Reverse-Engineering Banks’ Financial Strength Ratings Using Logical Analysis of Data¹

Peter L. HAMMER²*, Alexander KOGAN³**, Miguel A. LEJEUNE⁴***

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 for 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 for 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, 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, (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

¹ Acknowledgment: The authors are grateful to the editors and referees for their very insightful comments that have been instrumental in improving this manuscript. 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 research was essentially completed.
³ kogan@rutgers.edu; Rutgers Business School, Rutgers University, 180 University Avenue, Newark, NJ, USA
⁴ mlejeune@gwu.edu; The George Washington University, Washington, DC, USA
existing (proprietary and opaque) rating system we mean the close agreement of its ratings with those of the existing system. By its generalizability we mean its accuracy in rating those banks which were not used in developing the system.

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

This study reveals the weakness of the results obtained with multiple linear regression and 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 (yet more robust) 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 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 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 that 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 Logical Analysis of Data (LAD) 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, 2008). Skepticism about rating agencies finds its root in their inability to spot some of the largest financial collapses of the decade (Enron Corp, WorldCom Inc, Lehman Brothers) which some observers attribute to the presence of conflicts of interest (Financial Stability Forum, 2008) induced by the issuer-pay model, in which the rated company pays the rating agency a fee for the issued rating. A report of the Securities and Exchange Commission (2003) indicates that 90% of Moody’s and Fitch’s revenues come from fees paid by the issuers. Moreover, in many instances, key credit analysts in rating agencies participated directly in fee negotiation with issuers. The SEC reports (Reuters, 2008) that often the rating agency employees structuring a financial security were the same who were rating the product! The issuer-
pay model creates moral hazard situations and has led to the so-called complacency effect, which makes raters, especially in a non-crisis period, very reluctant to downgrade an obligor (Schwarcz, 2008) and is viewed as one of the main reasons for the subprime financial crisis. The development of an objective accurate internal rating model would make banks less dependent on the whims of the rating agencies.

Second, increasingly intense international market environment and changes in the regulatory framework driven by the Financial Stability Forum and the Basel Committee on Banking Supervision (Basel II) called forth incentives for banks to improve their credit risk rating systems. While the Basel Committee initially (i.e., in the 1988 Capital Accord) favored the ratings provided by external credit rating agencies, the Basel II Accord (Basel Committee on Banking Supervision, 2006) is actually a risk-focused regulatory framework under which a bank can develop its own risk model to calculate the regulatory capital for its credit portfolio. Compliance with the Basel II Accord requires the internal risk models (rating, probability of default, loss given default, exposure at default) to be cross-validated and accepted by the legislator, and qualifies banks for the adoption of the Internal Rating Based approach (Basel Committee on Banking Supervision, 2001, 2006) to calculate their capital provisions. The Financial Stability Forum endorsed the principles for banking supervision defined by the Basel Committee of Banking Supervision (2006) as crucial for having sound financial systems and as deserving “priority implementation”, while Jarrow (2007) judges the revised Basel II requirements insufficient to determine capital requirements for banks. The insolvency of an obligor is usually understood as happening when a fair evaluation of its assets is less than its liabilities. In the current crisis, however, many commentators claim that a key issue faced by financial institutions is of a liquidity nature, in the sense that they are "unable to meet their liabilities as they come due" (Schwarcz, 2008). The current financial turmoil has led the Basel Committee on Banking Supervision (2008), in line with the Financial Stability Forum's prudential oversight recommendations (2008), to strengthen its principles for banks' liquidity risk management. Other measures will probably follow, with the aim of complementing the Basel II requirements and forcing financial institutions to keep a robust cushion of high quality liquid assets in order to survive periods of severe liquidity stress.

Third, an internal rating system provides autonomy to a bank’s management to execute credit risk policies in accordance with the bank’s own core business goals and could potentially make credit operations more transparent and efficient. Consider for example the securitization process which suffers from the evident over-reliance by lenders on the FICO scoring model. A rule of thumb in the lending market is to approve a loan request if the FICO score of the applicant is above 620 (Avery et al., 1996; Freddie Mac, 2007). Keys et al. (2008) have observed that (i) loans to borrowers whose FICO score is just above the 620 threshold (AT) are much more likely to be securitized than these to borrowers whose FICO score is just below the 620 threshold (BT), and (ii) a portfolio of AT loans has a default rate that is 10 to 25% higher than
the default rate of a portfolio of BT loans. This indicates that lenders use the 620 FICO score as a sufficient condition for securitization purposes. More importantly, it suggests that this should not be the case and that borrowers with a FICO score above 620 should also be subjected to a more thorough screening process. The development of an objective, internally derived model accounting for the impact of "hard" (i.e., FICO scores) as well as "soft" (i.e., future income stability of the borrower) measures would allow the application of more thorough and homogeneous decision rules (for borrowers above and below the 620 threshold).

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

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

Credit risk rating systems play a pervasive role in the operations of banks (Tracey and Carey, 2000; Grunert et al., 2005) and strongly impact the operational risk of financial institutions. Operational risk is defined by the Basel Bank of International Settlements and its Committee on Banking Supervision (2006) as “the risk of loss resulting from inadequate or failed internal processes, people, and systems or from external events”. Jarrow (2008) distinguishes two main types of operational risk. 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. The credit rating system is first applied to rate customers and is a critical input for the credit pre-approval process. The decision to pre-approve or deny the granting of the credit is based on a binary decision model that uses the rating of the customer as the key explanatory variable. If the loan 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). This illustrates the use of the credit risk rating system across the entire loan origination process (rating assignment, credit approval, and underwriting decisions) and shows its impact on the operational risks of Type one (Jarrow, 2008).

The credit risk rating also impacts investment decisions. For instance, the Department of Labor introduced in 1988 a regulation stipulating that pension fund investment in asset-backed securities is restricted to those rated A or better. Investment-grade bond mutual funds can hold at most 5% of assets in non-investment grade bonds and must sell any security whose rating falls below B (Kisgen, 2006, 2007). In
some countries, debt issuers must receive a rating from an accredited agency if they intend to raise public money and to be listed on a stock exchange (Dale and Thomas, 1991). Some funds are forced to sell holdings that lose an investment grade rating regardless of price. The value of a bond (e.g., muni) whose risk rating moves from an investment to a speculative rating category can decrease sharply. There exists ample evidence that rating downgrades are accompanied by declines in stock and bond prices, while upgrades result in increases in security prices (Hand et al., 1992; Kisgen, 2007). The reliance upon an internal rating system anticipating such external rating modifications is thus particularly valuable.

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 credit risk ratings assigned by the lending officers affect the volume of the approved loans or credits as well as the compensation of the officers who could receive high bonuses for arranging deals or investments that later fail. 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). Moral hazard situations also arise from the originate-and-distribute securitization model, widely regarded as the "dominant explanation" (Gorton, 2009) for the current financial panic, in which the interests of originators (e.g., a company that "tranches" the pool of mortgages, assigns a rating to each tranche, and issues collateralized mortgage obligations (CMOs)) and investors (e.g., another financial company buying the CMOs) are not necessarily aligned.

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, the rating is used to calculate the bank's exposure associated with the credit, and to calculate the amount of economic and regulatory capital that a bank must keep to hedge against a possible default 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 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. 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).
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 influence the operations of the bank 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 obligor can be carried out through frequent supervisory examinations. However, on-site examination operations are extremely time-consuming and costly. Hence, the reliance upon a highly predictive rating system for the evaluation of the solvency of an obligor would not constrain banks to resorting to on-site visits and is a primary lever to lower the costs incurred by such operations (Derviz and Podpiera, 2004).

It was demonstrated (Jankowitsch et al., 2007) that, in a competitive framework, a poor prediction 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. More accurate rating systems decrease the negative effects of adverse selection (Jankowitsch et al., 2007) and tend to make banks more conservative, reducing the number of approved loans by 4% and the default rate by 18% (Stein and Jordão, 2003; Stein, 2005).

As for the effects of a better rating system on the loan pricing operations and the revenues they generate, 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, and that switching from a rating system with medium accuracy to one with high accuracy increases the annual rate of return by about 15 additional basis points. Other studies (Stein and Jordão, 2003; Stein, 2005) confirm this. They show how the adoption of a rating system with a five-point higher accuracy ratio allows a mid-sized bank with $50 billion of assets to increase its annual profits by $2.1 to $4.8 million. In the data set we use in this paper, this represents an average annual profit increase of 3.4% for the mid-sized banks with assets ranging between $40 and $60 billion. The monetary gains that can be obtained from a better rating model provide banks with a "strong incentive to invest into their rating system" (Jankowitsch et al., 2007). A rating system with superior accuracy increases the economic performance, the extent of which has been demonstrated (Jankowitsch et al., 2007) to be particularly marked in very competitive markets, in markets where customers are aware of their actual credit quality, for banks with weak lending portfolios, and for banks with high loss given default rates, and thus would have been magnified in the current crisis.

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 capital allocation of a bank for the coverage of the operational risk is estimated to be 25% and 35%, by 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 often exceeds that for market risk (see de Fontnouvelle et al., 2006 and 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, 2006). 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.

3. CREDIT RISK RATINGS OF BANKS

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

3.1. Limitations, Significance and Challenges

As it will be described in more details in the next sections, the proposed rating method is based on the learning of the ratings disclosed by the Fitch rating agency. Even if the vast majority of their ratings are justified, rating agencies have failed in predicting some major cases of creditworthiness deterioration, and have been severely criticized (see Section 2.1). Thus, the proposed rating method has some limitations. Its
effective use requires the a priori selection of a set of reliable ratings that are not plagued by moral hazard, which, in its turn, rests on the definition of what constitutes a reliable rating.

One might, therefore, wonder about the utility of the proposed rating model that replicates the "cleaned" (i.e., exonerated from the obvious rating mistakes) Fitch rating system. The contribution of such a rating model is nonetheless substantial in that it enables the bank to rate its counterparts as frequently as it desires. It gives the bank the possibility to rerun its model on an ongoing basis, to assess, in a proactive fashion, the solvency of the financial institutions, and to possibly anticipate the rating modifications of the agencies, which would be extremely valuable for the investment decisions of the bank (see Section 2.2). This is in sharp contrast to the reactive approach of waiting for rating agencies to modify their ratings. Indeed, rating agencies are not properly incentivized to revise ratings, especially downward, on a timely basis (complacency effect).

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

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

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

3.2. Prior Research on Banks’ Creditworthiness

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

Fitch Ratings (2006) analyzed a bank rating method based on joint probability analysis, which includes as input variables the probability of the bank failing and the probability of the potential supporter (i.e., a sovereign state) defaulting at the same time, and, if there is no such simultaneous default, the probability of the supporter being willing or not to provide such support. Fitch Ratings comes to the conclusion that 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.

3.3. External Bank Ratings

In this sub-section, we shall briefly describe the Fitch Individual Bank rating system. Fitch publishes over 1,000 international bank ratings worldwide, and it is generally viewed as the leading agency for rating bank credit quality. Fitch provides long- and short-term credit ratings, which are viewed as an opinion on the ability of an entity to meet financial commitments (interest, preferred dividends, or repayment of principal) on a timely basis (Fitch Ratings, 2001). These ratings are assigned to sovereigns and corporations, including banks, and are comparable worldwide.
Fitch provides a specialized rating scale for banks using *individual* and *support* ratings. Support ratings comprise 5 rating categories and assess the likelihood for a banking institution to receive support either from the owners or the governmental authorities should it run into difficulties. While the availability of support is a critical characteristic of a bank, it does not describe completely the banks’ solvability in case of adverse situations. It is worth noting for instance that even though a bank can have a state guarantee of support, its marketable obligations might drop. That is why, as a complement to a support rating, Fitch also provides an individual bank rating, which allows a credit quality evaluation separately from any consideration of outside support. It is purported to assess how a bank would be viewed if it were entirely independent and could not rely on external support.

For banks in investment grade countries, support and individual bank rating scales provide sufficient scope for differentiation. However, the scope for differentiation between banks in non-investment grade countries can be more limited. Indeed, the rating of an obligor located in a given country is limited from above by the rating of that country (Ferri et al., 1999). In countries with weak country risk rating, it is thus possible that a number of banks’ ratings may be restricted by the sovereign’s foreign currency rating and will be bunched together at the sovereign ceiling. That is why *national* ratings were developed to provide a greater degree of differentiation between issuers in countries subject to this bunching effect. National ratings are an assessment of credit quality relative to the rating of the “best” credit risk in the country. The national rating scale has 22 different risk categories as opposed to the 9 categories of the individual rating scale. Table 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.

### 3.4. Explanatory Variables

Based on the references mentioned above, approximately 40 parameters have been provisionally considered as potentially significant predictors of the banks’ creditworthiness. After the elimination of those parameters for which not all the data were available, we have restricted our attention to a set of 14 financial variables (loans, other earning assets, total earning assets, non-earning assets, 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*;
- **performance**: (i) profit efficiency: *return on average assets; return on average equity; ratio of interest income to average assets; net interest margin; ratio of other operating income to average assets*; (ii) cost efficiency: *cost to income ratio; ratio of non-interest expenses to average assets*;
- **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). Similarly, Bloomberg has developed a standard cardinal scale for comparing Moody’s, S&P’s and Fitch ratings.

### 3.5. 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**

<table>
<thead>
<tr>
<th>Rating Categories</th>
<th>A</th>
<th>A/B</th>
<th>B</th>
<th>B/C</th>
<th>C</th>
<th>C/D</th>
<th>D</th>
<th>D/E</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Banks</td>
<td>19</td>
<td>60</td>
<td>203</td>
<td>129</td>
<td>121</td>
<td>77</td>
<td>93</td>
<td>66</td>
<td>32</td>
</tr>
<tr>
<td>Percentage of banks</td>
<td>2.375%</td>
<td>7.5%</td>
<td>25.375%</td>
<td>16.125%</td>
<td>15.125%</td>
<td>9.625%</td>
<td>11.625%</td>
<td>8.25%</td>
<td>4%</td>
</tr>
</tbody>
</table>

### 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: $d_j = |\overline{y}_j - r_j|$, $j \in N$. Now we can calculate the discrepancy count $n_k = |\{j:d_j = k, j \in N\}|, k = 0, 1, \ldots, 8$, 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 $k$, $k=0,1,\ldots,8$. 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**

<table>
<thead>
<tr>
<th>$k$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>27.75%</td>
<td>33.13%</td>
<td>30.25%</td>
<td>7.50%</td>
<td>1%</td>
<td>0.25%</td>
<td>0.13%</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Test Sets: Average over 10 2-Folding Experiments</td>
<td>21.75%</td>
<td>27.38%</td>
<td>24.25%</td>
<td>20.38%</td>
<td>5.50%</td>
<td>0.50%</td>
<td>0.25%</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 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:

$$
\begin{align*}
\log \left( \frac{p_1}{1-p_1} \right) &= \alpha_1 + \sum_{i=1}^{k} \beta_i x_i \\
\log \left( \frac{p_1 + p_2}{1-p_1-p_2} \right) &= \alpha_2 + \sum_{i=1}^{k} \beta_i x_i \\
\vdots \\
\log \left( \frac{p_1 + p_2 + \ldots + p_k}{1-p_1-p_2-\ldots-p_k} \right) &= \alpha_k + \sum_{i=1}^{k} \beta_i x_i
\end{align*}
$$

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

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 $\alpha_k$ differ across credit risk rating categories, while the regression coefficients $\beta_j$ 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**

<table>
<thead>
<tr>
<th>(k)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
<td>36.50%</td>
<td>37.50%</td>
<td>17.13%</td>
<td>6%</td>
<td>2.25%</td>
<td>0.25%</td>
<td>0.38%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Test Sets: Average over 10 2-Folding Experiments</td>
<td>30.75%</td>
<td>38.88%</td>
<td>17.75%</td>
<td>8.38%</td>
<td>3.50%</td>
<td>0.50%</td>
<td>0.25%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

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).

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

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

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

The objective of LAD is to discover a binary-valued function \( f \) depending on the \( n \) input variables, which provides discrimination between positive and negative observations, and which closely approximates the actual one. This function \( f \) is constructed as a weighed sum of patterns.

In order to clarify how such a function \( f \) is found we start by transforming the original dataset to one in which the variables can only take the values 0 and 1. We achieve this goal by using indicator variables 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 some limitations on the set of patterns to be generated, by restricting their degrees (to low values), their prevalence (to high values), and their homogeneity (to high values). Several algorithms have been developed for the efficient extraction of (relatively small) subsets of positive and negative patterns corresponding to the above criteria and sufficient for classifying the observations in the dataset (Boros et al., 2000). Such collections of positive and negative patterns are called models. A model is supposed to include sufficiently many positive (negative) patterns to guarantee that each of the positive (negative) observations in the dataset is “covered” by (i.e., satisfies the conditions of) at least one of the positive (negative) patterns in the model. 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 \( \omega \), 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>
<thead>
<tr>
<th>Table 4: Classification Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observation Classes</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Positive</td>
</tr>
<tr>
<td>Negative</td>
</tr>
</tbody>
</table>

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:

$$Q = \frac{1}{2}(a + d) + \frac{1}{2}(e+f). \quad (2)$$

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 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 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>
<thead>
<tr>
<th>Observation Class</th>
<th>Rating Category</th>
<th>Discriminant Values / Category Averages</th>
<th>Discriminant Values / Class Weighted Averages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>A</td>
<td>0.353</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>A/B</td>
<td>0.323</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.303</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>B/C</td>
<td>0.136</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>0.005</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C/D</td>
<td>-0.122</td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>D</td>
<td>-0.258</td>
<td>-0.265</td>
</tr>
<tr>
<td></td>
<td>D/E</td>
<td>-0.264</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>-0.273</td>
<td></td>
</tr>
</tbody>
</table>

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

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 \( \delta_i \), and find values of \( \delta_i \) for which such a partitioning exists and which are “as close as possible” to the values \( d_i \). As it is often the case, we interpret “as close as possible” as minimizing the mean square approximation error. If we denote by \( j(i) \) the rating category of bank \( i \), by \( \epsilon \) an infinitesimal positive number, by \( N \) the set of banks, then the determination of the cutpoints \( x_j \) and of the adjusted discriminant values \( \delta_i \) can be modeled as follows:

\[
\begin{align*}
\text{minimize} & \quad \sum_{i \in N} (\delta_i - d_i)^2 \\
\text{subject to} & \quad \delta_i \leq x_{j(i)}, \quad i \in N \\
& \quad x_{j(i) - 1} \leq \delta_i - \epsilon, \quad i \in N \\
& \quad -1 = x_0 \leq x_1 \leq x_2 \leq \ldots \leq x_8 \leq x_9 = 1 \\
& \quad -1 \leq \delta_i \leq 1, \quad i \in N
\end{align*}
\]

(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 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 allows 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 $n_k (k = 0, \ldots, 8)$ 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, C, 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>
<thead>
<tr>
<th>( k )</th>
<th>( N = {A, A/B, B, D, D/E, E} )</th>
<th>( N = {B/C, C, C/D} )</th>
<th>( N = {A, A/B, B/C, C, C/D, D, D/E, E} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_k )</td>
<td>( n_k/</td>
<td>N</td>
<td>)</td>
</tr>
<tr>
<td>0</td>
<td>144</td>
<td>30.44%</td>
<td>90</td>
</tr>
<tr>
<td>1</td>
<td>254</td>
<td>53.71%</td>
<td>168</td>
</tr>
<tr>
<td>2</td>
<td>51</td>
<td>10.78%</td>
<td>57</td>
</tr>
<tr>
<td>3</td>
<td>21</td>
<td>4.44%</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0.63%</td>
<td>2</td>
</tr>
<tr>
<td>5-6-7-8</td>
<td>0</td>
<td>0.00%</td>
<td>0</td>
</tr>
</tbody>
</table>

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           | 0.13           | 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              | 0.01           | 0.01           | 0.02           | 0              | 0              | 0.01           | 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.

The comparison of Table 3 and Table 7 provides another confirmation that the LAD method outperforms the ordered logistic regression one when they are applied to rate banks not used in the construction of the model. Indeed, it can be seen that the LAD rating is, on average, within one (two) rating
category (categories) of the Fitch rating for 81.30% (93.80%) of the banks, while this statistic is equal to 69.63% (87.38%) for the comparison between the Fitch and the ordered logistic regression ratings. We also observe that, on average, the ordered logistic regression (respectively, LAD) ratings perfectly match the Fitch ones for 30.75% (respectively, 29.80%) of the banks. We refer the reader to Section 8 for a discussion of other advantages that the LAD approach has over the order logistic regression approach.

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>
<thead>
<tr>
<th>K</th>
<th>Asia</th>
<th>Europe</th>
<th>North America</th>
<th>South America</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>22.12%</td>
<td>34.87%</td>
<td>34.45%</td>
<td>16.82%</td>
</tr>
<tr>
<td>1</td>
<td>53.11%</td>
<td>45.93%</td>
<td>51.92%</td>
<td>61.90%</td>
</tr>
<tr>
<td>2</td>
<td>17.42%</td>
<td>15.80%</td>
<td>11.03%</td>
<td>8.87%</td>
</tr>
<tr>
<td>3</td>
<td>7.35%</td>
<td>3.40%</td>
<td>1.18%</td>
<td>8.35%</td>
</tr>
<tr>
<td>4</td>
<td>0.00%</td>
<td>0.00%</td>
<td>1.42%</td>
<td>4.06%</td>
</tr>
<tr>
<td>5 – 6 - 7 – 8</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Average Difference Between Fitch and LAD Ratings</td>
<td>1.10</td>
<td>0.88</td>
<td>0.83</td>
<td>1.21</td>
</tr>
</tbody>
</table>

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

- **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.
Table 10: Analysis of Discrepancies per Country Risk Rating Category

<table>
<thead>
<tr>
<th>$k$</th>
<th>Investment-Grade</th>
<th>Speculative-Grade</th>
<th>Default-Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>30.07%</td>
<td>30.26%</td>
<td>19.25%</td>
</tr>
<tr>
<td>1</td>
<td>53.12%</td>
<td>45.86%</td>
<td>62.50%</td>
</tr>
<tr>
<td>2</td>
<td>12.34%</td>
<td>15.79%</td>
<td>18.25%</td>
</tr>
<tr>
<td>3</td>
<td>3.58%</td>
<td>7.43%</td>
<td>0.00%</td>
</tr>
<tr>
<td>4</td>
<td>0.89%</td>
<td>0.66%</td>
<td>0.00%</td>
</tr>
<tr>
<td>5 - 6 - 7 - 8</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Average Difference Between Fitch and LAD Ratings</td>
<td>0.92</td>
<td>1.02</td>
<td>0.99</td>
</tr>
</tbody>
</table>

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. The results confirm the capacity of non-statistical models to derive highly accurate prediction models, as acknowledged in the literature (see, e.g., de Servigny, Renault, 2004).

8. BROADER IMPACT OF THE PROPOSED METHODOLOGY

Provided that a reliable set of observations is selected, the proposed rating methodology can be applied to construct an accurate and objective LAD-based bank credit risk rating model. An LAD rating model allows the financial institution to rate its counterparts on a proactive basis, at the frequency it desires, without having to wait for sometimes untimely rating modification announcements from Fitch, and responds to the needs motivating the development of internal credit rating systems described in Section 2.1.

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

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

Beyond its contribution to the improvement of a bank’s operations associated with the evaluation of obligors’ solvency (Section 2), the proposed methodology has the potential to contribute to the resolution of a number of related economic and managerial questions faced by other actors in the credit sector.

First, the proposed rating method can be very helpful not only for banks, but for other types of companies as well. Insurance companies, one of the major players in the credit risk market (providing insurance coverage, buying banks’ credit risk derivatives, and therefore being indirectly exposed to many credit risks) also need to assess the risk of their operations. The exposure of insurers to high levels of uncertainty and insolvency, combined with the trend for insurers to expand their business frontier to include new insurance areas, has led insurance supervisory authorities across the world to reform the solvency system of insurance companies (Florez-Lopez, 2007). This has led the European supervisors to initiate the Solvency II project in 2002 and to develop directives regarding the financial resources, supervisory review and market discipline of the sector (European Commission Internal Market and Services, 2005). The proposed approach, therefore, could be very valuable for deriving credit risk rating models satisfying the conditions of the Solvency II directives.

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

The applicability of the proposed model as an early warning system is extremely important in view of (i) the financial costs of a bank crisis: some examples of the magnitude of these costs are the so-called Savings and Loan crisis which resulted in $123.8 billion (2.1% of the 1990 GDP) of losses for the U.S. taxpayers (Curry and Shibut, 2000); the 1997-2002 Turkish bank crisis which led to restructuring costs of $16.9 billion (Canbas et al., 2005); the current financial crisis, the losses of which are predicted by the International Monetary Fund (2008) to approach $1 trillion; (ii) the steady increase of bank failures over the last 30 years in developing as well as in highly developed economies (Basel Committee on Banking Supervision, 2004); (iii) the risk of spillover across the whole financial system and economy. Indeed, the prevalence of high interbank linkages significantly increases the risk of spillover, which could result in a
bank crisis. Thus, an insolvent bank unable to honor its obligations could precipitate financial distress for its counterparts. Besides, other cascading effects must also be considered that could exacerbate cyclical recessions and result in more severe financial crises (Basel Committee on Banking Supervision, 2004). The bankruptcy of a bank can provoke depositors from other banks to withdraw their funds, depleting banks’ capital. The current mortgage crisis provides an illustration that a downward business cycle can cause distress for companies, rendering many loans delinquent and causing banks to further reduce business lending, and that financial crises can extend to other sectors of the economy as the availability of credit may be disrupted. The possible contagion effect through the banking sector, and possibly through other sectors or economies, highlights the importance of developing predictive rating models as tools to appraise ex-ante the magnitude of the risks and to allow the adoption of proactive measures to manage these risks and alleviate the consequences.

9. CONCLUDING REMARKS AND FURTHER RESEARCH

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

This study is devoted to the problem of reverse-engineering the Fitch bank credit ratings, a problem, which -- in spite of its important managerial implications -- is generally overlooked in the extant literature. We present 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. 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. Hence, the rating system can be used for multiple purposes (pre-approval, determination of pricing policies, etc.) in the credit granting decision process 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 risk rating systems provides managers with a powerful tool to help them decide whether or not, and under which conditions, a credit must be granted. Moreover, it can be employed to develop internal rating systems that satisfy the Internal Rating Based requirements, and are Basel 2 compliant. The construction of a credit rating system with the above properties: (i) can be used to mitigate the financial and operational risk (Jankowitsch et al., 2007, Stein, 2005) in a financial institution; and (ii) has a broad applicability scope, since it can be used by other entities (insurance companies, regulators) and for pursuing other objectives (financial stability, early warning system).

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

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

REFERENCES
Alexe S. 2007. LAD - Datascope V2.1 Software Package.
Basel Committee on Banking Supervision. 2001. The Internal Ratings Based Approach: http://www.bis.org/publ/bcbsca05.pdf


