledge base. The knowledge base itself is easy to interpret and consists of sets of modular relationships that fit together to derive an overall assessment of creditworthiness. The representation seeks to encapsulate best practice in underwriting and allows the incorporation of both qualitative and quantitative data into the analysis.

4. Analysis of the fit between the underwriting task and MRA

4.1. Statistical models

A common alternative modeling scheme for credit analysis is the use of statistical models of various sorts. These models can be potentially very powerful for predicting default and are often optimized to do this specifically. For this comparison, we consider only the state of the art in statistical models to rationalize our comparisons.

Statistical models are typically designed to predict default or to predict agency ratings. The most sophisticated of these take advantage of modeling techniques that control for data problems and non-linearity as well as the complexity of interactions (c.f., Dwyer, Kocagil, & Stein, 20040). Nonetheless, many statistical models deal poorly with missing data. Although well designed quantitative models can provide a level of explanatory functionality for their outputs in the form of driving factors, marginal effects, etc. for the input variables, they do not typically give deep insight into the credit process through these tools.

Most quantitative models are derived through statistical optimization and are thus not amenable to ad–hoc adjustment. Rather, this needs to be done by re-optimizing the model using new data. The compact formulaic representation of statistical models makes them ideal for embedding in other systems and they are computationally efficient which gives them very fast response speed.

Table 7
An assessment of the use of statistical methods

Attribute	Ideal solution	STAT	MEETS
Accuracy	High	High	✓✓
Explainability	High	Moderate	
Tolerance for noisy data	Moderate	Moderate	✓
Tolerance for sparse data	Moderate	Low	
Tolerance for complexity	High	MOD–High	✓
Response speed	Moderate	High	✓✓
Flexibility	High	Low	
Embedability	Moderate	High	✓✓

The biggest advantage to using a high quality statistical model is the accuracy of its predictions, which is typically high compared to alternatives (including expert systems). Also, the objective nature of statistical models makes them well suited to use as benchmarks for transactions between firms.

Table 7 provides an overview of the relative merits of using statistical approaches to credit underwriting.

4.2. Scorecards

Like statistical models, scorecards provide a framework based on the assessment of a small number of financial ratios. The mathematical model used for scorecard construction is typically parameterized either based on expert judgment or statistically. Like some statistical models, scorecards are often constrained to consider (additively) a small number of factors without interactions. Scorecards have been very successful in high volume domains (consumer lending, revolving consumer credit, etc.) where there is a significant number of data records on borrowers, a large degree of uncertainty in the predictability of default, and a few indicators can provide sufficient information on the likelihood of default. This when combined with the very large number of borrowers in typical portfolios, and the very small relative exposure per borrower creates a less demanding set of requirements for model performance. The main distinction between scorecards and more involved statistical models is one of accuracy and sophistication.

Typically, scorecards require the user to provide a series of inputs. Each input is assigned a value and these are summed to result in a score. The institution uses the score to help determine whether to lend to the borrower. Table 8, below, provides a summary of this discussion.

4.3. Summary

Summarizing the above analysis, some clear patterns emerge. We discuss these in the paragraphs that follow and give an overview in Table 9.

As we have defined the underwriting problem, particularly in situations where historical data are not available and market prices do not exist or in which including subjective factors is a requirement, MRA is the favored of the three approaches. Its major weakness is its somewhat lower accuracy and relatively slow response speed. This is more a function of the time required to input the necessary data than of the computation time. However, the reward for undertaking this detailed data gathering effort is that it provides much deeper insight into the drivers of credit than other approaches and allows much more flexibility for organizations to shape it to reflect their internal standards and credit culture.

Table 8
An assessment of the scorecard approach

Attribute	Ideal Solution	SC	MEETS
Accuracy	High	Moderate	✓
Explainability	High	Low	
Tolerance for noisy data	Moderate	Low	
Tolerance for sparse data	Moderate	Low	
Tolerance for complexity	High	Moderate	
Response speed	Moderate	High	✓✓
Flexibility	High	Low	
Embedability	Moderate	High	✓✓

Table 9
A comparison MRA, statistical models and scorecards

Attribute	Ideal solution	MRA	STAT	SC
Accuracy	High	✓	✓✓	✓
Explainability	High	✓		
Tolerance for noisy data	Moderate	✓	✓	
Tolerance for sparse data	Moderate	✓		
Tolerance for complexity	High	✓	✓	
Response speed	Moderate		✓✓	✓✓
Flexibility	High	✓		✓
Embedability	High	✓	✓✓	✓✓

Interestingly, using our analysis of the underwriting problem, we can also draw conclusions about the other two approaches. If quick, objective, high precision default prediction is the primary goal then other techniques cannot compete with statistical models, which clearly distinguish themselves. This is often the requirement for trading, securitization, portfolio management applications as well as for use as an input into the underwriting process. On the other hand, if complex modeling is not required, but flexibility, simplicity and speed matters, scorecard approaches suggest themselves. These are often the requirements for higher volume, lower exposure underwriting applications.

A significant consideration, and currently an open issue, is the relative performance of systems that do incorporate subjective factors relative to those that do not. It is clearly reasonable to assume that a lender that enjoys a privileged relationship with a specific borrower will likely have access to credit information not readily available to most market participants (and thus not reflected fully in market prices or broad accounting statements). In such situations, a highly experienced analyst may be able to use this information and his or her own knowledge to extend analysis beyond ‘what the numbers show’ and achieve higher levels of precision in the analysis of the individual firm.

What is less clear, however, is whether this advantage is one that applies to most borrowers in an institution’s portfolio being analyzed by the typical analyst. That is, on balance, does the advantage the institution gets from applying overrides and judgments in cases where the information provides discriminatory power exceed the costs of overriding consistent quantitative assessments in cases where there may be no such additional information content?

Because the evidence on this question is still mixed, many institutions opt for hybrid approaches in situations where there is sufficient data. Although they take a number of forms, such approaches typically are structured so that analysts’ judgments are constrained to move a particular internal rating only a notch or two away from the output of a quantitative model. In other variants, quantitative models are embedded in more flexible systems that permit analysts to adjust the inputs to the models based on their expertise. Of course, in situations where market data or historical information is not available, these approaches are not viable and pure expert systems provide one of the only reasonable approaches.

5. Conclusions

We have described an expert system for credit underwriting called Moody's KMV Risk Advisor™ (MRA). MRA has been in use in large financial institutions for many years. The system is implemented using knowledge representation that is rich enough to perform both numerical and subjective analyses and powerful enough to combine these into composite assessments for the analyst.

MRA allows an analyst to represent business expertise in a manner with which they are familiar using standard business concepts combined and weighted together using expert judgment. The system treats uncertainty in a consistent manner. It can present its evaluations in a form that shows graphically both its assessment and the uncertainty associated with it. MRA also provides support for handling incomplete information and thus can still perform analyses when some information is unavailable.

Importantly, MRA allows experts within an organization to modify the knowledge base, thus permitting the system to more closely conform with and support the internal credit culture and best practices of the firm. It is equally important to note, however, that this flexibility generally precludes the outputs of the system from being used outside the organization. The very attributes that allow extensive customization of the knowledge base for specific credit environments prevents two organizations from being able to objectively use the measure as a basis for transactions since they cannot use the (differently) customized systems as a common basis for comparison.

Overall the knowledge base technology described has demonstrated itself to be well suited to support the credit risk assessment process. In particular, by supporting weighting and interpolation functions in addition to standard functions such as slope and trend, it provides a representation sufficient to model the expertise required. Also, the flexibility of the product means that it can be tailored to become an integral part of the institution's lending procedures.

References

Arnold, J. H. (1989). Assessing credit risk in a complex world. *Commercial Lending Review*, *4*(3).

Arora, N., Bohn, J. R., & Zhu, F. (2005). *Reduced form vs. structural models of credit risk: A case study of three models*. San Francisco, CA: Moody's KMV.

Basel. (2000). Range of practice in banks' internal ratings systems. Basel: Basel Committee on Banking Supervision.

Basel. (2004). Basel Committee on Banking. International convergence of capital measurement and capital standards (a revised framework). Basel: Bank for International Settlements, 2004.

Dhar, V., & Stein, R. (1997). *Seven methods for transforming corporate data into business intelligence*. Englewood Cliffs, NJ: Prentice Hall.

Duda, R., Gaschnig, J., & Hart, P. (1979). Model design in the prospector consultant system for mineral exploration. In D. Michie (Ed.), *Expert systems in the micro-electronic age* (pp. 153–167). Edinburgh: Edinburgh University Press.

Duda, R. O., Hart, P. E., Nilsson, N. J., & Sutherland, G. L. (1978). Semantic network representations in rule-based inference systems. In D. A. Watterman, & F. Hayes-Roth (Eds.), *Pattern-directed inference systems* (pp. 203–221). New York: Academic Press.

Duda, R. O., Hart, P. E., Reboh, R., Reiter, J., & Risch, T. (1987). Syntel: Using a functional language for financial risk management. *IEEE Expert*, *2*(3), 18–31.

Duda, R. O., & Reboh, R. (1984). AI and decision making: The PROSPECTOR experience. In W. Reitman (Ed.), *Artificial intelligence applications for business* (pp. 111–147). Norwood, NJ: Ablex.

Douglas, W. D., Kocagil, A., & Stein, R. M. (2004). *The Moody's KMV EDF RiskCalc v3.1 model: Next-generation technology for predicting private firm credit risk*. New York: Moody's KMV.

Foster, G. (1984). *Financial statement analysis*. Englewood Cliffs, NJ: Prentice-Hall.

Hart, P. E. (1989). Syntel(TM): An architecture for financial applications. In H. Schorr, & A. Rappaport (Eds.), *Innovative applications of artificial intelligence* (pp. 62–70). Melno Park, CA: AAAI Press.

Kumra, R. (1996). *A functional network to assess credit risk*. London: Intelligent Systems for Finance and Commerce.

Moody's KMV. (2004). Guide to business analysis, Moody's KMV: Reigate.

Reboh, R., & Risch, T. (1987). Syntel (TM): Knowledge programming using functional representations. AAAI-86: Fifth national conference on artificial intelligence: Philadelphia, PA.

Risch, T., Reboh, R., Hart, P. E., & Duda, R. O. (1988). A functional approach to integrating database and expert systems. *Central American Common Market*, *31*(12), 1424–1437.

Stein, R. M. (2002). *Benchmarking default prediction models: Pitfalls and remedies in model validation*. New York: Moody's KMV.

Stein, R. M. (2005). The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. *Journal of Banking and Finance*, *20*(5), 1213–1236.

Stein, R. M., Kocagil, Ahmet. E., Bohn, Jeff, & Akhavein, Jalal (2003). *Systematic and idiosyncratic risk in middle-market default prediction: A study of the performance of the RiskCalc and PFMTM models*. New York: Moody's KMV.

Stimpson, D. (Ed.). (1991). *Global credit analysis: Moody's investor's service*. London: IFR Publishing.

Treacy, W. F., & Carey, M. S. (1998). Credit risk rating at large US Banks. Federal Reserve Bulletin (November).