$\mathcal{I}^2\mathcal{SDS}$
The Institute for Integrating Statistics in Decision Sciences

A Logical Analysis of Banks' Financial Strength Ratings

Peter L. Hammer
RUTCOR
Rutgers University

Alexander Kogan
Rutgers Business School
Rutgers University

Miguel A. Lejeune
Department of Decision Sciences
The George Washington University

A Logical Analysis of Banks' Financial Strength Ratings[1]

Peter L. HAMMER[2,*], Alexander KOGAN[3,*,**], Miguel A. LEJEUNE[4,***]

Abstract: We evaluate the creditworthiness of banks using statistical, as well as combinatorics-, optimization-, and logic-based methodologies. We reverse-engineer the Fitch risk ratings of banks using ordered logistic regression, support vector machine, and Logical Analysis of Data (LAD). The LAD ratings are shown to be the most accurate and most successfully cross-validated. The study shows that the LAD rating approach is (i) objective, (ii) transparent, and (iii) generalizable. It can be used to build internal rating systems that (iv) have varying levels of granularity, and (v) are Basel 2 compliant, allowing for their use in the decisions pertaining to the determination of the amount of regulatory capital.

Keywords: credit risk rating, bank operations, bank creditworthiness, Logical Analysis of Data, combinatorial pattern extraction

1. INTRODUCTION

Information-intensive organizations such as banks have yet to find optimal ways to exploit the increased availability of financial data (de Servigny and Renault, 2004; Wang and Weigend, 2004). Data mining and machine learning, in particular statistical (Jain et al., 2000) and combinatorial pattern recognition (Hammer and Bonates, 2006), provide a wealth of opportunities for the credit rating and scoring field, which lags behind the state-of-the-art methodological developments (Galindo and Tamayo, 2000; de Servigny and Renault, 2004; Huang et al., 2004). In this paper, we use the novel combinatorial and logic-based techniques of Logical Analysis of Data (LAD) (Hammer, 1986; Crama et al., 1988; Boros et al., 2000; Alexe et al., 2007) to develop credit risk rating models for evaluating the creditworthiness of banks.

The objective of this paper is to reverse-engineer the Fitch bank ratings to produce an (*i*) *objective*, (*ii*) *transparent*, (*iii*) *accurate*, and (*iv*) *generalizable* bank rating system. By the *objectivity* of a rating system we mean its reliance only on measurable characteristics of the rated banks. By its *transparency* we mean its formal explicit specification. By the *accuracy* of a rating system which is based on a widely used existing (proprietary and opaque) rating system we mean the close agreement of its ratings with those of the existing system. By its *generalizability* we mean its accuracy in rating those banks which were not used in

developing the system.

In this study, we shall: *(i)* identify a set of variables which can be used to accurately replicate the Fitch bank ratings; *(ii)* generate combinatorial patterns characterizing banks having high ratings and those having low ratings; *(iii)* construct a model to discriminate between banks with high and low ratings using combinatorial optimization techniques and the identified patterns; *(iv)* define an accurate and predictive bank rating system using the discriminant values provided by the constructed model; *(v)* cross-validate the proposed rating system.

This study reveals the weakness of the results obtained with multiple linear regression and support vector machines, and shows that ordered logistic regression can provide excellent results in reverse-engineering a bank rating system. These results are not only matched but exceeded by utilizing the substantively different methodology of LAD, whose results, while remarkably similar, are shown to be much more robust compared to those given by ordered logistic regression. In view of the essential differences in techniques, the conformity of bank ratings provided by LAD and ordered logistic regression strongly reinforces the validity of both the obtained results and of these rating methods.

It is worth noting the additional advantages provided by the LAD approach. First, the LAD credit risk model, as opposed to the ordered logistic regression one, does not assume that the effect of the variables used as predictors is the same on each bank and on each bank rating category. This feature is particularly relevant, since Kick and Koetter (2007) have shown that the credit risk importance of banks' financial structure differs across bank rating categories. Second, while the ordered logistic regression approach can be used only to construct a rating system that has the same number of rating categories as the benchmarked system, the LAD approach can generate rating systems with varying granularity levels: *(i)* a binary classification model to be used for the pre-approval operations; *(ii)* a model with the same granularity as the benchmarked rating model; *(iii)* a model with a different granularity than that of the benchmarked rating model, to allow the bank to refine its pricing policies and allocation of regulatory capital.

We show that the LAD-based model cross-validates extremely well, and therefore is highly generalizable. Thus, this approach can be used by financial institutions to develop Basel-compliant rating models. The accuracy of the predictions is a particularly strong achievement in view of the opaqueness of

the financial sector (Morgan, 2002). Financial institutions are highly leveraged and hold assets (e.g., structured financial securities), the risks of which fluctuate significantly and are very difficult to evaluate. Moreover, the proposed model is a cross-country one (see Table 1) whose prediction accuracy is verified for financial institutions spread across the world. In the literature, as reported in Arena (2008), most bank rating and failure models have been developed for a particular country (Germany: Kick and Koetter (2007), Czech Republic: Derviz and Podpiera (2004), Turkey: Canbas et al. (2005), Brazil: Barnhill and Souto (2008)) or continent (Tabakis and Vinci (2002)), thus not allowing for comparison within a common framework.

The structure of this paper is the following. Section 2 motivates the development of internal credit risk rating systems. In Section 3, we present the data used in this study. Section 4 provides a general overview of the Logical Analysis of Data (LAD) methodology used for reverse-engineering bank ratings. Section 5 develops an LAD model for discriminating banks with high and low ratings and presents a procedure for mapping the LAD numerical values reflecting the creditworthiness of banks to credit risk ratings. Section 6 assesses the accuracy and generalizability of the LAD model. Section 7 compares the LAD approach with other classification methods. Section 8 describes the broader impacts of the forward-looking LAD credit risk rating system. Finally, concluding remarks are presented in Section 9.

2. MOTIVATION FOR DEVELOPING INTERNAL CREDIT RISK RATING SYSTEMS

Although external credit risk rating systems (i.e., developed by rating agencies such as Moody's, Fitch, S&P's) continue to play a fundamental role in risk management practices, banks have been allocating increasing amounts of resources to the development of more accurate and granular risk rating models (Jankowitsch et al., 2007). Some of the main reasons behind this trend are the following ones.

First, the work of rating agencies has recently come under intense criticism (Financial Stability Forum, 2008). Skepticism about rating agencies finds its root in their inability to spot some of the largest financial collapses of the decade (Enron Corp, WorldCom Inc, Lehman Brothers) which some observers attribute to the presence of conflicts of interest (Financial Stability Forum, 2008) induced by the issuer-pay model, in which the rated company pays the rating agency a fee for the issued rating. A report of the Securities and Exchange Commission (2003) indicates that 90% of Moody's and Fitch's revenues come from fees paid by the issuers. Moreover, in many instances, key credit analysts in rating agencies participated directly in fee negotiation with issuers. The SEC reports (Reuters, 2008) that often the rating agency employees structuring a financial security were the same who were rating the product! The issuer-

pay model creates moral hazard situations and has led to the so-called complacency effect, which makes raters, especially in a non-crisis period, very reluctant to downgrade an obligor (Schwarcz, 2008) and is viewed as one of the main reasons for the subprime financial crisis. The development of an objective accurate internal rating model would make banks less dependent on the whims of the rating agencies.

Second, increasingly intense international market environment and changes in the regulatory framework driven by the Financial Stability Forum and the Basel Committee on Banking Supervision (Basel II) called forth incentives for banks to improve their credit risk rating systems. While the Basel Committee initially (i.e., in the 1988 Capital Accord) favored the ratings provided by external credit rating agencies, the Basel II Accord (Basel Committee on Banking Supervision, 2006) is actually a risk-focused regulatory framework under which a bank can develop its own risk model to calculate the regulatory capital for its credit portfolio. Compliance with the Basel II Accord requires the internal risk models (rating, probability of default, loss given default, exposure at default) to be cross-validated and accepted by the legislator, and qualifies banks for the adoption of the Internal Rating Based approach (Basel Committee on Banking Supervision, 2001, 2006) to calculate their capital provisions. The Financial Stability Forum endorsed the principles for banking supervision defined by the Basel Committee of Banking Supervision (2006) as crucial for having sound financial systems and as deserving "priority implementation". The current financial turmoil has led the Basel Committee on Banking Supervision (2008), in line with the Financial Stability Forum's prudential oversight recommendations (2008), to strengthen its principles for banks' liquidity management.

Third, an internal rating system provides autonomy to a bank's management to execute credit risk policies in accordance with the bank's own core business goals and could potentially make credit operations more transparent and efficient. The development of an objective, internally derived model would allow the application of more thorough and homogeneous decision rules.

The above discussion provides strong arguments for the development of internal models. However, it must be noted that internally developed models sometimes suffer from impaired data quality (i.e., bank mergers sometimes result in a series of redundant, inconsistent data marts populated by complex legacy systems), limited size of data sets available within the institution, as well as from moral hazard situations preventing their objective use (originate-and-distribute model), and sampling bias (May, 1998). Indeed, internal rating systems are constructed only on the basis of the loans that have been approved in the past by the institution, but are intended to be used for all future loan applications.

3. CREDIT RISK RATINGS OF BANKS

Credit risk ratings are derived for individual, sovereign as well as corporate obligors across industries. In this paper, we focus on the financial sector, a key category of corporate borrowers, and we construct a forward-looking rating system for the classification of banks with respect to their risk of defaulting over a given time horizon. Our approach is generalizable to other types of borrowers by the modification of the set of explanatory variables and the selection of appropriate data sets for model inference. As the financial sector typically assigns credit ratings for a one-year horizon (Treacy and Carey, 2000; Grunert et al., 2005), we analyze how a set of financial variables measured at year t (calendar year 2000) can be used to predict the credit risk rating in year $t+1$ (calendar year 2001).

3.1. Limitations, Significance and Challenges

As it will be described in more details in the next sections, the proposed rating method is based on the learning of the ratings disclosed by the Fitch rating agency. Even if the vast majority of their ratings are justified, rating agencies have failed in predicting some major cases of creditworthiness deterioration, and have been severely criticized (see Section 2). Thus, the proposed rating method has some limitations. Its effective use requires the a priori selection of a set of reliable ratings that are not plagued by moral hazard, which, in its turn, rests on the definition of what constitutes a reliable rating.

One might, therefore, wonder about the utility of the proposed rating model that replicates the "cleaned" (i.e., exonerated from the obvious rating mistakes) Fitch rating system. The contribution of such a rating model is nonetheless substantial in that it enables the bank to rate its counterparts as frequently as it desires. It gives the bank the possibility to rerun its model on an ongoing basis, to assess, in a proactive fashion, the solvency of the financial institutions, and to possibly anticipate the rating modifications of the agencies, which would be extremely valuable for the investment decisions of the bank. This is in sharp contrast to the reactive approach of waiting for rating agencies to modify their ratings. Indeed, rating agencies are not properly incentivized to revise ratings, especially downward, on a timely basis (complacency effect).

The above discussion raises the question of the robustness of a learning or classification method with respect to the possible presence of noisy observations (i.e., incorrect Fitch ratings) in the data used for the derivation of the model. The machine learning field has investigated the impact of noisy data on the

learning process. It was shown (Angluin and Laird, 1988; Kearns, 1998) that the presence of a limited number of noisy observations has, in general, a limited effect on the overall quality of the classifier. In particular, Zhu and Wu (2004) have demonstrated that "class" noise (i.e., incorrect rating category given to a bank) has a much smaller effect on the accuracy of the model inferred by the learning method than "attribute" noise (i.e., incorrect values of the attributes, say return on equity, describing the bank). This would suggest that the presence of a few incorrect ratings in the data set we use does not preclude the derivation of an accurate rating system.

The opacity of and the leverage across financial institutions make the construction of accurate credit risk rating systems for this sector particularly difficult. This explains why the main rating agencies (Moody's, Fitch, S&P) disagree much more often about the ratings given to banks than about those given to entities in other sectors (Tabakis and Vinci, 2002). The rating migration volatility of banks is historically significantly higher than it is for corporations and countries, and banks tend to have higher default rates than corporations (de Servigny and Renault, 2004). Another distinguishing characteristic of the banking sector is the external support (i.e., from governments) that banks receive and which the other corporate sectors do not (Fitch Ratings, 2006). This shows the difficulty of the task at hand and explains the repeated calls from Federal Reserve and the Federal Deposit Insurance Corporation (King et al., 2004) for the development of efficient models to appraise the creditworthiness of financial institutions and their risk of failure.

In the next sub-section, we review the rather scant literature on the construction of banks' credit risk rating models. This will allow us to identify different types of proposed models and support the selection of explanatory variables for our models.

3.2. Prior Research on Banks' Creditworthiness

Poon et al. (1999) use factor analysis to extract the three so-called "banks' intrinsic safety" factors from a set of 100 bank-specific accounting and financial variables reflecting profitability, efficiency, asset composition, interest composition, interest coverage, leverage, and risk. The authors evaluate several ordered logit models in which the dependent variable is Moody's bank financial strength ratings and the explanatory variables are some of the above three factors. About 130 banks are considered, and the accuracies for the several models range from 21% to 71%. Huang et al. (2004) use a back-propagation neural network to evaluate the creditworthiness of US and Taiwanese banks, and claim that the lower accuracy of the statistical methods is due to the fact that the multivariate normality assumption for

independent variables is very often violated in financial data sets. They observe that the most predictive variables differ with the considered location (US or Taiwan) of the banks. Fitch Ratings (2006) analyzed a bank rating method based on joint probability analysis, which includes as input variables the probability of the bank failing and the probability of the potential supporter (i.e., a sovereign state) defaulting at the same time, and, if there is no such simultaneous default, the probability of the supporter being willing or not to provide such support. Fitch Ratings comes to the conclusion that the a method is conceptually valid, but insists on the difficulty of constructing such a robust model, in view of the scarcity of the available empirical data. Kick and Koetter (2007) select a set of financial ratios to predict the credit risk level of German banks. They demonstrate that the various levels/categories of bank credit risk have different sensitivities to predictors, and reject the use of binary or ordered logit regression models to assess the bank distress level.

Regardless of the type (i.e., statistical, machine-learning) of model, it appears that financial ratios constructed on the basis of accounting data and reflecting the quality of the assets, as well as the bank's profitability and liquidity, are the key predictors. We refer the reader to Ravi Kumar and Ravi (2007) and the references therein for a detailed review of models derived with the goal of evaluating the financial strength of banks and the risk for a bank to go bankrupt.

3.3. External Bank Ratings

In this sub-section, we shall briefly describe the Fitch Individual Bank rating system. Fitch publishes over 1,000 international bank ratings worldwide, and it is generally viewed as the leading agency for rating bank credit quality. Fitch provides long- and short-term credit ratings, which are viewed as an opinion on the ability of an entity to meet financial commitments (interest, preferred dividends, or repayment of principal) on a timely basis (Fitch Ratings, 2001). These ratings are assigned to sovereigns and corporations, including banks, and are comparable worldwide.

Fitch provides a specialized rating scale for banks using *individual* and *support* ratings. Support ratings comprise 5 rating categories and assess the likelihood for a banking institution to receive support either from the owners or the governmental authorities should it run into difficulties. While the availability of support is a critical characteristic of a bank, it does not describe completely the banks' solvability in case of adverse situations. It is worth noting for instance that even though a bank can have a state guarantee of support, its marketable obligations might drop. That is why, as a complement to a support rating, Fitch also provides an individual bank rating, which allows a credit quality evaluation separately from any

consideration of outside support. It is purported to assess how a bank would be viewed if it were entirely independent and could not rely on external support.

For banks in investment grade countries, support and individual bank rating scales provide sufficient scope for differentiation. However, the scope for differentiation between banks in non-investment grade countries can be more limited. Indeed, the rating of an obligor located in a given country is limited from above by the rating of that country (Ferri et al., 1999). In countries with weak country risk rating, it is thus possible that a number of banks' ratings may be restricted by the sovereign's foreign currency rating and will be bunched together at the sovereign ceiling. That is why *national* ratings were developed to provide a greater degree of differentiation between issuers in countries subject to this bunching effect. National ratings are an assessment of credit quality relative to the rating of the "best" credit risk in the country. The national rating scale has 22 different risk categories as opposed to the 9 categories of the individual rating scale. Table 7 contains a detailed description of the rating categories characterizing the Fitch individual bank credit rating system. The individual bank credit ratings will be used in the remaining part of this paper, since these ratings are comparable across different countries, as contrasted with the national ratings, which are not.

3.4. Explanatory Variables

Based on the references mentioned above, approximately 40 parameters have been provisionally considered as potentially significant predictors of the banks' creditworthiness. After the elimination of those parameters for which not all the data were available, we have restricted our attention to a set of 14 financial variables (loans, other earning assets, total earning assets, non-earning assets, total assets, net interest revenue, customer and short-term funding, overheads, equity, net income, operating income, profit before tax, total liabilities and equity, other operating income) and 9 representative financial ratios describing: (a) **asset quality:** ratio of equity to total assets; (b) **performance**: (i) profit efficiency: return on average assets; return on average equity; ratio of interest revenue to average assets; net interest margin; ratio of other operating income to average assets; (ii) cost efficiency: cost to income ratio; ratio of non-interest expenses to average assets; (c) **liquidity**: ratio of net loans to total assets.

To verify whether the deletion of the variables with missing data introduces or not any bias, we verify whether the data are "missing completely at random" (MCAR) which, as defined by Rubin (1976), happens when the probability of obtaining a particular pattern of missing data is not dependent on the values that are missing and when the probability of obtaining the missing data pattern in the sample is not

dependent on the observed data. We carry out, using the SPSS 16.0 statistical software, Little's MCAR test (Little, 1988). The MCAR test is a chi-square test whose output (p-value) indicates whether or not missing values are randomly distributed across all observations. The MCAR test splits the observations into two groups (i.e., groups with and without missing data) and compares mean differences on the explanatory variables to establish that the two groups do not differ significantly. In our case, the p-value is equal to 0.991, indicating that the MCAR test is not significant and attesting that the missing data are randomly spread across the observations. We also note that, after the deletion, the variables remaining in the data set include those which the literature depicts as being predictive for estimating banks' creditworthiness and which reflect the key asset quality, operations and liquidity criteria.

The values of these variables were collected at the end of 2000 and are disclosed in the database called Bankscope (Fitch Ratings, Bureau Van Dijk, 2009), which is the largest existing bank database. As an additional (24th) variable, we use the Standard & Poor's (S&P) risk rating of the country where the bank is located. The S&P country risk rating scale comprises twenty-two categories (from AAA to D). We convert these categorical ratings into a numerical scale, assigning the largest numerical value (21) to the countries with the highest rating (AAA). Similar numerical conversions of country risk ratings are used by Ferri et al. (1999). Bloomberg also uses a standard cardinal scale for comparing Moody's, S&P's and Fitch ratings.

3.5. Observations

Our dataset consists of 800 banks rated by Fitch and operating in 70 different countries. The values of the ratings were collected at the end of 2001. Table 1 provides the geographic distribution of the banks included in the dataset. Table 2 lists the percentage of the banks in the dataset in each rating category. One can see that the extremal rating categories (A and E) comprise a very small number of banks (19 and 32 respectively). The majority of the banks in the dataset have received intermediate ratings.

Table 1: Geographic Distribution of Banks

Origin	Western Europe	Eastern Europe	Canada and USA	Developing Latin American countries	Middle East	Hong-Kong, Japan, Singapore	Developing Asian countries	Other: Oceania - Africa - Israel
Number of banks	247	51	198	45	44	55	145	6 -6 -3

Table 2: Distribution of Banks in Rating Categories

Rating Categories	A	A/B	B	B/C	C	C/D	D	D/E	E
Percentage of banks	2.375%	7.5%	25.375%	16.125%	15.125%	9.625%	11.625%	8.25%	4%

The first question to address in reverse-engineering a bank rating system concerns the choice of explanatory variables to be used in constructing the model. More precisely, the question to be answered is whether the amount of information contained in the selected 24 variables is sufficient for correctly classifying the 800 banks in the dataset. We provide below an affirmative answer to this question.

4. LOGICAL ANALYSIS OF DATA – AN OVERVIEW

The *logical analysis of data* (LAD) is a combinatorics-, optimization-, and Boolean logic-based methodology for analyzing archives of observations. Initially created for the classification of binary data (Hammer 1986, Crama et al., 1988), LAD was later extended (Boros et al., 1997; Hammer et al., 2006) from datasets having only binary variables to datasets which contain numerical variables. LAD distinguishes itself from other classification methods and data mining algorithms by the fact that it generates and analyzes exhaustively a major subset of those combinations of variables which can describe the *positive* or *negative* nature of observations (e.g., to describe solvent or insolvent banks, healthy or sick patients, etc.), and uses optimization techniques (Ryoo, Jang, 2009) to extract models constructed with the help of a limited number of significant combinatorial patterns generated in this way (Boros et al., 2000).

Rule-based classification or rating systems such as the one we construct with the LAD method are very flexible and can represent very complex decision rules. This greater flexibility can potentially cause some overfitting issues. That is why we shall perform an extensive cross-validation of the rating models we propose. We sketch below the basic concepts of LAD, referring the reader for a more detailed description to Boros et al. (2000), or to the recent surveys of Hammer and Bonates (2006) and Alexe et al. (2007).

In LAD, as in most of the other data analysis methods, each observation is assumed to be represented by an n-dimensional real-valued vector. For the observations in the given dataset, besides the values of the n components of this vector, an additional binary (0,1) value is also specified; this additional value is called the *class* of the observation, with the convention that 0 is associated to negative observations, and 1 to positive ones.

The objective of LAD is to discover a binary-valued function f depending on the n input variables, which provides discrimination between positive and negative observations, and which closely approximates the actual one. This function f is constructed as a weighed sum of patterns. In order to clarify how such a function f is found we start by transforming the original dataset to one in which the variables can only take the values 0 and 1. We achieve this goal by using *indicator variables*. Each indicator variable shows whether

10

the value of a numerical variable does or does not exceed a specified level, called a *cutpoint*. For example, one of the (two) indicator variables associated to the numerical variable *ratio of costs to incomes* shows whether the ratio of costs to incomes did or did not exceed 71.92. The selection of the cutpoints is achieved by solving an associated set covering problem (Boros et al., 1997). By associating an indicator variable to each cutpoint, the dataset is *binarized*.

Positive (negative) patterns are combinatorial rules which impose upper and lower bounds on the values of a subset of input variables, such that *(i)* a sufficiently high proportion of the positive (negative) observations in the dataset satisfy the conditions imposed by the pattern, and *(ii)* a sufficiently high proportion of the negative (positive) observations violate at least one of the conditions of the pattern.

In order to give an example of a pattern in the bank rating dataset, let us first define as *positive* those banks whose ratings are B or higher, and as *negative* those banks whose ratings are D or lower. As an example of a *positive pattern* in our dataset, we mention the pattern requiring the simultaneous fulfillment of the following three conditions: *(i)* the country risk rating is A+, AA-, AA, AA+, AAA; *(ii)* the ratio of the costs to incomes is at most equal to 71.92, and *(iii)* the return on equity is strictly larger than 11.82%. It can be seen that these three conditions are satisfied by 70.57% of the banks rated B or higher, and by none of the banks rated D, D/E or E.

The following terminology will be useful in this paper. The *degree* of a pattern is the number of variables the values of which are bounded in the definition of the pattern. The *prevalence* of a positive (negative) pattern is the proportion of positive (negative) observations covered by it. The *homogeneity* of a positive (negative) pattern is the proportion of positive (negative) observations among those covered by it. The pattern in the above example has degree 3, prevalence 70.57% and homogeneity 100%.

The first step in applying LAD to a dataset is to generate the *pandect*, i.e., the collection of all patterns in a dataset. Because of the enormous redundancy in this set, we impose some limitations on the set of patterns to be generated, by restricting their degrees (to low values), their prevalence (to high values), and their homogeneity (to high values). Several algorithms have been developed for the efficient extraction of (relatively small) subsets of positive and negative patterns corresponding to the above criteria and sufficient for classifying the observations in the dataset (Boros et al., 2000). Such collections of positive and negative patterns are called *models.* A model is supposed to include sufficiently many positive (negative) patterns to guarantee that each of the positive (negative) observations in the dataset is "covered" by (i.e., satisfies the

conditions of) at least one of the positive (negative) patterns in the model. Good models tend to minimize the number of points in the dataset covered simultaneously by both positive and negative patterns in the model.

A LAD model can be used for classification in the following way. An observation (whether it is contained or not in the given dataset) which satisfies the conditions of some of the positive (negative) patterns in the model, but which does not satisfy the conditions of any of the negative (positive) patterns in the model, is classified as *positive (negative)*. An observation satisfying both positive and negative patterns in the model is classified with the help of a discriminant that assigns specific weights to the patterns in the model (Boros et al., 2000). More precisely, if p and q represent the number of positive and negative patterns in a model, and if h and k represent the numbers of those positive, respectively negative patterns in the model which cover a new observation ω, then the value of the discriminant $\Delta(\omega) = h/p - k/q$, and the corresponding classification is determined by the sign of this expression. Finally, an observation for which $\Delta(\omega) = 0$ is left *unclassified,* since the model either does not provide enough evidence, or provides conflicting evidence for its classification. Fortunately it has been seen in all the real-life problems considered that the number of unclassified observations is extremely small (usually less than 1%). We represent the results of classifying the set of all observations in the form of a *classification matrix* (Table 3).

Table 3: Classification Matrix

Observation Classes	Classification of Observations		
	Positive	Negative	Unclassified
Positive	a	c	e
Negative	b	d	f

The value *a* (resp., *d*) represents the percentage of positive (negative) observations that are correctly classified. The value *c* (resp., *b*) is the percentage of positive (negative) observations that are misclassified. The value *e* (resp., *f*) is the percentage of positive (negative) observations that remain unclassified. Clearly, *a+c+e=100%* and *b+d+f=100%.* Given that about half of the unclassified observations can be classified correctly by assigning their class randomly with equal probability, the quality of the classification is defined (see Alexe et al., 2008 for a detailed description) by:

$$Q = \tfrac{1}{2}\,[(a + d) + \tfrac{1}{2}\,(e+f)]. \tag{1}$$

5. LAD MODEL FOR BANKS' CREDITWORTHINESS

The rationale for using the LAD method to construct bank ratings is fourfold. First, it allows the representation of higher-order interaction between variables. This feature is especially attractive in view of

the complexity of financial systems and the inter-connectivity between financial institutions. Second, the LAD approach allows every explanatory variable to have distinct effects on different rating categories. Third, it enables the derivation of rating systems with a varying, user-specified number of rating categories, thus permitting its application in multiple banking operations. Fourth, the LAD method is not constrained by any limitation on the number of variables (i.e., capable of accommodating a larger number of variables than that of observations), their nature (numerical, categorical), or the relationships (i.e., correlation, higher order dependencies) between variables that can be used for the derivation of the patterns used in the rating system.

Since LAD is a classification methodology, it is natural to first associate to the bank rating problem a related classification problem, with the expectation that the resulting LAD model can be successfully utilized for establishing an objective and transparent bank rating system.

We define as *positive observations* the banks which have been rated by Fitch as A, A/B or B, and as *negative observations* those whose Fitch rating is D, D/E or E. In the binarization process, cutpoints, obtained through the solution of a set covering problem, are introduced for the 24 variables described in Section 3.4, and a collection of powerful patterns is extracted from the binarized problem. The patterns are then used to generate a discriminant value for each bank that reflects its creditworthiness. The binarization and the generation of patterns are carried out using the LAD – Datascope 2.01 software package (Alexe, 2007) whose algorithmic procedures are described in Alexe and Hammer (2006).

In order to map the numerical values of the LAD discriminant to the nine bank rating categories of Fitch (A, A/B, …, E), we shall attempt to partition the interval of the discriminant values into nine sub-intervals corresponding to the nine categories. We shall assume that this partitioning is defined by *thresholds* x_j such that $-1 = x_0 \leq x_1 \leq x_2 \leq \ldots\ldots \leq x_8 \leq x_9 = 1$, where j indexes the rating categories (with $j=1$ corresponding to E, and $j=9$ corresponding to A). Ideally, a bank i should be rated j if its discriminant value falls between x_{j-1} and x_j (e.g., it should be rated A if its value falls between x_8 and x_9).

For a set of banks used to determine the x_i thresholds, such a partitioning may not exist. Therefore, in order to take "noisiness" into account, we shall map the LAD discriminant value d_i of bank i to an *adjusted discriminant value* δ_i, and find values of δ_i for which such a partitioning exists and which are "as close as possible" to the values d_i. As it is often the case, we interpret "as close as possible" as minimizing the mean square approximation error. If we denote by $j(i)$ the rating category of bank i, by ε an infinitesimal

positive number, by N the set of banks, then the determination of the thresholds x_j and of the adjusted discriminant values δ_i can be modeled as follows:

$$
\begin{aligned}
minimize \quad & \sum_{i \in N} (\delta_i - d_i)^2 \\
subject\ to \quad & \delta_i \leq x_{j(i)}, \ i \in N \\
& x_{j(i)-1} \leq \delta_i - \varepsilon, \ i \in N \\
& -1 = x_0 \leq x_1 \leq x_2, \ldots, \leq x_j, \ldots, x_8 \leq x_9 = 1 \\
& -1 \leq \delta_i \leq 1, \ i \in N
\end{aligned}
\tag{2}
$$

To solve the convex nonlinear problem above, we use in our numerical experiments the NLP solver Lancelot (Conn et al., 1990). The approach described above is very similar to the convex cost closure problem that can be used to determine adjustments of the observations minimizing the value of the deviation penalty function, while satisfying the ranking order constraints.

The LAD discriminant derived above and the values of the thresholds x_i determined by solving problem (2) define the LAD ratings of banks (whether those banks are in the training sample or not): the rating of a bank is determined by the particular sub-interval, defined by the x_i thresholds, which contains that bank's LAD discriminant value.

6. ANALYSIS OF THE LAD APPROACH

This section consists of three main parts. First, we discuss the fitted model obtained by considering the whole data set. Second, we perform an extensive cross-validation procedure to verify the presence/absence of overfitting (Hjorth, 1994). Finally, we derive additional insights (critical variables for evaluating the credit risk of a bank, classification quality per continent, rating categories, etc.) about the LAD rating model.

6.1. Fitted LAD Model

6.1.1. Accuracy and Robustness of the LAD Discriminant

To derive the fitted LAD classification model for bank ratings, we use the 473 banks whose ratings are A, A/B, B (positive observations) and D, D/E or E (negative observations). The 327 banks rated B/C, C, C/D are not used in this derivation. The fitted LAD classification model (see Table 12) is very parsimonious, consisting of only ten positive and nine negative patterns (P1,…, P10, respectively, N1,…,N9) such that every positive (resp. negative) observation in the model is covered by at least one positive (resp., negative) pattern included in the model, and is built on a *support set* of only 17 out of the 24

14

original variables. The other 7 numerical variables (total earning assets, total assets, customer and short-term funding, equity, total liabilities and equity, net income, and operating income), have also been binarized, but the algorithmic pattern generation method did not incorporate any of them in any of the patterns included in the LAD model.

Table 8 provides all the cutpoints used in pattern and model construction. For example, two cutpoints (24.8 and 111.97) are used to binarize the numerical variable "Profit before Tax" (PbT), i.e., two binary indicator variables replace PbT, one indicating whether PbT exceeds 24.8, and the other indicating whether PbT exceeds 111.97.

All the patterns in the model are of degree at most 3, have perfect homogeneity (100%), and very substantial prevalence (averaging 31.1% for the positive, and 28.5% for the negative patterns). As an example of a powerful negative pattern, consider the one defined by the following two conditions: *(i)* the country risk rating is strictly lower than A, and *(ii)* the profit before tax is at most equal to €111.97 millions. These conditions describe a *negative pattern*, since none of the positive observations (i.e., banks rated A, A/B or B) satisfy both of them, while no less than 69.11% of the negative observations (i.e., those banks rated D, D/E or E) do satisfy both conditions. This pattern has degree 2, prevalence 69.11%, and homogeneity 100% (since none of the positive observations satisfy its defining conditions).

The derived fitted model has a 100% classification accuracy (1). The perfect fit of the model may be due to overfitting. The true measure of the model's accuracy and robustness can only be obtained by cross-validation and is investigated in Section 6.2.

As a second measure of accuracy of the model, we examine the correlation between the values of the discriminant of the model (ranging between -1 and +1) and the bank ratings (represented on their numerical scale). Although this experiment included all the banks in the dataset (i.e., not only those rated A, A/B, B, D, D/E or E, which are used in creating the LAD model, but also those rated B/C, C or C/D, which are not used at all in the learning process), the correlation turns out to be 82.05% -- reconfirming the high predictive value of the LAD model.

Finally, as an additional check, we separately calculate the average discriminant values for the nine rating categories. The results are presented in Table 9 (left side), and show clearly the discriminating power of the LAD model. Interesting conclusions one can derive from this table are the following:

- The positive observations have higher average discriminant values than the unclassified ones, which, in

their turn, have higher average discriminant values than the negative ones.

- The average discriminant values are monotonically decreasing with the rating categories, and, although the model was not "taught" to make distinctions between the categories A, A/B, and B (and between D, D/E, and E), the average discriminant values drop by about 8% from one category to the next.

- Even in the case of the "unclassified" observations, which are not used in deriving the LAD model, the average discriminant value for category C is lower than that for category B/C, and the average discriminant value for category C/D is higher than that of category D.

6.1.2. Robustness of the LAD Rating Model and Conformity with Fitch Ratings

To partition the range of LAD discriminant values into rating sub-intervals, all 800 banks in the dataset are used to determine the x_i thresholds by solving problem (2). Then we use the original (unadjusted) LAD discriminant value of each bank to determine its rating category. We denote by n_k ($k = 0,\ldots,8$) the number of banks whose rating category determined in this way differs from the actual Fitch rating by exactly k categories, indicating the goodness-of-fit of the proposed rating system. We call the set of values n_k ($k = 0,\ldots,8$) the *discrepancy counts*.

In Table 11, we report the percentage n_k/N (normalized discrepancy count) of banks for which the difference between the LAD and the Fitch ratings is equal to k, $k=0,1,\ldots,8$. The discrepancy summary presented in Table 11 demonstrates a high goodness-of-fit of the proposed model. More than 95% of the banks are rated within at most two categories of their actual Fitch rating, with about 30% of the banks receiving exactly the same rating as in the Fitch rating system, and another 52% being off by exactly one category. The simplest reflection of the very high degree of coincidence between the LAD and the Fitch ratings is the fact that the weighted average distances $D_j = \sum_{k=1}^{8} k \; n_k$ between the two ratings are

- 0.911 for the categories A, A/B, B, D, D/E and E,

- 0.979 for the categories B/C, C and C/D, and

- 0.939 for all banks in the sample (categories A, A/B, B, B/C, C, C/D, D, D/E, E).

Note that while the rating thresholds (see Table 10) were determined using all the banks, the LAD discriminant was derived using only the banks rated A, A/B, B, D, D/E and E. It is interesting that the goodness-of-fit of the ratings calculated separately for the banks rated by Fitch as B/C, C, and C/D (i.e., those banks which were not used in deriving the LAD model) is very close to the goodness-of-fit for the

banks actually used (i.e., those rated by Fitch as A, A/B, B, D, D/E, and E) for deriving the LAD model. This finding indicates the stability of the proposed rating system and its appropriateness for rating "new" banks, i.e., banks which are not rated by agencies or banks the rater has not dealt with before. The Spearman rank correlation between the LAD and the Fitch ratings is equal to 84.19%.

6.2. Cross-Validation Experiments for LAD Approach

In order to cross-validate the model, we apply 10 times the two-folding cross-validation procedure (Hjorth, 1994). In each two-folding experiment, the observations are randomly split into two approximately equal subsets, and we derive a specific set of cutpoints and an LAD model by only considering the banks included in the training set of that experiment and rated A, A/B, B, C/D, D, and D/E. Then all the banks of the training set (including those rated B/C, C and C/D) in this experiment are used to determine, with the help of the convex problem (2), the rating thresholds for the experiment. Finally, the derived LAD model and the intervals determined by these rating thresholds are used to determine ratings for the banks in the testing set. Those banks are not used, in any stage of the proposed approach (determination of cutpoints, patterns, rating thresholds), to derive the LAD model. In the second half of the experiment, the roles of the two subsets are reversed, i.e., the set formerly used for testing is now used for training, and the one formerly used for training becomes the test set. The final estimates are calculated as the averages of the results obtained on the 20 testing sets in these experiments.

6.2.1. Accuracy and Robustness of Cross-Validated LAD Discriminants

The average classification quality (1) on the testing sets of the 20 models obtained as described above is 95.47%. It is remarkable that the standard deviation in the 20 experiments is only 0.03. The high accuracy and low standard deviation indicate high predictive value and demonstrate the robustness of the LAD classification system.

We also evaluate the stability of the correlation between the LAD discriminant values and the bank ratings, using the results of the two-fold cross-validation experiments described above. The average value of the correlation coefficient is 81.21%, with standard deviation of 0.03, showing the stability of the close association between the discriminant values and the original bank ratings.

6.2.2. Robustness of the Cross-Validated LAD Rating Models and Conformity with Fitch Ratings

In order to systematically evaluate the robustness of the proposed rating system, we use 10x2-

folding procedure described above to calculate the average discrepancy counts of the bank ratings (as predicted on the testing sets). The average normalized discrepancy counts over the 20 cross-validation experiments are given in Table 11. The fact that the difference between the Fitch and the LAD ratings is, on average, only 0.975, is an extremely strong indicator of the LAD model's stability and the absence of overfitting. This result is particularly significant in view of the occasional reports in the financial data mining literature that the high fit of machine learning methods, such as support vector machine, is achieved at the risk of overfitting (Huang et al, 2004; Galindo and Tamayo, 2000). Clearly, the LAD-based combinatorial rating approach is not subject to this overfitting problem. Additionally, we note that the average over the 20 cross-validation experiments Spearman rank correlation between the LAD and the Fitch ratings is equal to 84.19%.

6.3. Specific Features of LAD Model

While the focus in LAD is on discovering how the interactions between the values of small groups of variables (as expressed in patterns) affect the outcome (i.e., the bank ratings), one can also use the LAD model to learn about the importance of individual variables. A natural measure of importance of a variable in an LAD model is the frequency of its appearance in the model's patterns. The three most important variables in the 19 patterns constituting the LAD model are the credit risk rating of the country where the bank is located, the return on average total assets, and the return on average equity. The importance of the country risk rating variable, which appears in 14 of the 19 patterns, can be explained by the fact that credit rating agencies are often reluctant to give an entity a better credit risk rating than that of the country where it is located. That is why the country risk rating is sometimes referred to as the "sovereign ceiling" or the "pivot of all other country's ratings" (Ferri et al., 1999). The country risk rating was also found to be an important predictive variable for bank ratings by Poon et al. (1999). The return on average equity variable appears in six patterns, while the return on average assets variable is involved in five patterns. These two ratios, respectively representing the efficiency of assets in generating profits, and that of shareholders' equity in generating profits, are critical indicators of a company's prosperity, and are key predictors (Sarkar and Sriram, 2001) which auditors use to evaluate the wealth of a bank. The return on average equity is also found significant for predicting the rating of US banks by Huang et al. (2004).

Below, we analyze the classification quality of the LAD model according to:

- Continents: Table 4 shows that the LAD performs best for European and North American banks. The

higher average difference between the Fitch and LAD Ratings is to be taken cautiously, since we have only

29 South American banks in the dataset (3.625%).

Table 4: Analysis of Discrepancies per Continent

k	Asia	Europe	North America	South America
0	22.12%	34.87%	34.45%	16.82%
1	53.11%	45.93%	51.92%	61.90%
2	17.42%	15.80%	11.03%	8.87%
3	7.35%	3.40%	1.18%	8.35%
4	0.00%	0.00%	1.42%	4.06%
5 - 6 - 7 − 8	0.00%	0.00%	0.00%	0.00%
Average Difference between Fitch and LAD Ratings	1.10	0.88	0.83	1.21

• Fitch's rating categories (Table 5): the explanation for the higher average differences between Fitch and

LAD Ratings for the banks which have the extreme Fitch ratings is twofold. First, the number of banks in

those categories [19 rated A (2.375%), 66 rated D/E (8.25%), and 32 rated E (4%)] is limited, which may

lead to a higher variability of the average rating difference. Second, the maximum number of rating

categories by which the LAD and the Fitch ratings can differ is higher for the extreme than for the

intermediate rating categories.

Table 5: Analysis of Discrepancies per Fitch Rating Category

k	A	A/B	B	B/C	C	C/D	D	D/E	E
0	26.32%	45.15%	32.98%	35.82%	24.32%	24.31%	17.56%	34.95%	9.37%
1	36.84%	34.87%	43.12%	43.03%	57.40%	40.26%	69.62%	31.82%	25.00%
2	21.05%	19.98%	23.90%	18.90%	15.45%	29.31%	12.82%	24.24%	21.87%
3	15.79%	0.00%	0.00%	1.34%	2.83%	6.12%	0.00%	8.99%	31.25%
4	0.00%	0.00%	0.00%	0.91%	0.00%	0.00%	0.00%	0.00%	12.51%
5 − 6 − 7 - 8	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Average Difference between Fitch and LAD Ratings	1.263	0.748	0.909	0.885	0.968	1.172	0.9526	1.073	2.125

• Banks' country risk rating (Table 6): we consider the credit risk rating of the country where the bank is

located, and we allocate banks to three subgroups depending on whether their country has an investment-

grade (BBB- or higher), a speculative-grade (from BB+ to B-), or a default-grade (CCC+ or lower) rating.

Table 6: Analysis of Discrepancies per Country Risk Rating Category

k	Investment-Grade	Speculative-Grade	Default-Grade
0	30.07%	30.26%	19.25%
1	53.12%	45.86%	62.50%
2	12.34%	15.79%	18.25%
3	3.58%	7.43%	0.00%
4	0.89%	0.66%	0.00%
5 - 6 − 7 − 8	0.00%	0.00%	0.00%
Average Difference between Fitch and LAD Ratings	0.92	1.02	0.99

Table 6 shows that the LAD rating model performs equally well for every level of creditworthiness of the country in which a bank is located. The results confirm the capacity of non-statistical models to derive highly accurate prediction models, as acknowledged in the literature (see, e.g., de Servigny, Renault, 2004).

7. COMPARISON WITH OTHER CLASSIFICATION METHODS

7.1. Description of Other Classification Methods

In this Section, we compare the results obtained with the LAD approach to those obtained with the multiple linear regression (MLR), the ordered logistic regression (OLR), and the support vector machine (SVM) methods.

The standard econometric technique of multiple linear regression can be used to model the dependency of bank ratings on the independent variables described in the previous section. The dependent variable y is the numerical scale of the Fitch individual bank ratings as presented in Table 7.

A $(k+1)$-category ordered logistic model has the form:

$$
\begin{aligned}
&\log\left(\frac{p_1}{1-p_1}\right) = \alpha_1 + \sum_{i \in I} \beta_i x_i \\
&\vdots \\
&\log\left(\frac{p_1 + p_2 + \cdots + p_k}{1 - p_1 - p_2 - \cdots - p_k}\right) = \alpha_k + \sum_{i \in I} \beta_i x_i
\end{aligned}
\tag{3}
$$

with $p_1 + p_2 + \cdots + p_k + p_{k+1} = 1$. We denote by p_i the individual probability of category i, by

$$\log\left(\sum_{i=1}^{r} p_i \Big/ \left(1 - \sum_{i=1}^{r} p_i\right)\right),$$

$r=1...k$ the logarithm of odds, which indicates the log-odds of lower rather than higher scores when all independent variables equal 0, by x the vector of independent variables, by β the vector of logistic coefficients (slope parameters), and by α_j the intercepts that satisfy the constraints: $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_{k-1} \leq \alpha_k$. The regression coefficients, derived with a maximum likelihood approach, are used to predict the values of the logit-transformed probability that the dependent variable falls into one category rather than another.

Ordered logit fits a set of cutoff points. If there are $(k+1)$ categories associated with the dependent variable, it will find k cutoff values r_1 to r_k such that if the fitted value of logit(p) is below r_1, the dependent variable is predicted to take value 0, if the fitted value of logit(p) is between r_1 and r_2, the dependent variable is predicted to take value 1, and so on. Finally, note that, the ordered logit method fits a common slopes cumulative model, since in (3), only the intercepts α_k differ across credit risk rating categories, while the regression coefficients β_i are identical for all categories. Stated differently, it means that the model assumes

that the effects of the 24 variables are the same on all banks regardless of their rating category.

Support vector machine (SVM) algorithms are highly regarded supervised learning methods involving the construction of one or multiple hyperplanes in a high dimensional space that are used for binary classification purposes. In this paper, we use the WEKA implementation (Hall et al., 2009) of Platt's SVM algorithm (Platt, 1998). It is a sequential minimal optimization (SMO) algorithm for training a support vector classifier that reduces the solution of a large quadratic program to a series of small quadratic programs that are solved analytically. We utilize the version of the SMO algorithm that provides as its output not just a binary classification, but also a calibrated posterior probability, obtained by learning the parameters of an additional sigmoid function to map the SVM outputs into probabilities (Platt, 1999). These output probabilities are similar to the LAD discriminant values in that they indicate the margins of binary classification, and therefore they are utilized in exactly the same way as described in Section 5. Substituting SVM output probabilities for LAD discriminant values there, only the banks rated A, A/B, B, D, D/E or E are used to derive the support vector machine model, and then all the banks are used in solving the optimization problem (2) to determine the x_i thresholds that define the bank rating sub-intervals. Note that we have tested several kernels available through WEKA and selected the one providing the best results.

7.2. Analysis of Fitted Models

Applying the SMO classification model obtained as described in the above section to the classification of all the banks rated A, A/B, B, D, D/E or E, we find the classification quality (1) of the model to be 96.27%, while the correlation coefficient between the SVM output probability values and the Fitch numerical ratings is 75.59%. As for the LAD approach, we compute the average output probability values for the nine rating categories (Table 9). As opposed to the LAD discriminant values, the SVM ones are not monotonically decreasing with the rating categories. The average output probability value for category E (resp., D/E) is larger than those (resp., the one) for categories D/E and D (resp. category D).

Table 11 reports the normalized discrepancy counts obtained with the fitted MLR, OLR, and SVM models. For the MLR model (the ratings of which can be decimal values), we calculate the discrepancy count as follows: $n_k = \left| \left\{ j : d_j = k, j \in N \right\} \right|, k = 0,1,...,8$, where $d_j = | \bar{y}_j - r_j |$, $j \in N$ is the difference between Fitch's rating r_j for bank j and the rounded MLR rating $\bar{y}_j$ (which is the integer value closest to $\hat{y}_j$ denoting bank j's MLR rating). The percentage of banks precisely classified ($k = 0$) with the fitted MLR (resp., OLR,

SVM) rating model is equal to 27.75% (resp., 36.50%, 17%), while weighted average distance between the fitted MLR (resp., OLR, SVM) ratings and the Fitch ones is equal to 1.22 (resp., 1.02, 1.695).

In addition to the discrepancy counts, we use the following statistics to evaluate the accuracy of the fitted rating models obtained with the multiple linear regression (MLR), the ordered logistic regression (OLR) and the support vector machine (SVM) methods:

- The correlation coefficient between the Fitch numerical ratings and the fitted MLR (resp., OLR and SVM) ratings is equal to 74.09% (resp., 80.19%, 75.59%);

- The Spearman rank correlation coefficient between the Fitch numerical ratings and the fitted OLR (resp., SVM) ratings is equal to 75.60% (resp., 76.57%);

7.3. Analysis of the Cross-Validation Experiments

We carry out the cross-validation experiments for the MLR, OLR, and SVM using the same 10 2-folding procedure used for the LAD approach.

The average (taken over the 20 SVM models) classification quality (1) is 94.98% with a standard deviation of 0.06, while the average correlation coefficient between the SVM output probability values and the Fitch numerical ratings is 74.85%, with a standard deviation of 0.03. Regardless of the classification metric used, the LAD discriminant performs better than the SVM one (both for the fitted model and for the cross-validated experiments).

To evaluate the robustness of the ratings obtained with the MLR, OLR, and SVM methods, we analyze the cross-validation results:

- The average correlation coefficient between the Fitch numerical ratings and the cross-validated MLR (resp., OLR, SVM) ratings is equal to 71.84% (resp., 74.64%, 73.38%), with standard deviation of 14.5% (resp., 6.5%, 4.9%);

- The average Spearman rank correlation coefficient between the Fitch numerical ratings and the cross-validated OLR (resp., SVM) ratings is equal to 74.12% (resp., 73.98%).

Table 11 reports the average (over the 20 cross-validation experiments) normalized discrepancy counts obtained with the MLR, OLR, and the SVM methods. The average number of precise classifications ($k = 0$) on the test sets with the MLR method is only 21.75%. On the average, the difference between the actual and predicted rating categories for the banks in the testing sets is 1.63. The average percentage of

banks precisely classified with the OLR model is almost equal to 31%. On the average, the difference between the actual and predicted rating categories for the banks in the testing sets is 1.175. The average number of precise classifications with the SVM method is roughly 17%. The average difference between the actual and predicted rating categories for the banks in the testing sets is 1.7.

The results above show that the LAD model clearly outperforms the other three ones. The MLR and the SVM models have only limited generalizability. This negative conclusion is very much in agreement with Morgan's (2002) observations according to which the main rating agencies have serious difficulties in evaluating the ratings of banks, as manifested in the frequency of their divergent evaluations.

A further analysis of the results displayed in Table 11 provides another confirmation that the LAD method outperforms the MLR, OLR and SVM ones, in particular when they are applied to rate banks not used in the construction of the model. Indeed, it can be seen that the LAD rating is, on average, within one (two) rating category (categories) of the Fitch rating for 81.30% (93.80%) of the banks, while this statistic is 69.625% (87.375%) for the comparison between the Fitch and OLR ratings. We also observe that, on average, the OLR (respectively, LAD, MLR and SVM) ratings perfectly match the Fitch ones for 30.75% (resp., 29.80%, 21.75% and 16.875%) of the banks.

The OLR rating method appears to be the one providing the closest results to those obtained with the LAD method. To further evaluate the robustness and extendibility of the LAD rating model versus the OLR rating model, we compare the predicted results of 20 experiments (10 2-fold cross-validation runs) to check whether the rating discrepancy between the LAD and the Fitch ratings, and that between the OLR and the Fitch ratings, differ from each other in a significant way. Since the average difference between the Fitch and OLR ratings is 1.175, while it is only 0.975 for the comparison of Fitch and LAD, and the paired *t*-test indicates that we can reject, at the highest statistical confidence level (99.9%), the null hypothesis according to which these two rating discrepancies do not differ, we can conclude that the LAD rating approach statistically outperforms the OLR approach with respect to the rating discrepancy criterion. To verify that the statistical significance of the better performance of the LAD method does not rest on the parametric assumptions of the *t*-test, we also use Wilcoxon's signed-rank test as a non-parametric alternative to check whether the rating discrepancy between the predicted LAD and the Fitch ratings is significantly smaller than the one between the predicted OLR and the Fitch ratings. Again, at the confidence level of 99.9% the null

hypothesis can be rejected, indicating the statistical significance of the better LAD rating predictions.

8.   BROADER IMPACT OF THE PROPOSED METHODOLOGY

Provided that a reliable set of observations is selected, the proposed rating methodology can be applied to construct an accurate and objective LAD-based bank credit risk rating model. An LAD rating model allows the financial institution to rate its counterparts on a proactive basis, at the frequency it desires, without having to wait for sometimes untimely rating modification announcements from Fitch, and responds to the needs motivating the development of internal credit rating systems described in Section 2.

Two other points concerning the model and its use for the evaluation of banks' creditworthiness warrant mentioning. First, the current crisis is said to be primarily due to the inability of financial institutions to pay their liabilities when they are due (Schwarcz, 2008). New regulations (Basel Committee on Banking Supervision, 2008; Financial Stability Forum, 2008) have been and will be introduced to limit the liquidity risk. Our rating model uses as one of its predictors a key liquidity ratio (i.e., ratio of net loans to total assets) which Fannie Mae leadership admitted to have ignored (Bielski, 2008). The incorporation of additional liquidity variables or ratios would not detrimentally affect the proposed LAD method. Second, our model has been constructed on the basis of a wide and international data set composed of 800 banks whose inclusion is not based on the acceptance or not, by a certain financial institution, of past credit applications. This precludes the sampling bias issue, and the objectivity of the model is not compromised by moral hazard situations as those arising in the context of an internally developed model.

As mentioned in Section 3, the applicability of our credit risk rating methodology is not confined to bank obligors. The application of our approach to other types of obligors (individuals, countries, etc.) should account for the following points. The set of variables used to derive the ratings should be adjusted for each category of obligor. For example, it is not possible to use balance-sheet variables to assess the solvency of an individual obligor or of a country. Moreover, if some variables (i.e., return on equity) are considered as informative on the creditworthiness of an obligor regardless of the industrial sector in which it is active, others are sector specific (e.g., production indices or export levels can be significant for appraising the creditworthiness of companies in the manufacturing or consumer goods sectors, but not that of a bank).

The proposed rating method can be very helpful not only for banks, but for other types of companies as well. Insurance companies, one of the major players in the credit risk market (providing insurance

coverage, buying banks' credit risk derivatives, and therefore being indirectly exposed to many credit risks) also need to assess the risk of their operations. The exposure of insurers to high levels of uncertainty and insolvency, combined with the trend for insurers to expand their business frontier to include new insurance areas, has led insurance supervisory authorities across the world to reform the solvency system of insurance companies (Florez-Lopez, 2007). This has led the European supervisors to initiate the Solvency II project in 2002 and to develop directives regarding the financial resources, supervisory review and market discipline of the sector (European Commission Internal Market and Services, 2005). This approach, thus, could be very valuable for deriving credit risk rating models satisfying the conditions of the Solvency II directives.

Additionally, the contribution of the proposed LAD method towards the central banks' objective of reaching financial stability, defined as "the smooth functioning of the key elements that make up the financial system" (Oosterloo et al., 2007), could be noteworthy. The combinatorial nature of the LAD model makes it possible to capture higher-order interactions governing complex systems, which highlights the model's potential for being used as an early warning system (Estrella et al., 2002) for the detection of weak banks and the risk of a systemic bank crisis (King et al., 2004; Barnhill and Souto, 2008). In particular, the LAD model makes possible a refined evaluation of the financial situation of counterparties and a better understanding of the interconnectivity between financial institutions.

The applicability of the proposed model as an early warning system is extremely important in view of *(i)* the financial *costs* of a bank crisis: some examples of the magnitude of these costs are the so-called Savings and Loan crisis which resulted in $123.8 billion (2.1% of the 1990 GDP) of losses for the U.S. taxpayers (Curry and Shibut, 2000); the 1997-2002 Turkish bank crisis which led to restructuring costs of $16.9 billion (Canbas et al., 2005); the current financial crisis, the losses of which are predicted by the International Monetary Fund (2008) to approach $1 trillion; *(ii)* the steady *increase* of bank failures over the last 30 years in developing as well as in highly developed economies (Basel Committee on Banking Supervision, 2004); *(iii)* the risk of *spillover* across the whole financial system and economy. Indeed, the prevalence of high interbank linkages significantly increases the risk of spillover, which could result in a bank crisis. Thus, an insolvent bank unable to honor its obligations could precipitate financial distress for its counterparts. Besides, other cascading effects must also be considered that could exacerbate cyclical recessions and result in more severe financial crises (Basel Committee on Banking Supervision, 2004). The bankruptcy of a bank can provoke depositors from other banks to withdraw their funds, depleting banks'

25

capital. The current mortgage crisis provides an illustration that a downward business cycle can cause distress for companies, rendering many loans delinquent and causing banks to further reduce business lending, and that financial crises can extend to other sectors of the economy as the availability of credit may be disrupted. The possible contagion effect through the banking sector, and possibly through other sectors or economies, highlights the importance of developing predictive rating models as tools to appraise ex-ante the magnitude of the risks and to adopt proactive measures to manage the risks and alleviate the consequences.

9. CONCLUDING REMARKS AND FURTHER RESEARCH

The evaluation of the creditworthiness of banks and other financial organizations is very challenging (due to the opaqueness of the banking sector and the higher variability of its creditworthiness) and is extremely important (due to a growing number of banks going bankrupt, the magnitude of losses caused by such bankruptcies, the cascading effects that the failure or solvency issues of a bank can have on the whole financial system and economy). Credit risk rating systems play a fundamental role in the banks' operations pertaining to loan approval, management reporting, pricing, determination of the covenants and collaterals of the credit line, limit setting, and loan loss provisioning, among others. A common link between the above operations is the credit risk rating which affects each and every decision and operation of the financial institution throughout the life cycle of the granted credit.

This study is devoted to the problem of reverse-engineering the Fitch bank credit ratings, a problem, which -- in spite of its important managerial implications -- is generally overlooked in the extant literature. We present four approaches to address this problem, the first three being statistical (MLR, OLR) and machine learning (SVM) methods, while the last one (LAD) is a combinatorial pattern extraction method, which identifies strong combinatorial patterns distinguishing banks with high and low ratings. These patterns constitute the core of the rating model developed here for assessing the credit risk of banks.

The study shows then that ordered logistic regression and the LAD method can provide superior results (as compared to MLR and SVM) in reverse-engineering a popular bank rating system. It appears that, in spite of the widely differing nature of the two approaches, their results are in a remarkable agreement. In view of the essential differences in techniques, the conformity of bank ratings provided by LAD and by ordered logistic regression strongly reinforces the validity of these rating methods, and identifies financial variables that are key for evaluating the creditworthiness of banks.

Comparing the LAD and the OLR ratings with the Fitch ratings, and considering the associated

26

classification accuracy (i.e., the average difference between the ratings provided by these two approaches on one hand, and by the Fitch ratings on the other hand), we can see that the LAD method outperforms the OLR method. This result is very strong, since the critical component of the LAD rating system – the LAD discriminant – is derived utilizing only information about whether a bank's rating is "high" or "low", without the exact specification of the bank's rating category. Moreover, the LAD approach uses only a fraction of the observations in the dataset, since none of the banks to which Fitch assigns one of its three intermediate rating categories is used to derive the LAD model. As a contrast, the OLR model needs an extended input, requiring the knowledge of the precise Fitch rating category to which each bank belongs, and uses all the banks in the dataset to derive the rating model. The higher classification accuracy of LAD appears even more clearly when performing cross-validation and applying the LAD model derived from information about the banks in the training set to those in the testing set.

Besides its higher accuracy and robustness, the critical advantage of the rating system constructed using LAD is that it does not assume or constrain the effect of the predictor variables (e.g., ratio of equity to total assets) to be the same for each rating category. This is a major difference with the OLR model, and it is particularly important in view of the recent finding of Kick and Koetter (2007) that the individual impact of each banks' balance-sheet item differs across banks' credit risk rating categories.

Additionally, the proposed LAD method can actually be used to generate a rating system which exhibits the number of rating categories desired by its user. The varying granularity property is very appealing for risk managers in that it enables the LAD methodology to be used for different bank operations. The LAD-based rating model can take the form of a binary classification model, and can be used at the pre-approval stage (credit screening operation) to discriminate banks to which a credit line cannot be extended from those to which the granting of credit can be considered. Finer credit risk rating systems can be constructed to allow risk managers to further differentiate their customers and to tailor accordingly their credit pricing policies. The monetary value of a more granular rating system stems from allowing for a substantial decrease in the amount of regulatory capital (Jankowitsch et al., 2007).

The study also shows that the LAD-based approach to reverse-engineering bank ratings is: *(i) objective, (ii) transparent*, *(iii) generalizable,* and it provides a model that is *(iv) parsimonious* and *(v) robust*. The proposed method to construct risk rating systems provides managers with a powerful tool to help them decide whether or not, and under which conditions, a credit must be granted. Moreover, it can be

employed to develop internal rating systems that satisfy the Internal Rating Based requirements, and are Basel 2 compliant. The construction of a credit rating system with the above properties: *(i)* can be used to mitigate the financial and operational risk (Jankowitsch et al., 2007, Stein, 2005) in a financial institution; and *(ii)* has a broad applicability scope, since it can be used by other entities (insurance companies, regulators) and for pursuing other objectives (financial stability, early warning system).

Additional empirical research will be needed to check how the current rating system would have performed during the current crisis, and to assess whether the higher flexibility (varying number of categories, etc.) of the LAD-based approach generates added value, and if yes, its extent. New regulations on internal rating systems are likely to be imposed in the near future. It will be interesting to verify how this will affect the set of variables used in the construction of the LAD ratings.

Another avenue for future research concerns the so-called CAMELS rating system. The CAMELS rating is a US supervisory rating (Lopez, 1999) that evaluates the overall condition of a bank based on its financial statements and on-site examination by regulators (i.e., Federal Reserve, Office of the Comptroller of the Currency, Federal Deposit Insurance Corporation). The acronym CAMELS refers to the six components that are evaluated: (C) capital adequacy; (A) asset quality; (M) management; (E) earnings, (L) liquidity, and (S) sensitivity to market risk. There are five rating categories (from 1 to 5) with 1 being the strongest and 5 being the weakest. These ratings are not released to the public but only to the top management of the bank to avoid a run on a bank with a weak CAMELS rating. An interesting research direction would be to verify the predictive power of the CAMELS ratings for the financial meltdown and to derive the LAD-based ratings on their basis.

REFERENCES

Alexe S. 2007. *LAD - Datascope V2.1 Software Package*.

Alexe G., Alexe S., Bonates T.O., Kogan A. 2007. Logical Analysis of Data – the Vision of Peter L. Hammer. *Annals of Mathematics and Artificial Intelligence* **49**, 265-312.

Alexe G., Alexe S., Kogan A., Hammer P.L. 2008. Comprehensive vs. Comprehensible Classifiers in Logical Analysis of Data. *Discrete Applied Mathematics* 156, 870-882.

Alexe S., Hammer P.L. 2006. Accelerated Algorithm for Pattern Detection in Logical Analysis of Data. *Discrete Applied Mathematics* **154**, 1050-1063.

Angluin D., Laird P. 1988. Learning from Noisy Examples. *Machine Learning* **2**, 343-370.

Arena M. 2008. Bank Failures and Bank Fundamentals: A Comparative Analysis of Latin America and East Asia during the Nineties Using Bank-Level Data. *Journal of Banking & Finance* **32**, 299-310.

Barnhill T.M. Jr., Souto M.R. 2008. Systemic Bank Risk in Brazil: An Assessment of Correlated Market, Credit, Sovereign, and Inter-Bank Risk in an Environment with Stochastic Volatilities and Correlations: http://ideas.repec.org/p/zbw/bubdp2/7323.html

Basel Committee on Banking Supervision. 2001. The Internal Ratings Based Approach: http://www.bis.org/publ/bcbsca05.pdf

Basel Committee on Banking Supervision. 2004. Bank Failures in Mature Economies. *Working Paper* **13**

Basel Committee on Banking Supervision. 2006. International Convergence of Capital Measurement and Capital Standards: A Revised Framework (Basel II): http://www.bis.org/publ/bcbs128.htm

Basel Committee on Banking Supervision. 2008. Principles for Sound Liquidity Risk Management and Supervision: http://www.bis.org/publ/bcbs138.pdf

Bielski L. 2008. Credit Risk Management Sunk by Crisis? *ABA Banking Journal.* **December 2008**, 31-34.

Boros E., Hammer P.L., Ibaraki T., Kogan A. 1997. Logical Analysis of Numerical Data. *Mathematical Programming* **79**, 163-190.

Boros E., Hammer P.L., Ibaraki T., Kogan A., Mayoraz E., Muchnik I. 2000. An Implementation of Logical Analysis of Data. *IEEE Transactions on Knowledge and Data Engineering* **12** (2), 292-306.

Canbas S., Cabuk A., Kilic S.B. 2005. Prediction of Commercial Bank Failure via Multivariate Statistical Analysis of Financial Structures: The Turkish Case. *European Journal of Operational Research* **166**, 528-546.

Conn A., Gould N., Toint P. 1990. An Introduction to the Structure of Large Scale Nonlinear Optimization Problems the LANCELOT Project. In: *Computing Methods in Applied Sciences and Engineering.* Eds: Glowinski R., Lichnewsky A., SIAM, Philadelphia, PA, 42-54.

Crama Y., Hammer P.L., Ibaraki T. 1988. Cause-Effect Relationships and Partially Defined Boolean Functions. *Annals of Operations Research* **16**, 299-326.

Curry T., Shibut L. 2000. Cost of the S&L Crisis. *FDIC Banking Review* **2** (2).

Derviz A., Podpiera J. 2004. Predicting Bank CAMEL and S&P Ratings: The Case of the Czech Republic. *Working Paper Series of the Czech National Bank*.

Estrella A., Park S., Peristiani S. 2002. Capital Ratios and Credit Ratings as Predictors for Bank Failures. *Federal Reserve Bank of New York: Economic Policy Review*, 33-52.

European Commission Internal Market and Services. 2005. *Solvency II Roadmap—towards a Framework Directive*, MARKT/2502/05, European Commission Internal Market and Services DG, Brussels, March.

Ferri G., Liu L-G., Stiglitz J. 1999. The Procyclical Role of Rating Agencies: Evidence from the East Asian Crisis. *Economic Notes* **3**, 335-355.

Financial Stability Forum. 2008. *Report of the Financial Stability Forum on Enhancing Market and Institutional Resilience:* http://www.fsforum.org/publications/r_0804.pdf

Fitch Ratings. 2001. Fitch Simplifies Bank Rating Scales. *Technical Report.*

Fitch Ratings. 2006. The Role of Support and Joint Probability Analysis in Bank Ratings. *Fitch Report*.

Fitch Ratings, Bureau Van Dijk. 2009. The Fitch Universal Format on BankScope. *White Paper*: http://www.bvd.co.uk/bankscope/bankscope.pdf

Florez-Lopez R. 2007. Modelling of Insurers' Rating Determinants. An Application of Machine Learning Techniques and Statistical Models. *European Journal of Operational Research* **183**, 1488-1512.

Galindo J., Tamayo P. 2000. Credit Risk assessment Using Statistical and Machine Learning/Basic Methodology and Risk Modeling Applications. *Computational Economics* **15**, 107-143.

Grunert J., Norden L., Weber M. 2005. The Role of Non-Financial Factors in Internal Credit Ratings. *Journal of Banking & Finance* **29**, 509-531.

Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.W. 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explorations* 11 (1), 10-18.

Hammer P.L. 1986. Partially Defined Boolean Functions and Cause-Effect Relationships. *International Conference on Multi-Attribute Decision Making Via OR-Based Expert Systems*. Passau, Germany.

Hammer P.L., Bonates T.O. 2006. Logical Analysis of Data: From Combinatorial Optimization to Medical Applications. *Annals of Operations Research* **148** (1), 203-225.

Hammer P.L., Kogan A., Lejeune M.A. 2006. Modeling Country Risk Ratings Using Partial Orders. *European Journal of Operational Research* **175** (2), 836-859.

Hjorth J.S.U. 1994. *Computer Intensive Statistical Methods Validation, Model Selection, and Bootstrap*. London, Chapman & Hall.

Huang Z., Chen H., Hsu C.-J., Chen W.-H., Wu S. 2004. Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support Systems* **37**, 543-558.

International Monetary Fund - International Capital Markets Department on Market Developments and Issues. 2008. *Global Financial Stability Report.*

Jain K., Duin R, Mayo J. 2000. Statistical Pattern Recognition: A Review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 4-37.

Jankowitsch R., Pichler S. Schwaiger W.S.A. 2007. Modelling the Economic Value of Credit Rating Systems. *Journal of Banking & Finance* **31**, 181-198.

Kearns M. 1998. Efficient Noise-Tolerant Learning from Statistical Queries. *Journal of the Association for Computing Machinery* **45** (6), 983-1006.

Kick T., Koetter M. 2007. Slippery Slopes of Stress: Ordered Failure Events in German Banking. *Journal of Financial Stability* **3**, 132-148.

King T.B., Nuxoll D.A., Yeager T.H. 2004. Are the Causes of Bank Distress Changing? Can Researchers Keep Up? *Federal Reserve Bank of St. Louis, Supervisory Policy Analysis Working Paper* **2004-07**.

Little R.J.A. 1988. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* **83**, 1198-1202.

Lopez J.A. 1999. Using CAMELS Ratings to Monitor Bank Conditions. *Federal Reserve Bank of San Francisco Economic Letter* **99-19**.

Morgan D.P. 2002. Rating Banks: Risk and Uncertainty in an Opaque Industry. *The American Economic Review* **92** (4), 874-888.

Oosterloo S., de Haan J., Jong-A-Pin R. 2007. Financial Stability Reviews: A First Empirical Analysis. *Journal of Financial Stability* **2**, 337-355.

Poon W.P.H., Firth M., Fung H.-G. 1999. A Multivariate Analysis of the Determinants of Moody's Bank Financial Strength Ratings. *Journal of International Financial Markets, Institutions & Money* **9** (3), 267-283.

Platt J.C. 1998. Fast Training of Support Vector Machines using Sequential Minimal Optimization. In: *Advances in Kernel Methods - Support Vector Learning*. Eds: Scholkopf B., Burges C., Smola A.J. MIT Press, 185-208.

Platt J.C. 1999. Probabilities for SV Machines. In: *Advances in Large Margin Classifiers*. Eds: Smola A.J., Bartlett P.J., Scholkopf B., Schuurmans D. MIT Press, 61-73.

Ravi Kumar P., Ravi V. 2007. Bankruptcy Prediction in Banks and Firms via Statistical and Intelligent Techniques – A Review. *European Journal of Operational Research* **180**, 1-28.

Reuters. 2008. S.E.C. Criticizes Ratings Agencies' Conflicts of Interest. *New York Times*, July 8, 2008: http://dealbook.blogs.nytimes.com/2008/07/08/sec-criticizes-ratings-agencies-conflicts-of-interest/

Rubin D. B. 1976. Inference and Missing Data. *Biometrika* **61**, 581-592.

Ryoo H. S., Jang I.-Y. 2009. MILP Approach to Pattern Generation in Logical Analysis of Data. *Discrete Applied Mathematics* **157** (4), 749-761.

Sarkar S., Sriram R.S. 2001. Bayesian Models for Early Warning of Bank Failures. *Management Science* **47** (11), 1457–1475

Schwarcz S. 2008. Understanding the 'Subprime' Financial Crisis: http://ssrn.com/abstract=1288687

De Servigny A., Renault O. 2004. *Measuring and Managing Credit Risk*. McGraw-Hill, New York, NY.

Stein R.M. 2005. The Relationship between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan Pricing. *Journal of Banking and Finance* **29**, 1213-1236.

Tabakis E., Vinci A. 2002. Analyzing and Combining Multiple Credit Assessments of Financial Institutions. *European Central Bank – Working Paper Series*.

Treacy W.F., Carey M.S. 2000. Credit Risk Rating Systems at Large US Banks. *Journal of Banking & Finance* **24** (1-2), 167-201.

Securities and Exchange Commission. 2003. *Report on the Role and Function of Credit Rating Agencies in the Operation of the Securities Markets*. http://www.sec.gov/news/studies/credratingreport0103.pdf

Wang H., Weigend A.S. 2004. Data Mining for Financial Decision Making. *Decision Support Systems* **37**, 457-460.

Zhu X., Wu X. 2004. Class Noise vs. Attribute Noise: A Quantitative Study of their Impacts. *Artificial Intelligence Review* **22**, 177-210.

APPENDIX

Table 7: Fitch Individual Rating System (Fitch Ratings, 2001)

Category	Numerical Scale	Description
A	9	A very strong bank. Characteristics may include outstanding profitability and balance sheet integrity, franchise, management, operating environment, or prospects.
B	7	A strong bank. There are no major concerns regarding the bank. Characteristics may include strong profitability and balance sheet integrity, franchise, management, operating environment or prospects.
C	5	An adequate bank which, however, possesses one or more troublesome aspects. There may be some concerns regarding its profitability, balance sheet integrity, franchise, management, operating environment or prospects.
D	3	A bank which has weaknesses of internal and/or external origin. There are concerns regarding its profitability, management, balance sheet integrity, franchise, operating environment or prospects.
E	1	A bank with very serious problems which either requires or is likely to require external support.

In addition, Fitch uses four gradations between the five ratings above: A/B, B/C, C/D and D/E, the corresponding numerical values of which being respectively 8, 6, 4 and 2. This conversion of the Fitch individual bank ratings into a numerical scale is not specific to us. Poon et al. (1999) proceed similarly for Moody's bank strength financial ratings.

Table 8: Cutpoints

Numerical Variables	Cutpoints		Numerical Variables	Cutpoints		Numerical Variables	Cutpoints
Country Risk Rating	11.5 , 16 , 19.5 , 20		Overhead	127 , 846		Non Int Exp / Avg Assets	2.77 , 3.71 , 4.93
Other Earning Assets	1661		Profit before Tax	24.8 , 111.97		Return on Average Assets	0.30 , 0.80

Loans	3135	Net Interest Revenue	816 , 2150	Return on Average Equity	4.91 , 11.82 , 15.85 , 19.23
Non-Earning Assets	364	Equity / Total Assets	4.90 , 6.28 , 9.38	Cost to Income Ratio	64.12 , 71.92
Other Operating Income	47 , 155	Net Interest Margin	1.87 , 341, 4.45	Net Loans / Total Assets	44.95 , 66.50
Other Operating Inc / Avg Assets	1.28 , 1.86	Net Int Rev / Avg Assets	2.42		

Table 9: Discriminant/Output Probability Values

		LAD		SVM	
Observation Class	Rating Category	Discriminant Values / Category Averages	Discriminant Values / Class Weighted Averages	Probability Values / Category Averages	Probability Values / Class Weighted Averages
	A	0.353		0.985	
	A/B	0.323	0.326	0.956	0.948
	B	0.303		0.942	
	B/C	0.136		0.756	
	C	0.005	0.006	0.496	0.545
	C/D	-0.122		0.269	
	D	-0.258		0.059	
	D/E	-0.264	-0.265	0.067	0.078
	E	-0.273		0.158	

Table 10: Rating Thresholds

i	0	1	2	3	4	5	6	7	8	9
x_i	-1	-0.338	-0.263	-0.218	0.002	0.116	0.277	0.351	0.407	1

Table 11: Discrepancy Analysis

Method j	Model	k									D_j
		0	1	2	3	4	5	6	7	8	
		$n_k/|N|$									
LAD	Fitted Model	29.25%	52.75%	13.50%	3.87%	0.63%	0.00%	29.25%	0.00%	0.00%	0.939
	Average over 10 2-Folding Experiments	29.80%	51.50%	12.50%	4.40%	1.20%	0.60%	0.00%	0.00%	0.00%	0.975
MLR	Fitted Model	27.75%	33.13%	30.25%	7.50%	1%	0.25%	0.13%	0%	0.00%	1.22
	Average over 10 2-Folding Experiments	21.75%	27.38%	24.25%	20.38%	5.50%	0.50%	0.25%	0%	0%	1.63
OLR	Fitted Model	36.50%	37.50%	17.13%	6%	2.25%	0.25%	0.38%	0.00%	0.00%	1.02
	Average over 10 2-Folding Experiments	30.75%	38.875%	17.75%	8.375%	3.50%	0.50%	0.25%	0.00%	0.00%	1.175
SVM	Fitted Model	17.00%	31.375%	30.25%	12.25%	6.00%	2.375%	0.375%	0.25%	0.125%	1.695
	Average over 10 2-Folding Experiments	16.875%	34.375%	31.125%	8.875%	1.875%	2.75%	3.5%	0.5%	0.125%	1.7

Table 12: LAD Model

Patterns	Country Risk rating	Loans	Other Earning Assets	Non-Earning Assets	Net Interest Revenue	Other Operating Income	Overheads	Profit before Tax	Equity / Total Assets	Net Interest Margin	Net Int Rev / Avg Assets	Oth Op Inc / Avg Assets	Non Int Exp / Avg Assets	Return on Average Assets (ROAA)	Return on Average Equity (ROAE)	Cost to Income Ratio	Net Loans / Total Assets
P1	>16														>11.82	≤71.92	
P2	>19.5											>1.28			>4.91		
P3	>11.5					>127									>15.85		
P4	>16		>1661								>2.42						
P5	>19.5								>4.9				≤3.71				
P6	>19.5									>1.87			≤4.93				
P7	>19.5							>24.8								≤11.82	
P8	>16								≤6.28					>0.30			
P9	>16	>3135					≤127										
P10								>111.97					≤3.71				≤44.95
N1	≤16				≤47												
N2	≤11.5									>1.872							
N3	≤20													≤0.30			≤66.5
N4	≤16														≤19.23	≤64.12	
N5						>127			≤4.9					≤0.30			
N6				≤364									>2.77	≤0.80			
N7						>127						≤1.86				>71.92	
N8	≤19.5				≤2150										≤11.82		
N9					≤155			≤24.8		≤4.45							

34