$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2008-14

September 1, 2008

Reverse-Engineering Banks' Financial Strength Ratings Using Logical Data Analysis

Peter L. Hammer
RUTCOR
Rutgers University, USA

Alexander Kogan
Rutgers Business School
Rutgers University, USA

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

Reverse-Engineering Banks' Financial Strength Ratings Using Logical Analysis of Data[1]

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

Abstract: We study the problem of evaluating the creditworthiness of banks using statistical, as well as combinatorics-, optimization-, and logic-based methodologies. We reverse-engineer the Fitch credit risk ratings of banks using ordered logistic regression and Logical Analysis of Data (LAD). It is shown that LAD provides the most accurate rating model. The obtained ratings are successfully cross-validated, and the derived model is used to identify the financial variables most important for bank ratings. The study also 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, allowing their use in the banks' operations related to the credit granting decision process (pre-approval, determination of pricing policies), and *(v)* are Basel 2 compliant, allowing their use in the financing and provisional decisions pertaining to the determination of the amount of regulatory capital. The impact on the reduction of the credit and operational risk, the extended scope, and the monetary value of such a risk rating system are discussed.

Keywords: credit risk rating, bank creditworthiness, Logical Analysis of Data, combinatorial pattern extraction, Basel II, operational efficiency

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*,

[1] Acknowledgment: The authors are grateful to the editors and referees for their very insightful comments. The authors express their appreciation to Dr. Sorin Alexe for providing the LAD – Datascope 2.01 software package to carry out the computational experiments.

[2] Peter L. Hammer tragically died in a car accident on December 27, 2006, after this manuscript was essentially completed.

[3] kogan@rutgers.edu; Rutgers Business School, Rutgers University, 180 University Avenue, Newark, NJ, USA
[4] mlejeune@gwu.edu; George Washington University, DC, USA

(*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. On the other hand, it also shows that ordered logistic regression can provide excellent results in reverse-engineering a bank rating system. Moreover, the study demonstrates that the substantively different methodology of LAD can be utilized in reverse-engineering a superior bank rating system, which turns out to provide remarkably similar results 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 higher discrimination power, i.e. with higher 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 internal, 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 11 in the electronic companion) whose prediction accuracy is verified for financial institutions spread across the world. In the existing 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 describes the interface between credit risk rating systems and banks' operations, and the impact rating systems have on the banks' profitability and their credit and operational risk. In Section 3, we present the data used in this study. Section 4 contains the results of estimating a multiple linear regression model and an ordered logistic regression model, as well as the evaluation of their accuracy and generalizability. Section 5 provides a general overview of the concepts and techniques of the logical analysis of data methodology used for reverse-engineering bank ratings. Section 6 develops an LAD model for discriminating banks with high and low ratings, assesses the accuracy and generalizability of the LAD model, presents a procedure for mapping the LAD numerical values reflecting the creditworthiness of a bank to a credit risk rating, and discusses the key distinctive features of the LAD rating methodology. Section 7 describes the results of evaluating the conformity of the LAD bank ratings with the original Fitch ratings. Section 8 analyzes the extended scope of a forward-looking credit risk rating system. Finally, concluding remarks are presented in Section 9.

2. BANKS' OPERATIONS AND CREDIT RISK RATING SYSTEMS

2.1. Motivation for 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, 2007). 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) which some explain by the presence of conflicts of interest (Financial Stability Forum, 2007, 2008). In his March 2006 testimony before the Senate Banking Committee, Jeff Diermeyer, the president of the CFA Institute, regretted that rating agencies "have been reluctant to embrace any type of regulation over the services they provide" (Wall Street Letter, 2006).

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 ratings 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 to be accepted by legislator, and qualifies banks for the adoption of the Internal Rating Based approach (Basel Committee on Banking Supervision, 2001, 2006a) to calculate their capital provisions. The Financial Stability Forum endorsed the principles for banking supervision defined by the Basel Committee of Banking Supervision (2006b) as crucial for having sound financial systems and as deserving “priority implementation”.

Third, an internal rating system provides autonomy to a bank’s management to execute credit risk policies in accord with the bank’s own core business and goals. Aside from the Basel II requirements, the reliance on internally developed credit risk models makes credit operations more transparent and efficient.

2.2. Use of credit risk rating system in banks’ operations

Credit risk rating systems play a pervasive role in the operations of a bank (Tracey and Carey, 2000; Grunert et al., 2005). We describe below the role played by internally developed credit risk rating systems in the loan approval, management reporting, pricing, limit setting, and loan loss provisioning operations of financial institutions. Clearly, the credit risk rating system has a tremendous impact on the operational risk of a financial institution, which is defined by the Basel Bank of International Settlements and its Committee on Banking Supervision (2006a) as “the risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events", Jarrow (2008) partitions operational risk into two main types. Type one concerns the risk of a loss due to the firm's operating system (i.e., a failure in a transaction, loan, or investment), while Type two is the risk of a loss due to managerial incentives and is an agency cost.

In credit pre-approval operations, a binary decision model using the credit risk rating of the obligor as the key explanatory variable is typically utilized to pre-approve or deny the granting of the credit. If the case is pre-approved, the credit risk rating is subsequently used to define the conditions (underwriting limits, maturity, interest rate, covenants, collaterals) under which the final approval can be given (Treacy and Carey, 2000). Evidently, the rating assignment, approval, and underwriting processes are closely intertwined: the credit risk rating is reviewed at each stage of the underwriting or credit approval operations. This illustrates the impact that an accurate credit risk rating system has on the operational risks of Type one (Jarrow, 2008). Credit risk ratings are also used to report to senior management and trustees various key metrics such as risk positions, loan loss reserves, economic capital allocation, profitability, and employee compensation (Treacy and Carey, 2000).

A transparent and accurate credit risk rating system is a powerful tool to hedge against certain moral hazard situations and the resulting operational risk. Some financial institutions do not have a clearly specified credit risk rating model, and allow lending officers to assign credit ratings based on their judgment. The lending officers are in charge of the marketing of banking services, and their performance (and therefore compensation) is determined with respect to the "profitability" of the relationships between the bank and its customers. Clearly, the credit risk ratings assigned by the lending officers will affect the volume of the approved loans or credits as well as the compensation of the officer who thus may have an incentive to assign ratings in a way that is not consistent with the employer's interests. Hence, an objective and transparent credit risk rating system can reduce the occurrence of such perverse incentive situations and mitigate the operational risk of Type two as defined by Jarrow (2008).

There is clearly a continuum between the assignment of a rating, the credit approval, pricing and monitoring operations, and financial, i.e., capital provisioning, decisions. Once a rating is assigned to a potential borrower and the subsequent credit approval decisions and pricing operations have been carried out, additional financial decisions are to be taken and operations executed. The expected loss of the granted credit is computed and is then used to calculate the amount of economic and regulatory capital that a bank must keep to hedge against possible defaults of the borrower. The expected loss of a credit facility is an increasing function of the probability of default of the borrower to which the credit line is granted, as well as the exposure at default and the loss given default associated with the credit facility. Since there is usually a mapping between credit risk rating and the borrower's probability of default, and provided that the rating of a borrower is known to be a key predictor in assessing the recovery rate associated with a credit facility granted to this borrower, the importance of the credit risk rating in calculating the amounts of regulatory and economic capital is evident.

2.3. Impact of credit risk rating system on banks' operations and profitability

It is important to note that credit risk ratings are not used on a ”point-in-time” basis, but rather are used and influence the operations of the financial institution throughout the life cycle of the granted credit. Once a rating is assigned and the credit is granted, the credit supervisor becomes responsible for the monitoring of the credit and for modifying the rating promptly if the creditworthiness of the borrower changes (Treacy and Carey, 2000). The monitoring of the performance and the early detection of a possible worsening of the supervised bank can be carried out through frequent supervisory examinations (Gilbert, 1993). However, such on-site examination operations are extremely time-consuming and costly. Hence, the reliance upon a highly predictive rating system for bank creditworthiness or failure would not constrain banks to resorting to on-site examinations and is a primary lever to lower the costs incurred by such operations (Derviz and Podpiera, 2004).

An improvement of the prediction accuracy of a rating system decreases the potential effects of adverse selection (Jankowitsch et al., 2007) and significantly enhances the pricing operations and the revenues they generate (Stein, 2003; Stein and Jordão, 2003). It was demonstrated (Jankowitsch et al., 2007) that, in a competitive framework, a poor statistical power of a bank's rating system lessens the economic performance because of adverse selection. Customers with a better credit quality than that assessed by the bank will migrate to another institution whose rating system recognizes their better credit quality, and the bank will be left with a portfolio of customers with a credit quality lower than estimated. A rating system with superior accuracy has a strong positive impact on economic performance, the extent of which depends mainly on the competitiveness of the market environment and the customers' awareness of their actual credit quality. As for the effects of a better rating system on the customers' selection and loan pricing operations, Jankowitsch et al. (2007) show that the switch from a rating system with low accuracy to one with medium accuracy increases the annual rate of return on average by 30 to 40 basis points. The effect is particularly strong in very competitive markets and with high loss given default rates. The monetary impact of an objective and highly predictive credit risk rating system is enormous as shown by Moody's KMV (Stein, 2003; Stein and Jordão, 2003) and Jankowitsch et al. (2007).

Another way to understand the critical importance of a credit risk rating system is to look at the magnitude of the provisions made by financial institutions to cover potential operational losses. The Basel Committee's Risk Management Group (2003) found that, on average, banks allocate 15% of their capital to hedge against operational risk exposure, while the allocation of the total financial risk of a bank for the coverage of the operational risk is estimated to be 20%, 25%, and 35%, by Crouhy et al. (2001), Jorion (2000), and Cruz (2002), respectively. de Fontnouvelle et al. (2005) report that internationally active banks allocate annually roughly \$2-\$7 billion for operational risk, and that the capital charge for operational risk will often exceed the charge for market risk **(**see also de Fontnouvelle et al., 2006; Chapelle et al., 2008 for an analysis of the impact of operational risk).

As mentioned above, the bank's rating system is used for calculating the expected loss and the regulatory capital requirements set by the Basel II Approach (Basel Committee on Banking Supervision, 2006a). Jankowitsch et al. (2007) show that the economic value of a better credit risk rating system is also driven by its effect on the reduction of the amount of regulatory capital that is needed to meet the requirements imposed by Basel. Also, a more accurate rating system makes it possible to introduce additional rating categories, and a finer grained rating system will lead to lower capital requirements due to the concavity of the regulatory capital formula. Empirical tests (Jankowitsch et al., 2007) show that the switch from a five-category (which is the number of rating categories required by the Basel II standardized approach) rating system to one containing ten categories allows an average reduction of the amount of regulatory capital of about 20 basis points and can reach up to 31 basis points. Similarly, in earlier studies, Moody's KMV (Stein, 2003; Stein and Jordão, 2003) shows that rating models with higher predictive capability can generate dramatic rise in the performance of a bank. In particular, they consider a bank with \$50 billion of assets for which a five-point-higher accuracy ratio (Engelmann et al., 2003) of its credit risk rating system is shown to trigger an increased profitability ranging between \$2.1 and \$4.8 million per year.

3. BANKS' CREDIT RISK RATINGS: PRIOR RESEARCH, EXPLANATORY VARIABLES, DATA

Our purpose is to construct a forward-looking rating system allowing for the classification of banks with respect to their risk of defaulting over a given time horizon. 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).

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) to develop 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.

3.1. 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. Griffiths and Beynon (2005) use machine learning to identify the variables accounted for in Moody's Bank Financial Strength Rating.

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 such 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 financial predictors, and reject the use of binary or ordered logit regression models to assess the bank distress/health 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.

3.2. 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 12 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.

3.3. 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, net interest revenue, customer and short-term funding, overheads, equity, net income, total liability and equity, operating income) and 9 representative financial ratios. These ratios are defined in the Electronic Companion and describe:

- **asset quality:** *ratio of equity to total assets;*
- **operations**: *net interest margin; ratio of interest income to average assets*; *ratio of other operating income to average assets; ratio of non-interest expenses to average assets; return on average assets (ROAA)*; *return on average equity (ROAE)*; *cost to income ratio;*
- **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, which is the largest existing bank database. As an additional variable, we use in this study the S&P risk rating of the country where the bank is located. The S&P country risk rating scale comprises twenty-two different 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) and Sy (2004). Moreover, Bloomberg, a major provider of financial data, developed a standard cardinal scale for comparing Moody's, S&P's and Fitch ratings.

3.4. Observations

Our dataset consists of eight hundred banks rated by Fitch and operating in 70 different countries. The values of the ratings were collected at the end of 2001. Table 11 provides the geographic distribution of the banks included in the dataset. Table 1 lists the number and 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: Distribution of Banks in Rating Categories

Rating Categories	**A**	**A/B**	**B**	**B/C**	**C**	**C/D**	**D**	**D/E**	**E**
Number of Banks	19	60	203	129	121	77	93	66	32
Percentage of banks	2.375%	7.5%	25.375%	16.125%	15.125%	9.625%	11.625%	8.25%	4%

4. REGRESSION MODELS

The first issue 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. Moreover, the model built in this section will serve later as a benchmark to evaluate the qualities of other rating models.

4.1. Multiple Linear Regression Model

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 12. The 24 independent variables described in the previous section are denoted by $x_1,\ldots,x_{24}$. The estimation of the standard regression model gives the results shown in Table 13. It can be seen that while the regression model is statistically significant, its goodness-of-fit as measured by the R-square is equal to only 54.9%, which is not sufficiently high to be used as an accurate model of bank ratings.

To assess the generalizability and predictive value of the regression model, we use the technique of 2-fold cross-validation (see e.g., Hjorth, 1994), i.e., we randomly partition the set of 800 banks into 2 disjoint subsets of 400 banks each. First, we remove one of the subsets, estimate the regression model on the remaining 50% of the data, and then evaluate the R-square of the estimated regression model on the removed subset. Then we reverse the role of the two subsets, repeat the calculations, and calculate the average of the two R-squares. Finally, the 2-fold cross-validation process is repeated 10 times, using each time another randomly generated equipartition of the dataset. The average R-square obtained in these 10 2-fold cross-validation experiments turns out to be 0.5161 with the standard deviation of 0.145. The results above show that while the regression model does not seem to be overfitting the data, it has only a 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.

Denoting by $\hat{y}_j$ bank j's rating predicted using the regression model, by $\overline{y_j}$ the rounding of $\hat{y}_j$ to its closest integer value and by r_j the rating given by Fitch to bank j, we compute the difference d_j between the rounded regression rating and Fitch's rating: $d_j = |\,\hat{y} - r_j|,\ \ j \in N$. Now we can calculate the *discrepancy count* $n_k = \left|\{j \mid d_j = k, j \in N\}\right|, k = 0,1,\ldots,8$, which represents the number of banks whose integer-rounded rating category predicted by a rating model differs from the actual Fitch rating by exactly k categories. In Table 2, we report the percentage n_k/N of banks for which the difference between the integer-rounded regression rating and Fitch's rating is equal to *k, k=0,1,...,8*. Row 2 and Row 3 in Table 2 report these numbers for the whole data set and for the testing sets, respectively.

Using the regression model constructed on the whole data set, the percentage of banks precisely classified (k = 0) is equal to 27.75%, while the average number of precise classifications on the test sets is much lower, amounting only to 21.75%. On the average, the difference between the actual and predicted rating categories for the banks in the training set is 1.22, while this difference for the test sets is 1.63.

Table 2: Analysis of Discrepancies

k	0	1	2	3	4	5	6	7	8
Data Set	27.75%	33.13%	30.25%	7.50%	1%	0.25%	0.13%	0	0
Test Sets: Average over 10 2-Folding Experiments	21.75%	27.38%	24.25%	20.38%	5.50%	0.50%	0.25%	0	0

4.2. Ordered Logistic Regression Model

The ordered logistic regression model fits a common slopes cumulative model, which is a parallel lines regression model based on the cumulative probabilities of the response categories rather than on their individual probabilities. A k-category ordered logistic model has the form:

$$\begin{aligned}
&\log\left(\frac{p_1}{1-p_1}\right)=\alpha_1+\sum_{i\in I}\beta_i x_i\\
&\log\left(\frac{p_1+p_2}{1-p_1-p_2}\right)=\alpha_2+\sum_{i\in I}\beta_i x_i\\
&\vdots\\
&\log\left(\frac{p_1+p_2+\ldots+p_k}{1-p_1-p_2-\ldots-p_k}\right)=\alpha_k+\sum_{i\in I}\beta_i x_i
\end{aligned} \tag{1}$$

with $p_1+p_2+\ldots+p_k=1$. We denote by p_i the individual probability of category i, by $\log\left(\sum_{i=1}^{r}p_i/(1-\sum_{i=1}^{r}p_i)\right)$, 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) that are category-invariant, and by α_j the intercepts which are category-specific and satisfy the constraints: $\alpha_1 \le \alpha_2 \le \ldots \le \alpha_{k-1} \le \alpha_k$.

Like dichotomous logistic regression, an ordered logistic regression model is estimated using maximum likelihood methods, to find the best set of regression coefficients to predict the values of the logit-transformed probability that the dependent variable falls into one category rather than another. Logistic regression assumes that if the fitted probability p is greater than 0.5, the dependent variable should have the value 1 rather than 0. Ordered logit doesn't have such a fixed assumption; it does instead fit 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, in the ordered logistic regression model (1), 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.

The most useful tool to assess the quality of the constructed ordered logistic regression model is by comparing actual group membership with the membership predicted by the model, as summarized in the analysis of discrepancies table. Below, we provide the analysis of discrepancies for the problem at hand, in which we have 9 discrepancy levels ($k=0,1,\ldots,8$) and 24 independent variables ($i=1,\ldots,24$). As we did for the linear regression model, we apply the technique of 2-fold cross-validation, and repeat it 10 times.

Table 3: Analysis of Discrepancies

k	0	1	2	3	4	5	6	7	8
Data Set	36.50%	37.50%	17.13%	6%	2.25%	0.25%	0.38%	0%	0%
Test Sets: Average over 10 2-Folding Experiments	30.75%	38.88%	17.75%	8.38%	3.50%	0.50%	0.25%	0	0

It can be seen in Table 3 that the percentage of banks precisely classified (k = 0) with the ordered logistic model built using the entire data set is equal to about 37%, while the average number of precise classifications on the test sets is lower, amounting to almost 31%. On the average, the difference between the actual and predicted rating categories for the banks in the data set is 1.02, while this difference for the test sets is 1.175. The Spearman rank correlation between the ordered logistic regression ratings and the Fitch ratings is equal to 75.60%.

Clearly, the ratings provided by this model are of good quality, thus proving the appropriateness of the selected set of 24 variables for the rating of bank creditworthiness. In the next sections, we shall describe a new type of rating model in which the rating of each bank is accompanied by a justification of why this bank has not been given a higher or a lower rating. Before constructing this new model, we shall briefly outline the logical analysis of data methodology on which this rating model will be based.

5. 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., 2004; 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 to extract models constructed with the help of a limited number of significant combinatorial patterns generated in this way (Boros et al., 2000). We sketch below very briefly the basic concepts of LAD, referring the reader for a more detailed description to Boros et al. (2000), or to the recent surveys (Hammer and Bonates, 2006; 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, beside the values of the n components of this vector, an additional binary (0, 1) value is also specified; this additional value is called the *output* or the *class* of the observation, with the convention that 0 is associated to negative observations, and 1 to the positive ones.

The purpose of LAD is to discover a binary-valued function f depending on the n input variables, which provides a 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* which show whether the values the variables take in a particular observation are "large" or "small"; more precisely, each indicator variable shows whether the value of a numerical variable does or does not exceed a specified level, called a *cutpoint*. For example, to the numerical variable *ratio of costs to incomes*, we associate in our banking model an indicator variable showing 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 with ratings of 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 a number of 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. Furthermore, good models tend to minimize the number of points in the dataset covered simultaneously by both positive and negative patterns in the model. It will be seen in the next section that the model we have found for the bank rating problem consists of 10 positive and 9 negative patterns, and that every negative or positive observation in the model is covered by some of the positive, respectively negative, patterns included 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 a dataset in the form of a *classification matrix* (Table 4).

Table 4: Classification Matrix

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

The value a (respectively d) represents the percentage of positive (negative) observations that are correctly classified. The value c (respectively b) is the percentage of positive (negative) observations that are misclassified. The value e (respectively f) is the percentage of positive (negative) observations that remain unclassified. Clearly, $a+c+e=100\%$ and $b+d+f=100\%$. The quality of the classification is defined by:

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

6. LAD MODEL FOR BANK RATINGS

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 recall that we have defined 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 of D, D/E or E. The binarization (Section 6.1) and the LAD model construction (Section 6.2) were carried out using the LAD – Datascope 2.01 software package (Alexe, 2007) whose algorithmic procedures are described in Alexe and Hammer (2006).

6.1. Binarization

In the binarization process, cutpoints were introduced for the 17 of the 24 numerical variables shown in Table 14. 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 14 provides all the cutpoints used in pattern and model construction. The dataset used to derive these cutpoints includes the banks rated A, A/B, B, D, D/E, and E. 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.

The first step of applying the LAD technique to the problem binarized with the help of these variable cutpoints was the identification of a collection of powerful patterns. One example of such a powerful positive pattern was shown in the previous section. 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 profits before tax are 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 49.11%, and homogeneity 100% (since none of the positive observations satisfy its defining conditions).

6.2. Construction and Description of the LAD Model

The model we have developed for bank ratings, shown in Table 15, is very parsimonious, consisting of only ten positive and nine negative patterns (P1,…, P10, respectively, N1,…,N9), and is built on a *support set* of only 17 out of the 24 original variables. 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).

The availability and processing of data have been major obstacles in the way of using credit risk rating models. As noted by Treacy and Carey (2000), extracting data about the profitability of past credits from archival files can be prohibitively costly. Until recently, many banks did not maintain such data sets and were heavily dependent on qualitative judgments. It is only after the currency crises of the early 1990s and the requirements imposed by the Basel Accord that financial institutions have seen an incentive in collecting the necessary data and maintaining the databases. The move toward an accentuated reliance on rating models is based on the assumption that models produce more consistent ratings and that, over the long haul, operating costs will diminish since less labor will be required to produce ratings. The proposed model will reinforce the above two incentives to develop and rely upon credit risk models in bank operations:

- the accuracy and predictive ability of the proposed model will guarantee dependable ratings,
- its parsimony will alleviate the costs of extracting and maintaining large data sets, will result in leaner and more standardized loan approval operations and in faster decisions, thus reducing the overall operating costs.

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).

6.3. Accuracy and Robustness of the LAD Model

Applying the LAD model described in the previous section to the classification of all the 473 banks whose ratings are A, A/B, B, D, D/E or E, we find the accuracy of the model to be 100%. In order to cross-validate the model's accuracy, we use 10 two-folding cross-validation experiments; in each two-folding cross validation experiment, the observations are randomly split into two approximately equal subsets. The cutpoints are computed and a model is constructed using only one of the two subsets of observations (in contrast to Section 6.1), and it is applied for classifying the observations in the other subset. 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 average accuracy of the 20 models obtained this way 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 proposed classification system.

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. 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.

Finally, as an additional check, we separately calculate the average discriminant values for the nine rating categories. The results are presented in Table 5, 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 similarly 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.

Table 5: Average Discriminant Values

Observation Class	Rating Category	Discriminant Values / Category Averages	Discriminant Values / Class Weighted Averages
Positive	A	0.353	0.326
	A/B	0.323	
	B	0.303	
Unclassified	B/C	0.136	0.006
	C	0.005	
	C/D	-0.122	
Negative	D	-0.258	-0265
	D/E	-0.264	
	E	-0.273	

6.4. From LAD Discriminant Values to Ratings

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 *cutpoints* x_i such that $-1 = x_0 \leq x_1 \leq x_2 \leq \ldots\ldots\ldots \leq x_8 \leq x_9 = 1$, where i indexes the rating categories (with 1 corresponding to E, and 9 corresponding to A). Ideally, a bank should be rated i if its discriminant value falls between x_i and x_{i+1} (e.g., it should be rated A if its value falls between x_8 and x_9).

In reality, such a partitioning may not exist. Therefore, in order to take "noisiness" into account, we shall replace the LAD discriminant values d_i of bank i by 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 and by N the set of banks considered, then the determination of the cutpoints x_j and of the adjusted discriminant values δ_i can be modeled as follows:

$$\boxed{\begin{aligned} &\textit{minimize} \quad \sum_{i\in N}(\delta_i - d_i)^2 \\ &\textit{subject to} \quad \delta_i \leq x_{j(i)+1},\ i\in N \\ &\qquad\qquad x_{j(i)} < \delta_i,\ i \in N \\ &\qquad\qquad -1 = x_0 \leq x_1 \leq x_2,\ldots,\leq x_j,\ldots,x_8 \leq x_9 = 1 \\ &\qquad\qquad -1 \leq \delta_i \leq 1,\ i\in N \end{aligned}} \tag{3}$$

To solve the convex nonlinear problem above, we use in our numerical experiments the NLP solver Lancelot. The approach described above is very similar to the convex cost closure problem (Hochbaum and Queyranne, 2003) 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 cutpoints x_i determined by solving problem (3) can be used not only for rating banks which are in the training sample, but even those which are not. In this case, the bank rating is determined by the particular sub-interval containing the LAD discriminant value. The values of the cutpoints x_i determined by solving the problem (3) for all the 800 banks in the dataset are presented in Table 16.

6.5. Advantages of the LAD Approach for Credit Risk Rating Systems

Let us now emphasize two distinguishing features of the LAD approach that the credit risk literature recognizes as critical. The first one pertains to the possibility offered by this approach to construct credit risk rating systems of varying granularity. The proposed LAD method can actually be used to generate a rating system which exhibits the number of rating categories desired by its user. This contrasts with ordered logistic regression models in which the number of categories is fixed and equal to the number of categories in the external rating system (i.e., Fitch).

The varying granularity property is very appealing for risk managers in that it allows 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. The LAD rating approach can also be used to derive models with higher granularity (i.e., more than 9 rating categories), which are especially valuable for pricing operations and the determination of the conditions and covenants of the granted credit line. Finer credit risk rating systems allow risk managers to further differentiate their customers and to tailor accordingly their credit pricing policies. As mentioned in Section 2, the monetary value of a more granular rating system is very significant, allowing for a substantial decrease in the amount of regulatory capital (Jankowitsch et al., 2007).

The second 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 ordered logistic regression 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.

7. CONFORMITY OF FITCH AND LAD BANK RATINGS

In order to evaluate how well the original LAD discriminant values fit in the identified rating sub-intervals, we use the original (unadjusted) LAD discriminant value of each bank to determine its rating category. We recall that n_k (k = 0,…,8) represent 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. While the rating cutpoints were determined using all the banks in the sample, the LAD discriminant was derived only from the banks rated A, A/B, B, D, D/E and E. Therefore, the discrepancy counts should be calculated separately for the banks rated A, A/B, B, D, D/E and E, and for the banks rated B/C, C and C/D.

The discrepancy summary presented in Table 6 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 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).

It is interesting to remark 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%, and is higher than that between the ordered logistic regression and the Fitch ratings (75.60%).

Table 6 : Discrepancy Analysis

k	*N = {A, A/B, B, D, D/E, E}*		*N = {B/C, C, C/D}*		*N = {A, A/B, B, B/C, C, C/D, D, D/E, E}*	
	n_k	$n_k/\|N\|$	n_k	$n_k/\|N\|$	n_k	$n_k/\|N\|$
0	144	30.44%	90	27.52%	234	29.25%
1	254	53.71%	168	51.38%	422	52.75%
2	51	10.78%	57	17.43%	108	13.50%
3	21	4.44%	10	3.06%	31	3.87%
4	3	0.63%	2	0.61%	5	0.63%
5-6-7-8	0	0.00%	0	0.00%	0	0.00%

In order to systematically evaluate the robustness of the proposed rating system, we use again the cross-validation technique described in Section 3. We apply 10 times the two-folding procedure to derive the average discrepancy counts of the bank ratings predicted for the testing sets. More specifically, in each two-folding experiment, 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 optimization problem (3), the rating cutpoints for the experiment. Finally, the derived LAD model and the intervals determined by these rating cutpoints are used to determine ratings for the banks in the testing set. The accuracy of this rating is then evaluated using the discrepancy counts. The average discrepancy counts (over the two folds) for each of the 10 cross-validation runs are given in Table 7, along with the average discrepancy counts over the 10 experiments.

The fact that (if we include in our calculation every category of banks in the sample, whether it was used or not in deriving the LAD model) the difference between the Fitch and the LAD ratings is, on average, only 0.976, 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.

We conduct 20 experiments (10 times 2-folding) to evaluate the robustness and extendability of the ordered logistic regression rating model (Table 3) and the LAD rating model (Table 7).

Table 7 : Cross-Validated Discrepancy Analysis

k	Exp 1 $n_k/\|N\|$	Exp 2 $n_k/\|N\|$	Exp 3 $n_k/\|N\|$	Exp 4 $n_k/\|N\|$	Exp 5 $n_k/\|N\|$	Exp 6 $n_k/\|N\|$	Exp 7 $n_k/\|N\|$	Exp 8 $n_k/\|N\|$	Exp 9 $n_k/\|N\|$	Exp 10 $n_k/\|N\|$	Average $n_k/\|N\|$
0	0.29	0.29	0.31	0.27	0.32	0.3	0.32	0.31	0.29	0.28	29.80%
1	0.56	0.53	0.51	0.54	0.52	0.49	0.48	0.5	0.51	0.51	51.50%
2	0.09	0.14	0.1	0.13	0.09	0.14	0.14	0.12	0.14	0.15	12.50%
3	0.04	0.03	0.08	0.05	0.04	0.03	0.03	0.05	0.05	0.05	4.40%
4	0.01	0.01	0	0.01	0.02	0.03	0.01	0.02	0.01	0	1.20%
5	0.01	0	0	0	0.01	0.01	0.02	0	0	0.01	0.60%
6-7-8	0	0	0	0	0	0	0	0	0	0	0.00%
Average Difference Between Fitch and LAD Ratings											0.976 categories

These 20 experiments can also be used to check whether the rating discrepancy between the LAD and the Fitch ratings, and that between the ordered logistic regression and the Fitch ratings, differ from each other in a significant way under the assumptions that the paired differences are independent and identically normally distributed. A paired *t*-test indicates that we can reject, with the highest statistical confidence level (99.99%), the null hypothesis according to which the two rating discrepancies described above do not differ. Hence, we can conclude that the LAD rating approach statistically outperforms the ordered logistic regression approach with respect to the rating discrepancy criterion.

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

• Continents: Table 8 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 8: 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 9): 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 9: 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 10): 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 category. Table 10 shows that the LAD rating model performs equally well for every level of creditworthiness of the country in which a bank is located.

Table 10: 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

We report below the degree of agreement among the LAD ratings, the ratings provided by ordered logistic regression (OLR), and the Fitch ratings:

- in 12% of the observations, the LAD, the OLR, and the Fitch ratings are identical;
- in 17.5% of the observations, the LAD and the Fitch ratings are identical, but differ from OLR;
- in 24.875% of the observations, the Fitch and the OLR ratings are identical, but differ from LAD;
- in 17.5% of the observations, the LAD and the OLR ratings are identical; but differ from Fitch;
- in 28.125% of the observations, the LAD, the OLR and the Fitch ratings all differ.

These 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).

8. BROADER IMPACT OF THE PROPOSED METHODOLOGY

The scope of applications of the proposed methodology is not limited to its use by banks to evaluate the creditworthiness of their counterparts. First, banks could use the LAD method to assess the risk of default presented by other types of obligors (individuals, countries, etc.).

Second, the proposed rating method is also very helpful for other types of companies. 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 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). The proposed approach, therefore, could be very valuable for deriving credit risk rating models satisfying the conditions of the Solvency II directives.

Third, the proposed LAD method could be beneficial for central banks and financial regulatory bodies to ensure financial stability defined as "the smooth functioning of the key elements that make up the financial system" (Oosterloo et al., 2007). More precisely, the forward-looking feature of our bank credit risk rating system makes it possible to use it as an early warning system (Estrella et al., 2002) for the detection of weak banks (i.e., those banks whose liquidity or solvency are or will be impaired until a major improvement in their resources materializes (Financial Stability Forum, 2002)), and the risk of systemic bank crisis (King et al., 2004; Barnhill and Souto, 2008). The early warning will allow the regulatory body to take preventive measures that will preserve the value of the bank's assets with minimal disruption to its operations and minimal resolution costs. In that respect, the Bank of International Settlements (Basel Committee on Banking Supervision, 2004) has been emphasizing the value of more accurate credit risk models so that problematic banks could be identified early enough.

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 (Curry and Shibut (2000) estimate that the so-called Savings and Loan crisis cost the U.S. taxpayers about $123.8 billion, 2.1% of 1990 GDP; from 1997 to 2002, twenty Turkish banks experienced such financial difficulties that they were transferred to the Savings Deposit Insurance Funds; the resulting restructuring costs amounted to about $16.9 billion (Canbas et al., 2005); the International Monetary Fund (2008) predicts in its Global Financial Stability Report that financial losses due to the current credit crisis might 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 in 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' capital. The current mortgage crisis provides an illustration that a downward business cycle can cause companies' distress, 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 risk and to allow the adoption of proactive measures to manage such risks and alleviate the consequences.

9. CONCLUDING REMARKS

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 three approaches to address this problem, the first two being statistical methods (multiple linear regression and ordered logistic regression), while the third 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 starts by demonstrating the inadequacy of the results obtained using multiple linear regression. It shows then that ordered logistic regression and the LAD method can provide superior results 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, with the correlation level between the ratings of LAD and those of ordered logistic regression exceeding 81%. Moreover, it is shown that both rating systems are in close agreement with the Fitch ratings, with the stability and robustness of this agreement being demonstrated by cross-validation. 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 ordered logistic regression ratings with the Fitch ratings, and considering the associated 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 ordered logistic regression 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 ordered logistic regression 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 proposed LAD method presents two other critical advantages (Jankowitsch et al., 2007; Kick and Koetter, 2007) over the ordered logistic regression method. The first one pertains to the capacity to construct rating models that have the granularity desired by the user of the model, and generates major savings through the decrease in the amount of regulatory capital. The second advantage stems from the fact that the LAD model does not assume that an explanatory variable has the same effect on all banks and on all rating categories.

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 credit risk rating systems provides managers with a very practical and powerful tool to help them decide whether or not, and on which conditions, a credit must be granted. This approach can be used for different purposes (pre-approval, determination of pricing policies, etc.). Moreover, it can yield rating models with varying levels of granularity that can be used at different stages in the credit granting decision process, and 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 has: *(i)* tremendous *monetary value* (see Jankowitsch et al. (2007) and Moody's KMV studies (Stein, 2003; Stein and Jordão, 2003)), allowing the mitigation of the financial and operational risk in a financial institution; and *(ii)* a very broad *applicability scope,* since it can be used by other entities (insurance companies, regulators) and to pursue other objectives (financial stability, early warning system).

REFERENCES

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

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

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. *Working Paper*.

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. 2006a. International Convergence of Capital Measurement and Capital Standards: A Revised Framework (Basel II). http://www.bis.org/publ/bcbsca.htm

Basel Committee on Banking Supervision. 2006b. Core Principles for Effective Banking Supervision. http://www.bis.org/publ/bcbs129.pdf

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.

Chapelle A., Crama Y., Hubner G., Peters J.-P. 2008. Practical Methods for Measuring and Managing Operational Risk in the Financial Sector: A Clinical Study. *Journal of Banking & Finance* **32**, 1049-1061.

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

Crouhy M., Galai D., Mark R. 2001. *Risk Management*. McGraw-Hill, New York.

Cruz M. G. 2002. *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons, New York.

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*.

Engelmann B., Hayden E., Tasche, D. 2003. Testing Rating Accuracy. *Risk* 16 (1), 82-86.

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. 2002. Supervisory Guidance on Dealing with Weak Banks.

Financial Stability Forum. 2007. FSF Working Group on Market and Institutional Resilience. *Preliminary Report to the G7 Finance Ministers and Central Bank Governors.*

Financial Stability Forum. 2008. *Report of the Financial Stability Forum on Enhancing Market and Institutional Resilience.*

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*.

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.

de Fontnouvelle P., Dejesus-Rueff V., Jordan J.S., Rosengren E.S. 2006. Capital and Risk: New Evidence on Implications of Large Operational Losses. *Journal of Money, Credit, and Banking* **38** (7), 1819-1846.

de Fontnouvelle P., Garrity V., Chu S., Rosengren E. 2005. The Potential Impact of Explicit Basel II Operational Risk Capital Charges on the Competitive Environment of Processing Banks in the United States. *Working Paper, Federal Reserve Bank of Boston*.

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

Gilbert R.A. 1993. Implications of Annual Examinations for the Bank Insurance Fund. *Federal Reserve Bank of St. Louis Review* **75** (1), 35-52.

Griffiths B., Beynon M.J. 2005. Expositing Stages of VPRS Analysis in an Expert System: Application with Bank Credit Ratings. *Expert Systems with Applications* **29**, 879-888.

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

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

Hammer P.L., Liu Y., Simeone B., Szedmák S. 2004. Saturated Systems of Homogeneous Boxes and the Logical Analysis of Numerical Data. *Discrete Applied Mathematics* **144** (1-2), 103-109.

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

Hochbaum D.S., Queyranne M. 2003. The Convex Cost Closure Problem. *SIAM Journal of Discrete Mathematics* **37**, 843-862.

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.

Jarrow R.A. 2008. Operational Risk. *Journal of Banking and Finance* **32**, 870–879.

Jorion P. 2000. *Value-at-Risk: the New Benchmark for Managing Financial Risk.* 2nd Edition. McGraw-Hill.

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.

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.

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.

Risk Management Group. 2003. The 2002 Loss Data Collection Exercise for Operational Risk: Summary of the Data Collected. *Report to Basel Committee on Banking Supervision.*

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

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

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

Stein R.M. 2003. Power, Profitability and Prices. Why Powerful Models Increase Profits and How to Define a Lending Cutoff if You Must. *Moody's KMV Technical Report* **030124**.

Stein R.M., Jordão. 2003. What is a More Powerful Model Worth? *Moody's KMV Technical Report*.

Sy A.N.R. 2004. Rating the Rating Agencies: Anticipating Currency Crises or Debt Crises? *Journal of Banking & Finance* **28** (11), 2845-2867.

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.

Wall Street Letter. 2006. CFA To Senate: Follow Our Lead On Credit Rating: http://www.wallstreetletter.com/pdf/WSL030606.pdf.

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

ELECTRONIC COMPANION

Table 11: Geographic Distribution of Banks

Origin	Number of banks	Countries
Western Europe	247	Andorra (3), Austria (1), Belgium (8), Denmark (4), Finland (4), France (29), Germany (31), Greece (5), Iceland (2), Ireland (5), Italy (30), Luxembourg (2), The Netherlands (12), Norway (9), Portugal (10), Spain (42), Sweden (6), Switzerland (5), UK (39)
Eastern Europe	51	Croatia (2), Czech Republic (3), Hungary (3), Latvia (1), Lithuania (4), Poland (7), , Romania (2), Russia (19), Slovak Republic (3), Slovenia (6), Ukraine (2)
Canada and USA	198	Canada (6), USA (192)
Developing Latin American countries	45	Argentina (2), Bermuda (2), Brazil (10), Chile (5), Colombia (2), Dominican Republic (4), El Salvador (2), Mexico (6), Panama (2), Peru (1), Venezuela (9)
Middle East	44	Bahrain (5), Cyprus (2), Egypt (2), Kuwait (4), Lebanon (2), Malta (1), Saudi Arabia (8), Turkey (14), UAE (6)
Hong-Kong, Japan, Singapore	55	Hong-Kong (18), Japan (34), Singapore (3)
Developing Asian countries	145	Azerbaijan (1), China (16), India (32), Indonesia (9), Kazakhstan (5), Malaysia (8), Pakistan (4), Philippines (14), South Korea (12), Taiwan (31), Thailand (10), Vietnam (3)
Oceania	6	Australia (6)
Africa	6	South Africa (6)
Israel	3	Israel (3)

Table 12: 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 gradations among these five ratings: 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 13 : Regression results

The estimation of the standard regression model is given by: $y = a + \sum_{i=1}^{24} \beta_i x_i + \varepsilon$.

	Beta	*t*-test	*p*-value
(Constant)		1.111	.267
x_1	.686	24.313	.000
x_2	-.430	-3.991	.000
x_3	-.188	-1.817	.070
x_4	.001	-.024	0.897
x_5	.046	.699	.485
x_6	.011	-.091	0.905
x_7	.286	2.230	.026
x_8	-.119	-.822	.412
x_9	.001	-.091	0.907
x_{10}	.214	1.367	.172
x_{11}	-.086	-.380	.704
x_{12}	.084	.319	.750
x_{13}	-.330	-1.146	.252
x_{14}	.546	2.369	.018
x_{15}	.001	-0.090	.000
x_{16}	.021	.627	.531
x_{17}	-.063	-.391	.696
x_{18}	.309	1.966	.050
x_{19}	.283	2.967	.003
x_{20}	-.299	-2.840	.005
x_{21}	-.028	-.553	.580
x_{22}	.095	1.766	.078
x_{23}	.026	.594	.552
x_{24}	-.007	-.224	.823

Definition of financial ratios used

We use as predictor variables 9 financial ratios providing an evaluation of:

- **asset quality:** *ratio of equity to total assets*: as equity is a cushion against asset malfunction, this ratio measures the amount of protection afforded to the bank by the Equity they invested in it. The higher this figure the more protection there is.
- **operations**:
  - *net interest margin*: this is the net interest income expressed as a percentage of earning assets. The higher this figure the cheaper the funding or the higher the margin the bank is commanding. Higher margins and profitability are desirable as long as the asset quality is being maintained.
  - *ratio of interest income to average asset*: this is the net interest income as a percentage of the total balance sheet.
  - *Ratio of other operating income to average assets*: when compared to the above ratio, this indicates to what extent fees and other income represent a greater percentage of earnings of the bank. As long as this is not volatile trading income it can be seen as a lower risk form of income. The higher this figure is the better.

- *Ratio of non interest expenses to average assets:* non interest expenses or overheads plus provisions give a measure of the cost side of the banks performance relative to the assets invested.
- *return on average assets (ROAA)*: it compares the efficiency and operational performance of banks as it looks at the returns generated from the assets financed by the bank.
- *return on average equity (ROAE)*: it measures of the return on shareholder funds. Obviously here the higher the figure the better but one should be careful in putting too much weight on this ratio as it may be at the expense of an over leveraged balance sheet.
- *the cost to income ratio:* it measures the overheads or costs of running the bank, the major element of which is normally salaries, as percentage of income generated before provisions. It can be distorted by high net income from associates or volatile trading income.

- **liquidity**:
  - *ratio of net loans to total assets:* it indicates what percentage of the assets of the bank are tied up in loans. The higher this ratio the less liquid the bank is.

Table 14: 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 15: 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							

Table 16: Rating Cutpoints

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