WORKING GROUP ON RISK ASSESSMENT

Credit Risk Assessment Revisited

Methodological Issues and Practical Implications

2007

EUROPEAN COMMITTEE OF CENTRAL BALANCE SHEET DATA OFFICES

WORKING GROUP ON RISK ASSESSMENT

Credit Risk Assessment Revisited

Methodological Issues and Practical Implications

2007

EUROPEAN COMMITTEE OF CENTRAL BALANCE SHEET DATA OFFICES

Imprint
Publisher and editor: European Committee of Central Balance Sheet Data Offices.
Printed by: Oesterreichische Nationalbank, 1090 Vienna, Austria.

MEMBERS OF THE WORKING GROUP

Franz Richter, Oesterreichische Nationalbank, CHAIRMAN

Laura Auria, Deutsche Bundesbank

Mireille Bardos, Banque de France *(until summer 2007)*

Francois Coppens, National Bank of Belgium

Liliana Toledo Falcón, Banco de España

Martin Fassbender, Deutsche Bundesbank

George van Gastel, National Bank of Belgium

Susana Filipa Lima, Banco de Portugal

Javier Maycas, Banco de España

Romulus Mircea, National Bank of Romania

Cyrille Stevant, Banque de France *(since fall 2007)*

Gianni Tessiore, Centrale dei Bilanci

Cedric Traversaz, Banque de France

Gerhard Winkler, Oesterreichische Nationalbank

ACKNOWLEDGEMENTS

We are indebted to Federico J. Sáez y Pérez de la Torre from Banco de España and Homero Alexandre Goncalves from Banco de Portugal, who all shared their insights and provided many helpful comments.

We also wish to thank Mr. Peter Buchegger, Ms. Birgit Fromm-Leichtfried, Ms. Claudia Neumayer, Mr. Reinhard Petschnigg, Ms. Eva-Maria Springauf and Ms. Sabine Wukovits from Oesterreichische Nationalbank for their help in editing and publishing this paper as well as respondents from the European Committee of Central Balance Sheet Data Offices for their support.

PREFACE

F. VARETTO / PRESIDENT OF ECCBSO

The **European Committee of Central Balance Sheet Data Offices (ECCBSO)** brings together central banks, statistical offices and similar institutions within the European Union, that posses large datasets of financial statements stemming from non-financial enterprises. The objective of the ECCBSO is an exchange of views in order to improve the analysis of financial statements, particularly for assessing the creditworthiness of enterprises and for statistical purposes. The goal of the ECCBSO is achieved through several Working Groups dedicated to specific tasks. In co-operation with the European Commission the ECCBSO set up the **BACH Database**, which contains aggregated harmonised financial statements statistics to support macro-economic analysis. In addition, the ECCBSO developed the **European Sector Reference** Database which provides quartiles statistics useful for micro-economic comparisons. Both databases are publicly available on internet. More detailed information on the Committee's activities will soon be available on the **website** which is currently in process of being launched (www.eccbso.org).

The research of the **Working Group on Risk Assessment** of the ECCBSO has been dedicated (since its creation in the nineties of the last century) to credit risk modelling and measurement, recognising that this field has become a timely, important and widespread matter. In all areas of finance, research on credit risk modelling and measurement is extensive in both, industry and academia.

In Europe, this trend is driven by: progress in risk measurement and management techniques, growth and integration of credit markets, the adoption of the revised Basel Capital Accord and the new collateral framework of the Eurosystem. Within this context the Working Group has the following fundamental aims:

- to analyze and demonstrate the importance of financial accounting data in the assessment of corporate creditworthiness
- to study and further develop different methodologies for credit risk assessment and model validation
- to empirically investigate and compare the advantages and disadvantages of various measurement and validation techniques from a practical, mostly central-bank,

perspective, making use of the aforementioned ECCBSO's datasets of financial statements from non-financial enterprises.

The collection of papers contained in this publication shows well the spirit behind the work of this group. The papers do not only address concrete and timely problems in credit risk assessment from a purely methodological side but also study their practical implications. I thank the group for its great work and hope that this publication will be read by academics, risk managers, regulators, and others who might find the problems and solutions, intellectually challenging and practically useful.

TABLE OF CONTENTS

I. INTRODUCTION

F. RICHTER / OESTERREICHISCHE NATIONALBANK

Credit risk has been one of the most active areas of recent financial research. In pursuit of its mission, the Working Group on Risk Assessment of the ECCBSO aims at contributing to this distinct field of research. The papers collected in this publication answer important scientific questions and suggest prospective paths for new research. Furthermore, they contain implementation methods and provide solutions to practical problems in credit risk measurement. In most of the papers, empirical examples are carried out from the special perspective of a central bank as most of the participats stem from such institutions. Due to their responsibility for implementing monetary policy and the involvement of many of these institutions in banking supervision, central banks have a natural interest in accurately measuring and managing credit risk. With respect to the topics being addressed, the papers could be classified into three different groups:

I.1 DEFINING THE DEFAULT EVENT

In credit risk modelling, the probability of default (PD) is one of the key parameters to be estimated. Research on different techniques for the estimation of the probability of default is extensive. It appears, however, that too little attention is paid to the different possible definitions of default in practice, although a clear understanding of the definition of default is crucial for a correct interpretation of any estimate of a PD. In his contribution, Mr. Traversaz addresses exactly this topic. He provides an overview of the various definitions of default used in the field of credit risk assessment and highlights their differences. Furthermore, he gives a practical example when he explains how Banque de France recently changed its definition of default from the original “narrow” one to a broader one. His findings demonstrate the importance of transparency concerning the default definition being used when publishing default statistics or any other information related to the assessment of the quality of a rating system.

I.2 MODELLING CREDIT RISK

Two interesting papers provide empiricial evidence on credit scoring and default prediction in central banks. In her contribution, Ms. Toledo Falcon explains in great detail how Banco de Espana is currently developing logit models in order to enlarge its methodological spectrum used for the in house credit risk assessment. She discusses the merits of logit models and interprets the results obtained from her model, as well as the model's power.

The second paper on credit scoring was contributed by Ms. Lauria Auria and Mr. Moro. The authors discuss the advantages and disadvantages of support vector machines (SVMs) as a new promising non-linear, non-parametric classification technique, which can be used for credit scoring. After a basic review of the SVMs and their advantages and disadvantages on a theoretical basis, empirical results of an SVM model for credit scoring (calibrated using data of Deutsche Bundesbank) are presented and discussed.

Whereas the two previous papers are concerned with default prediction of firms, in their contribution, Mr. Tessiore and Mr. Favale study the different parameters of credit risk of Project finance. Under Basel II, Project finance (PF) is one of five sub-classes of specialized lending (SL) within the corporate asset class. In their paper, they suggest a quantitative method based on Monte Carlo (MC) simulations of future cash flows (of the project involved with a Project Finance operation) that allows for an analytical estimation of the PD, the LGD and the EL. Furthermore, they explain how their approach has been implemented by Centrale dei Bilanci.

Ms. Bardos rounds off the discussion on credit risk modelling when she explains in great detail what is at stake when estimating the probability of default using a scoring function. In her paper she presents some credit scoring construction principles, which increase the quality of the tool and the accuracy of default probability. She discusses some arguments for the choice of a model and concentrates on the comparison between Fisher linear discriminant analysis (LDA) and logistic regression (LOGIT). Furthermore, she shows how to determine stable risk classes and how to derive accurate estimates of the probabilities of default. In her empirical examples she uses data of the Banque de France.

I.3 VALIDATING CREDIT RISK MODELS

The last contribution in this collection of papers is concerned with validation. The aims of the paper of Mr. Coppens and Mr. Winkler are twofold;

a. First, they attempt to express the threshold of a single “A” rating (as issued by major international rating agencies) in terms of annualised probabilities of default. They use public data from Standard & Poor’s and Moody’s to construct confidence intervals for the level of probability of default to be associated with the single “A” rating. The focus on the single A rating level is not accidental. Single A is the credit quality level at which the Eurosystem considers financial assets as eligible collateral for its monetary policy operations.

b. Second, they attempt to review various existing validation models that enable the analyst to check the ability of a credit assessment system to forecast future default events.

Within this context, the paper proposes a simple mechanism for comparison of the performance of major rating agencies and other credit assessment systems. The aim is to provide a simple validation yardstick to help monitoring the performance of the different credit assessment systems, more specifically, those participating in the assessment of eligible collateral for Eurosystem’s monetary policy operations. Their analysis extends the earlier works of Blochwitz and Hohl (2001) and Tiomo (2004). Mr. Blochwitz has been a former member of the Working Group on Risk Assessment. Thanks to his contributions, the first “Traffic Light Approach” could be developed for the Eurosystem. This approach has been extended by Mr. Coppens and Mr. Winkler and is now being used by the Eurosystem to monitor the performance of credit rating systems in the context of the European Credit Assessment Framework.

II. DIFFERENT DEFINITIONS OF DEFAULT: ISSUES FOR COMPARISON AND DISCLOSURE OF RATING PERFORMANCE – AN OVERVIEW WITH A SPECIAL FOCUS ON UNPAID TRADE BILLS IN FRANCE

C. TRAVERSAZ / BANQUE DE FRANCE

II.1. SYNOPSIS

When it comes to credit risk assessment, the probability of default (PD) plays an essential role. In particular, it has an important influence when designing risk classes or drawing a comparison between different rating scales. At the same time less attention is paid to the different definitions of default which are a basic input factor to calculate these PDs.

This article deals with the various definitions of default used in the field of credit risk assessment. Firstly, it shows that different definitions exist and that the information available differs significantly from one player to another. Secondly, it describes the particular case of Banque de France which has recently added a broader definition of default to its original "narrow" definition. A comparative analysis of the properties of these two definitions highlights the difficulty of comparing PDs which are based on different definitions of default, even when applied to a common portfolio. These findings lead to the emphasis of the importance of transparency when publishing default statistics and more generally any information related to the assessment of the quality of a rating system.

II.2. INTRODUCTION

When introducing the Basel 2 framework, the Basel Committee has come up with a new definition of default. According to the reference text[1] each bank which intends to adopt the Internal Rating Based (IRB) approach has to use this definition of default when calculating its minimum capital requirements for credit risk. According to this definition, a default is deemed to have occurred when "the bank considers that the obligor is unlikely

[1] Basel 2 Committee - *International Convergence of Capital Measurement and Capital Standards*.

to pay[2] its credit obligations to the banking group in full, without recourse by the bank to actions such as realising security (if held)" or when "the obligor is past due more than 90 days on any material credit obligation to the banking group (…)".
This clarification was necessary not only for obtaining a coherent framework but also for triggering a convergence process of all the different definitions of default which are still in use by the different players in the credit risk assessment field (rating agencies, commercial banks, national central banks and others). The wide set of definitions currently used often makes it extremely difficult to identify and measure exactly the link among them. The reasons for these difficulties are of different nature like access to information, interpretability of definitions, scope of companies (listed on the Stock Exchange, small companies …) etc.
Actually, each player in the rating sector has used the data available in the best possible way, whereas this data can differ strongly from central banks to commercial banks or to international rating agencies. Even among European national central banks the available information differs according to domestic legal and organizational specificities. In order to use the available data in the most efficient way, each player tries to take into account the proprietary information to design its rating scale and, sometimes, its definition of default.
This has been achieved by Banque de France (BdF) when we decided to include unpaid trade bills[3] in our method. French Law has put under BdF responsibility collection and management[4] of payment default data on trade bills. Due to the high success rate of predicting risk, it had been decided to use this information for the design of some grades of our rating scale[5], in particular the worse ones.

As a result BdF added to its failure definition based on judicial proceedings (information publicly available) a default definition (called Banque de France's definition of default) with a larger scope than the definition based on the occurrence of judicial proceedings, using unpaid trade bills.

The first part of this article is dedicated to the different definitions of default which are currently used by players who assess credit risk. The second part deals with how BdF designed its particular definition of default, the issues faced hereby and the comparison of the two definitions of default used in Banque de France.

[2] This notion is explained in details in the original text.
[3] A bill of exchange is an unconditional order in writing, addressed by one company to another, requiring the company to whom it is addressed to pay on demand, or at a fixed and determinable future time, a sum certain in money to the order of a specified company, or to bearer. It is a negotiable instrument, similar to a post-dated cheque, which is usually sold at a discount.
[4] In accordance with the amended regulations 86.08 and 86.09 of the Banking and Financial Regulations Committee
[5] Banque de France rating grades are from the best to the worse : 3++, 3+, 3, 4+, 4, 5+, 5, 6, 8, 9, P. Rating P is awarded in case of legal proceedings and ratings 8 and 9 when unpaid trade bills occur (depending on the cumulated amount).

II.3. DIFFERENT PLAYERS, DIFFERENT DEFINITIONS OF DEFAULT

The world of credit risk assessment is a wide one. This is why so many credit assessment institutions can be found all over the world and why each player focuses on a specific population. This specification can be caused by location (National Central Banks for example), portfolio limits (commercial banks), whether a company is publicly listed or not, has recourse to bond and other traded securities markets or not (international agencies) etc. Due to these different scopes, each player uses the information he has access to in order to design the most appropriate rating scale. As a result a variety of definitions of default are in use nowadays.

A comparison of all these definitions of default showed that the most common input is related to failure events. As bankruptcy and insolvency are public information and as these events are the last step in the falling down process of a company, almost all players which assess credit risk use these judicial events to design their worse grades.
For example, Banque de France's worst grade is defined exactly in this way: the P rating is related to companies undergoing legal proceedings (turnaround procedure or judicial liquidation). Moreover, this "failure" definition is currently the most widely used within Eurosystem National Central Banks which run an In-house Credit Assessment System. However, even though the failure event may appear as the greatest common divisor of all the definitions of default in use, one must not forget that even failure events may differ from one country to another because of differences between judicial procedures.

Many other players use what could be named "default on marketable debt instruments". This is namely the approach which is implemented by international rating agencies. The advantage of this method is that this information (non payment on marketable debt instruments) is public. On the other hand it only applies to big companies and, moreover, such a default may occur rather late in the financial difficulty process of a company as it has an influence on market share or bond value.

The third group of players which can be easily identified are mainly commercial banks. Due to their business activity they have access to defaults on the bank loans that are part of their loans' portfolio and they generally use this information to design their definition of default. More precisely, banks who want to adopt the IRB approach have to use it because it is part of the Basel 2 definition of default ("the obligor is past due more than 90 days on any material credit obligation to the banking group (…)"[6]). Some National Central Banks in the Eurosystem gather this information from commercial banks as well. This is the case, for example, in Austria and in Portugal.

Last, but not least, some players use different definitions or a combination of the previously mentioned definitions, always depending on the availability of information. For example, Banque de France uses failure and unpaid trade bills.

Looking at all these definitions the decision to promote way in the Basel 2 framework a definition of default which aims at including all the possible default events while being precise enough to be applied in a harmonised way by all credit institutions brings a major improvement. Nevertheless, it is interesting to point out that this definition of default has to be used (in the Basel 2 framework) only by commercial banks which adopt the IRB approach. External Credit Assessment Institution (ECAI), the institutions in charge of assessing credit risk for banks adopting the Standardised Approach, have no obligation to adopt the Basel 2 definition of default, although the guidelines set up by the European Committee of Banking Supervisors encourages converging towards this definition. As a result it is the role of the supervisors, given all the data concerning default rates and the definition of default used by the institutions, to map the rating scale of each ECAI to the so called "reference scale"[7].

One of the reasons why a harmonised default-definition for all the ECAI has not been required is that an ECAI can be any player assessing credit risk (Rating Agencies, commercial firms such as credit bureaus, National Central Banks…), provided that this player matches the requirements sets by the CEBS guidelines and by the national supervisors. As each player is interested in a specific population (e.g. large vs. small companies, quoted vs. unquoted …) and thus uses a definition of default which fits with the own portfolio, the determination of a harmonised default-definition for all the players in the market does not seem practicable at least over the short term.

Nevertheless national supervisors have to take into account the specific definition of each applicant: "the Capital Requirements Directive requires competent authorities to consider qualitative factors such as the pool of issuers covered by the ECAI, the range of credit

[6] Basel 2 Committee - *International Convergence of Capital Measurement and Capital Standards*.
[7] The reference scale is based on default rates (AAA-AA is equivalent to a PD of 0.10%, A to 0.25%, BBB to 1.00%, BB to 7.50% and B to 20.00%). See Annex 2, Internal Convergence of Capital Measurement and Capital Standards – Basel Committee on Banking Supervision.

assessments that it assigns, the meaning of each credit assessment, **and the ECAI's definition of default**"[8].

The following second part of this article illustrates how Banque de France dealt with the difficulty of comparing different definitions of default when introducing a second definition. The latter was implemented to have a better understanding on the impact of widening the original definition of default (based only on legal proceedings).

II.4. Designing a new definition of default: Banque de France`s experience

Currently Banque de France uses two different definitions of default. The "narrow" one ("BdF failure") is based on legal proceedings (turnaround procedure or judicial liquidation). The "broad" one ("BdF default"), which includes the first one, is based in addition on unpaid trade bills.

"BdF failure" definition: a company is said to be in "failure" if there is a judicial procedure (turnaround procedure or liquidation) against it → the company is rated as P. **"BdF default" definition:** a company is said to "default" if the previous condition applies to the company (a judicial procedure) or if the company gets a rating of 9 because of trade bill payment incidents declared by one or several credit institutions.

Box 1: BdF failure & default definition

Deriving from the SIT[9] (Interbank Teleclearing System) Banque de France is managing a database (called CIPE) which registers all the incidents concerning payments of trade bills. This database is fed automatically by all the banks located in France through dematerialised automated online reporting systems. When an incident on a payment of a trade bill occurs, the following information is registered: supplier, customer, amount, expected date for payment and the reason for the non payment. Reasons for non payment are divided into two main categories. The first one is "Impossibility to Pay" (IP) which includes for example reasons such as "insufficient deposit" or "judicial decision".

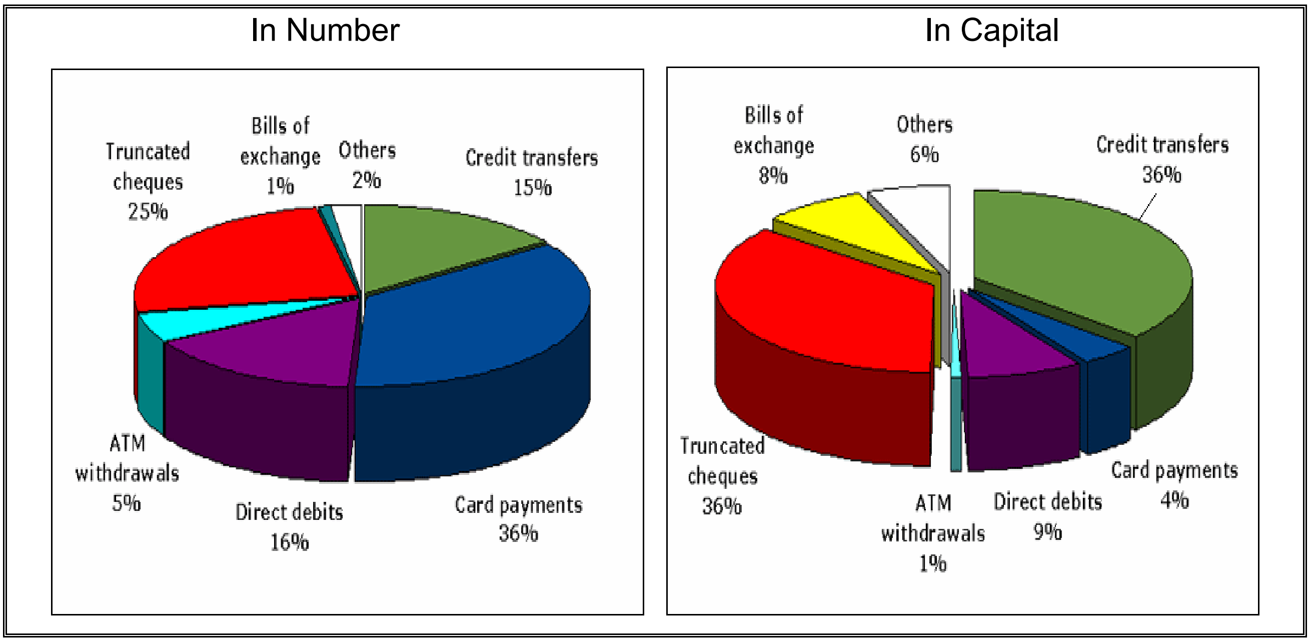
[8] C.E.B.S. – Guidelines on the recognition of External Credit Assessment Institutions – 20 January 2006 – Article 147.
[9] France has three interbank payment systems, two for large-value payments (cash transfers, settlements for delivery of securities,...) and one for retail payments (cheques, bank cards, credit transfers, direct debits, interbank payment orders (TIP), bills of exchange, ...). The two large-value systems are the Transfers Banque de France (TBF) system, which forms part of TARGET, the Trans-European Automated Real-time Gross settlement Express Transfer system, and the Paris Net Settlement (PNS) system. The Interbank Teleclearing System (SIT) handles retail payments, clearing interbank exchanges of cashless payment instruments.

The second one may be identified as “Claim Contestation” (CC); it regroups reasons such as “disagreement on the amount of the claim” or “on the date”, “incomplete payment” (see below), etc.

- **“Impossibility to Pay” reasons:**
  - holder dead
  - request for prorogation
  - without sufficient funds
  - will be paid by subrogation
  - insufficient deposit
  - no order to pay
  - judicial decision
  - objection to payment on the account
- **“Claim Contestation” reasons:**
  - received by error: already paid
  - late complaint
  - disputed drawing
  - disputed amount
  - disputed date
  - incomplete payment
  - objection to payment by the debtor

Box 2: Reasons for unpaid trade bills

In France trade bills of exchange are a commonly used means of payment. In 2006 they represented 1% in number but 8% in value of the 12.3 billion transactions, worth a total of € 5.075 billion, processed by the SIT[10]. Furthermore, for small and medium enterprises trade credit is a major source of business financing.



Source: www.gsit.fr

Exhibit 1: Operations processed via the SIT in 2006

[10] Source: www.gsit.fr

II.4.1. Banque de France's rating: "Cotation Banque de France"

The database of Banque de France (FIBEN) records information on a broad population of natural and legal persons, including sole traders. The branch network[11] collects a wide range of information on these populations.

The information gathered on each natural person or legal entity is analysed and then issued as a succinct overall assessment that can be easily interpreted by the different users. The rating is an overall assessment of a company's ability to meet its financial commitments over a three year horizon. The rating process is not an automatic one. Instead this rating is established on the basis of numerous factors like assessing the company's financial position (profitability and balance sheet structure), its expected development, the assessment of management, shareholders, the economic environment, affiliated companies or those with which the company has close financial commercial ties and the existence of payment incidents or legal proceedings.

The credit rating is represented by a number which can be followed by a plus. The different ratings, ranging from best to worst, are: 3++, 3+, 3, 4+, 4, 5+, 5, 6, 8, 9 and P. The P rating refers to companies undergoing legal proceedings while ratings 8 and 9 indicate the existence of payment incidents on trade bills. More precisely, the difference between 8 and 9 depends both on the amount (different thresholds) and on the nature of unpaid trade bills (IP are much more risky than CC). In order to ensure a balanced assessment, the decision on the rating is taken by the analyst after, among others, having consulted a short complementary report drafted by the banker that has reported the default event.

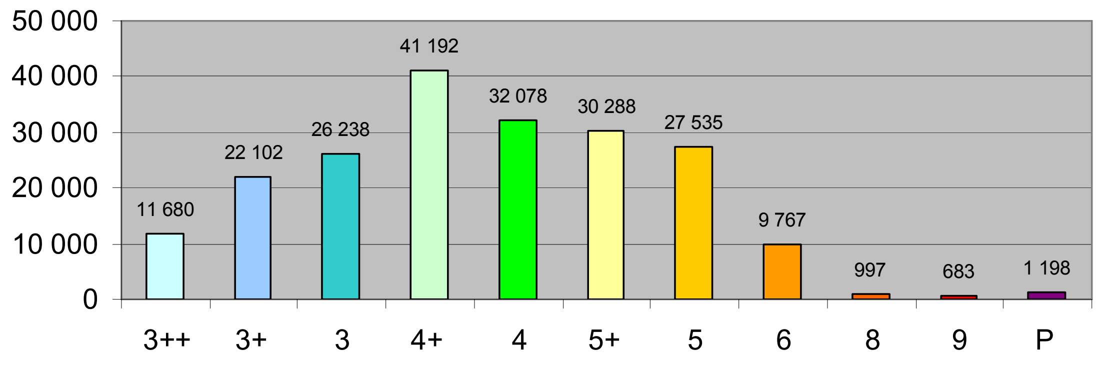
On December 31st 2006 the database included about 5.2 million companies of which about 220 000 were rated according to this methodology.

[11] The branch network of Banque de France is constituted by 96 departmental branches and 21 economic centres.

Signification of Banque de France's ratings

3++: Excellent
3+: very good
3: Good
4+: quite good
4: Acceptable
5+: fairly poor
5: Poor
6: very poor
8: in difficulty, in view of rather frequent reported payment incidents
9: compromised, when reported payment incidents indicate severe cash flow difficulties
P: initiation of legal proceedings (turnaround procedure or judicial liquidation)

Numbers of companies rated after a study of their financial statements the January 1st, 2006.

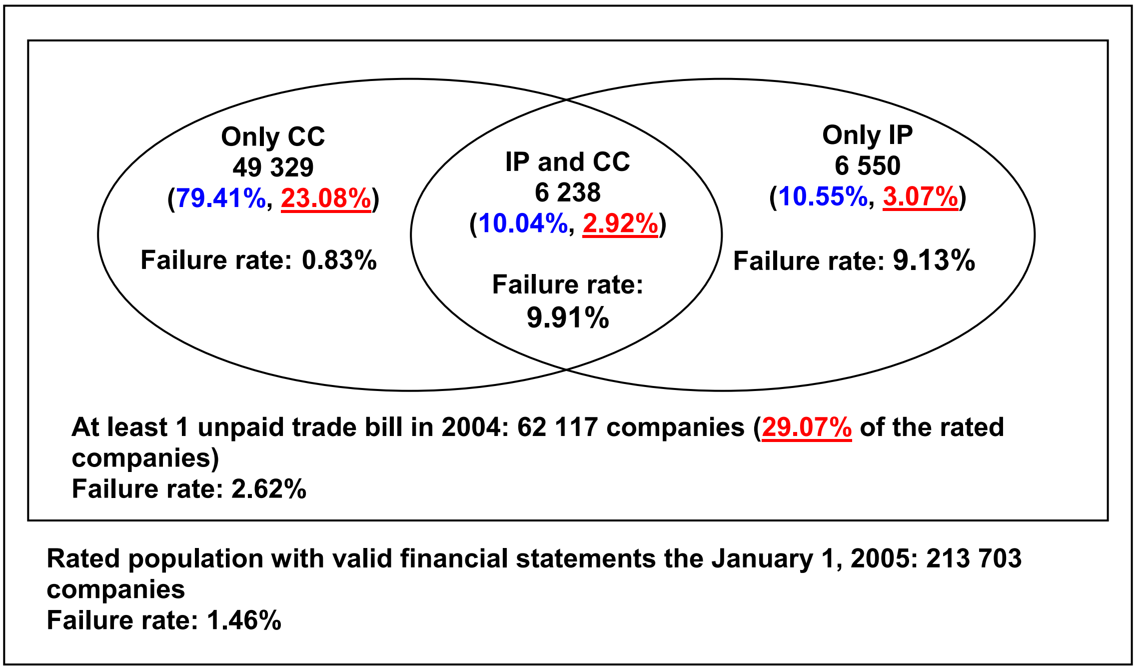


Source: Banque de France

Exhibit 2: Signification of Banque de France`s ratings

II.4.2. Assessing the predictive power of unpaid trade bills

The first step before integrating a new element into a methodology is to check its efficiency. This can either be done by comparing the failure rates obtained for each category of companies (with or without unpaid trade bills) or, to be even more accurate, by separating companies depending on the reasons of unpaid trade bills. The results of this preliminary work are shown in the Chart 2 below.



(**X%**, **Y%**) :
- X% of the companies with at least 1 unpaid trade bill in 2004
- Y% of the companies rated the January 1, 2005

Failure: judicial proceedings (turnaround procedure or judicial liquidation)

CC: unpaid trade bill with a “claim contestation” reason.

IP: unpaid trade bill with an “impossibility to pay” reason.

29% of all the companies were concerned by at least one unpaid trade bill in 2004.

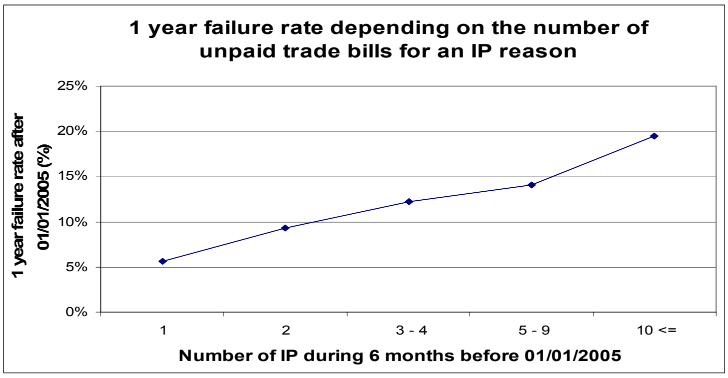
Source: Banque de France

Exhibit 3: Failure rates depending on the population

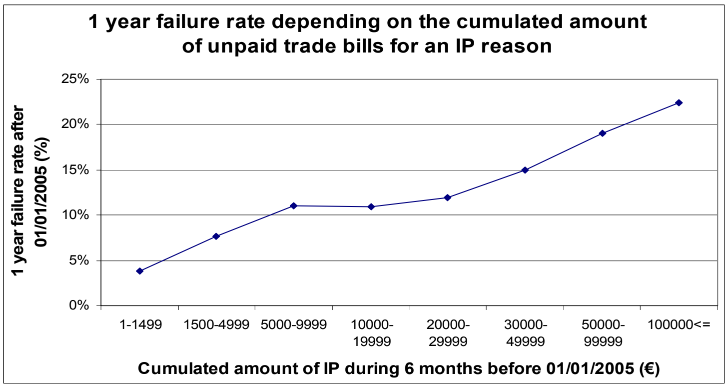
Looking at the results obtained, at least two conclusions may be drawn. The first one is that taking into account the reason of non payment is absolutely crucial. The second is that, according to failure rates obtained for unpaid trade bills with an “IP” reason (Impossibility to Pay), this element (IP) has a predictive power to detect failure.

II.4.3. Designing rating grades taking into account unpaid trade bills

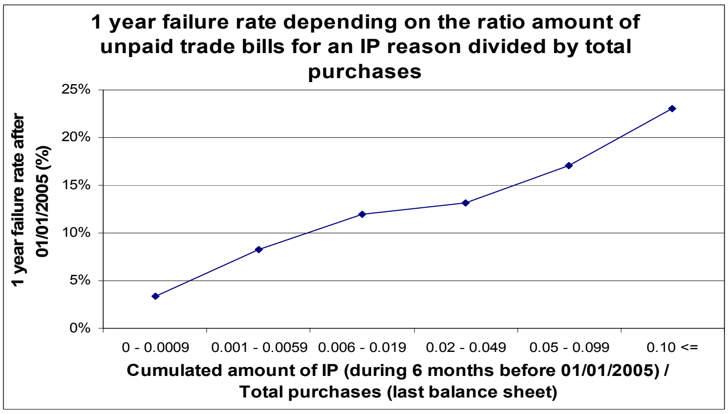
After having defined a predictive element to detect failure as a starting point, the next step was to learn more about the relationship between the repartition of the unpaid trade bills (in numbers, in amounts and relatively to the purchasing expenses) and the failure rates. Results are shown in Exhibit 4.



Number of IP	Nb	1 year failures	1 year failure rates
1	2 789	158	5,67%
2	1 195	111	9,29%
3 - 4	1 106	135	12,21%
5 - 9	1 254	177	14,11%
≥ 10	1 492	291	19,50%



Cumulated amount of IP during last 6 months (€)	Nb	1 year failures	1 year failure rates
1 - 1 499	1 721	66	3,83%
1 500 - 4 999	1 106	85	7,69%
5 000 - 9 999	834	92	11,03%
10 000 - 19 999	1 173	128	10,91%
20 000 - 29 999	753	90	11,95%
30 000 - 49 999	854	128	14,99%
50 000 - 99 999	869	165	18,99%
≥ 100 000	526	118	22,43%



Cumulated amount of IP / Purchases (%)	Nb	1 year failures	1 year failure rates
0 - 0.09	1 459	49	3,36%
0.1 - 0.59	1 395	115	8,24%
0.6 - 1,9	1 597	191	11,96%
2 - 4,9	1 433	189	13,19%
5 - 9,9	674	115	17,06%
≥ 10	378	87	23,02%

Source: Banque de France

Exhibit 4: Repartitions of the unpaid trade bills depending on failure rates

Given the results of this analysis Banque de France defined two different thresholds and consequently created two additional grades based on unpaid trade bills. These grades are slightly better than “P” (P stands for a failure event) and have been named “8” and “9”.

II.4.4. Designing a new definition of default

Mainly for stability reasons (actually, we did not want a volatile grade with companies just coming and leaving; see chart 4 below) it had been decided to define only grade 9 as additional default rating class (beside “P”).

After introducing the two new rating grades it was necessary to observe after a time span of one year the impact of the change in the rating scale on the portfolio by means of a transition matrix.

	December 31st, 2006														
January 1st, 2006 / %	3++	3+	3	4+	4	5+	5	6	7	8	9	P	0	Number	Default (9+P)
3++	67,6	17,4	5,0	3,6	1,2	0,3	0,1	0,0	0,0	0,0	0,0	0,0	4,6	11 815	0,01
3+	10,6	55,2	16,1	8,9	3,4	0,6	0,4	0,1	0,0	0,0	0,0	0,0	4,6	21 459	0,03
3	1,7	15,2	53,1	16,5	6,5	1,0	0,6	0,1	0,0	0,0	0,0	0,1	5,3	27 316	0,11
4+	0,8	3,5	12,7	50,9	15,7	4,5	2,7	0,3	0,1	0,1	0,0	0,5	8,1	45 528	0,51
4	0,1	0,7	3,5	22,7	44,6	10,4	6,6	0,8	0,2	0,3	0,1	0,9	9,2	33 921	1,03
5+	0,0	0,1	0,4	8,0	13,0	50,1	12,8	2,3	0,3	0,5	0,3	1,4	10,7	30 918	1,7
5	0,0	0,1	0,2	3,5	7,4	18,9	47,2	6,4	0,7	0,9	0,5	3,2	11,0	26 069	3,7
6	0,0	0,0	0,2	1,0	2,5	8,8	18,9	55,1	0,7	0,7	0,3	4,3	7,5	9 099	4,54
8	0,0	0,0	0,1	0,6	4,5	8,8	24,7	9,2	5,9	12,4	12,3	17,5	4,1	887	29,76
9	0,0	0,0	0,0	0,5	1,8	3,6	12,9	8,7	6,0	15,6	25,4	22,8	2,7	552	48,19
P	0,0	0,0	0,0	0,0	0,1	0,0	48,2	0,4	0,4	0,5	0,0	45,7	4,7	963	45,69
Total	5,3	9,6	12,3	19,8	14,9	13,1	10,8	3,8	0,3	0,4	0,3	1,4	8,1	208 527	1,67

The transition matrix shows the migration of the rated companies during a period of one year. For example, 17.4% of the companies rated 3++ by January 1st 2006, were rated 3+ by December 31st 2006.

Source: “The Banque de France rating – a performance evaluation (failure and default rates, transition matrixes)”[12]

Exhibit 5: Transition matrix

[12] This report can be downloaded on Banque de France website: http://www.baqnue-france.fr/fr/instit/telechar/services/echelle.pdf

The transition matrix proves that introducing grades 8 and 9 into the rating scale has no significant impact on the best grades but affects mostly the worse ones. This is a first clear indication that default and failure rates do not differ too much for best grades.

Rating	Number of companies	1 year failure	
		Number	Rate (%)
3++	11 635	0	0,00%
3+	21 156	2	0,01%
3	26 418	14	0,05%
4+	44 418	138	0,31%
4	33 306	230	0,69%
5+	31 168	405	1,30%
5	26 874	892	3,32%
6	9 584	495	5,16%
8	877	163	18,59%
9	500	95	19,00%
TOTAL	**205 936**	**2 434**	**1,18%**

Companies rated after a study of 2004 financial statements

Source: "The Banque de France rating – a performance evaluation (failure and default rates, transition matrixes)"[13]

Exhibit 6: Failure rates for each grade

Failure rates confirm that introducing unpaid trade bills has allowed increasing the discriminatory power of the rating scale. Actually, one grade (initially grade 6) has been split into three grades with increasing failure rates. For example, assuming that companies rated 8 or 9 would have been rated 6 (the worse grade before grades 8 and 9 based on unpaid trade bills were created), it had been made possible to separate a grade with a failure rate of 6.87% ($= \frac{495+163+95}{9584+877+500}$) into three grades with failure rates from 5.16% to 19.00%.

[13] This report can be downloaded on Banque de France website: http://www.baqnue-france.fr/fr/instit/telechar/services/echelle.pdf

II.4.5. Comparing default and failure rates

The next chart shows a comparison between the two definitions of default ("BdF failure" and "BdF default"). The figures confirm that default and failure rates are very similar for the best rating grades. On the other hand results can differ significantly as far as worse grades are concerned. However, this was expected when enlarging the "BdF failure" definition of default, assuming that Banque de France's methodology is not only able to predict failure but also to predict emerging financial difficulties.

Rating	Number of companies	1 year failure		1 year default	
		Number	Rate (%)	Number	Rate (%)
3++	11 635	0	0,00%	2	0,02%
3+	21 156	2	0,01%	2	0,01%
3	26 418	14	0,05%	16	0,06%
4+	44 418	138	0,31%	168	0,38%
4	33 306	230	0,69%	301	0,90%
5+	31 168	405	1,30%	508	1,63%
5	26 874	892	3,32%	1 050	3,91%
6	9 584	495	5,16%	564	5,88%
8	877	163	18,59%	347	39,57%
9	500	95	19,00%	500	100,00%
TOTAL	**205 936**	**2 434**	**1,18%**	**3 458**	**1,68%**

Companies rated after a study of 2004 financial statements

Source: "The Banque de France rating – a performance evaluation (failure and default rates, transition matrixes)"[14]

Exhibit 7: Default and failure rates

II.5. SUMMARY AND CONCLUSION

This article deals with the relevance of default definitions particularly as there are many definitions in use by the different market players (rating agencies, commercial banks, national central banks, etc.) and as there is an ongoing discussion process triggered by national and pan-European regulations in the context of Basel II and the European Directive[15].

[14] This report can be downloaded on Banque de France website: http://www.baqnue-france.fr/fr/instit/telechar/services/echelle.pdf
[15] European Directive 2006/48/CE of the European Parliament and of the Council of June 14, 2006, relating to the taking up and pursuit of the business of credit institutions (Banking. Directive).

Comparing the two definitions of default used in Banque de France suggests that there is an unambiguous correlation between default definition and default rates. However the results obtained indicate that the impact of the definition of default on default rates is quite complex as there is no simple way to switch from one definition to another, for example with a linear coefficient.

As a first consequence, comparing ratings from different rating issuers which do not exactly use the same definition of default is far from being straightforward. As a second implication, the publication of default rates should be accompanied with a comprehensive description of the methodology used, and any change in the methodology should be disclosed. This effort in improving transparency would encourage sharing best practices and helping reduce information asymmetries between credit assessment institutions and rating users.

III. LOGIT MODELS TO ASSESS CREDIT RISK

L. TOLEDO FALCÓN / BANCO DE ESPAÑA

III.1. INTRODUCTION

The rating of a firm can be defined as the mapping of the expected probability of default into a discrete number of quality classes or rating categories (Krahnen and Weber, 2001). For the quantification of the expected likelihood of future default, rating systems are based on models that may combine qualitative elements with quantitative financial and economic information.

Banco de España (BdE) carries out an in-house credit assessment of non-financial companies in order to decide which of them can be considered eligible collateral for monetary policy purposes. At the moment, an expert-based system is used, which allows considering very relevant qualitative aspects. As this method is highly time-and-resource-consuming, this limits the amount of firms that can be analyzed, which is around one hundred. In addition to that a bias of big enterprises has to be observed.

With the aim of enlarging the methodological spectrum of BdE, as well as of producing estimated probabilities of default[16], BdE has been developing a new methodology which is to be put into practice in the immediate future. In this sense, logit models will be used as additional tools in the in-house credit risk assessment process.

In this respect, a very important point, that must be stressed, is that the objective is to select a set of very good companies by minimizing the error of misclassifying defaulting companies as sound ones (type I error), while at the same time being aware that some companies, due to lack of data, will not be examinable. In addition to this, the decision to be taken on each single firm, based on the probability of default estimated by a logit model, will be whether such firm is eligible or not. At this moment there are no plans to produce ratings that allow classifying firms in different homogeneous groups. To do so, an additional step would be required to map probabilities of default into ratings.

[16] As explained in section 3, the default definition used here is past due more than ninety days on any credit obligation.

The objective of this article is to present the methodology that the Operations Department of BdE is following in the estimation of logit models, as well as some general results for the Spanish sample. To do so, section 2 briefly presents the general concepts and some particular details of the methodology applied. Section 3 describes the data, which undergo a univariate analysis in section 4. Section 5 presents the results for some multivariate models and, finally, section 6 draws the main conclusions from the analysis.

III.2. METHODOLOGY

Since the late 60s, many studies have tried to demonstrate the usefulness of accounting information periodically published by firms to detect default proximity. Since the pioneering papers by Beaver (1966) and Altman (1968 and 1977), a lot of studies have tested a wide range of variables and proposed many different methodologies. Every rating model starts from a definition of failure which defines two sub-samples of firms: failure and non-failure firms. The next step is to use one or several ratios commonly used to evaluate firm situation to discriminate between both groups. The most used techniques have been the statistical ones, especially multiple discriminant analysis and conditional probability models (logit and probit). More recently, iterative partitions, artificial neural networks and support-vector machines have also been applied.

In this paper logit models are discussed, which were introduced for default prediction purposes by Ohlson (1980) and Zavgren (1985), and which have many positive qualities:

- they do not assume multivariate normality;
- they are transparent when evaluating the meaning of coefficients, this is, the importance of each variable;
- they allow to obtain a direct estimation of probability of failure (default);
- according to the literature, they show good predictive results when compared to other techniques;
- they allow taking into account non-linear relations among variables; which has proven to be very useful in BdE´s sample; and
- they work well with qualitative explanatory variables.

Logit analysis is a conditional probability technique which allows studying the relationship between a series of characteristics of an individual and the probability that this individual belongs to one of two groups defined *a priori*. A binary variable Y showing the status of the firm at the end of a certain period is explained by a set of factors X according to (1):

$$P(Y_i = 1/X_i) = F[X_i, \mathrm{B}] \quad ; \quad P(Y_i = 0/X_i) = 1 - F[X_i, \mathrm{B}] \qquad (1)$$

where:

Y_i : binary failure variable for firm i, taking value "1" if the firm fails and "0" otherwise

X_i : values for firm i for the set of J independent variables

B : set of parameters

The logit model, whose general equation is given by (2), guarantees that $F[X_i, B]$ is constrained to interval [0,1]. A direct consequence of such specification is expression (3) [17], which defines the score.

$$p_i = P(Y_i = 1/X_i) = \frac{1}{1 + e^{-\alpha - \sum_j \beta_j x_{ji}}} \qquad (2)$$

$$Z_i = \log it(p_i) = \ln \frac{p_i}{1 - p_i} = \alpha + \sum_j \beta_j x_{ji} \qquad (3)$$

where:

β_j : parameter of variable j

x_{ji} : value of variable j for firm i

Z_i : score of firm i

[17] Although this is the usual way of presenting a logit model, with variables entering the equation in a linear way; this type of methodology also allows controlling for non-linear relationships among variables. In non-linear logits, the corresponding quadratic terms are also part of the equation.

Parameters α and β are estimated by maximum likelihood for a joint probability given by (4), where *N* is the number of firms in the sample. A positive sign in the estimated coefficient β^*_j implies a positive relation between variable *j* and the probability of failure.

$$P\left(Y_1, Y_2, \ldots, Y_N / X\right) = \prod_{i=1}^{N} p_i^{Y_i} \left(1 - p_i\right)^{1-Y_i} \qquad (4)$$

As explained above, the endogenous variable (*Y*) in a logit model is a binary variable taking value 1 if the firm fails and 0 otherwise. The concept of failure, however, embraces a wide range of status, including discontinuities affecting the firm, the owner or the business; economic and financial failure and legal insolvency. Most studies use liquidation, bankruptcy or default as a definition for failure. Here, the default definition used by the Central Credit Register (CIR) of BdE is applied, which coincides with the Basel II definition of default.

The selection of the independent variables (*X*) is crucial for the model's performance, since they must be significant and relevant to differentiate between "good" and "bad" firms. Firstly, a wide set of variables was considered. This range of variables includes the main economic and financial ratios of the firm which, according to the related literature and to BdE's experience, have a strong discriminating power for credit risk. Other non-financial factors, both quantitative and qualitative, were also considered to be potentially relevant. Some of the analysed variables are discussed - as described in detail in Table 1 in the Annex - and can be divided into four groups according to the aspect of the firm they try to measure:

- *Solvency*: financial autonomy, adjusted financial autonomy, total leverage, financial leverage, financial expenses coverage (EBIT), financial expenses coverage (EBITDA) and repayment capacity.
- *Profitability*: economic return, financial return, operating margin and resource generation capacity.
- *Liquidity*: acid test, general liquidity, short term debt to total debt, short term interest-bearing debt to total debt, short term bank debt to total debt and net liquidity to liabilities.
- *Other*: firm age, size and growth; public sector or financial institutions stockholding, group membership, economic sector and macroeconomic environment.

As section 4 states, these variables went through a univariate analysis to determine, for each group, which of them are the best in predicting default. This is a very useful step which helps to reduce the number of variables to be taken into account in the multivariate analysis after having identified colinearity.

In order to be consistent with the firm assessment process operating in the Firms Assessment Unit, where the assessment for a given year uses information up to the previous year, the time horizon for the prediction will be one year. Then, default in year *t* will be predicted according to the available information in year *t-1*.

III.3. DATA

A frequent and important problem faced when trying to construct a rating model is the lack of the necessary data. It is essential to find information on firms' financial statements as well as on default, which usually is property of companies and/or credit institutions. For Spanish borrowers, both types of data can be obtained from BdE's internal databases: the Central Balance Sheet Office (CB) and the CIR, respectively.

III.3.1. Independent variables

Accounting information and some qualitative data were obtained from BdE's CB[18]. It provided information from:

- its annual database (CBA), where firms fill in a comprehensive questionnaire on a voluntary basis; and
- the Commercial Register (CBB), where all firms must deposit their financial statements every year[19].

[18] It must be noted that all information used in this study stems from individual accounts.
[19] Both databases are complementary. This is, if an observation is included in CBA database, then it is not part of CBB database.

The information available is much more comprehensive for those firms in CBA, that provide a wider split-up for some entries of financial statements, as well as qualitative information. This allows computing more ratios for CBA observations than for CBB ones. Such limitation mainly affects debt composition, which is a potentially relevant issue to credit quality assessment. However, the possibility of obviating CBB information due to its “simplicity” must be ruled out, as this database provides the vast majority of observations. CBB information is essential to the attainment of a representative sample of the universe of Spanish firms, since CBA database is biased towards big enterprises (as we will see below, the average sizes of both databases are quite different according to all size criteria considered).

Therefore, two possible solutions were faced:

- To estimate a single model for the whole sample considering just those variables which can be computed for both databases; which in practical terms means those variables that can be computed for CBB.
- To estimate two different models, one for each database. Then, more variables would be considered in the construction of CBA model than in the CBB one.

The first option would imply rejecting most ratios that capture debt composition, which *a priori* are expected to be very relevant to credit quality assessment. Consequently, the second option was chosen, so a differentiation between CBA model and CBB model will be shown.

The initial sample included all observations in both databases for period 1991-2004[20]. After a previous treatment to detect and refine errors, the result was 144,973 observations in CBA and 1,197,427 in CBB (see Table 1).

	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	TOTAL
CBA	7,223	7,213	7,797	8,960	9,404	9,842	10,087	9,963	10,085	10,582	11,141	12,948	14,327	15,401	144,973
CBB	673	1,472	9,283	16,472	42,940	61,065	81,470	91,394	83,029	97,675	118,157	162,660	195,471	235,666	1,197,427

Note: not all observations have data for every ratio.

Table 1: Number of observations

[20] In the initial data upload, a filter in terms of size was applied in order to eliminate firms without activity. To be precise, firms simultaneously fulfilling the following two conditions were dropped: 1) number of employees = 0, and 2) turnover lower than 100,000 euros. The aim was identifying and eliminating those firms without real activity which are registered for “other reasons” (no employees and low turnover); while keeping head offices (no employees but significant turnover) and small enterprises (low turnover and number of employees different from zero).

Table 2 shows the sectoral distribution of observations for both databases according to the CB-26 classification criterion followed by the CB, which considers twenty six different categories or subsectors. As demonstrated, all groups have a significant number of observations. However, this was not the sectoral classification used. Since the economic sector was going to be included in the model through dummy variables, a narrower aggregation resulting in fewer categories was desirable. The grouping criterion, which is also depicted in table 2, considers five economics sectors (and, consequently, five dummy variables): energy, industry, market services, construction and others.

SECTOR	CB-26 SUBSECTOR	CBA		CBB	
1. ENERGY	1. Energy products extraction	315	0,1%	513	0,0%
	2. Mineral extraction, except energy products	765	0,3%	3.913	0,2%
	4. Oil refining and nuclear fuel treatment	129	0,0%	74	0,0%
	17. Electric energy, gas and water production and distribution	1.421	0,5%	1.675	0,1%
	18. Water collecting, treatment and distribution	875	0,3%	1.062	0,0%
	Total energy	**3.505**	**1,3%**	**7.237**	**0,3%**
2. INDUSTRY	3. Food, drink and tobacco	9.430	3,4%	26.800	1,2%
	5. Chemical industry	5.065	1,8%	6.541	0,3%
	6. Other non-metal mineral products	4.242	1,5%	13.208	0,6%
	7. Metallurgy and metal products manufacture	6.375	2,3%	43.809	2,0%
	8. Machinery and mechanical equipment construction	3.864	1,4%	13.653	0,6%
	9. Electric, electronic and optical equipment and materials	3.270	1,2%	9.447	0,4%
	10. Transport material manufacture	2.464	0,9%	3.385	0,2%
	11. Textile and clothing industry	5.023	1,8%	24.307	1,1%
	12. Leather and shoe industry	1.213	0,4%	10.172	0,5%
	13. Wood and cork industry	1.692	0,6%	16.429	0,7%
	14. Paper, edition and graphic arts	4.739	1,7%	26.318	1,2%
	15. Rubber transformation and plastic materials	2.697	1,0%	9.354	0,4%
	16. Various manufacturing industries	2.908	1,0%	23.062	1,0%
	Total industry	**52.982**	**19,1%**	**226.485**	**10,3%**
3. MARKET SERVICES	20. Trade; automobile, motorcycle, moped and personal goods for household use repairs	35.817	12,9%	358.005	16,2%
	21. Transport, storage and communications	7.371	2,7%	58.404	2,6%
	24. Hotel and catering industry	4.475	1,6%	60.839	2,8%
	25. Real estate and renting. Business services	22.104	7,9%	198.916	9,0%
	Total market services	**69.767**	**25,1%**	**676.164**	**30,6%**
4. CONSTRUCTION	**19. Construction**	**11.222**	4,0%	**157.992**	7,2%
5. OTHERS	26. Other services in CB	4.103	1,5%	68.745	3,1%
	22. Agriculture, ranching, hunting and forestry	2.198	0,8%	28.750	1,3%
	23. Fishing	550	0,2%	1.879	0,1%
	Total others	**6.851**	**2,5%**	**99.374**	**4,5%**
MISSING VALUES	**Missing values**	**646**	**0,2%**	**30.175**	**1,4%**
	TOTAL	**278.078**	**100,0%**	**2.206.687**	**100,0%**

Table 2: Sectorial distribution (number of observations)

Table 3 shows the descriptive statistics for the different variables and databases. In general, firms in both databases are quite different. Among the main characteristics CBA firms show, in general, better financial autonomy, leverage, repayment capacity and return ratios, apart from higher size and growth rate; while CBB firms have a lower percentage of short term debt on total debt.

It has to be mentioned that, during the previous treatment of data, outlying values were corrected to prevent biases in the analysis. Outliers were defined here as those values which were more than two standard deviations away -above or below- from the population average. Once identified, there are different alternative procedures to treat them: (i) dropping those observations showing outlying values for some variables, (ii) treating outliers as missing values; or (iii) replacing outliers with “extreme values”, this is, with ±2 standard deviations from the mean. In this study, the last option was considered to be the best, as it allows keeping the element “very high/very low value” for the affected variable, which is a very useful input to the model; without losing the rest of the information for that observation.

	OBSERVATIONS		**AVERAGE**		**STANDARD DEVIATION**		**MINIMUM**		**MAXIMUM**	
	CBA	**CBB**	**CBA**	**CBB**	**CBA**	**CBB**	**CBA**	**CBB**	**CBA**	**CBB**
Financial autonomy	114,239	1,181,931	36.89	21.01	205.58	230.33	-54,500.00	-111,299.60	100.61	50,700.05
Adjusted financial autonomy	114,239	1,181,931	37.43	21.01	202.52	230.33	-54,500.00	-111,299.60	4,664.77	50,700.05
Total leverage	114,239	832,531	545.11	638.71	8,709.45	10,292.04	-430,350.00	-1,683,222.00	1,615,608.00	1,652,362.00
Financial leverage	114,239	-	257.39	-	4,878.93	-	-270,400.00	-	685,457.60	-
Financial expenses coverage–EBIT	99,811	955,520	23.09	10.82	850.26	125.61	-5,017.00	-14,157.00	230,953.00	18,421.00
Financial exp. coverage–EBITDA	99,817	1,029,961	25.49	16.81	342.57	140.15	-5,057.00	-11,245.00	45,248.00	21,936.00
Repayment capacity	99,811	540,466	12.79	-1,122.76	404.52	7,465,449.00	-37,741.00	$-1.29\cdot10^{9}$	73,246.00	$2.22\cdot10^{9}$
Economic return	99,811	990,269	7.58	4.68	16.30	26.98	-1,657.14	-5,046.67	1,136.31	5,011.43
Financial return	99,811	720,684	20.56	$-3.08\cdot10^{7}$	642.04	$2.59\cdot10^{10}$	-99,805.98	$-2.20\cdot10^{13}$	74,121.08	$5.94\cdot10^{10}$
Operating margin	99,814	990,293	-0.24	-0.44	984.19	840.18	-139,700.00	-230,800.00	215,100.00	469,300.00
Resource generation capacity	99,817	1,073,503	0.54	0.03	69.03	1.99	-2,172.33	-614.00	18,778.05	754.00
Acid Test	114,239	879,601	2.16	1.22	48.57	22.06	-73.13	-6,864.50	14,351.13	11,457.00
General liquidity	114,239	1,196,106	2.82	2.49	49.31	40.75	-80.64	-6,874.00	14,351.13	16,231.00
Short term debt / total debt	114,341	842,746	80.83	65.80	24.58	47.39	-241.36	-6,816.68	141.67	15,349.44
S.t. interest-bearing debt/total debt	114,239	-	20.19	-	22.27	-	-22.48	-	950.00	-
Short term bank debt / total debt	114,322	-	17.25	-	20.59	-	-18.41	-	800.00	-
Net liquidity / liabilities	114,239	-	-4.65	-	162.31	-	-53,600.00	-	325.28	-
Age	106,507	-	22.12	-	20.99	-	0.00	-	98.21	-
Size (turnover)	144,973	1,197,427	21,931.25	724.48	63,533.11	3,690.67	0.00	0.10	501,761.80	2,595,844.00
Size (total assets)	114,239	1,197,427	33,818.15	565.91	141,735.90	918.11	0.00	-0.10	1,328,777.00	6,510.40
Size (number of employees)	144,906	1,166,703	123.88	7.27	328.93	7.77	0.00	0.10	2,673.48	34.32
Growth (turnover)	144,960	1,180,377	100,828.20	28.89	2,230,055.00	163.83	-100.00	-99.99	$7.13\cdot10^{7}$	2,552.08
Growth (total assets)	114,239	1,197,351	120,203.60	19.95	3,022,850.00	66.74	-100.00	-99.91	$1.06\cdot10^{8}$	575.23
Public sector stockholding	114,239	-	0.04	-	0.19	-	0.00	-	1.00	-
Financial institutions stockholding	114,239	-	0.02	-	0.15	-	0.00	-	1.00	-
Group membership	144,973	1,197,427	0.85	0.79	0.36	0.40	0.00	0.00	1.00	1.00
Energy sector	144,327	1,167,252	0.02	0.01	0.15	0.08	0.00	0.00	1.00	1.00
Industry sector	144,327	1,167,252	0.37	0.19	0.48	0.40	0.00	0.00	1.00	1.00
Market services sector	144,327	1,167,252	0.48	0.58	0.50	0.49	0.00	0.00	1.00	1.00
Construction sector	144,327	1,167,252	0.08	0.14	0.27	0.34	0.00	0.00	1.00	1.00
Other sectors	144,327	1,167,252	0.05	0.09	0.21	0.28	0.00	0.00	1.00	1.00
GDP growth	144,973	1,197,427	3.02	3.41	1.38	0.90	-1.03	-1.03	5.06	5.06
GDP growth t+1	144,973	1,197,427	3.11	3.50	1.35	0.76	-1.03	-1.03	5.06	5.06

Table 3: Descriptive statistics

III.3.2. Dependent variable

BdE's CIR collects data on lending by financial institutions on a monthly basis according to the following criteria: direct loans to resident borrowers will have to be reported to the CIR if the institution's overall business in Spain is 6,000 euros or more or if business in any other country is 60,000 or more. Direct loans to non-residents are reported from 300,000 euros.

The definition of default corresponds to the criteria used by the BdE's Firm Assessment Unit, which coincides with the CIR and Basel II definitions. The borrower´s payments must be past due more than 90 days on any credit obligation[21] (interest or principal). In addition to this, two conditions have to be met: (i) the unpaid amount must be higher than 30.000 euros and (ii) the percentage of unpaid debt must be higher than 1% of the total debt. The objective was to eliminate so-called "technical defaults": defaults that occur because of different reasons not related to the borrower's credit quality, which usually match low amount defaults[22].

The distribution of the default variable generated according to these conditions in the training sample is shown in table 4. It can be observed that the default rate, which has progressively declined during the sample period for both databases, is always lower for CBB firms. The reasons for this are not very clear *a priori*; since, from the descriptive statistics presented above, no straightforward conclusion can be drawn regarding the credit quality of firms in both databases. One potential reason could be the differences in firm size among samples. In principle, the expected sign of the relationship between default and firm size should be negative (the higher the size, the lower the probability of default), as big firms are supposed to have better access to funds in financial markets. This fact has been confirmed by other works. However, informational imperfections can make such sign get positive for a certain sample, since information about defaults of small enterprises is likely to be less available than for big ones. According to this explanation, the larger size of CBA firms could justify their higher percentages of default[23].

[21] In this respect, we considered commercial and financial credit and euro-denominated debt instruments.
[22] It must be mentioned here that different definitions of default were analysed considering various limits, but the results of the models in terms of accuracy, number of eligible firms and significant variables were quite similar.
[23] This issue will come up again later on when the results of the estimations are presented.

In addition to this, another possible reason for the positive sign has to do with the default definition used here, which is far from legal proceedings or bankruptcy. In this sense, banks permissiveness regarding big firms' 90-day (or similar short horizons) defaults might make such behaviour a common practice; while small firms would not be allowed by banks to do so.

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
CBA	4.40	5.07	4.55	3.09	3.01	1.88	1.34	1.14	1.19	0.75	0.88	0.63	0.49	0.38
CBB	2.97	2.45	2.35	1.81	1.74	1.29	0.93	0.75	0.60	0.63	0.64	0.45	0.49	0.24

Table 4: Default rate (%)

As mentioned in the introduction, the purpose of this article is to illustrate the general methodology that BdE is working on. The steps followed in the estimation were identical for CBA and CBB models, so, for brevity reasons, once the available data have been described, all presented results and comments from now on will only refer to the CBA model.

III.4. PRESELECTION OF FACTORS: UNIVARIATE ANALYSIS

A previous individual analysis of factors is a key issue to help deciding which of them to introduce in the multivariate model. This analysis consists of checking some aspects of variables´ statistical relationship with default, such as the sign of that relationship, its monotonicity or the variable's predictive power of default. In addition to this, the correlations among variables must also be taken into account when deciding the composition of the multivariate model.

III.4.1. Type of relationship between factors and default

To analyze the type of relationship between factors and default, a graph was constructed for each variable showing in the x-axis that variable's percentiles and in the y-axis the observed default frequency for each percentile. Such type of graphs allows coming to preliminary conclusions not only about the sign of the relationship, but also about its type (linear, quadratic, monotonic, etc).

The first step was to check whether the sign of the statistical relationship between default and every single factor was the expected one. Graphs 1 and 2 show, for instance, the cases of total leverage and financial return, respectively. While default seems to increase as total leverage rises, it falls as financial return increases; which are in both cases reasonable relationships. As for the type of relationship, the one between default and total leverage appears to be quite linear, while there seems to be some non-linear component in the relationship with financial return.

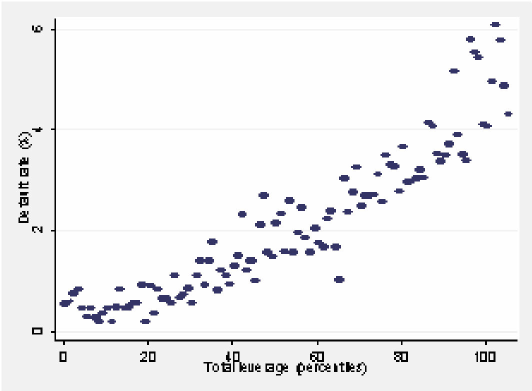


Exhibit 1: Total leverage vs. Default

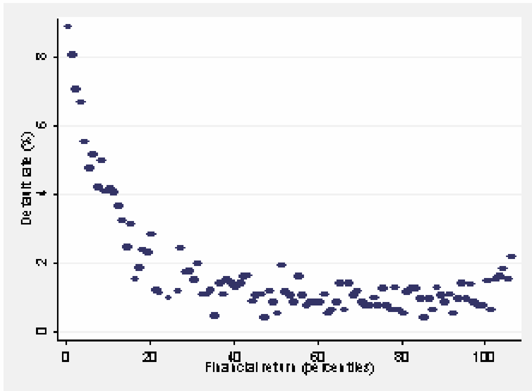


Exhibit 2: Financial return vs. default

The second step was to study the monotonic character of these relationships, which is a desirable property for the good functioning of the final model. The aforementioned graphs confirm that the observed relationships are, in general, monotonic.

Although for most variables the observed relationships are as expected, like in the examples presented above, the case of size variables is an exception. For every definition of size considered the observed sign is slightly positive, while the expected sign is negative. Exhibit 3 and 4 show the cases of total assets and number of employees. Two possible causes for this can be, as mentioned in the previous section, informational imperfections affecting small firm defaults or the specific default definition used.

Anyway, these univariate relationships can be hiding interactions with other variables. Therefore, to come to real conclusions about the sign and type of relationship with default, it is necessary to wait for the results of the multivariate estimation, where the individual effects of every variable are controlled and the meaning of coefficients is easier to interpret.

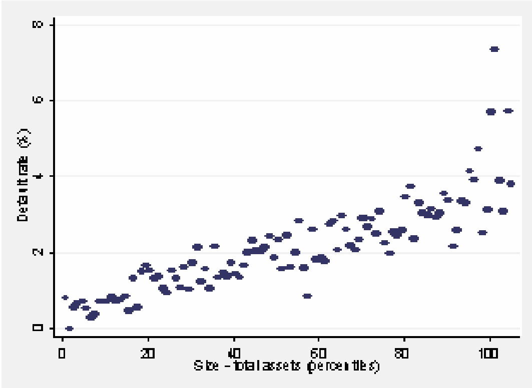


Exhibit 3: Total assets vs. default

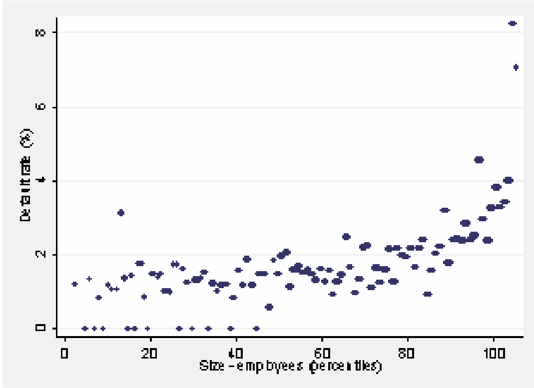


Exhibit 4: Number of employees vs. default

III.4.2. Predictive power of individual factors

Once the type, sign and monotonicity of each variable's relationship with default has been studied, it must be determined which factors have a higher explanatory power. For this purpose we use the ROC curve (Receiver Operating Characteristic), which measures the ability of a variable or model (combination of variables) to correctly classify the dependent variable for a certain sample. Such curve is described in detail in Box 1.

Box 1: ROC curves

The theory behind ROC analysis comes from statistical decision theory and was originally used during World War II to solve some practical problems with radar signals. More recently it has spread into other fields, such as the evaluation of the ability of discrete-choice econometric models to correctly classify individuals.

The predictive power of a discrete-choice model such as the *logit* is measured through its sensibility and its specificity. The **sensibility** (or fraction of true positives -FTP-) is the probability of correctly classifying and individual whose observed situation is "default", while the **specificity** is the probability of correctly classifying an individual whose observed situation is "no default". Given a model to predict default as a function of a series of individuals' characteristics, if the sample data are presented in a contingency table showing the model result (probability of default generated by the model higher or lower than a given threshold) and the observed situation ("default" or "not default"), then estimating the sensibility and specificity of the model is straightforward.

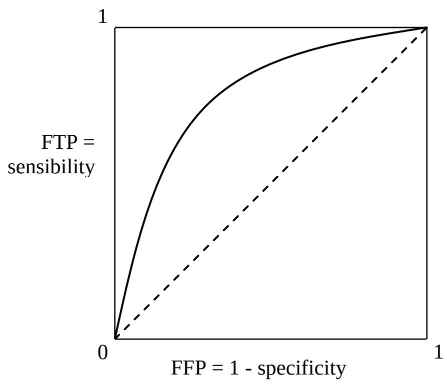
		Observed situation	
		Default	**No default**
Model prediction	**Default** (probability of default higher than threshold)	True positive (TP)	False positive (FP)
	No default (probability of default lower than threshold)	False negative (FN)	True negative (TN)
		TP+FN	TN+FP

Sensibility and specificity can be easily estimated for this model -for a given threshold- as:

Sensibility = TP/(TP+FN) = FTP (fraction of true positives) = 1- FFN (fraction of false negatives)

Specificity = TN/(TN+FP) = FTN (fraction of true negatives) = 1-FFP (fraction of false positives)

The ROC curve is the graphic representation of the relationship between the sensibility and the specificity of a model for all possible thresholds.



Through this representation of the pairs (1-specificity, sensibility) obtained for each potential value of the threshold, the ROC curve gives us a global representation of diagnostic accuracy. ROC curve has a positive slope, reflecting the existing *trade-off* between sensibility and specificity (the only way to achieve a higher sensibility is by reducing specificity). Should discrimination be perfect (100% sensibility and 100% specificity), the ROC curve would go through the upper-left corner. But if the model is not able to discriminate among groups, the ROC curve becomes the diagonal connecting the lower-left and the upper-right corners. This is, predictive accuracy increases as the curve shifts from the diagonal towards that corner, and the area below the ROC curve can be used as an index for the model's global accuracy. Maximum accuracy would then correspond to an area equal to 1 and random prediction to an area equal to 0.5, while areas below 0.5 would indicate negative prediction; this is, there is a relationship between predicted values and truth, but it is contrary to the expected one. Then, ROC curves are useful for assessing a model's global performance (area below the curve), as well as for comparing models (comparison amongst curves) or probability thresholds (comparison amongst different points on a curve).

For every continuous variable considered, a univariate logit model was estimated to predict default, from which the corresponding ROC curves were derived. The results are presented in table 5. This methodology, however, is not applicable to discrete variables. The evaluation of such factors' predictive power was done by analyzing the respective graphs showing average default rate for each possible value of the dummy variable. In this sense, significant differences among default rates for different values of the dummy variable would suggest that such a factor is potentially relevant to the prediction of default.

SOLVENCY		PROFITABILITY		LIQUIDITY		OTHERS	
Financial autonomy	0.7021	Economic return	0.6295	Acid test	0.6339	Age	0.5559
Adjusted financial autonomy	0.6999	Financial return	0.6685	General liquidity	0.6243	Size (turnover)	0.5515
Total leverage	0.6913	Operating margin	0.5660	Short term debt/total debt	0.6224	Size (total assets)	0.6433
Financial leverage	0.7356	Resource gen. capacity	0.4576	S.t. interest-bearing debt/total debt	0.6296	Size (employees)	0.6058
Fin. exp. coverage-EBIT	0.7093			Short term bank debt / total debt	0.6412	Growth (turnover)	0.4482
Fin. exp. coverage-EBITDA	0.7037			Net liquidity/liabilities	0.6701	Growth (total assets)	0.5491
Repayment capacity	0.3494					GDP growth t	0.6684
						GDP growth t+1	0.6594

Table 5: Area below ROC curve

According to the results in table 5, the most powerful factors in terms of default prediction seem to be those having to do with solvency. In fact, some of them even reach ROC area values above 0.7. Financial leverage, for instance, appears as a very good individual predictor of default, with a ROC area of 0.74. As for profitability ratios, they show in general a lower predictive ability. Among them, financial and economic returns stand out from the others. As regards liquidity variables, net liquidity on liabilities is the best performing one, followed by short term bank debt on total debt and the acid test. Also GDP growth seems to play a significant role in default prediction, as well of size in terms of total assets. Finally, the analysis of the graphs constructed for discrete variables revealed that there seem to be significant differences in default rates due to financial institutions stockholding, group membership and economic sector.

However, it must be pointed out that these preliminary results just suggest which variables have the best predictive power when used individually. The best multivariate model will not necessarily include these ones, as each ratio's predictive capacity also depends on its interaction with other factors in the model.

In addition to this, it is not only the individual predictive power that has to be taken into account when deciding which factors to include in the multivariate model, but also other aspects such as: what is the ratio really measuring, whether the sign of its estimated coefficient is sensible, etc. Then, a ratio that *a priori* does not seem to be very useful can be included in the final model or *vice versa*.

III.5. MODEL SELECTION: MULTIVARIATE ANALYSIS

Starting from the univariate analysis and after checking the colinearity relationships among variables, numerous models including different groups of variables were tested using forward stepwise selection. This process meant not only deciding which particular variables to choose from the set previously defined, but also trying different model designs. In this sense, some examples of the main possibilities tested are: linear *vs* non-linear models, including previous default or not, models including the economic sector through dummy variables *vs* different models by sector; and models including size variables *vs* different models by size.

Among the main conclusions from the numerous tentative estimations we must underline that non-linear logits, which also include the quadratic terms of variables (both squares and cross products), get significantly better results in terms of predictive power than linear ones. This fact is consistent with previous literature, as well as with international rating agencies´ experience, and confirms the preliminary results obtained from the univariate analysis, which indicated some type of non-linear behaviour in certain variables. In addition to this, the models perform even better when previous default[24] is included as an additional factor; and macroeconomic environment was also found to be very significant in default prediction. In this respect, it must be noted that various variables capturing macroeconomics were tested (interest rates, stock indexes, consumer price indexes, gross domestic product (GDP), etc), but the best performing one was GDP growth.

[24] Previous default was defined as a dummy variable taking value 1 if the firm had defaulted the previous year and zero otherwise.

As for the role of the economic sector, only the construction sector was found to perform better with its own specific model than with a general one for all sectors. This is a natural result, since construction firms generally present financial structures which are significantly different from firms in other economic sectors.

The best performing models for CBA database reach ROC areas around 0.92, which is a good figure according to previous literature. Table 6 presents, as an example, the estimation results[25] of a general model for all sectors; which therefore includes the economic sector through dummy variables.

Number of obs	90792					
Log likelihood	-4775.1431					
Default	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
Previous default	4.948075	.0664364	74.48	0.000	4.817862	5.078288
Financial autonomy (F1)	-.0080397	.0009478	-8.48	0.000	-.0098973	-.006182
Total leverage (F2)	7.12e-06	2.50e-06	2.85	0.004	2.23e-06	.000012
Economic return (F3)	-.0220226	.0028432	-7.75	0.000	-.0275951	-.0164501
S.t. bank debt/total debt (F4)	.0359708	.0053597	6.71	0.000	.0254659	.0464757
Net liquidity/liabilities (F5)	-.0049072	.0019983	-2.46	0.014	-.0088236	-.0009903
Size (total assets) (F6)	3.90e-06	5.46e-07	7.14	0.000	2.83e-06	4.97e-06
Energy sector	-4.17434	.1846912	-22.60	0.000	-4.536328	-3.812352
Industry sector	-4.37221	.0939173	-46.55	0.000	-4.556285	-4.188136
Market services sector	-4.62329	.0879246	-52.58	0.000	-4.795619	-4.450961
Construction sector	-4.340172	.1278889	-33.94	0.000	-4.590829	-4.089514
Other sectors	-4.250317	.1502603	-28.29	0.000	-4.544822	-3.955812
GDP growth t	-.1703206	.0176309	-9.66	0.000	-.2048766	-.1357646
Squares						
$F1^2$	-3.24e-07	4.59e-08	-7.06	0.000	-4.14e-07	-2.34e-07
$F2^2$	-7.96e-12	2.93e-12	-2.71	0.007	-1.37e-11	-2.21e-12
$F4^2$	-.0003472	.0000805	-4.31	0.000	-.000505	-.0001894
$F5^2$	-.0000527	7.68e-06	-6.86	0.000	-.0000677	-.0000376
$F6^2$	-2.92e-12	4.89e-13	-5.97	0.000	-3.88e-12	-1.96e-12
F1·F2	.0000272	6.16e-06	4.42	0.000	.0000152	.0000393
F1·F3	-.0001676	.000027	-6.21	0.000	-.0002206	-.0001147
F1·F4	-.0002911	.0000469	-6.20	0.000	-.0003831	-.0001991
F2·F3	8.73e-08	4.69e-08	1.86	0.063	-4.72e-09	1.79e-07
F5·F6	4.01e-08	8.19e-09	4.89	0.000	2.40e-08	5.61e-08
Area under ROC curve	0.9036					

Table 6: General model for CBA

In this case, the model predicts probability of default based on: previous default, five financial ratios (financial autonomy, total leverage, economic return, short term bank debt on total debt and net liquidity on liabilities), size, the five sectoral dummies, GDP growth and the squares and crossed products of some of these variables[26]. All signs except the one of size are as expected. Then, even when controlling for other variables' effects in a multivariate model, size continues showing a "wrong" sign.

[25] This estimation used period 1991-2003 so as to keep 2004 data for validation.

[26] Given the good predictive results achieved by the models, non-linear behaviours seem to be well captured by quadratic terms, so cubic terms were not tested.

In order to try to solve the potential informational imperfections commented above, which could be the reason for this, other definitions of default for different time horizons were considered, but the result was the same. Besides, dropping the size variable from the model worsened results in terms of predictive accuracy, so the decision was to keep it. Anyway, the value of its estimated coefficient -and therefore its impact on predicted probability of default- is very low.

This design results in a ROC area of 0.90, which means nothing but a global measure of the model's accuracy. As explained in the box about ROC curves, the percentage of errors in terms of "false positives" or "false negatives" will depend on the chosen threshold. After ordering the observations by probability of default, which is the model's output, a decision has to be made which specific value of probability of default to set as a benchmark. This will determine the number of eligible (estimated probability of default lower than the threshold) and non-eligible (estimated probability of default higher than the threshold) firms and, consequently, the number of classification errors; so it represents a crucial decision. If the threshold is very low, most observed defaults will be correctly classified (high sensibility), but a lot of false positives will arise; e.g., many firms which did not default will be classified as "non eligible". On the other hand, if we allow for a higher threshold, more firms will be considered eligible and false positives will be reduced, but false negatives will increase. For the purpose of collateral assessment, the aim is to minimize the number of false negatives subject to getting a significant amount of collateral, this is, a significant number of eligible enterprises. In this respect, the percentage of false negatives will serve as input to the internal ECAF performance monitoring process, which comprises both an annual rule and a multi-period assessment.

See for example the case of the model in table 6. Table 7 shows its predictive results in 2004 for four different thresholds. For a benchmark of 0.10%, which is the Eurosystem's credit quality threshold[27], this model classifies 124 firms as eligible. None of them defaulted in the subsequent twelve months. The same happens for thresholds 0.20% and 0.30%: the percentage of observed defaults detected by the model is 100%. For a threshold probability equal to 0.40%, however, this model would correctly classify 98% of observed defaults (56 firms) and would be wrong in the remaining 2% of the cases (1 firm).

[27] The Eurosystem's credit quality threshold is defined in terms of a "single A" credit assessment (which means a minimum long-term rating of A- by Fitch or Standard & Poor's, or A3 by Moody's), which is considered equivalent to a probability of default over a one-year horizon of 0.10%.

On the other hand, for a 0.40% threshold the model would correctly classify 27% of firms which did not default (2,448 firms), and would fail in 73% of them (6,514 firms). In other words, for that given threshold the model would classify 2,449 firms as eligible in 2004 and only one of them defaulted during the subsequent year; which means that 0.04% of the eligible database defaulted. This is the relevant error for the annual back-testing rule of the ECAF performance monitoring process, and this figure fulfils the conditions to pass such test[28]. In fact, for thresholds 0.10%, 0.20% and 0.30%, as table 7 shows, the error in terms of false negatives and, consequently, of defaults within the eligible database, would be zero. Then, for 2004 this system would comply with ECAF rules for all thresholds considered here.

p*=0.10%		**OBSERVED Default**	**OBSERVED Not default**	**TOTAL**
PREDICTED	**Default (not eligible)**	57	8,838	8,895
	Not default (eligible)	0	124	124
	TOTAL	57	8,962	9,019

Percentage of defaults among eligible firms: 0.00%

p*=0.20%		**OBSERVED Default**	**OBSERVED Not default**	**TOTAL**
PREDICTED	**Default (not eligible)**	57	8,349	8,406
	Not default (eligible)	0	613	613
	TOTAL	57	8,962	9,019

Percentage of defaults among eligible firms: 0.00%

p*=0.30%		**OBSERVED Default**	**OBSERVED Not default**	**TOTAL**
PREDICTED	**Default (not eligible)**	57	7,518	7,575
	Not default (eligible)	0	1,444	1,444
	TOTAL	57	8,962	9,019

Percentage of defaults among eligible firms: 0.00%

p*=0.40%		**OBSERVED Default**	**OBSERVED Not default**	**TOTAL**
PREDICTED	**Default (not eligible)**	56	6,514	6,570
	Not default (eligible)	1	2,448	2,449
	TOTAL	57	8,962	9,019

Percentage of defaults among eligible firms: 0.04%

Table 7: Predictive results for general model in 2004 (number of firms)

If we replicated this exercise for 1991-2004 predictions -where 99,811 observations were assigned a predicted probability of default-, we would get that the percentage of defaults within the eligible database would be 0% for a 0.10% threshold (1,478 eligible observations). For higher thresholds, the number of eligible observations would obviously increase (8,071 eligible observations for a 0.20% benchmark, 19,058 for a 0.30% and 31,720 for a 0.40%) and the system would still be acceptable in terms of the ECAF performance monitoring process.

[28] These conditions are part of the ECAF internal procedures.

In order to get an idea of how this model discriminates between eligible and not eligible firms, table 8 compares their general characteristics by showing the average ratios of eligible (8,071) and not eligible observations (91,740) for the whole sample and a threshold equal to 0.10%.

Variable	Average eligible	Average not-eligible
Financial autonomy	69.53	24.73
Adjusted financial autonomy	70.24	25.00
Total leverage	153.11	1,198.69
Financial leverage	28.76	593.15
Financial expenses coverage - EBIT	305.95	35.18
Financial expenses coverage - EBITDA	254.17	46.37
Repayment capacity	0.62	16.40
Economic return	40.37	6.00
Financial return	69.30	51.18
Operating margin	20.65	-5.41
Resource generation capacity	0.07	0.45
Acid Test	12.63	5.01
General liquidity	13.27	5.90
Short term debt / total debt	94.34	75.15
Short term interest-bearing debt / total debt	12.69	17.89
Short term bank debt / total debt	9.87	15.27
Net liquidity / liabilities	38.67	-8.45
Age	18.77	21.58
Size (turnover)	46,061.04	29,508.49
Size (total assets)	127,817.50	46,889.78
Size (employees)	154.19	153.42
Growth (turnover)	199.40	598.59
Growth (total assets)	15.63	198.48
Public sector stockholding	0.03	0.04
Financial institutions stockholding	0.06	0.02
Group membership	0.91	0.89

Table 8: Comparison of average ratios: General model in 1991-2004

As explained above, many different model designs were tested, among which sectoral models played an important role. It has already been mentioned that only the construction sector performed slightly better when using its own sectoral model than applying a general model for all sectors. To briefly illustrate an additional example of estimations, a few results of a sectoral construction model are commented below.

The composition and results of the specific model for the construction sector are very similar to the mentioned general one, although with some variations[29]. In particular, the variables on which the prediction is based for the construction sector are: previous default, four financial ratios (financial autonomy, total leverage, financial return and short term bank debt on total debt), size, GDP growth and the squares of some of these variables. The results in terms of signs of the variables and of predictive accuracy are almost the same as for the general model, reaching a ROC area of 0.90 and almost equal figures for the percentage of defaults within the eligible database for the different thresholds (the results in terms of the ECAF monitoring process are almost identical).

The main difference to be pointed out was the higher number of eligible enterprises, most of them coming from the construction sector. This is a logical result, since construction firms generally present high leverage ratios because of the type of activity they develop; so its functioning is better modelled by their own specific model than by the general one, which penalizes them by comparison with the other four sectors. For example, while for a threshold of 0.10% the general model classified 12 4 firms as eligible in 2004, this model produced 136 eligible enterprises. The figures for the whole sample 1991-2004 (99,811 observations) are, for the same threshold, 1,478 eligible observations for the general model and 1,576 for the sectoral specification.

III.6. SUMMARY AND CONCLUSION

BdE is working on a new methodology based on logit models to be incorporated in the near future to its in-house credit assessment system. The purpose is twofold: (i) enlarging the methodological spectrum of BdE, presently consisting of an expert-based assessment; and (ii) producing estimated probabilities of default.

The relevant information needed for the estimation of such models -i.e. dependent and independent variables- is available from internal BdE databases. A wide range of financial ratios and other factors has been considered as potential determinant of default. From the preliminary analysis it can be said that, in general, the relationship between default and these variables behaves as expected by theoretical economic reasoning.

[29] In this case, the specific sectoral model as applied to the construction firms and the general model presented before was applied to the other four sectors after a re-estimation without the observations belonging to the construction sector.

In particular, most signs fit in the expected ones and the relationships seem to be monotonic. Only the case of size variables represents an exception, as the sign appears to be slightly positive whereas the expected one is negative. Two possible explanations for this have been put forward. The first one is based on the effect of informational imperfections affecting small firm defaults; which can result in higher default rates for big firms than for small firms in a given sample. The second one has to do with the default definition used and banks permissiveness to big firms defaults within a certain time period.

The univariate analysis pointed at solvency ratios as the most powerful factors for default prediction. Among them, financial leverage showed the highest value for the ROC area, reaching 0.74. After this, multivariate analysis was carried out testing a huge number of model designs. Although there is still much work to be done, several general conclusions can be drawn from it. Firstly, non-linear logits get significantly better results in terms of predictive power than linear ones. Secondly, the models perform even better when previous default is included as an additional factor. Thirdly, macroeconomic environment was also found to be very significant in default prediction, with GDP growth as the best performing factor among those capturing macroeconomics. As for the role of the economic sector, only the construction sector was found to perform better with its own specific model than with a general one for all sectors. An important point in this respect is that, although the composition of the construction sectoral model is quite similar to the general one, the sector-specific model produces a higher number of eligible enterprises without incurring in higher errors.

Regarding particular results, some models reach ROC areas around 0.92, which represents a good figure according to previous literature. In addition to this, the best performing models would allow BdE to significantly increase the eligible database while continuing easily fulfilling the ECAF performance monitoring process. The models presented here, for instance, would have been in the green zone during the analyzed period if the threshold probability had been set equal to 0.10%, the Eurosystem’s credit quality benchmark; and even for thresholds of 0.20% or 0.30%. In this case, these systems are reliable and safe tools for collateral selection.

Financial autonomy (%)	*equity× 100 / liabilities*
Adjusted financial autonomy (%)	*equity× 100 / (liabilities - reserves for risks and expenses)*
Total leverage (%)	*loan funds× 100 / equity*
Financial leverage (%)	*(long term loan funds + short term cost-bearing resources)× 100 /*
Financial expenses coverage - EBIT	*(net operating result + financial income) / financial expenses*
Financial expenses coverage - EBITDA	*gross operating result / financial expenses*
Repayment capacity (years)	*total debt / ordinary generated resources*
Economic return (%)	*(net operating result + financial income + positive translation differences)× 100 / total assets average*
Financial return (%)	*ordinary result× (1-t)× 100 / equity average*
Operating margin (%)	*net operating result× 100 / net turnover*
Resource generation capacity (veces)	*generated resources / operating income*
Acid Test (times)	*current assets without stock / short term loan funds*
General liquidity (times)	*current assets / current liabilities*
Short term debt / total debt (%)	*short term debt× 100 / total debt*
Short term interest-bearing debt / total	*short term interest-bearing debt× 100 / total debt*
Short term bank debt / total debt (%)	*short term bank debt× 100 / total debt*
Net liquidity / liabilities (%)	*(working capital - NOF)× 100 / liabilities*
Age (years)	Firm age
Size	**Turnover (thousand euros)**
	Total assets (thousand euros)
	Number of employees (number)
Growth	**Turnover growth (%)**
	Balance-sheet growth (%)
Public sector stockholding	Dummy equal to 1 if public sector direct participation in capital > 0
Financial institutions stockholding	Dummy equal to 1 if financial institutions participation in capital >0
Group membership	Dummy equal to 1 if the firm is a member of a group
Sector	**Energy sector** (dummy equal to 1 if the firm belongs to the
	Industry sector (dummy equal to 1 if the firm belongs to the
	Market services sector (dummy equal to 1 if the firm belongs to
	Construction sector (dummy equal to 1 if the firm belongs to the
	Other sectors (dummy equal to 1 if the firm belongs to any other
Macroeconomic environment	**GDP growth year *t*** (%)
	GDP growth year *t+1*(%)

Table 9: ANNEX – Definition of exogenous variables

REFERENCES

C. ZAVGREN: Assessing the vulnerability to failure of American industrial firms: a logistic analysis: in Journal of Business Finance and Accounting, No.12 (1). 1985.

E. ALTMAN: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy: in The Journal of Finance, vol. XXIII, No. 4. 1968.

E. ALTMAN/ R.HADELMANN/ P. NARAYANAN: ZETA analysis: a new model to identify bankruptcy risk of corporations: in Journal of Banking and Finance, No. 1. 1977.

J. OHLSON: Financial ratios and the probabilistic prediction of bankruptcy: in Journal of Accounting Research, No. 18. 1980.

J.P. KRAHNEN/ M. WEBER: Generally accepted rating principles - a primer: in Journal of Banking and Finance, No. 25. 2001.

W.H. BEAVER: Financial ratios as predictors of failure: in Journal of Accounting Research, January. 1966.

IV. ADVANTAGES AND DISADVANTAGES OF SUPPORT VECTOR MACHINES (SVMS)

L. AURIA / R. MORO

DEUTSCHE BUNDESBANK / DEUTSCHES INSTITUT FÜR WIRTSCHAFTSFORSCHUNG

IV.1. INTRODUCTION

There is a plenty of statistical techniques, which aim at solving binary classification tasks such as the assessment of the credit standing of enterprises. The most popular techniques include traditional statistical methods like linear Discriminant Analysis (DA) and Logit or Probit Models and non-parametric statistical models like Neural Networks. SVMs are a new promising non-linear, non-parametric classification technique, which already showed good results in the medical diagnostics, optical character recognition, electric load forecasting and other fields. Applied to solvency analysis, the common objective of all these classification techniques is to develop a function, which can accurately separate the space of solvent and insolvent companies, by benchmarking their score values. The score reduces the information contained in the balance sheet of a company to a one-dimensional summary indicator, which is a function of some predictors, usually financial ratios. Another aim of solvency analysis is to match the different score values with the related probability of default (PD) within a certain period. This aspect is especially important in the Eurosystem, where credit scoring is performed with the target of classifying the eligibility of company credit liabilities as a collateral for central bank refinancing operations, since the concept of eligibility is related to a benchmark value in terms of the annual PD.

The selection of a classification technique for credit scoring is a challenging problem, because an appropriate choice given the available data can significantly help improving the accuracy in credit scoring practice. On the other hand, this decision should not be seen as an “either / or” choice, since different classification techniques can be integrated, thus enhancing the performance of a whole credit scoring system.

In the following paper SVMs are presented as a possible classification technique for credit scoring. After a review of the basics of SVMs and of their advantages and disadvantages on a theoretical basis, the empirical results of an SVM model for credit scoring are presented.

IV.2. BASICS OF SVMS

SVMs are a new technique suitable for binary classification tasks, which is related to and contains elements of non-parametric applied statistics, neural networks and machine learning. Like classical techniques, SVMs also classify a company as solvent or insolvent according to its score value, which is a function of selected financial ratios. But this function is neither linear nor parametric. The formal basics of SVMs will be subsequently briefly explained. The case of a linear SVM, where the score function is still linear and parametric, will first be introduced, in order to clarify the concept of margin maximisation in a simplified context. Afterwards the SVM will be made non-linear and non-parametric by introducing a kernel. As explained further, it is this characteristic that makes SVMs a useful tool for credit scoring, in the case the distributional assumptions about available input data can not be made or their relation to the PD is non-monotone.

IV.2.1. Margin Maximization

Assume, there is a new company *j*, which has to be classified as solvent or insolvent according to the SVM score. In the **case of a linear SVM** the score looks like a DA or Logit score, which is a linear combination of relevant financial ratios $x_j = (x_{j1}, x_{j2}, \ldots x_{jd})$, where x_j is a vector with *d* financial ratios and x_{jk} is the value of the financial ratio number *k* for company *j, k=1,…,d*. So z_j , the score of company *j*, can be expressed as:

$$z_j = w_1 x_{j1} + w_2 x_{j2} + \ldots + w_d x_{jd} + b \qquad (1)$$

In a compact form:

$$z_j = x_j^{\,T} w + b \qquad (1')$$

where:

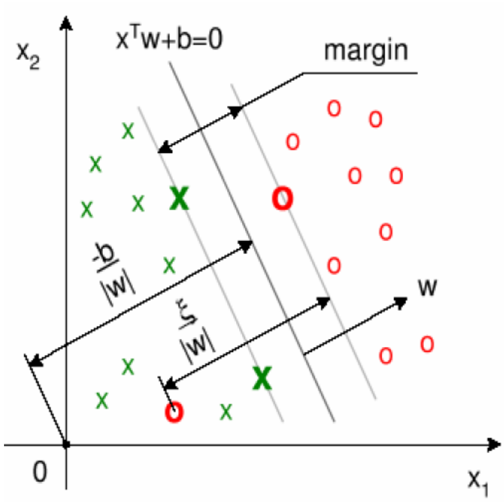
w is a vector which contains the weights of the *d* financial ratios and *b* is a constant. The comparison of the score with a benchmark value (which is equal to zero for a balanced sample) delivers the "forecast" of the class – solvent or insolvent – for company *j*.

In order to be able to use this decision rule for the classification of company *j*, the SVM has to learn the values of the score parameters *w* and *b* on a training sample. Assume this consists of a set of *n* companies $i = 1, 2, \ldots, n$. From a geometric point of view, calculating the values of the parameters *w* and *b* means looking for a hyperplane that best separates solvent from insolvent companies according to some criterion. The criterion used by SVMs is based on **margin maximization** between the two data classes of solvent and insolvent companies. The margin is the distance between the hyperplanes bounding each class, where in the hypothetical **perfectly separable** case no observation may lie. By maximising the margin, we search for the classification function that can most safely separate the classes of solvent and insolvent companies. Exhibit 1 represents a binary space with two input variables. Here crosses represent the solvent companies of the training sample and circles the insolvent ones. The threshold separating solvent and insolvent companies is the line in the middle between the two **margin boundaries**, which are canonically represented as $x^T w + b = 1$ and $x^T w + b = -1$. Then the margin is $2 / ||w||$, where $||w||$ is the norm of the vector *w*.

In a **non-perfectly separable** case the margin is "**soft**". This means that in-sample classification errors occur and also have to be minimized. Let ξ_i be a non-negative slack variable for in-sample misclassifications. In most cases $\xi_i = 0$, that means companies are being correctly classified. In the case of a positive ξ_i the company *i* of the training sample is being misclassified. A further criterion used by SVMs for calculating *w* and *b* is that all misclassifications of the training sample have to be minimized.

Let y_i be an indicator of the state of the company, where in the case of solvency $y_i = -1$ and in the case of insolvency $y_i = 1$. By imposing the constraint that **no observation may lie within the margin except some classification errors**, SVMs require that either $x_i^T w + b \geq 1 - \xi_i$ or $x_i^T w + b \leq -1 + \xi_i$, which can be summarized with:

$$y_i\left(x_i^T w + b\right) \geq 1 - \xi_i,\ \forall\ i = 1, \ldots, n. \tag{3}$$



Source: W. Härdle, R.A. Moro, D. Schäfer, March 2004, Rating Companies with Support Vector Machines, Discussion Paper Nr. 416, DIW Berlin.

Exhibit 1: Geometrical Representation of the SVM Margin

The optimization problem for the calculation of *w* and *b* can thus be expressed by:

$$\min_w \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \xi_i \tag{2}$$

$$s.t. \quad y_i\left(x_i^T w + b\right) \geq 1 - \xi_i\ , \tag{3}$$

$$\xi_i \geq 0 \tag{4}$$

In the first part of (2) we maximise the margin $2 / ||w||$ by minimizing $||w||^2 / 2$, where the square in the norm of *w* comes from the second term, which originally is the sum of in-sample misclassification errors $\xi_i\ / ||w||$ times the parameter *C*. Thus SVMs maximize the margin width while minimizing errors. This problem is quadratic i.e. convex.

C = “capacity” is a tuning parameter, which weighs in-sample classification errors and thus controls the generalisation ability of an SVM. The higher C, the higher is the weight given to in-sample misclassifications, the lower is the generalization of the machine. Low generalisation means that the machine may work well on the training set but would perform miserably on a new sample. Bad generalisation may be a result of overfitting on the training sample, for example, in the case that this sample shows some untypical and non-repeating data structure. By choosing a low C, the risk of overfitting an SVM on the training sample is reduced. It can be demonstrated that C is linked to the width of the margin. The smaller C, the wider is the margin, the more and larger in-sample classification errors are permitted.

Solving the above mentioned constrained optimization problem of calibrating an SVM means searching for the minimum of the following Lagrange function:

$$L(w,b,\xi;\alpha,v)=\frac{1}{2}w^T w+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left\{y_i\left(w^T x_i+b\right)-1+\xi_i\right\}-\sum_{i=1}^{n}v_i\xi_i, \qquad (5)$$

where $\alpha_i \geq 0$ are the Lagrange multipliers for the inequality constraint (3) and $v_i \geq 0$ are the Lagrange multipliers for the condition (4). This is a convex optimization problem with inequality constraints, which is solved my means of classical non-linear programming tools and the application of the **Kuhn-Tucker Sufficiency Theorem**. The solution of this optimisation problem is given by the saddle-point of the Lagrangian, minimised with respect to *w*, *b*, and ξ and maximised with respect to α and v. The entire task can be reduced to a convex quadratic programming problem in α_i. Thus, by calculating α_i, we solve our classifier construction problem and are able to calculate the parameters of the linear SVM model according to the following formulas:

$$w=\sum_{i=1}^{n}y_i\alpha_i x_i \qquad (6)$$

$$b=\frac{1}{2}\left(x_{+1}^{\mathrm{T}}+x_{-1}^{\mathrm{T}}\right)\cdot w \qquad (7)$$

As can be seen from (6), α_i, which must be non-negative, weighs different companies of the training sample. The companies, whose α_i are not equal to zero, are called **support vectors** and are the relevant ones for the calculation of *w*. Support vectors lie on the margin boundaries or, for non-perfectly separable data, within the margin. By this way, the complexity of calculations does not depend on the dimension of the input space but on the number of support vectors. Here x_{+1} and x_{-1} are any two support vectors belonging to different classes, which lie on the margin boundaries.

By substituting (6) into the score (1'), we obtain the score z_j as a function of the scalar product of the financial ratios of the company to be classified and the financial ratios of the support vectors in the training sample, of α_i, and of y_i. By comparing z_j with a benchmark value, we are able to estimate if a company has to be classified as solvent or insolvent.

$$\Rightarrow z_j = \sum_{i=1}^{n} y_i \alpha_i \langle x_i, x_j \rangle + b \qquad (8)$$

IV.2.2. Kernel-transformation

In the **case of a non-linear SVM**, the score of a company is computed by substituting the scalar product of the financial ratios with a kernel function.

$$z_j = \sum_{i=1}^{n} y_i \alpha_i \langle x_i, x_j \rangle + b \rightarrow z_j = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b, \qquad (8')$$

Kernels are symmetric, semi-positive definite functions satisfying the Mercer theorem. If this theorem is satisfied, it is ensured that there exists a (possibly) non-linear map Φ from the input space into some feature space, such that its inner product equals the kernel. The non-linear transformation Φ is only **implicitly** defined through the use of a kernel, since it only appears as an inner product.

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle . \qquad (9)$$

This explains how non-linear SVMs solve the classification problem: the input space is transformed by Φ into a feature space of a higher dimension, where it is easier to find a separating hyperplane. Thus the kernel can side-step the problem that data are non-linearly separable by **implicitly** mapping them into a feature space, in which the linear threshold can be used. Using a kernel is equivalent to solving a linear SVM in some new higher-dimensional feature space. The non-linear SVM score is thus a linear combination, but with new variables, which are derived through a kernel transformation of the prior financial ratios. The score function does not have a compact functional form, depending on the financial ratios but on some transformation of them, which we do not know, since it is only implicitly defined. It can be shown that the solution of the constrained optimisation problem for non-linear SVM is given by:

$$w = \sum_{i=1}^{n} y_i \alpha_i \Phi(x_i) \tag{6'}$$

$$b = -\frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_{i,} x_{+1}) + \sum_{i=1}^{n} \alpha_i y_i K(x_{i,} x_{-1}) \right) \tag{7'}$$

But, according to (7’) and (8’), we do not need to know the form of the function Φ in order to be able to calculate the score. Since for the calculation of the score (8) the input variables are used as a product, only the kernel function is needed in (8’). As a consequence, ***Φ* and *w* are not required for the solution of a non-linear SVM**.

One can choose among many types of kernel functions. In practice, many SVM models work with **stationary Gaussian kernels with an anisotropic radial basis.** The reason why is that they are very flexible and can build fast all possible relations between the financial ratios. For example linear transformations are a special case of Gaussian kernels.

$$K(x_i, x_j) = e^{-(x_j - x_i)^T r^{-2} \Sigma^{-1} (x_j - x_i)/2} \tag{10}$$

Here Σ is the variance-covariance matrix of all financial ratios of the training set. This kernel first transforms the "anisotropic" data to the same scale for all variables. This is the meaning of "isotropic". So there is no risk that financial ratios with greater numeric ranges dominate those with smaller ranges. The only parameter which has to be chosen when using Gaussian kernels is *r*, which controls the radial basis of the kernel. This reduces the complexity of model selection. The higher is *r*, the smoother is the threshold which separates solvent from insolvent companies.[30]

Gaussian kernels non-linearly map the data space into a higher dimensional space. Actually the definition of a Gaussian process by specifying the covariance function (depending on the distance of the company to be evaluated from each company of the training sample) **avoids explicit definition of the function class of the transformation**. There are many possible decompositions of this covariance and thus also many possible transformation functions of the input financial ratios. Moreover each company shows its own covariance function, depending on its relative position within the training sample. That is why the kernel operates locally. The value of the kernel function depends on the distance between the financial ratios of the company *j* to be classified and respectively one company *i* of the training sample. This kernel is a normal density function up to a constant multiplier. x_i is the center of this kernel, like the mean is the center of a normal density function.

IV.3. What is the Point in Using SVMs as a Classification Technnique?

All classification techniques have advantages and disadvantages, which are more or less important according to the data which are analysed, and thus have a relative relevance. SVMs can be a useful tool for insolvency analysis, in the case of non-regularity in the data, for example when the data are not regularly distributed or have an unknown distribution. It can help evaluate information, i.e. financial ratios which should be transformed prior to entering the score of classical classification techniques.

30 By choosing different *r* values for different input values, it is possible to rescale outliers.

The **advantages of the SVM technique** can be summarised as follows:

- By introducing the kernel, SVMs gain flexibility in the choice of the form of the threshold separating solvent from insolvent companies, which does not have to be linear and even does not have to have the same functional form for all data, since its function is non-parametric and operates locally. As a consequence they can work with financial ratios, which show a non-monotone relation to the score and to the probability of default, or which are non-linearly dependent, and this without needing any specific work on each non-monotonous variable.
- Since the kernel **implicitly** contains a non-linear transformation, no assumptions about the functional form of the transformation, which makes data linearly separable, are necessary. The transformation occurs implicitly on a robust theoretical basis and human expertise judgement in advance is not needed.
- SVMs provide a good out-of-sample generalization, if the parameters C and r (in the case of a Gaussian kernel) are appropriately chosen. This means that, by choosing an appropriate generalization grade, SVMs can be robust, even when the training sample has some bias.
- SVMs deliver a unique solution, since the optimality problem is convex. This is an advantage compared to Neural Networks, which have multiple solutions associated with local minima and for this reason may not be robust over different samples.
- With the choice of an appropriate kernel, such as the Gaussian kernel, one can put more stress on the similarity between companies, because the more similar the financial structure of two companies is, the higher is the value of the kernel. Thus when classifying a new company, the values of its financial ratios are compared with the ones of the support vectors of the training sample which are more similar to this new company. This company is then classified according to with which group it has the greatest similarity.

Here are some examples where the SVM can help coping with non-linearity and non-monotonicity. One case is, when the coefficients of some financial ratios in equation (1), estimated with a linear parametric model, show a sign that does not correspond to the expected one according to theoretical economic reasoning. The reason for that may be that these financial ratios have a non-monotone relation to the PD and to the score. The unexpected sign of the coefficients depends on the fact, that data dominate or cover the part of the range, where the relation to the PD has the opposite sign. One of these financial ratios is typically the growth rate of a company, as pointed out by [10]. Also leverage may show non-monotonicity, since if a company primary works with its own capital, it may not exploit all its external financing opportunities properly. Another example may be the size of a company: small companies are expected to be more financially instable; but if a company has grown too fast or if it has become too static because of its dimension, the big size may become a disadvantage. Because of these characteristics, the above mentioned financial ratios are often sorted out, when selecting the risk assessment model according to a linear classification technique. Alternatively an appropriate evaluation of this information in linear techniques requires a transformation of the input variables, in order to make them monotone and linearly separable.[31]

A common **disadvantage of non-parametric techniques** such as SVMs is the lack of transparency of results. SVMs cannot represent the score of all companies as a simple parametric function of the financial ratios, since its dimension may be very high. It is neither a linear combination of single financial ratios nor has it another simple functional form. The weights of the financial ratios are not constant. Thus **the marginal contribution of each financial ratio to the score is variable**. Using a Gaussian kernel each company has its own weights according to the difference between the value of their own financial ratios and those of the support vectors of the training data sample.

[31] See [6] for an analysis of the univariate relation between the PD and single financial ratios as well as for possible transformations of input financial ratios in order to reach linearity.

Interpretation of results is however possible and can rely on graphical visualization, as well as on a local linear approximation of the score. The SVM threshold can be represented within a bi-dimensional graph for each pair of financial ratios. This visualization technique cuts and projects the multidimensional feature space as well as the multivariate threshold function separating solvent and insolvent companies on a bi-dimensional one, by fixing the values of the other financial ratios equal to the values of the company, which has to be classified. By this way, different companies will have different threshold projections. However, an analysis of these graphs gives an important input about the direction towards which the financial ratios of non-eligible companies should change, in order to reach eligibility.

The PD can represent a third dimension of the graph, by means of isoquants and colour coding. The approach chosen for the estimation of the PD can be based on empirical estimates or on a theoretical model. Since the relation between score and PD is monotone, a local linearization of the PD can be calculated for single companies by estimating the tangent curve to the isoquant of the score. For single companies this can offer interesting information about the factors influencing their financial solidity.

In the figure below the PD is estimated by means of a Gaussian kernel[32] on data belonging to the trade sector and then smoothed and monotonized by means of a Pool Adjacent Violator algorithm.[33] The pink curve represents the projection of the SVM threshold on a binary space with the two variables K21 (net income change) and K24 (net interest ratio), whereas all other variables are fixed at the level of company *j*. The blue curve represents the isoquant for the PD of company *j*, whose coordinates are marked by a triangle.

[32] This methodology is based on a non-parametric estimation of the PD and has the advantage that it delivers an individual PD for each company based on a continuous, smooth and monotonic function. This PD-function is computed on an empirical basis, so there is no need for a theoretical assumption about the form of a link function.
[33] See [11].

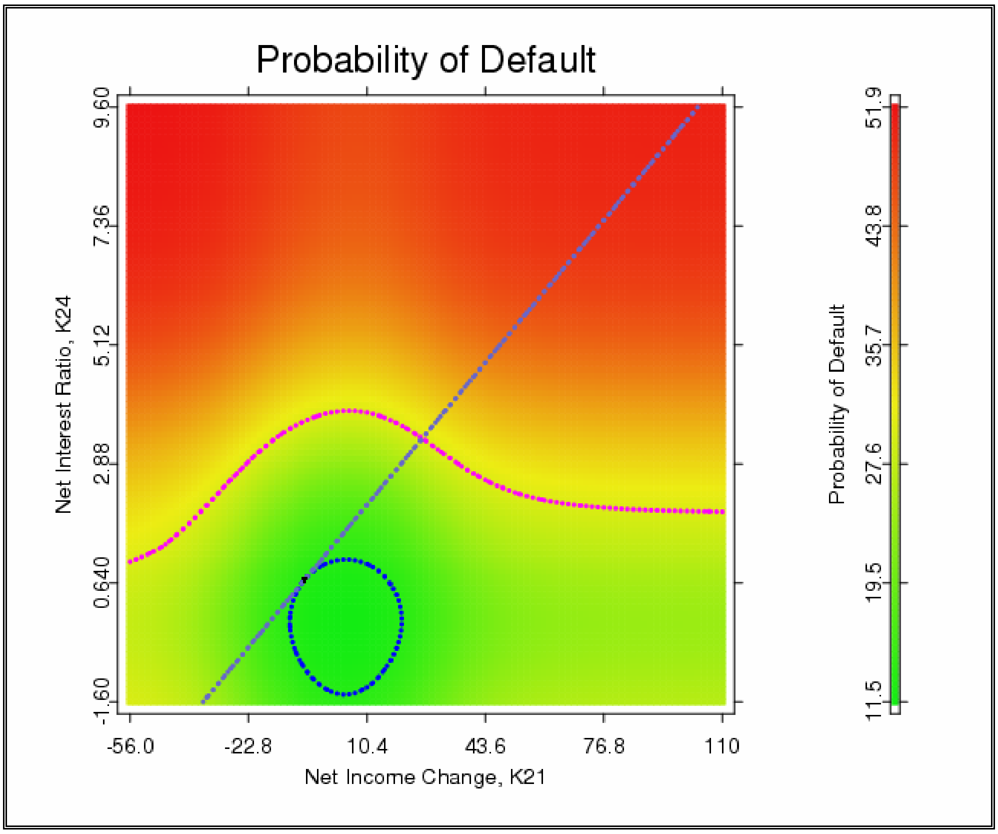


Exhibit 2: Graphical Visualization of the SVM Threshold and of a Local Linearization of the Score Function: Example of a Projection on a Bi-dimensional Graph with PD Colour Coding

The grey line corresponds to the linear approximation of the score or PD function projection for company *j*. One interesting result of this graphical analysis is that successful companies with a low PD often lie in a closed space. This implies that there exists an optimal combination area for the financial ratios being considered, outside of which the PD gets higher. If we consider the net income change, we notice that its influence on the PD is non-monotone. Both too low or too high growth rates imply a higher PD. This may indicate the existence of the optimal growth rate and suggest that above a certain rate a company may get into trouble; especially if the cost structure of the company is not optimal i.e. the net interest ratio is too high. If a company, however, lies in the optimal growth zone, it can also afford a higher net interest ratio.

IV.4. AN EMPIRICAL SVM MODEL FOR SOLVENCY ANALYSIS

In the following chapter, an empirical SVM model for solvency analysis on German data will be presented.[34] The estimation of score functions and their validation are based on balance sheets of solvent and insolvent companies. In doing so a **company is classified as insolvent if it is the subject of failure judicial proceeding**. The study is conducted over a long period, in order to construct durable scores that are resistant, as far as possible, to cyclical fluctuations. So the original data set consists of about 150.000 firm-year observations, spanning the time period from 1999 to 2005. The **forecast horizon** is three and a half years. That is, in each period a company is considered insolvent, if it is subject to legal proceedings within three and a half years since the observation date. Solvent companies are those that have not gone bankrupt within three and a half years after the observation date. With shorter term forecast horizons, such as one-year, data quality would be poor, since most companies do not file a balance sheet, if they are on the point of failure. Moreover, companies that go insolvent already show weakness three years before failure. In order to improve the accuracy of analysis, a different model was developed for each of the following three **sectors**: manufacturing, wholesale/retail trade and other companies. The three models for the different sectors were trained on data over the time period 1999-2001 and then validated out-of-time on data over the time period 2002-2005.

Two important points for the selection of an accurate SVM model are the choice of the input variables, i.e. of the financial ratios, which are considered in the score, as well as of the tuning parameters C and r (once a Gaussian kernel has been chosen).

sector	year							total	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**solv.**	**ins.**
manufacturing	6015	5436	4661	5202	5066	4513	698	30899	692
wholesale / retail trade	12806	11230	9209	8867	8016	7103	996	57210	1017
other	6596	6234	5252	5807	5646	5169	650	34643	711

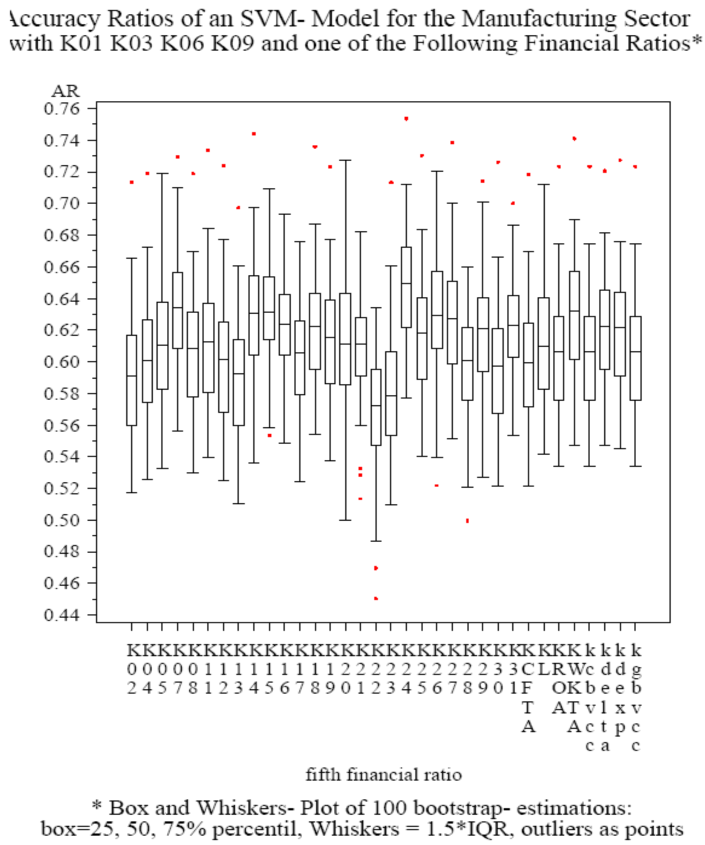
Table 1: Training and Validation Data Set Size – Without Missing Values

[34] The database belongs to the balance sheet pool of the "Deutsche Bundesbank".

The choice of the input variables has a decisive influence on the performance results and is not independent from the choice of the classification technique. These variables normally have to comply with the assumptions of the applied classification technique. Since the SVM needs no restrictions on the quality of input variables, it is free to choose them only according to the model accuracy performance. The **input variables selection methodology** applied in this paper is based on the following empirical tools.

The discriminative power of the models is measured on the basis of their accuracy ratio (AR) and percentage of correctly classified observations, which is a compact performance indicator, complementary to their error quotes. Since there is no assumption on the density distribution of the financial ratios, a robust comparison of these performance indicators has to be constructed on the basis of bootstrapping.

The different SVM models are estimated 100 times on 100 randomly selected training samples, which include all insolvent companies of the data pool and the same number of randomly selected solvent ones. Afterwards they are validated on 100 similarly selected validation samples. The model, which delivers the best median results over all training and validation samples, is the one which is chosen for the final calibration. A similar methodology is used for **choosing the optimal capacity *C* and the kernel-radius *r* of the SVM model**. That combination of C and r values is chosen, which delivers the highest median AR on 100 randomly selected training and validation samples.



Ehibit 3: Choice of the Financial Ratios of an SVM Model for the Manufacturing Sector: An Example for the Choice of the Fifth Input Variable

Our analysis first started by estimating the three SVM models on the basis of four financial ratios, which are presently being used by the “Bundesbank” for DA and which are expected to comply with its assumptions on linearity and monotonicity. By enhancing the model with further non-linearly separable variables a significant performance improvement in the SVM model was recorded. The new input variables were chosen out of a catalogue, which is summarized in Table 3, on the basis of a bootstrapping procedure by means of forward selection with an SVM model. Variables were added to the model sequentially until none of the remaining ones improved the median AR of the model. Exhibit 3 shows the AR distributions of different SVM models with 5 variables. According to these graphical results one should choose K24 as the fifth variable. As a result of this selection procedure, the median AR peaked with ten input variables (10FR) and then fell gradually.

Sector		
Manufacturing	**Wholesale/Retail Trade**	**Other**
K01: pre-tax profit margin	K01: pre-tax profit margin	K02: operating profit margin
K03: cash flow ratio	K04: capital recovery ratio,	K05: debt cover
K06: days receivable	K06: days receivable	K06: days receivable
		K07: days payable
K09: equity ratio adj.	K09: equity ratio adj.	K08: equity ratio
K11: net income ratio	K17: liquidity 3 (current assets to short debt)	K12:guarantee a.o. obligation ratio (leverage 1)
K15: liquidity 1		
K18: short term debt ratio	K18: short term debt ratio	K18: short term debt ratio
K24: net interest ratio	K21: net income change	K19: inventories ratio
	K24: net interest ratio	K21: net income change
K26: tangible asset growth	K31: days of inventories	K31: days of inventories
KWKTA: working capital to total assets	KL: leverage	KL: leverage

Table 2: Final Choice of the Input Variables Forward Selection Procedure

A univariate analysis of the relation between the single variables and the PD showed that most of these variables actually have a non-monotone relation to the PD, so that considering them in a linear score would require the aforementioned transformation. Especially growth variables as well as leverage and net interest ratio showed a typical non-monotone behaviour and were at the same time very helpful in enhancing the predictive power of the SVM.

Exhibit 4 summarizes the predictive results of the three final models, according to the above mentioned bootstrap procedure. Based on the procedure outlined above, the following values of the kernel tuning parameters were selected: r = 4 for the manufacturing and trade sector and r = 2.5 for other companies. This suggests that this sector is less homogeneous than the other two. The capacity of the SVM model was chosen as C = 10 for all the three sectors. It is interesting to notice, that the robustness of the results, measured by the spread of the ARs over different samples, became lower, when the number of financial ratios being considered grew. So there is a trade-off between the accuracy of the model and its robustness.

Variable	Name	Aspect	Q 0.01	median	Q 0.99	IQR	Relation to the PD
K01	Pre-tax profit (income) margin	profitability	-57.1	2.3	140.1	6.5	- n.m.
K02	Operating profit margin	profitability	-53	3.6	80.3	7.2	-
K03	Cash flow ratio (net income ratio)	liquidity	-38.1	5.1	173.8	10	-
K04	Capital recovery ratio	liquidity	-29.4	9.6	85.1	15	-
K05	Debt cover (debt repayment capability)	liquidity	-42	16	584	33	-
K06	Days receivable (accounts receivable collection period)	activity	0	29	222	34	+ n.m.
K07	Days payable (accounts payable collection period)	activity	0	20	274	30	+ n.m.
K08	Equity (capital) ratio	financing	-57	16.4	95.4	27.7	-
K09	Equity ratio adj. (own funds ratio)	financing	-55.8	20.7	96.3	31.1	-
K11	Net income ratio	profitability	-57.1	2.3	133.3	6.4	+/- n.m.
K12	guarantee a.o. obligation ratio (leverage 1)	leverage	0	0	279.2	11	-/+ n.m.
K13	Debt ratio	liquidity	-57.5	2.4	89.6	18.8	-/+ n.m.
K14	Liquidity ratio	liquidity	0	1.9	55.6	7.2	-
K15	Liquidity 1	liquidity	0	3.9	316.7	16.7	-
K16	Liquidity 2	liquidity	1	63.2	1200	65.8	- n.m.
K17	Liquidity 3	liquidity	2.3	116.1	1400	74.9	- n.m.
K18	Short term debt ratio	financing	0.2	44.3	98.4	40.4	+
K19	Inventories ratio	investment	0	23.8	82.6	35.6	+
K20	Fixed assets ownership ratio	leverage	-232.1	46.6	518.4	73.2	-/+ n.m.
K21	Net income change	growth	-60	1	133	17	-/+/- n.m.
K22	Own funds yield	profitability	-413.3	22.4	1578.6	55.2	+/- n.m.
K23	Capital yield	profitability	-24.7	7.1	61.8	10.2	-
K24	Net interest ratio	cost. structure	-11	1	50	1.9	+ n.m.
K25	Own funds/pension provision r.	financing	-56.6	20.3	96.1	32.4	-
K26	Tangible assets growth	growth	-0.2	13.9	100	23	-/+ n.m.
K27	Own funds/provisions ratio	financing	-53.6	27.3	98.8	36.9	-
K28	Tangible asset retirement	growth	0.1	19.3	98.7	18.7	-/+ n.m.
K29	Interest coverage ratio	cost structure	-2364	149.5	39274.3	551.3	n.m.
K30	Cash flow ratio	liquidity	-27.9	5.2	168	9.7	-
K31	Days of inventories	activity	0	41	376	59	+
K32	Current liabilities ratio	financing	0.2	59	96.9	47.1	+
KL	Leverage	leverage	1.4	67.2	100	39.3	+ n.m.
KWKTA	Working capital to total assets	liquidity	565.9	255430	51845562.1	865913	+/- n.m.
KROA	Return on assets	profitability	-42.1	0	51.7	4.8	n.m.
KCFTA	Cash flow to total assets	liquidity	-26.4	9	67.6	13.6	-
KGBVCC	Accounting practice, cut		-2	0	1.6	0	n.m.
KCBVCC	Accounting practice		-2.4	0	1.6	0	n.m.
KDEXP	Result of fuzzy expert system, cut		-2	0.8	2	2.8	-
KDELTA	Result of fuzzy expert system		-7.9	0.8	8.8	3.5	-

n.m.= non-monotone

+ = positive relation

- = negative relation

+ n.m.= non monotone relation, mostly positive

- n.m.= non monotone relation, mostly negative

+/- n.m. = non-monotone relation, first positive then negative

-/+ n.m. = non-monotone relation, first negative then positive

-/+/- n.m. = non-monotone relation, first negative, then positive then again negative

Table 3. The Catalogue of Financial Ratios – Univariate Summary Statistics and Relation to the PD[35]

[35] K1-K32 as well as KGBVCC and KDEXP are financial ratios belonging to the catalogue of the "Deutsche Bundesbank". See [4].

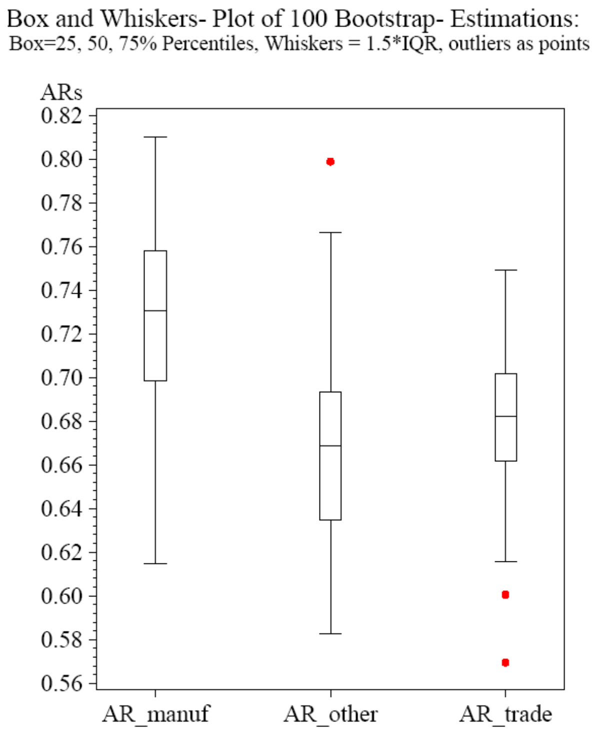


Exhibit 4: Predictive Results: ARs of the Final SVM Model after Bootstrapping

IV.5. SUMMARY AND CONCLUSION

SVMs can produce accurate and robust classification results on a sound theoretical basis, even when input data are non-monotone and non-linearly separable. So they can help to evaluate more relevant information in a convenient way. Since they linearize data on an implicit basis by means of kernel transformation, the accuracy of results does not rely on the quality of human expertise judgement for the optimal choice of the linearization function of non-linear input data. SVMs operate locally, so they are able to reflect in their score the features of single companies, comparing their input variables with the ones of companies in the training sample showing similar constellations of financial ratios. Although SVMs do not deliver a parametric score function, its local linear approximation can offer an important support for recognising the mechanisms linking different financial ratios with the final score of a company. For these reasons SVMs are regarded as a useful tool for effectively complementing the information gained from classical linear classification techniques.

REFERENCES

B. BAESENS/ T. VAN GESTEL/ S. VIANE/ M. STEPANOVA/ J. SUYENKS/J. VANTHIENEN: Benchmarking State-of-the-art Classification Algorithms for Credit Scoring: in Journal of the Operational Research Society , No. 1-9. 2003.

B. ENGELMANN/ E. HAYDEN/ D. TASCHE: Measuring the Discriminative Power of Rating Systems: in Deutsche Bundesbank Discussion Paper, Series 2: Banking and Financial Supervision. No 1. 2003.

B. SCHÖLKOPF/ A. SMOLA: Learning with Kernels -Support Vector Machines, Regularization, Optimization and Beyond, MIT Press, Cambridge, MA, 2002, http://www.learning-with-kernels.org. 2002

C.-W. HSU/ C.-C. CHANG/ C.J. LIN: A Practical Guide to Support Vector Classification: on http://www.csie.ntu.edu.tw. 2007.

Deutsche Bundesbank: How the Deutsche Bundesbank Assesses the Credit Standing of Enterprises in the Context of Refinancing German Credit Institutions, Markets Department, on: http://www.bundesbank.de/download/gm/gm_broschuere_bonitaetunternehmen_en.pdf. June 2004

E. FALKENSTEIN: Riskcalc for Private Companies: Moody's Default Model, Moody's Investor Service. 2000.

E. HAYDEN: Modeling an Accounting-Based Rating System for Austrian Firms,: A Dissertation in der Fakultät für Wirtschaftwissenschaften und Informatik. Universität Wien. Juni 2002.

E. MAMMEN: Estimating a Smooth Monotone Regression Function: in The Annals of Statistics. Vol. 19. No. 2: 724-740. June 1991.

N. CRISTIANINI/ J. SHAWE-TAYLOR: An Introduction to Support Vector Machines and Other Kernel-based Learning Methods: in Repr. 2006, Cambridge University Press, 2000.

T. VAN GESTEL/ B. BAESENS/ J. GARCIA/ P. VAN DIJCKE: A Support Vector Machine Approach to Credit SCORING: ON: HTTP://WWW.DEFAULTRISK.COM/PP_SCORE_25.HTM. 2007.

W. K. HÄRDLE/ R. A. MORO/ D. SCHÄFER: Rating Companies with Support Vector Machines: in DIW Discussion Paper. No. 416. 2004.

W. K. HÄRDLE/ R. A. MORO/ D. SCHÄFER: Support Vector Machines – Eine neue Methode zum Rating von Unternehmen: in DIW Wochenbericht. No. 49/04. 2004.

V. VAPNIK: The Nature of Statistical Learning Theory: in Springer. 2000.

V. PROJECT FINANCE – A MONTE CARLO APPROACH TO ESTIMATE PROBABILITY OF DEFAULT, LOSS GIVEN DEFAULT AND EXPECTED LOSS

G. TESSIORE / V. FAVALE
CENTRALE DEI BILANCI

V.1. SYNOPSIS

Under Basel II, Project finance (PF) is one of five sub-classes of specialized lending (SL) within the corporate asset class. Basel II proposes a pure *qualitative-judgmental* method - the “Supervisory Slotting Criteria Approach” - to evaluate Probability of Default (PD), Loss Given Default (LGD) and Expected Loss (EL) of a PF operation while allowing banks to develop their own Internal Rating System methodology.

In this paper, we suggest a *quantitative* method based on Monte Carlo (MC) simulations of future cash flows (of the project involved within a Project Finance operation) that allows for an analytical estimation of the PD, the LGD and the EL.

We suggest adopting two different MC simulations: a main simulation of future cash flows to estimate the PD of the project and several conditioned-to-default MC simulations of recovery rates to estimate the LGD.

We use the tool developed by the Centrale dei Bilanci for the evaluation of PF risk to show how to collect data from different projects in a standardized way, how to model micro and macroeconomic scenarios with “hierarchical” dependences and different stochastic distributions and, finally, how to manage specific risk events in compliance with Basel II requirements.

V.2. Introduction – What is Project Finance ?

Project Financing (PF) is a method of funding in which the lender looks primarily to the revenues generated by a single project, both as a source of repayment and as collateral for the exposure. This type of financing is usually for large, complex and expensive installations. In such transactions, the lender is usually paid solely or almost exclusively out of the money generated by the contracts for the facility's output, such as the electricity sold by a power plant. The borrower is usually an SPV (Special Purpose Vehicle) that is not permitted to perform any function other than developing, owning, and operating the installation. The consequence is that repayment depends primarily on the project's cash flow and on the collateral value of the project's assets.

There are many actors that are involved in a PF operation. We can summarize the relationship between all the actors and the SPV with the following scheme (Exhibit 1).

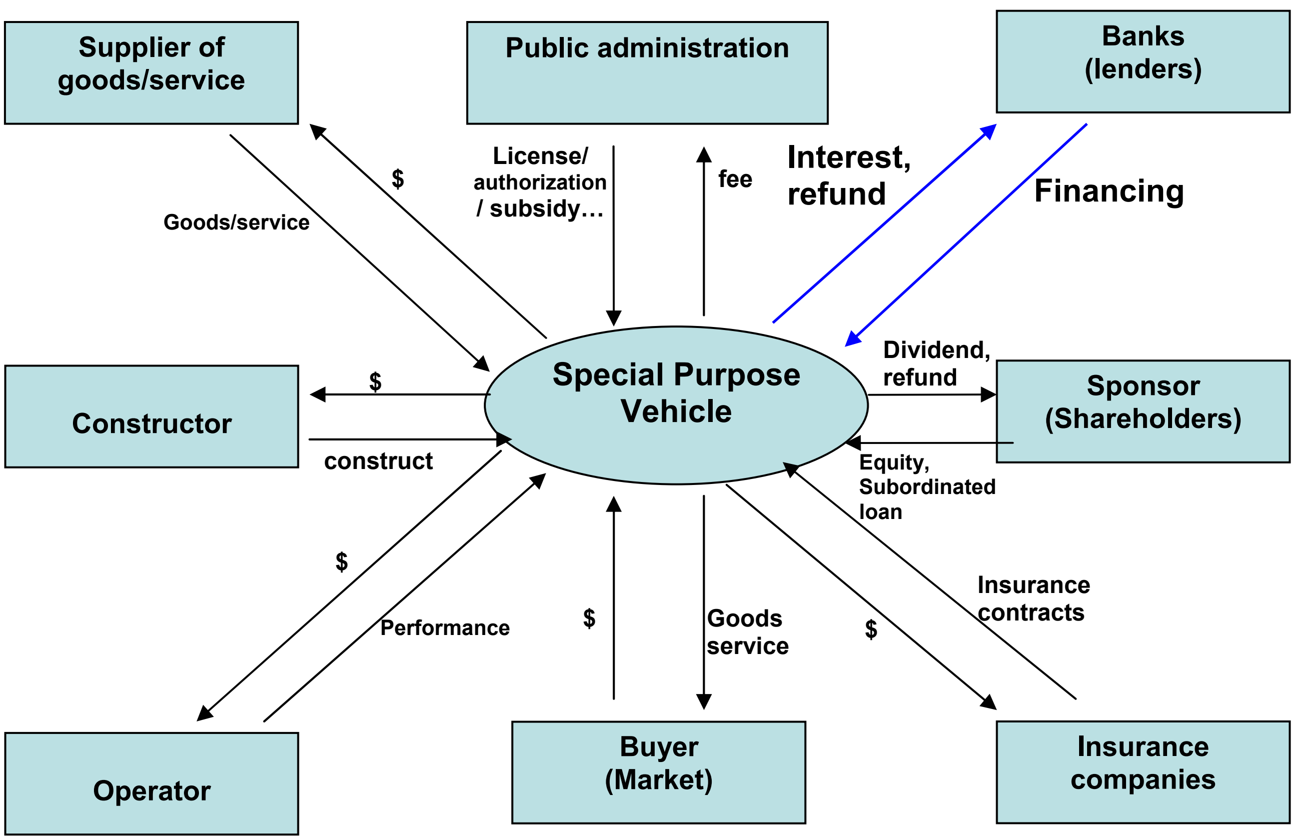


Exhibit 1: Relationship between all the actors and the SPV

Banks and the sponsor finance the SPV to get the project started. During the life of the project the positive cash flows generated by the project allow the SPV to refund loans from banks and the sponsor. If the cash flows are insufficient to refund the bank loans (also using ad-hoc reserves created during the life of the project), the project is in ***default***. From this moment, a *recovery procedure starts which normally is* based on a disposal of the assets on the market or on a disposal of the ongoing project[36].

The following paper presents the point-of-view of the banks and illustrates a specific method developed by Centrale dei Bilanci aiming at:

- collecting and standardizing the information of each PF operation using a common template available for all different kinds of projects;
- estimating the Probability of Default (PD) of a project in compliance with Basel II requirements and, finally,
- estimating the Loss Given Default (LGD) and the Expected Loss (EL) .

V.3. CENTRALE DEI BILANCI APPROACH TO EVALUATE PF OPERATIONS

Special Purpose Vehicles (SPVs) typically are incorporated "ad hoc" in order to manage a specific project. Therefore historical data are not available and it is very difficult to assess the credit quality of these companies. The SPV's historical features are not relevant anyhow because the refund depends on the future cash flows of the project.

Given that in PF operations default events are very rare, it is consequently not possible to create an extensive database to estimate a good statistical model.
Moreover, projects operate under different scenarios with specific and often complex contractual structures that make it hard to evaluate a new project using the experience of (few) similar projects. This means that comparing different projects, even of the same "family", could not be sufficient for a correct evaluation of the risk involved.

[36] If the project is large and difficult to sell on the market, normally banks restructure the debt in a new financing operation. In this case, however, it is important to evaluate the residual debt and the future value of the project.

The “term sheets” of PF operations provide for long-term planning; therefore, it is important to consider the whole life of the project to evaluate correctly the dynamics of future repayments. It would be unrealistic to establish an “average evaluation” suitable for the whole life of the project.

LGD depends on many factors: the (complex) collateral structure, the seniority of the debt, the residual value of the assets (or the future cash flows); as a consequence, it is very difficult to summarize these measures in a qualitative or – even more difficult - in a statistical model.

The complexity of the project makes a merely *qualitative* analysis inadequate. For these reasons, it is preferable, for the assessment of credit risk of a PF operation, to use a method that:

- bypasses the problem of a lack of historical data,
- is forward looking,
- is not a simple qualitative analysis, but that is also able to capture the complexity of the specific project,
- estimates a long term PD and its term structure,
- weighs both the microeconomic and macroeconomic environment of the project and, finally,
- estimates LGD in a quantitative and robust way.

Monte Carlo simulations of economical-financial scenarios are a suitable methodology to estimate the PD, the LGD and the EL for PF operations.

V.3.1. Introduction to the Monte Carlo approach

The main idea of Monte Carlo simulations is to generate large numbers of realistic future economical-financial scenarios based on a probabilistic model of the composite environment in which the project operates. Theoretically, the macro-economic scenario influences the micro-economic scenario of the specific project. The tool simulates both the macro- and the micro-economic scenarios. Given the simulated scenario, the tool computes the cash flows and the financial statements of the project over its life.
Aggregating the results of all the simulated scenarios, the tool estimates PD, LGD and EL of the project both from a short and a long-term perspective.

The model considers the figures of the project as *random variables* which depend on a series of *risk drivers* or on other characterstics of the project. The (hierarchical) dependence structure within the variables of the project and the stochastic distribution associated to each variable define the probabilistic model object of the MC simulation.

In our approach, we suggest to start analyzing the business plan of the project, i.e. the most likely scenario we should expect: the analyst enters the data of the business plan into the forms of the tool, namely data on investment, revenues, costs, financing structure, stocks, taxation etc. This step is essential because it allows different projects to be treated in a standardized way. An algorithm re-computes the balance sheets of the business plan in an automatic way to check if the data entry was correct.

Then, the process follows by defining the probabilistic model of the project. The analyst, for each item of the business plan, finds the risk drivers that can affect its value. The analyst defines the relationship between the items of the business plan and their risk drivers. These relationships can be defined by a mathematical formula or by a correlation coefficient. The relationships are always *causal*: this means that each variable is generated depending on the values assumed by the linked variables at a lower level. If the relationship is defined by a correlation coefficient, the probabilistic distribution of the variable is needed.

Most of MC methods are based on a copula method that generates all the variables at the same time, given a complete correlation matrix. In a PF operation, the list of variables and risk drivers can be very long so that it would be practically impossible to define such a huge correlation matrix.
In our approach, we use a “hierarchical” structure between variables thereby bypassing this problem and allowing the analyst to specify only the known relationships. All the other relationships are implicit in the hierarchical structure.

The formal validity of the hierarchical dependence structure is automatically checked by the tool.

V.3.2. Project phases and default definition

We distinguish two different typical phases during the life of a project:

- Construction period
- Operating period

During the construction period (that can last several years) the capital expenditure generates outflows – negative cash flows - financed by banks and shareholders by the PF operation. During the construction period the default event happens when the *investment exceeds the maximum amount of loan financing available.*

During the operating period the project generates positive cash flows that allow to refund the debt. The default event happens when the **DEBT SERVICE (interest + repayments) of the year T is greater than:**

Operating Cash Flow generated during year T
+ DSRA (Debt Service Reserve Account)[37] at the beginning of year T
+ Liquidity at the beginning of year T (plus credit revolving facility need during year T)
+ MRA (Maintenance Reserve Account)[38] at the beginning of year T

[37] DSRA is a cash reserve set aside to repay future debt, in the event that cash generated by operations is temporary insufficient.
[38] MRA is a reserve account that builds up cash balances sufficient to cover a project’s maintenance expenses.

The default condition is tested for each year of the simulated scenarios.

The first image in Exhibit 2 shows a default event occurring during the 3rd year of the operating period. The second image represents 1000 simulated scenarios of cash flows of the project, with a construction period of 4 years and an operating period of 15 years.

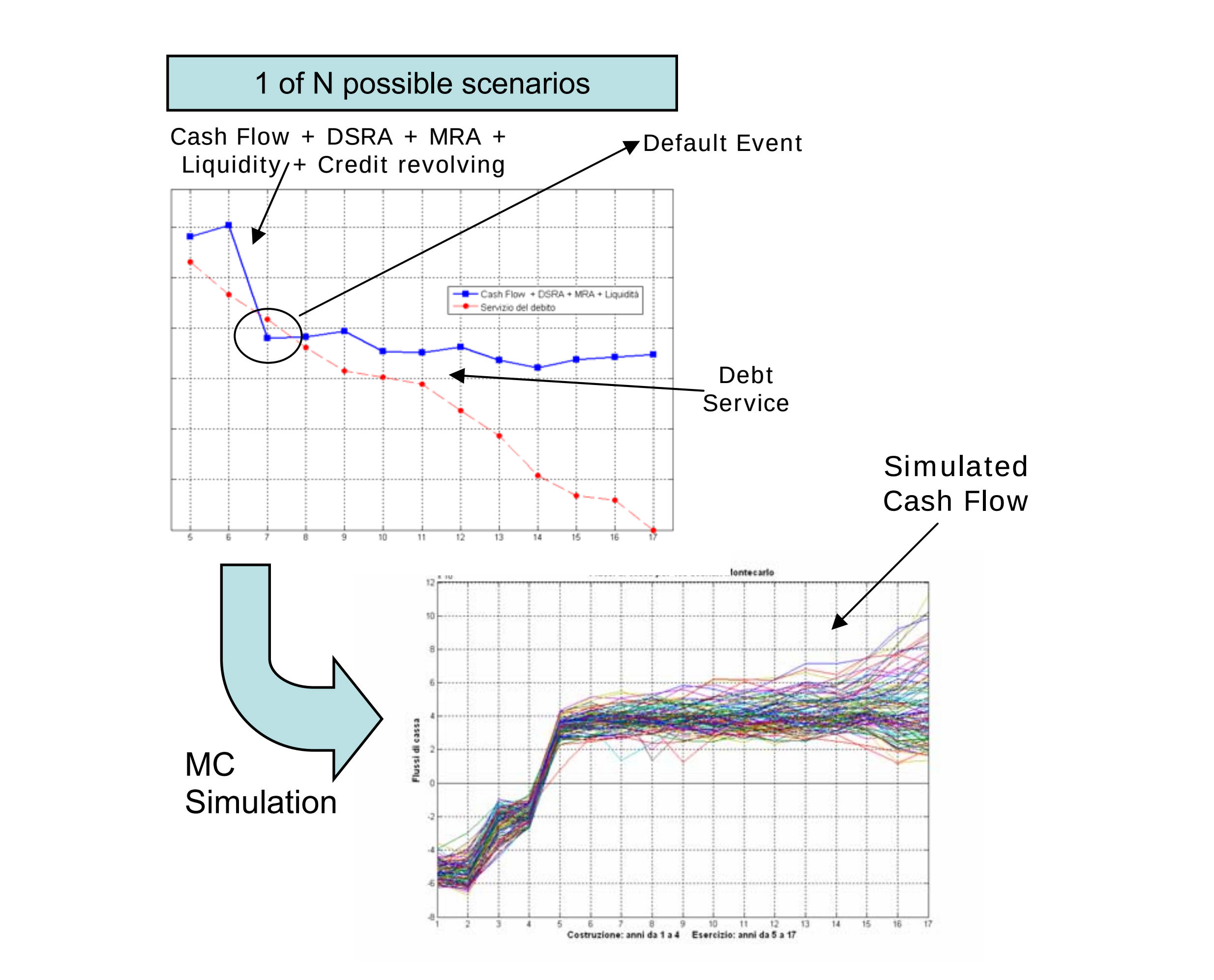


Exhibit 2: Simulated scenarios

The MC simulation generates n-thousand future possible scenarios. When the default happens, a second MC simulation starts allowing the estimation of the LGD.

V.4. IMPLEMENTATION OF THE METHOD IN CENTRALE DEI BILANCI'S TOOL

V.4.1. The process

The procedure of PD-LGD estimation could be summarized by the following steps:

- Integration of the main assumptions in the BASE CASE.
- Identification of the RISK DRIVERS
- Definition of stochastic structure
    - Relationship between variables and risk drivers
    - Probabilistic distributions

Data entry

- Monte Carlo simulation of the economical-financial scenarios
    - Financial statement computation
    - Cash flow/Debt service simulation
    - Testing of default event
    - Computation of PD, LGD and EL

Monte Carlo Simulation

- Reporting

V.4.2. Data entry

PF covers a wide range of projects, especially in the sector of public utilities (transportation, energy, oil & gas, water distribution) and construction (general contractors, motorways, public services).

Because of the different characteristics of the projects the data entry has been structured in the following 10 sections with complete flexibility within each of them:

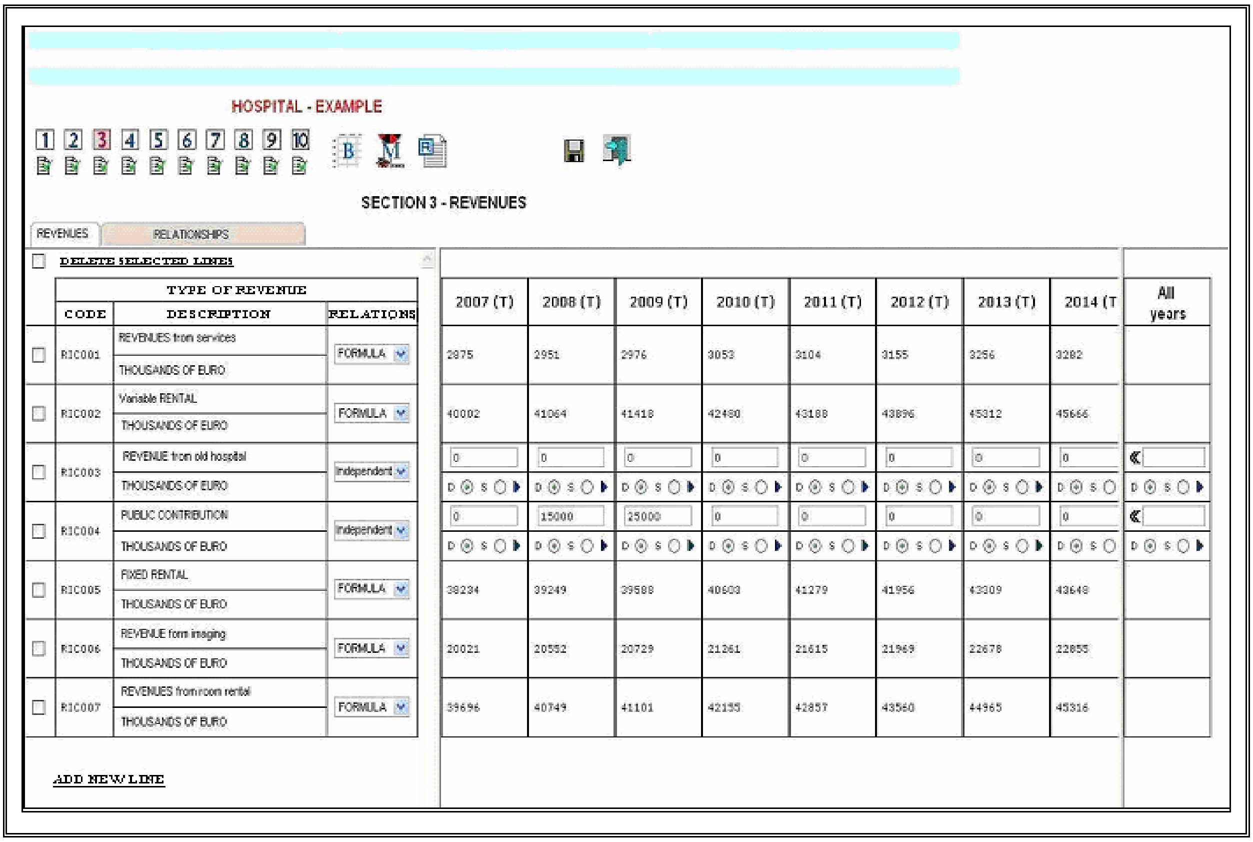
- Sections 1-7 collect both data on the *base case that will be used for key financial results* and information on the *stochastic structure* of the project being used in the MC simulation.
  - Section 1 – General characteristics of the project
  - Section 2 – Investment and construction data
  - Section 3 – Operating revenues
  - Section 4 – Operating costs
  - Section 5 – Working capital
  - Section 6 – Taxation
  - Section 7 – Funding structure
    - Bank loans:
      - Term loan (main terms of the loan)
      - Stand-by loan (available during construction period)
      - Revolving credit (debt that support term loan)
      - VAT financing
      - Other loans
    - Shareholders' funding
      - Subordinated loans
      - Equity
- Section 8 encompasses a list of risks that might affect the project. The structure of the risks is compliant with the requirements of the Basel II Slotting Criteria Approach, and includes a list of specific risk events to be considered in the PD estimation. These risks affect the base case values filled in section 2 to 7.
- Section 9 and 10 collect information on recovery values and collaterals. These sections are used in LGD estimation and only for default scenarios.

- o Section 9 – Recovery values (specifies the *rules* for computing the recovery values of an asset in an automatic way)
- o Section 10 - Collaterals

- The values filled in sections 2 to 7 refer to the following items:
- The base case assumptions (*mandatory*)
- The relationship between the value and other variable(s) (*optional*)
- The stochastic distribution of the value in each year (*optional*)

In each section, there is a sub-section where it is possible to specify values / relationship / formula / stochastic distributions of all the risk drivers linked to the main items of the financial statements.

In Exhibit 3 and Exhibit 4 we show an abstract from the data entry procedure: Exhibit 3 presents a list of seven different kinds of revenue (all customized by the analyst). Exhibit 4 presents a list of several risk drivers used to generate the values in Exhibit 3.


HOSPITAL - EXAMPLE

SECTION 3 - REVENUES

REVENUES | RELATIONSHIPS

DELETE SELECTED LINES

CODE	DESCRIPTION	RELATIONS	2007 (T)	2008 (T)	2009 (T)	2010 (T)	2011 (T)	2012 (T)	2013 (T)	2014 (T	All years
RIC001	REVENUES from services THOUSANDS OF EURO	FORMULA	2875	2951	2976	3053	3104	3155	3256	3282	
RIC002	Variable RENTAL THOUSANDS OF EURO	FORMULA	40002	41064	41418	42480	43188	43896	45312	45666	
RIC003	REVENUE from old hospital THOUSANDS OF EURO	Independent	0 D S	0 D S	0 D S	0 D S	0 D S	0 D S	0 D S	0 D S	D S
RIC004	PUBLIC CONTRIBUTION THOUSANDS OF EURO	Independent	0 D S	15000 D S	25000 D S	0 D S	0 D S	0 D S	0 D S	0 D S	D S
RIC005	FIXED RENTAL THOUSANDS OF EURO	FORMULA	38234	39249	39588	40603	41279	41956	43309	43648	
RIC006	REVENUE form imaging THOUSANDS OF EURO	FORMULA	20021	20552	20729	21261	21615	21969	22678	22855	
RIC007	REVENUES from room rental THOUSANDS OF EURO	FORMULA	39696	40749	41101	42155	42857	43560	44965	45316	

ADD NEW LINE

Exhibit 3: Revenues

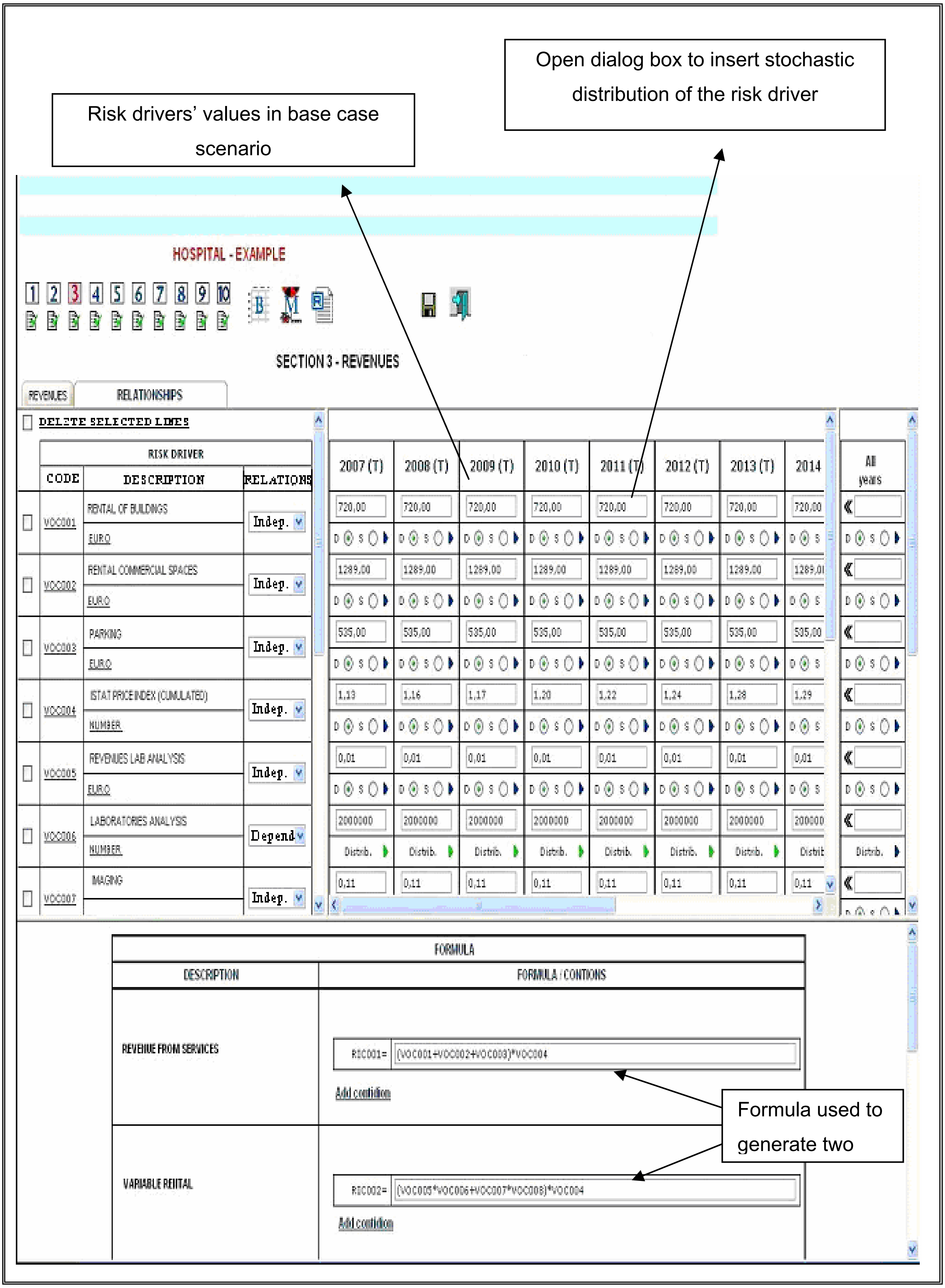


Exhibit 4: Risk drivers

The formulas to generate revenues are shown in Exhibit 4. In a specific dialog box (not shown here) it is possible to define the correlation between variables. The analyst enters only the known and direct correlation between the risk drivers and the items of the project.

This structure of the windows is common to sections 2 to 7 of data entry.

In section 8, information of specific risk events are collected. The risk events could be the same as suggested in the Basel II Slotting Criteria Approach or any other risk events that could affect the evaluation of the project. For each risk event the analyst have to specify the probability of the event and the effect of this event on the financial statemet. The risk's impact on simulated cash flow and, consequently, on the estimated PD of the project could be really significant.

In section 9, data on the method of estimation of LGD are collected. The recovery rates, interest rates and the durations of the recovery processes are also gathered in this section.

In section 10, we explore the collateral structure and the seniority of the debt (and the hierarchy of the collateral).

Section 9 and 10 are used in a default-depending Monte Carlo simulation for the LGD estimation.

V.4.3. The base engine

An automatic report is generated for each year of the base case (in PDF format) containing the following information:

- financial statement
- income statement
- cash flow statement
- economical ratios
    - profitability
    - financial structure
    - solvency
- some measure of risk, typical of PF operations
    - DSCR = Debt Service Cover Ratio (yearly)
    - LLCR = Loan Life Cover Ratio
    - PLCR = Project Life Cover Ratio
    - IRR = Internal Rate of Return

These measures are computed for the entire simulated scenario.

The algorithm that computes these measures is called “base engine” and can be summarized in the following flow chart (Exhibit 5).

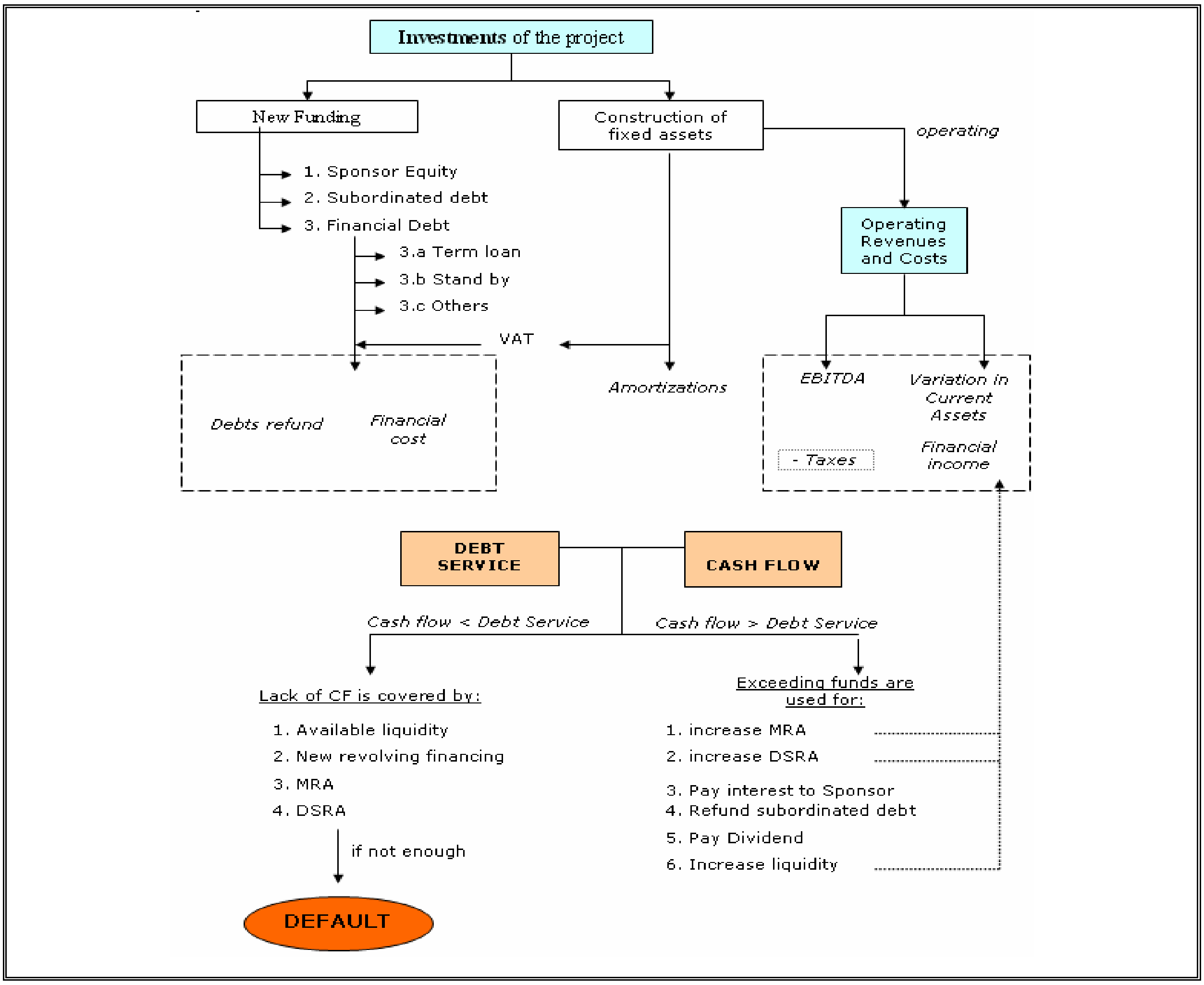


Exhibit 5: Base engine

During the construction period, the “investment scenario” establishes the amount of needed funding (by bank and shareholders).

During the operating periods, financial statements, operating cash flow and debt service are computed using inter alia revenues, expenses, changes in working capital, accounting reserves, taxes, issuing of revolving finance, repayment of financing and dividends.

Operating cash flow and debt service of the year are used to test the default definition.

The aggregated Debt Service actually paid on all the simulated scenarios is used in the PD and LGD estimation.

V.4.4. The Monte Carlo engine

The items of the financial statement are simulated depending on a list of risk drivers or other characteristics of the project.

The risk drivers are the key variables of the MC simulation. They could be:

- *microeconomic* (like quantity sold, price per piece, unit labor cost, …): microeconomic risk drivers describe the detailed characteristics of the project;
- *macroeconomic* (like internal or external inflation, GDP growth, oil price, exchange rate, …): macroeconomic risk drivers describe the macroeconomic scenario (and its evolution) which influences the microeconomic risk drivers or directly the individual items of the financial statement;

We think it is important to share the same risk drivers (and the same stochastic structure/relationship/statistical hypothesis) between different projects operating in a similar macroeconomic environment/scenario in order to achieve a consistent evaluation.

V.4.4.1. Relationship between variables

We define two kinds of relationships between the variables (risk drivers/items of the financial statements) of the project:

- *Formula:* a mathematical formula links the values of an item with some specifics risk drivers. It is possible to define a conditional structure for the formulas (for example to allow a discount policy depending on the quantity sold) and to use different formulas for each year.
- *Correlation:* a (linear) correlation coefficient summarizes the relationship between a risk driver with other risk drivers or items of the financial statement. This coefficient can be referred to as a relationship time T on T, time T on T-n and also time T on T-n of the same variable (autocorrelation), or any combination of this type of relationship.

Both relationships are always *hierarchical*: one variable in time T depends on one or more other variables observed in time T or T-n. For this reason, the correlation coefficient has to be considered as a target correlation that, given the observed values of the risk drivers already generated, influences the new values of the generating item of the project. As a first step, the values of the risk driver (depending on its stochastic distribution/relationship) are generated. *Then* the values of the item depending on the observed values of the linked risk driver are calculated.

Only *known* and *direct* dependences between variables are specified in the model. The other dependences are *implicit* in the defined hierarchical-stochastic structure. In this way we minimize the amount of information needed to generate the MC simulation bypassing the problem of specifying a complete correlation matrix between all the variables involved in the project.

Not all the variables have to be put in relationship with the other variables. Some variables can be considered *independent* from the other variables or even *constant*.

V.4.4.2. Probability distribution

All the risk drivers and characteristics of the project could be associated (if necessary) to a specific stochastic distribution. This distribution could be different for each year and for each variable.

If the variable is *constant* (i.e. if it does not change the value during the MC simulation), obviously the stochastic distribution is not required.

If the variable is *independent* (i.e. if it changes its value during the MC simulation *independently* from the other variables), the probability distribution is mandatory.

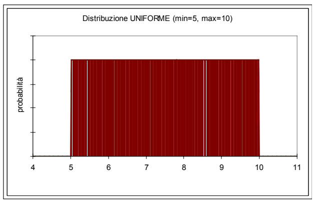
If the variable depends on other variables, one has to distinguish:

- if the relationship is defined by a formula; then it is not necessary to specify a stochastic distribution because the value is fully defined by the formula;
- if the relationship is specified by a *correlation coefficient; then* the distribution is mandatory.

The distributions we consider useful for MC simulation are the following (Exhibit 5 and 6).

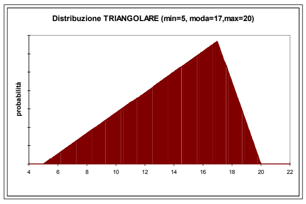
Uniform

Parameters: Min and Max



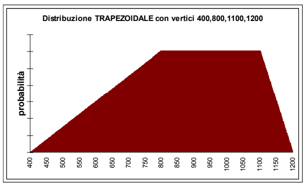
Triangular

Parameters: Min, Mode, Max



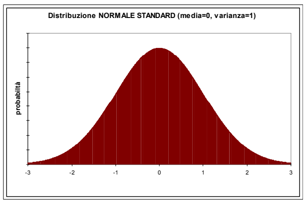
Trapezoidal

Parameters: Min, Mode 1, Model 2, Max



Normal

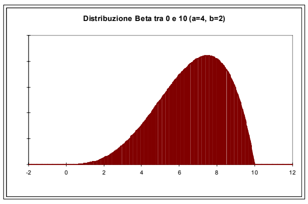
Parameters: Mean and Variance, Min and Max

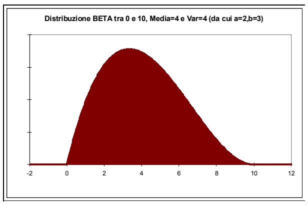


Beta with long left tail **Beta with long right tail** **Beta "U" shaped**

Parameters: Mean and Variance, Min and Max

(an automatic check on mean and variance verifies consistency with target shape)





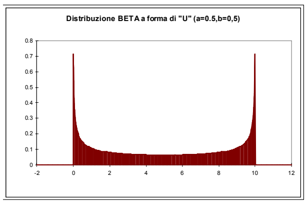
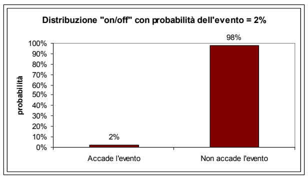


Exhibit 5: Continuous distributions

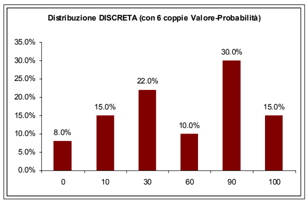
"on-off"

Parameter: probability of success



Discrete value-probability

Parameters: couples value-probability



Discrete uniform interval- probability

Parameters: range of uniform interval and correspondent probability

Exhibit 6: Discrete distributions

We use the following method to generate random values.

Define $X \sim f(\theta)$ a random variable with the **Probability distribution function** $f(\theta)$ known.

Given

- $p = F(x) = \textit{the Cumulative Distribution Function (CDF) of the r.v.}\ \mathrm{X},$
- $x = G(p) = F^{-1}(p)$ *the* inverse *of the* CDF
- $\mathbf{u} = \begin{bmatrix} \mathrm{u}_1 \\ \mathrm{u}_2 \\ \dots \\ \mathrm{u}_n \end{bmatrix}$ an array of uniform random values between 0 and 1

We generate the array

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$$

of random values distributed by F(x), with

$$\mathbf{x} = F^{-1}(\mathbf{u})$$

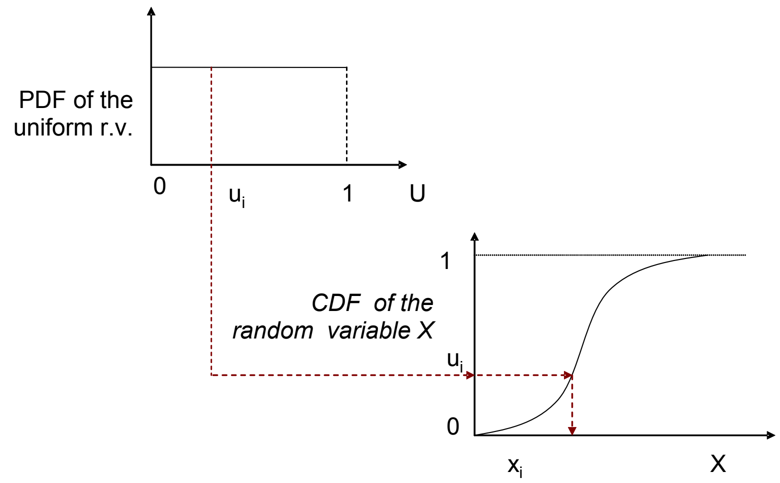


Exhibit 7: Probability distribution function vs. Cumulative Distribution function

This method allows generating **independent** random values given the marginal distribution desired.

In general, we follow this method to generate random values **depending on the desired correlation structure.**

Define:

$$
\begin{aligned}
&y = item\, of\ the\ financial\ statement \\
&x = risk\ driver \\
&\rho = correlation\ coefficient\ between\ y\ and\ x, where\ y\ depends\ on\ x \\
&F_y = Distribution\ Function\ of\ y \\
&F_x = Distribution\ Funcion\ of\ x
\end{aligned}
$$

$$
\begin{aligned}
&y = f(x \mid \rho, F_y, F_x) \\
&f(\bullet) \text{ is the link function defined univocally by } \rho, F_y \text{ and } F_x
\end{aligned}
$$

with *y* being a random variable generated depending on the observed values on x, given the target correlation and the marginal distributions of *x* and *y*.

V.4.4.3. Example

Suppose we would like to generate three variables with the following marginal distributions (constant each year):

V1 ~ Triangular (100,150,200)

V2 ~ Triangular (5, 20, 60)

V3 ~ Normal (50, 100, 0,100)

The hierarchical-relationship is:

#1) V1 (T) is auto correlated with V1(T-1) at 0.80

#2a) V2(T) is auto correlated with V2(T-1) at 0.40

#2b) V2(T) depends on V1(T) with a correlation of -0.40

#3a) V3(T) depends on V1(T) with a correlation of 0.50

#3b) V3(T) depends on V2(T) with a correlation of 0.40

Analysing the relationships[39], it turns out that V1 is the low-level variable because it depends only on itself at time T-1. V2 is the second-level variable because it depends on itself at time T-1 and on V1 at time T. V3 is the third-level variable because it depends on V1 and V2 at time T.

For this reason, the procedure first generates V1(T) depending on V1(T-1), then generates V2(T) depending on V2(T-1) and V1(T) and, finally, generates V3(T) depending on V1(T) and V2(T).

In this example we generate one thousand simulated values of V1, V2 and V3 using the MC engine[40] and check if the desired relationships and marginal distributions are satisfied.

The graphs below summarize the relationships and the marginal distribution observed after the MC simulation.

[39] In a real project, the relationships could be very complicated due the large number of variables involved. For this reason, a specific algorithm to check automatically the rank of the variables was developed. Thus there is no need for the analyst to enter the rank of the variables himself.

[40] If the three variables were risk drivers of a project that lasts for 10 years, the MC engine would simulate n-thousand series of 10 years of V1, V2 and V3. In this example, we generate directly one series of one thousand values.

Exhibit 8 shows the marginal distributions (histograms from top left to bottom right) and the cross correlations (scatter plots in the right upper triangular) between the three variables V1, V2 and V3 on the thousand simulated values. Note that observed marginal distributions have the shape of the desired marginal distributions and the distributions of the points in the scatter plots are in line with the desired cross correlations.

More in the details, the observed cross-correlation matrix is:

	V1	**V2**	**V3**
V1	1.0000	-0.5869	0.4909
V2	-0.5869	1.0000	0.3944
V3	0.4909	0.3944	1.0000

that it's very close to the target correlation that was:

	V1	**V2**	**V3**
V1	1.00	-0.60	0.50
V2	-0.60	1.00	0.40
V3	0.50	0.40	1.00

Also the autocorrelations of V1 (correlation between V1(T) and V1(T-1)) and V2 (correlation between V2(T) and V2(T-1)) are very close to the target correlation.
In fact the observed autocorrelation on V1 is 0.8078 (the target being 0.80) and the observed autocorrelation on V2 is 0.3823 (the target being 0.40).

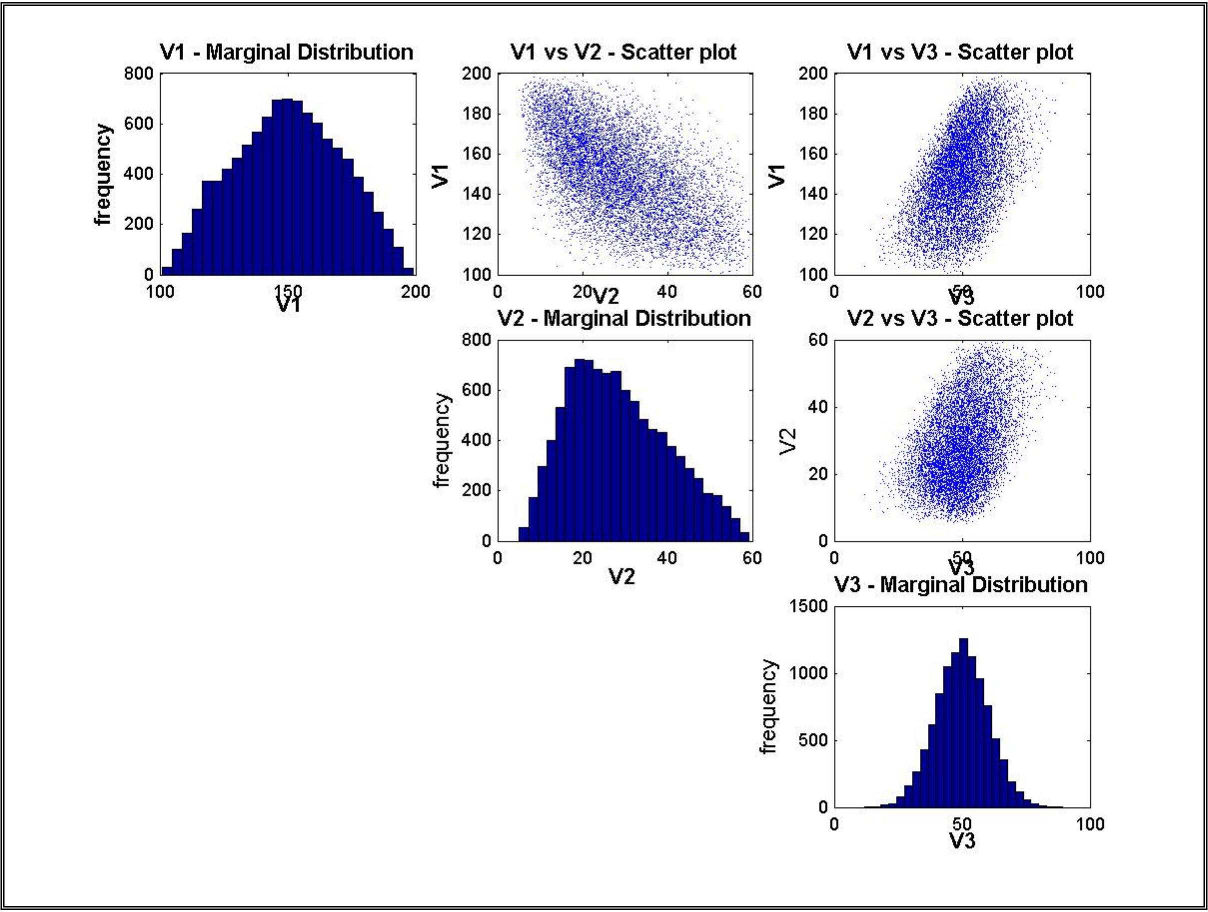


Exhibit 8: The marginal distributions

The graphs below (Exhibit 9) show the different behaviour of V1 and V2 over time.

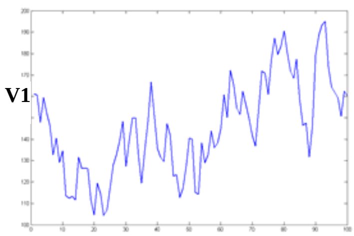
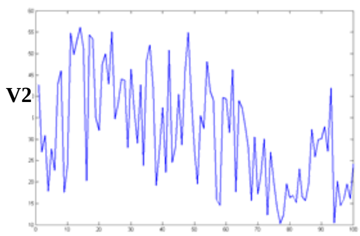
V1 – First 100 simulated observations **Autocorrelation: 0.8**	**V2 – First 100 simulated observations** **Autocorrelation: -0.60**
 **Time**	 **Time**

Exhibit 9: Different behaviour of V1 and V2

This simple example shows how it is possible to generate three variables with different marginal distributions and at the same time with the desired correlation time T on T (cross correlation) and the desired correlation time T on T-1 (autocorrelation). With the algorithm developed in the tool it is possible to manage all the combinations of cross/serial/auto correlations.

V.4.5. Probability of Default

In PF operations, the definition of Probability of Default (PD) is not unique, but depends on the time horizon we consider.

One can distinguish between:

- a short term PD: it is the probability that the project will default the next year.
- a long term PD: it can be considered as an *average annualized* long term PD. It includes the risk evaluation of the whole residual life of the project. We refer to this type of PD as the proper PD of the project

With an iterative method it is possible to compute at time T_0 the list of short term PDs and long term PDs for all following years. These PDs are conditional upon the project being still alive at the beginning of the year, given the information known in time T_0.

We define two different methods to estimate these PDs depending on which kind of information we decide to use and on the purpose of the PD.

V.4.5.1. First method – the frequency approach

This method uses only the "binary" information of the number of defaulted/non defaulted scenarios.
The short term PD of the year T is computed as the number of defaulted scenarios during the year T on the number of survived scenarios at the beginning of the year.
The long term PD is the yearly mean of the short term PDs, considering the life of the financing (that can be shorter than the life of the project).
In a similar way we compute the list of short and long term PDs for the residual life of the project.

V.4.5.2. Second method – the financial approach

This method considers the lender bank's point of view and uses the amounts of simulated future cash flows and debt repayments. The PDs computed with this method are annualized long term PDs.
The PD is computed as a function of the deviation of the internal rate of return (from the bank's point of view) of all simulated scenarios from the internal rate of return of non-defaulted scenarios. This annualized long term PD can be computed for each year over the whole life of the project. The list of future PDs can be considered the term structure of the PDs of the project.
This PD is consistent with the pricing policy of the bank.

V.4.6. Loss Given Default and Expected Loss

The Exposure at Default (EAD) and therefore the LGD depend primary on *when* the project defaults - at the beginning of the operating period the exposure is higher than towards the end of the operating period - and on the *recovery rate*.

We define two alternative recovery processes in case of default:

- **disposal of the assets of the project**: the recovery is predominantly realized through selling the assets on the market (given the values of the assets at the moment of default)
- **disposal of the ongoing project**: the market price is approximated by the present value of future cash flows

In both cases, the actual recovery values depend also on:

- the *duration* of the recovery process (that can be different among the assets)
- the *discount rates* during the recovery process (that can change during the recovery process)
- the structure and amount of *collaterals* (real or personal collateral's values may depend on specific drivers)
- the amount and seniority of the *debt*

to be taken into account in a specific MC simulation.

For these reasons, we cannot estimate the LGD with the main MC simulation of future cash flows, but we need a *default-dependent* MC simulation for each defaulted scenario. The goal is to estimate the average recovery value simulating all the variables involved in the recovery process.

The percentage of LGD is computed using a similar financial approach as the one used for the PD estimation: in this case, the internal rate of return from the bank's point of view is computed considering the effective recovery values on defaulted scenarios.

The Expected Loss (as a percentage) is the LGD times the PD of the project.

V.5. SUMMARY AND CONCLUSION

The Monte Carlo simulation discussed in this paper can be considered an appropriate way to evaluate the risk of project finance operations both for economic capital management purposes and for fulfilling regulatory requirements (Basel II).

The heterogeneity of different projects and the complexity of the term structure of the projects can be managed with a standardized and flexible data entry procedure where all data are classified in sections depending on their role in the business plan. In order to model the complex macro- and micro-economic environment, in which the project operates, and to take into account different scenarios, it is important to manage correlations/relationships time T on T, time T on T-n and autocorrelations time T on T-n simultaneously. Many different stochastic distributions are used to model a large number of different situations. When these relationships/stochastic distributions are not sufficient to manage all relevant risk events, a specific section providing for the risk events defined by Basel II is used.

If the probabilistic model defined by the analyst is based on realistic assumptions, the method potentially provides a robust PD estimation that can be used both for pricing the PF operations and for Basel II requirements.

The LGD, computed with a financial approach, requires a specific Monte Carlo simulation depending on the default event.

In order to have a fine estimation of the PD during the life of the project, it is important to review the assumptions used in the simulation each year. After the first year, for example, the analyst can enter the observed values for this year into the tool. The entered values thereby become “deterministic” or “fixed”,Upon this basis, the analyst can review the assumptions used for the next years. The MC simulation will treat the first year as constant and the following years as stochastic. In this way, the PD estimation will become more precise year by year.

VI. WHAT IS AT STAKE WHEN ESTIMATING THE PROBABILITY OF DEFAULT USING A SCORING FUNCTION ?

M. BARDOS / BANQUE DE FRANCE

VI.1. INTRODUCTION

> *"Statistical inference techniques, if not applied to the real world, will lose their importance and appear to be deductive exercises. Furthermore, it is my belief that in a statistical course emphasis should be given to both mathematical theory of statistics and to application of the theory to practical problems. A detailed discussion on the application of a statistical technique facilitates better understanding of the theory behind the technique." C. Radhakrishna RAO in* Linear Statistical Inference and Its Applications.

Most of the statistical studies on credit scoring focus on score construction. It is more unusual that they link the statistical techniques with a detailed analysis of the users' requirements regarding the properties of these tools. Concerning companies' failure the users are financial analysis experts or bankers in credit risk departments or banking supervisors.

The increasing need for better control of credit risk by banks has led to a stepping-up of research concerning credit scoring. In the context of the Basle II agreement, the International Banking Committee has stressed the importance of forecasting the *expected loss* (EL) and, using extreme quantiles, the *unexpected loss* (UL) for a population of companies, in particular for customers of each commercial bank. In order to do so, it is necessary to estimate the *default probability of each company at a given time horizon* (PD).

The objective of an accurate forecasting gives rise to several needed properties and questions. We stress what is at stake in the construction and the use of credit scores.

The responsibilities of Banque de France in the field of financial stability and the extensive coverage of its data files on companies have paved the way for developing a scoring system able to fullfil most of these needed properties.

This paper presents some credit scoring construction principles, which increase the quality of the tool and the accuracy of default probability. It does not cover the complete debate on model choice, but discusses some arguments regarding this choice and concentrates on the comparison between Fisher linear discriminant analysis (LDA) and logistic regression (LOGIT).

VI.2. THE REQUIRED PROPERTIES OF A CREDIT SCORING SYSTEM

Companies risk assessment quality relies on *the accuracy of estimates of default probability*. If individual diversified probabilities of failure cannot be built in most of the cases[41], however, in the case of large samples, it is possible to determine *homogeneous classes of risk*. Their homogeneity is one of the most important properties to aim at. If this requirement is met, their role is comparable to rating grades.
Such objectives give rise to several other questions regarding the properties of the built up score. These relate to:

- the stability over time of risk classes
- the independence of the risk measurement vis-à-vis the business cycle
- the stability of transition matrices
- the estimated correlation of risks

[41] Beside the fact that estimated probabilities on small samples are not relevant, many models lead straight to risk classes (Cf. CART method , Breiman and alii (1984)).

In order to tackle these issues, the quality of the construction process of the score is determinant. There are several sensitive stages of the process. One of the most important concerns the determination of the learning sample and test samples (the historical period in observation, the forecasting horizon and selection of variables). ((Cf. M. Bardos (2007, 1998), M. Bardos, S. Foulcher, E. Bataille (2004)).

The validation process and the determination of stable risk classes will also have implications on the frequency with which the tool should be updated, and the interaction between the business cycle, forecasting and revision.

These issues have increasingly been the subject of research and it appears that they are highly interdependent. Their impacts on the robustness and the effectiveness of the tool have set out the choices made at the Banque de France.

The aim of this article is to stress the importance of accurate estimation of default probabilities. The means for doing so are developed in the context of two mainly used discriminant techniques, Fisher linear discriminant analysis (LDA) and logistic regression (LOGIT). Appropriateness of the model to companies' accounting data, quality and interpretation capacity of the operating tools will be looked at. We will see that if LOGIT needs parametric assumptions, LDA can be presented in two ways according to the chosen decision rule: first as a distribution free model, second as a parametric model.

VI.3. THE MODELS

Supervised classification, also called dicriminant analysis, covers a large domain of techniques beyond the wellknown Fisher discriminant analysis. Detailed theoretical comparisons have been made in several studies: Hand (2006), Hristache, Delcroix, Patiléa (2004), Baesens and alii (2003), Bardos (2001b), Bardos, Zhu (1998), Thiria and alii (1997), McLachlan (1992), Gnanadesikan and alii (1989). Considerations related to the **suitability for companies' economic data** and **robustness over time** can be found in this litterature. The main models are assessed: Fisher's linear or quadratic discriminant analysis, logistical regression, and some non-parametric methods, such as Disqual[42], Decision Trees, Neural Networks, the Neighborhood method, the Kernel method, Support Vector Machines.

Discrimination models achieve opposition of a priori known groups. The aim in constructing a score may be confined to identify risk signals, and its construction needs to be based on a decision rule and consequently, for some models, on a decision threshold. However, if one wishes to obtain an operational tool, its practical use also requires knowledge of the probability of failure at a given horizon.

Implemented on companies' data, in order to separate failing companies from sound companies, the methods that result in linear combination of ratios are very robust[43] and are easy to analyse.

Indeed, corporate failure is a complex phenomenon for which the actual causal variables are difficult to access. The score functions therefore make use of symptoms such as descriptors of the company's situation before its failure. In other words, it is impossible to accurately define companies' failure processes, contrary to what occurs in other fields of application of discriminant analysis that are closer to physical science, such as shape recognition, where overlearning is easier to master and techniques such as neural networks are successfully applied.

[42] This method builts discriminant function on qualitative data. It has been created by G. Saporta. A recent application on companies'strategic data has been implemented by L. Lelogeais (2003)
[43] For example, quadratic formula or methods that necessitate the determination of thresholds for the explaining variables are generally less stable over time. Cf. M. Bardos (2001b), M. Bardos S. Foulcher E. Bataille (2004), M. Bardos W.H. Zhu (1998b)

It is, therefore, the very traditional linear discriminant analysis (LDA) of Fisher that is used at the Banque de France. Nevertheless this technique, used on ratios built with companies' accounting data, leads to functions very close to those obtained with a LOGIT model. We will explain why and present the reasons of our choice.

VI.3.1. Fisher linear discriminant analysis

Two decision rules can be implemented to estimate the LDA.

We consider D the group of failing companies, N the group of non-failing companies, $X = (X_1, X_2, ..., X_p)$ the vector of the p ratios of the firm *e*, μ^N and μ^D the means of *X* on each group, *T* the total variance-covariance matrix.

The *first decision rule* responds to a geometric criteria, the ***distance comparison:***

$d(X, \mu^D) \leq d(X, \mu^N) \Leftrightarrow$ "e is allocated to the group D"

Using the metric matrix T^{-1}, the rule becomes :

f(X) is negative $\Leftrightarrow$ "e is allocated to the group D"

where:

$f(X) = (\mu^N - \mu^D)' T^{-1} (X - \frac{\mu^N + \mu^D}{2})$ is the discriminant function.

This model does not require parametric assumptions, nevertheless the shape of data have to be rather regularly distributed (Saporta (1990)).

The *second decision rule* is the ***Bayesian rule of minimum expected cost of error***. In the case of multinormality and homoscedasticity of the probability distributions of the descriptors *X* on each group to be discriminated, it leads to the same discriminant function:

$$f(X) = (\mu^N - \mu^D)' T^{-1} (X - \frac{\mu^N + \mu^D}{2}).$$

But in this case the threshold of decision is $\ln\frac{C_{2/1}\pi_1}{C_{1/2}\pi_2}$ instead of 0.

$C_{i/j}$ is the cost of error, i.e. allocating to the group *i* a company which actually belongs to the group *j*, π_i is the *a priori* probability to belong to the group *i*.

One of the crucial advantages of this scoring function **F** is giving the **possibility of interpretation** by the mean of ratio contributions to the value of the score.

It is possible to rewrite the score as follows: $f(X) = \sum_j \alpha_j (X_j - p_j)$

where:

$\alpha = (\mu^N - \mu^D)'T^{-1}$ is the k coefficients vector of the function *f*, $p_j = (\mu_j^N + \mu_j^D)/2$ is the pivot value for the jth ratio. The expression $\alpha_j (X_j - p_j)$ is ***the contribution of ratio j to the score f(X)***.

The contributions can be interpreted in the following way: negative contributions to the score pinpoint weaknesses of the company, whereas positive contributions refer to strengths.

This decomposition of the score value as the sum of the contributions is extremely helpful to the financial analyst who assesses the company. Generally, this expert is not a statistician. He uses many kinds of information, the score value being one of them. The contributions help him deepening the company analysis by identifying its weak and sound points each year, and give him the opportunity to follow the evolution of these points.

VI.3.2. Logistic regression model

The logistic regression estimates the *a posteriori* probability under the following hypothesis:

$$p_i = P(Y_i = 1 / X_i) = \frac{1}{1+e^{-\beta-\alpha X_i}} \qquad 1-p_i = P(Y_i = 0 / X_i) = \frac{1}{1+e^{\beta+\alpha X_i}}$$

where:
$Y_i = 1$ if the company $i \in N$ and
$Y_i = 0$ if $i \in D$.

The likelihood is $\prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i}$ where *n* is the sample size, $n = n_D + n_N$.

The parameters α and β are estimated by the maximum of likelihood method.
p_i is the *a posteriori* probability of being sound.

Consequently, $logit\ p_i = \ln\frac{p_i}{1-p_i} = \beta + \alpha X_i$, and the ***decision rule*** can be:

"The company *i* is classified sound" $\Leftrightarrow$ $p_i > 1-p_i \Leftrightarrow \mathrm{log}it\ p_i > 0 \Leftrightarrow \beta + \alpha X_i > 0$

Another decision rule can be built on $\beta + \alpha X_i > K$. The introduction of the threshold *K* gives the opportunity to calibrate the decision according to the risk objective of the bank, quantified by error costs (Hand (1981)).

VI.3.3. Comparison of LDA and logit models

The much-debated comparison between Fisher LDA and LOGIT deserves further investigation – in terms of their theoretical properties [T. Amemiya, J. Powell (1983), A.W. Lo (1986), B. Efron (1975), Maddala (1999)], interpretability (the great advantage of Fisher's LDA: contributions of variables to the value of the score), sensitivity to the sampling plan of the logistical regression [G. Celeux, J.P. Nakache (1994)], estimation of the probability of failure (either via a theoretical formula or by use of Bayes' theorem on empirical distributions).

In the parametric context, the logistic regression has a wider hypothesis background than the Fisher linear discriminant analysis. As a matter of fact, the linearity of the logit corresponds to linearity of the quotient of log-likelihoods on each group:

$\frac{L_D\ (x)}{L_N\ (x)} = \alpha \mathrm{x} + \beta$ which is ***the fundamental hypothesis (H)*** of the logistic regression model.

In the case of multinormality and homoscedasticity of the probability distributions of the explanatory variables on each group to discriminate, Fisher linear discriminant analysis can be applied. At the same time, it provides the linearity of the quotient of the log likelihoods, the Fisher linear discriminant analysis consequently appearing as a particular case of logistic regression.

When the hypothesis (H) is verified by the data, the *a posteriori* probability is calculated by the same formula in both models.

$$p_i = P(Y_i = 1 / X_i) = \frac{1}{1 + e^{-\beta - \alpha X_i}} \quad \text{and } 1 - p_i = P(Y_i = 0 / X_i) = \frac{1}{1 + e^{\beta + \alpha X_i}}.$$

However, as a matter of fact, this hypothesis is generally not verified. Then, the use of these theoretical formulae are dangerous, since they do not fit to the data, and when the sample is large enough, it is by far better to estimate the probability of failure using the Bayes theorem applied to the empirical distributions of the score on each group as developped in VI.4.

VI.4. PROBABILITY OF FAILURE

The probability of failure provides a measure of the intensity of risk**.** It is much more informative than a decision threshold. Several crucial issues determine the quality of the tool:

- The forecasting horizon **must be consistent with the nature of the data.**
    - o There is by definition a lag of a few months between balance sheet variables and the time at which the company is assessed, and these variables describe what has occurred over the course of the past year; they are consequently better suited to a medium-term forecast than to a short-term one. Balance sheets undoubtedly provide useful and robust information, provided that the assessment and the forecasting horizon are well matched.
        - ▪ With a one-year horizon, it might be thought that it would be possible to create a short-term indicator, which, if it were to be re-estimated sufficiently often, would allow to track the conditions under which companies are operating. But such an indicator would then follow the business cycle closely.
        - ▪ However, this kind of perspective is very difficult to work with as frequent re-estimation in a changing environment is deemed to lead to functions that always lag the current situation.
    - o It has therefore been decided to work on a medium-term horizon with quantitative variables based on balance sheets, and which are submitted to a method of financial analysis whose quality is long established. Given that balance sheet structures are related to the sector to which a company belongs, scores are created according to the major sectors (industry, wholesale trade, retail trade, transport, construction, business services).

- An estimate of **a posteriori failure probabilities well suited to the empirical data** using Bayes' theorem is closely associated with the determination of risk classes. Robustness over time must be ensured for the average probability per risk class. The confidence interval of this average indicates the **accuracy** and provides a measure of what could happen in a worst case scenario.

The probability of failure for a company *e* for which the scores'value *s* belongs to the interval *r* can be written as follows:

$$P(e \in D / s \in r) = \frac{p(s \in r / e \in D)\pi_D}{p(s \in r)} = \frac{p_D \pi_D}{p_D \pi_D + p_N \pi_N}$$

The conditionnal probability on each group is:

$p_D = p(s \in r / e \in D)$ and $p_N = p(s \in r / e \in N)$

The *a priori* probability of faillure is π_D and $\pi_N = 1 - \pi_D$ is the probability of not failing.

Box 1: Bayes` theorem

Estimating probability may be associated with the theoretical model used or may be done on the basis of empirical distributions and Bayes' theorem (see Box 1). The choice between the two will depend on how representative the files are and how close to the assumptions of the model the data are.

Thanks to the importance and representativity of the Banque de France files, empirical distributions can be used in a very efficient way, thus coming closer to reality than theoretical formulas of models whose underlying assumptions are not completely satisfied. Furthermore the empirical-distributions-method allows controlling the accuracy of estimated probabilities and homogeneity of risk classes.

The algorithm of the estimation is the following: the *a posteriori* probability of failure is computed on small intervals of score for each year (in the table: 1999 to 2002); the average μ of these yearly results and the standard error of this average are computed for each class of risk. Some of the neighboured small intervals are progressively gathered with the aim of reducing the standard error and narrowing the confidence intervals of the average a posteriori probability of failure. Table 1 presents the recent results for manufacturing industry with the BDFI2 score computed on samples of about 40,000 french companies each year.

The last column presents the upper limit of the confidence interval, it represents the risk in the worst case scenario at a 99% level; with representative large samples this method gives probabilities very close to the reality, as shown by control studies. In order to ensure PD accuracy, confidence intervals must not overlap one another. This gives an upper limit to the number of classes: too many classes would implicate the overlapping of confidence intervals and a lack of differenciation between the classes.

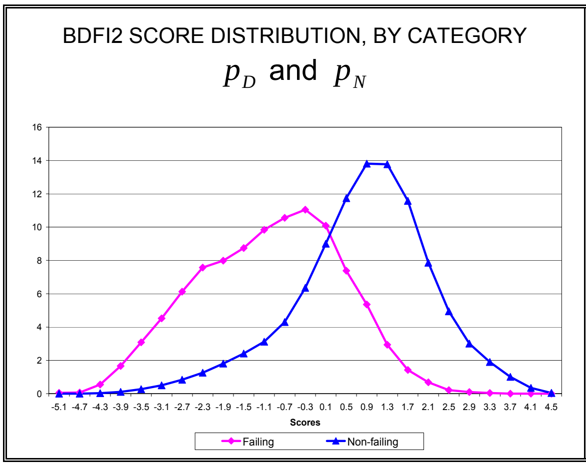


Exhibit 1: BDFI2 score distribution

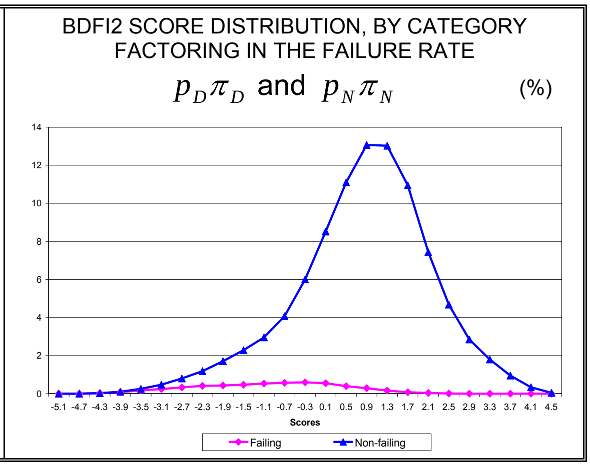


Exhibit 2: BDFI2 score distribution

Score Interval	Risk class	A posteriori probability calculated by year				Average μ	Standard deviationσ	Confidence interval	
		Proba 1999	Proba 2000	Proba 2001	Proba 2002			Inf	sup
BDFI2 <-2.4	1	44.65	46.94	44.21	43.11	**44.73**	1.61	42.31	47.14
-2.4 ≤ BDFI2<-1.8	2	34.23	32.81	33.89	30.91	**32.96**	1.49	30.72	35.2
-1.8 ≤ BDFI2<-0.8	3	22.6	22.9	23.57	23.2	**23.07**	0.42	22.44	23.69
-0.8 ≤ BDFI2<-0,3	4	17	18.11	18.98	17.96	**18.01**	0.81	16.79	19.23
-0,3 ≤ BDFI2<0	5	11.36	12.66	15.37	13.85	**13.31**	1.71	10.74	15.88
0 ≤ BDFI2<0,4	6	8.61	9.52	9.56	9.48	**9.29**	0.45	8.61	9.98
0,4 ≤ BDFI2<1,2	7	3.55	4.38	4.48	4.53	**4.24**	0.46	3.55	**4.93**
1,2 ≤ BDFI2<1,6	8	**1.92**	**1.69**	**2.14**	**2.13**	1.97	**0.21**	**1.65**	**2.29**
1,6 ≤ BDFI2<2,4	9	**0.65**	**0.8**	**0.93**	**1.02**	0.85	**0.16**	**0.6**	**1.09**
2,4 ≤ BDFI2	10	**0.31**	**0.36**	**0.33**	**0.31**	0.33	**0.02**	**0.29**	**0.36**

Source : Banque de France – Fiben November 2006

Table 1: A posteriori probability of failure by score interval

Three years horizon – Manufacturing sector 1999-2002

A prori probability ist the ***failure rate at three years horizon: 7.56%***

Several methods have been tested to estimate the confidence intervals: binomial model, central limit theorem approach, parametric or non parametric Bootstrap (Efron, Tibshirani (1993), Bardos (2001a), Hansen, Schuermann (2006)). Applied to Banque de France data, they lead to rather similar intervals: the tightest are given by the binomial model, the largest by bootstrap method. In table 1, the central limit theorem is applied.

A wide scale of *a posteriori* probabilities is obtained: from 0.33% for the safest companies to 44.73% for the riskiests. The number of risk classes is limited by the overlapping property. It is a guarantee of true differentiation between risk classes.

At-risk classes have *a posteriori* probabilities of failure much higher than the *a priori* probability (classes 1 to 5). The **neutral class**' *a posteriori* probability is close to its *a priori* probability (class 6). **Sound classes** have a *posteriori* probabilities much lower than their *a prior* probability (classes 7 to 10).

The risk classes building process is linked to the estimation of the *a posteriori* failure probability. The aim is to guarantee the homogeneity of PD inside each class by measuring PD accuracy and controlling medium term stability of the classes. The operational use of scores implies the necessity of yearly score quality controls. Besides good allocation rates, the stability of risk classes is tested by re-estimating the PDs and their confidence intervals in order to avoid overlapping (Bardos (2006)). Furthermore, a decrease of discriminating power would imply the re-estimation of the score function itself.

(%)

Risk class									
1	2	3	4	5	6	7	8	9	10
2.6	2.6	6.4	4.3	3.8	7.1	23.7	14.4	21.6	13.5

At-risk classes	Neutral	Sound classes
19.7	7.1	73.2

Table 2: Breakdown of manufacturing industry companies among risk classes (2005)

VI.4.1. A posteriori probability estimation by several methods

It must be stressed that very often the data do not verify the parametric assumptions of the chosen model. This holds for the logistic regression model as well as for the LDA in the case of multinormality and homoscedasticity of the descriptors on each group. The following example shows that this is a serious problem.

We estimate the discriminant function *L* with a logistic regression. In table 3 the *a posteriori* probability of failure is estimated with the Bayes theorem applied to empirical distributions as presented in VI.4 (see Box 1).

Score interval	**- 2,5**	**- 2**	**- 1,5**	**- 1**	**- 0,5**	**- 0**	**0,5**	**1**	**1,5**	**2**	**2,5**	
Failure Probability (%)	67.8	57.4	44.2	30.7	20.5	13.4	7.1	3.4	1.7	1.2	0.8	0.6
Breakdown of firms in the intervals (%)	1.6	0.6	1.3	2.4	5.4	9.8	13.5	16.6	15.1	12.6	8.9	12.2

Source : Banque de France – Centrale des bilans

Table 3: Probability of using empirical distributions of L on each group and the Bayes theorem on each score interval

In table 4 the *a posteriori* probability of failure is estimated with the theoretic formula of the model and the mean of this probability is calculated on the same intervals. The results in table 3 and table 4 are quite different. That is why, in case of large and representative data sets, the first method better suited to the real data should be preferred.

Score interval	**- 2,5**	**- 2**	**- 1,5**	**- 1**	**- 0,5**	**- 0**	**0,5**	**1**	**1,5**	**2**	**2,5**	
Average failure theoretical probability on each interval (%)	96.8	90.5	85.2	77.7	67.9	56.2	43.8	32.1	22.3	14.8	9.5	4.7
Breakdown of firms in the intervals (%)	1.6	0.6	1.3	2.4	5.4	9.8	13.5	16.6	15.1	12.6	8.9	12.2

Source : Banque de France – Centrale des bilans

Table 4: Probability of failure using the logistic model L theoretical formula

$L(X) = \beta + \alpha X$. For a company *i*, the failure probability is given by

$$1 - p_i = P(Y_i = 0 / X_i) = \frac{1}{1 + e^{\beta + \alpha X_i}}$$

These methods, implemented on new large samples, lead to similar findings (Kendaoui (2007)).

VI.4.2. Influence of Sampling

The method of sampling[44] is another important element in logistic regression estimation (Celeux, Nakache (1994)).

We consider the random variable T defined by:

T=1 if the observation is in the sample,

T=0 if not. If *X* is the vector of explanatory variables, we have:

$$P(Y=0/X=x,T=1)=$$
$$\frac{P(Y=0/X=x)P(T=1/X=x,Y=0)}{P(Y=0/X=x)P(T=1/X=x,Y=0)+P(Y=1/X=x)P(T=1/X=x,Y=1)}$$

As the sampling is independent of *X* :

$$P(Y=0/X=x,T=1)=$$
$$\frac{P(Y=0/X=x)P(T=1/Y=0)}{P(Y=0/X=x)P(T=1/Y=0)+P(Y=1/X=x)P(T=1/Y=1)} \text{ (equation (E))}$$

The influence of the sampling scheme on logistic regression is important, as shown below.

We consider two logistic regressions:

L estimated on a sample with an equal random sampling rate on each group D or N;

L' estimated on a sample with different random sampling rates on group D and on group N.

logit $\Pi(X)=\alpha X+\beta$ corresponds to the regression *L* and the *logit* $\Pi^*(X)$ corresponds to the regression *L'* .

It can be shown using equation (E) that:

[44] Of course only representative samples are considered here, otherwise the variables selection should be biased.

$$\textit{logit}\ \Pi^*(X) = \textit{logit}\ \Pi(X) + \ln\frac{n_N}{n_D} + \ln\frac{\pi_D}{\pi_N} = \alpha X + \beta + \ln\frac{n_N}{n_D}\frac{\pi_D}{\pi_N}$$

where:

n_N and n_D are the respective number of companies in the samples of groups N and D.

Consequently $L'(X) = \alpha X + \beta'$ where $\beta' = \beta + \ln\frac{n_N}{n_D}\frac{\pi_D}{\pi_N}$.

This shows that the difference between *L* and *L*' appears only in the constants *β* and *β'*, but not in the coefficients *α*.

In the Tables 4 and 5, the estimations concern the same companies. If the average failure probabilities by interval are the same, except for the queue values, the proportions of companies in each interval are very different. This is the consequence of very different distributions, those of L' (table 5) being translated from those of L (table 4) by $\ln\frac{n_N}{n_D}\frac{\pi_D}{\pi_N}$, and individual failure' probabilities of a company estimated by the two models differ in a significative way.

Score interval **- 2,5** **- 2** **- 1,5** **- 1** **- 0,5** **- 0** **0,5** **1** **1,5** **2** **2,5**

Average failure probability (%)	95.5	90.5	85.2	77.7	67.9	56.2	43.8	32.1	22.3	14.8	9.5	4.8
Breakdown of firms in the intervals (%)	6.7	6.2	10.8	14.6	16.4	14.6	11.7	8.6	5.2	2.8	1.4	1.0

Source : Banque de France – Centrale des bilans

Table 5: Probability of failure by risk class using the logistic model L' taking in account a sampling scheme with unequal random sampling rates

$$1 - p_i = P(Y_i = 0 / X_i) = \frac{1}{1 + e^{\beta' + \alpha X_i}}$$

VI.4.3. Other a posteriori probabilities on subpopulations

The BDFI2 score has been estimated on the whole manufacturing industry population. The failing companies group gathers companies whose failure will occur during the three years following the observed accounts date. It has led to compute a 3 year horizon probability of failure.

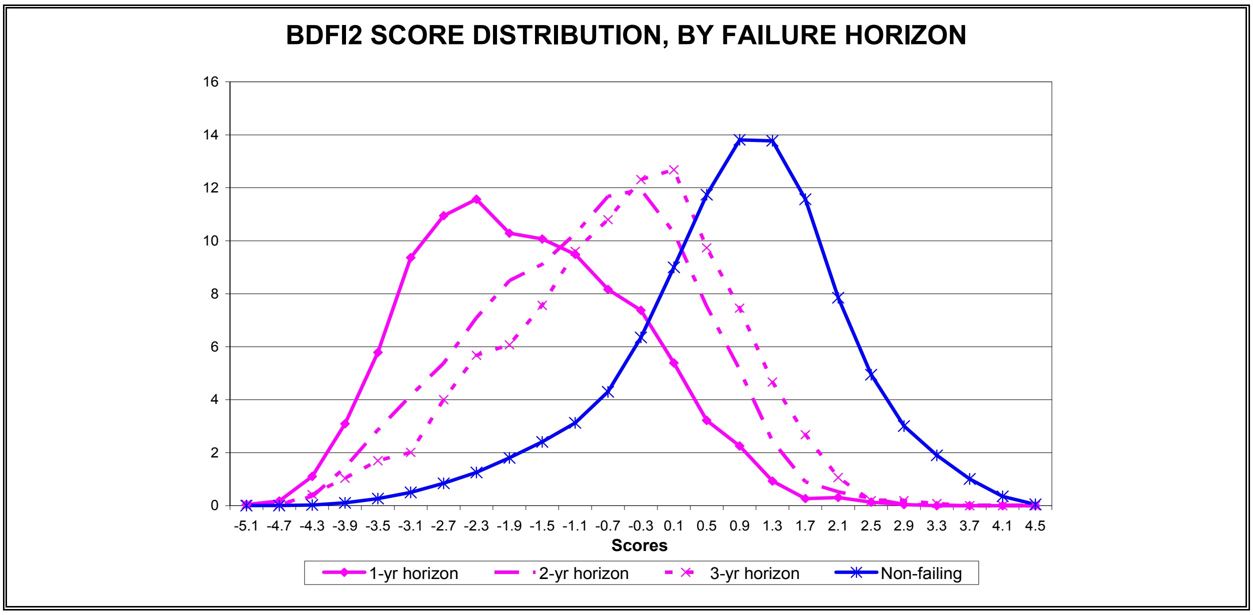
Two important questions arose:

- How to estimate a one year horizon probability of failure?
- Is it possible to take into consideration the size of the company to estimate its probability of failure?

VI.4.3.1. Probability at a one year horizon

In VI.4.1 it has been shown that a direct estimation of a score at a 1 year horizon of failure would be based on a too small sample of failing companies, due to the lack of accounting data when the failure is imminent. Such an estimation would be of poor quality.

Chart 3 leads to a solution of this problem. It shows that the closer to the failure the company is, the more negative its score is. In consequence, using the already estimated BDFI2 score, its distributions according to the horizon of failure give the possibility to oppose the one year horizon failing companies to the other ones.
Then risk classes and associated probabilities at a one year horizon can be estimated (Table 6). In order to obtain non overlapping confidence intervals, only six risk classes have been defined.



Source : Banque de France – Fiben

Exhibit 3: BDFI2 Score distribution by failure horizon

	Risk class	Proba 1999	Proba 2000	Proba 2001	Proba 2002	μ	σ	Inf	sup
BDFI2 <-2,4	1	23,48	26	21,85	22,97	**23,58**	1,75	20,95	26,21
-2,4 ≤ BDF I2 < -1,3	2	11,14	11,37	12,38	11,55	**11,61**	0,54	10,8	12,42
-1,3 ≤ BDF I2 < -0,2	3	5,63	4,93	6,16	5,51	**5,56**	0,51	4,8	6,32
-0,2 ≤ BDF I2 < 0,5	4	1,45	1,9	2,09	1,81	**1,81**	0,27	1,41	2,21
0,5 ≤ BDF I2 < 1,4	5	0,41	0,55	0,54	0,53	**0,51**	0,06	0,41	0,6
1,4 ≤ BDF I2	6	0,1	0,13	0,09	0,09	**0,1**	0,02	0,07	0,13

Neutral Class

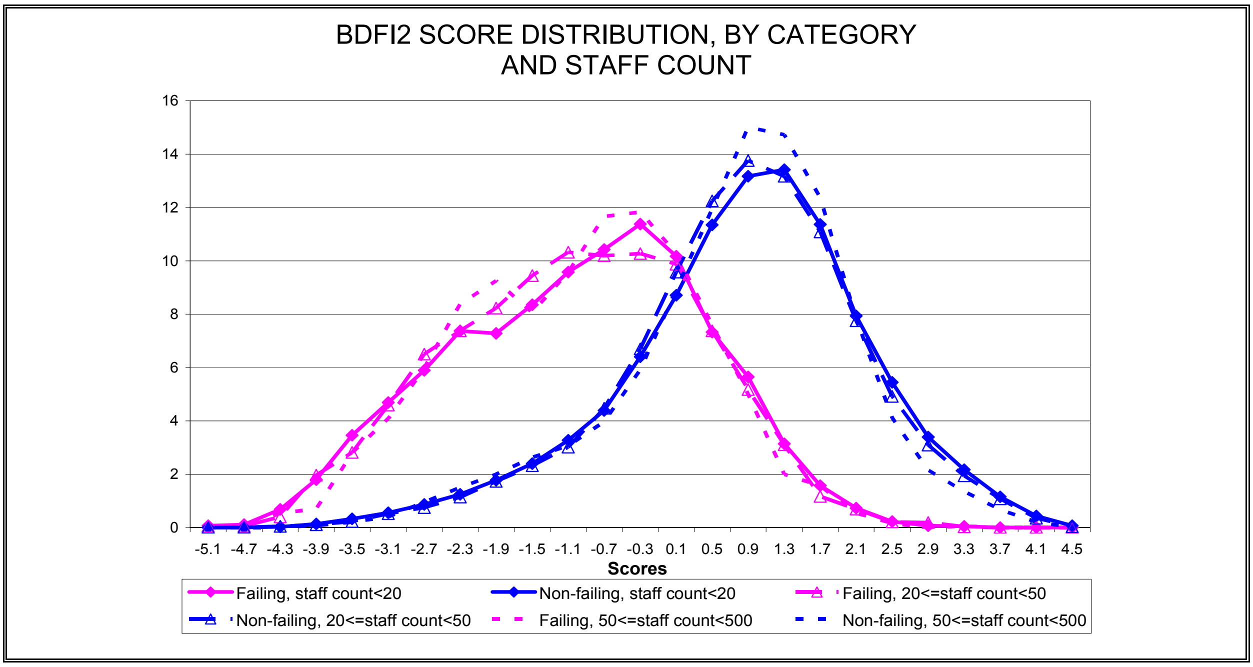
One year *a priori* probability = **2,28%**

Source : Banque de France – Fiben

Table 6: A posteriori probability of failure at a ***one year horizon***

VI.4.3.2. Probability according to the company size

It is well known that company size has an influence on failure risk. The first column of table 7 shows the decrease of failure rate as the size increases. Nevertheless the size has not been included in the list of score explanatory variables. The aim was to get a score function giving a synthetic view of the idiosyncratic financial situation of the company and to avoid to penalize systematically the company by its size. A way of taking into account the size is to study score distributions according to the size and to the category (failing and non failing companies). In chart 4, as the distributions only differ between categories, but not between sizes inside the same category, the score appears not to be size dependent. Then, it is possible to estimate the *a posteriori* probability of failure by size on the same risk classes than for the entire population. The results in table 7 stress the soundness of many small and medium enterprises, and at the same time, identify high risk SMEs.



Source : Banque de France – Fiben

Exhibit 4: BDFI2 Score distribution by category and staff count

(%)

Size	Failure rate [45]	Risk classes				
		Very sound	Sound	Neutral	Risky	Very Risky
Small companies	**9.86**	**0.88**	**4.53**	**11.77**	**24.21**	**44.86**
		[0.62 ; 1.14]	[3.49 ; 5.56]	[8.91; 14.63]	[21.86 ; 26.55]	[41.14; 48.57]
Medium companies	**7.22**	**0.61**	**3.28**	**10.81**	**21.0**	**39.65**
		[0.39; 0.83]	[2.90 ; 3.66]	[9.75 ; 11.88]	[19.26 ; 22.74]	[37.66 ; 41.64]
Large companies	**2.94**	**0.32**	**1.38**	**5.27**	**8.50**	**15.30**
		[0.10 ; 0.54]	[1.15 ; 1.61]	[4.64 ; 5.90]	[6.71; 10.28]	[11.44 ; 19.17]
All	**7.56**	**0.66**	**3.43**	**10.69**	**20.83**	**38.89**
		[0.51 ; 0.81]	[2.96 ; 3.91]	[9.43 ; 11.95]	[19.92 ; 21.74]	[36.90; 40.87]

Source : Banque de France – Fiben

In blue : confidence interval at 95 % level

Table 7: A posteriori probability of failure at a three year horizon and confidence intervals, by size and risk classes – Manufacturing Industry

VI.5. Summary and Conclusion

The scores constructed at the Banque de France cover a wide range of sectors. Applied to a representative sample of firms whose turnover exceeds EUR 0.75 million, they enable studying many issues related to credit risk and debated in the Basle Committee such as risk correlation, risk and business cycle, risk contagion, transition matrixes and companies' trajectories, risk concentration.

The score can help in individual risk diagnosis of companies, but it is not a substitute to a rating. A rating, based on expertise and human judgment, takes into account not only quantitative data, but also many qualitative information which could not enter in a modellisation. As the result of a scoring system can be an input in the risk assessment process, its quality must be as high as possible.

Deepening the study of discriminant analysis techniques it appears that the quality of a score is not necessary dependent on using either a LDA model or a LOGIT model. As a matter of fact it is a wrong idea, but unfortunately widely believed, that Fisher LDA requires multinormality and homoscedasticity of explanatory variables on each group. The geometric decision rule gives good results in many data configurations. The assumptions of logistic regression are semi parametric and should be verified before use.

[45]Failure rate at a three year horizon: it is taken as *a priori* probability of failure.

More determining is the quality of the selection process of relevant explanatory variables. If qualitative data have to be used, LOGIT will be preferred. To study a particular case, comparison between the logistic score value and the weights of each explanatory variable will give an idea of their respective importance. But with a Fisher LDA, the interpretation of each company case can be more precise by the means of score decomposition in ratio' contributions: it gives the company position vis-à-vis the means of failing and non failing groups. It delivers the weak and sound points of the company with accuracy.

In order to study individual risk, as well as risk for an entire population, the knowledge of each company's probability of failure is an essential instrument. The pertinence of the estimation method is at stake in risk forecasting. The proposed method, implementing the Bayes theorem on the empirical distributions of the score function, is generally better suited to the real data than the theoretical formula of the two studied models, LDA and LOGIT. Relevant determination of risk classes limits their number, and by the fact increases their pertinence in term of homogeneity and differentiation which are required properties by the Basle II agreement.

REFERENCES

E. I. ALTMAN/ A. SAUNDERS: Credit Risk Measurement: developments over the last 20 years: in Journal of Banking and Finance. No. 21:1721-1742. 1998.

T. AMEMIYA/ J. POWELI: A comparison of Logit model and normal discriminant analysis when the independent variables are binary,: in Karlin – Amemiya - Goodman, edition Studies in econometrics, time series and multivariate statistics, Academic Press New York . 1983.

T.W. ANDERSON: An introduction to multivariate statistical analysis, Wiley, Chapter 6 “Classification of observations. 1984.

BAESENS/ T VAN GESTEL, S VIAENE, M STEPANOVA, J SUYKENS, J VANTHIENEN: Benchmarking state-of-art classification algorithms for credit scoring: in Journal of the Operational Research Society. 2003.

M. BARDOS: What is at stake in the construction and the use of credit scores ?: in Computationnal Economics. 2007.

M. BARDOS : Banque de France scores : development, applications, and maintenance : in Banque de France Bulletin Diges. No. 151, july. 2006.

M. BARDOS : Recent developments in the Banque de France’s scoring method : in Banque de France Bulletin Digest. No. 93, september. 2001.

M. BARDOS : Analyse discriminante: application au risque et scoring financier, Dunod. 2001.

M. BARDOS: Detecting the risk of company failure: in The Journal of Banking and Finance. No. 22 . 1998.

M. BARDOS/ S. FOULCHER/ E. BATAILLE : Banque de France scores: method, results, practice : in Banque de France. 2004.

M. BARDOS, W.H.ZHU: Comparison of linear discriminant analysis and neural networks, application for the detection of company failures: in Bio-mimetic approaches in management science. Kluwer Academic Publishers. No. 77-100. 1998.

BASEL II agrement documentation on BIS website

L. BREIMAN/ J.H. FREIDMAN/ R.A. OHLSON/ C.J. STONE: Classification and regression trees: Edition Wadsworth International Group. 1984.

G. CELEUX/ J.P. NAKACHE : Analyse discriminante sur variables qualitatives : in Polytechnica. 1994.

B. EFRON: The efficiency of logistic regression compared to normal discriminant analysis: in Journal American Statistical Association. No. 70. 892-898. 1975.

B. EFRON/ R.J. TIBSHIRANI: An introduction to the Bootstrap: in ed. Chapman & Hall. 1993.

R. GNANADESIKAN and alii: Discriminant Analysis and Clustering: in Statistical Science, vol. 4. No. 1:34-69 . 1989a.

C. GOURIEROUX : Économétrie des variables qualitatives : in Économica. 1989b.

D.J. HAND: Classifier Technology and the Illusion of Progress: in Statistical Science, vol. 21. No. 1:1 – 14. 2006.

D.J. HAND: Discrimination and classification: in Wiley series in probability and mathematical statistics. 1981.

G. HANSON/ T. SCHUERMANN: Confidence Intervals for probabilities of default: in Journal of Banking and Finance. No. 30: 2281-2301. 2006.

HRISTACHE/ DELCROIX/ PATILÉA : On semi parametric m-estimation : in Journal of Statistic Planning and Inference. 2004.

L. KENDAOUI: Comparing Logistic Regression and Linear Discriminant Analysis: an application to the detection of firm failure. In: Proceedings of ASMDA. 2007.

G.J. MC LACHLAN: Discriminant Analysis and Statistical Pattern Recognition: in Wiley. 1992.

L. LEBART/ A. MORINEAU/ M. PIRON: Statistique Exploratoire Multidimensionnelle : in Dunod. 1998.

L. LELOGEAIS: Un score sur variables qualitatives pour la détection précoce des défaillances d’entreprises: in Bulletin de la Banque de France. No. 114. 2003.

A.W. LO: Logit versus discriminant analysis: in Journal of Econometrics. No. 31:151-178 . 1986.

G.S. MADDALA: Limited-dependent and qualitative variables in econometrics: in Cambridge University Press. 1999.

G. SAPORTA : Probabilités, analyse des données et statistique : in Technip. 1990.

S. THIRIA/ Y. LECHEVALLIER/ O. GASCUEL/ S. CANU: Statistique et méthodes neuronales. 1997.

Banque de France publications can be downloaded from the web site: www.banque-france.fr/publications

VII. COMPARING THE PERFORMANCE OF EXTERNAL CREDIT ASSESSMENT INSTITUTIONS TO OTHER RATING SOURCES BY MEANS OF PROBABILITIES OF DEFAULT WITH A CONCRETE APPLICATION TO THE EUROSYSTEM CREDIT ASSESSMENT FRAMEWORK

F. COPPENS / G. WINKLER

NATIONAL BANK OF BELGIUM / OESTERREICHISCHE NATIONALBANK

VII.1. INTRODUCTION

External credit assessment institutions, or rating agencies, play an important role in the markets for credit risk. By means of their credit ratings they provide information that is of crucial relevance for investors and the whole financial services industry as well as its regulators. Furthermore, they often act as a benchmark for other rating sources whose performance is compared with the externals´ either by themselves or by their supervisors. For both issues a clear understanding of the meaning of and the information inherent in external credit assessment institutions´ ratings is of the utmost importance.

In this context, the problem of assigning probabilities of default to certain rating grades has found considerable attention by many different market players. It is also essential for institutions like the Eurosystem to clarify what specific rating grades mean in terms of probabilities of default since the Eurosystem like most other central banks also partly relies on external credit institutions´ ratings in its monetary operations. Though it is well known that agencies´ ratings may to some extent also be dependent on the expected severity of loss in the event of default (e.g. Cantor and Falkenstein 2001), a clear relation between probabilities of default and rating grades definitely exists, and it has been the object of investigation of several earlier studies (Cantor and Falkenstein 2001, Blochwitz and Hohl 2001, Tiomo 2004, Jafry and Schuermann 2004 and Christensen et al. 2004).

Once having assigned probabilities of default to external credit assessment institutions´ rating grades, one may wish to compare the performance of external credit assessment institutions´ rating grades to the risk buckets of other rating sources by means of probabilities of default in a second stage. As a matter of fact, external credit assessment institutions are often regarded as a benchmark for all other types of rating providers (e.g. Hui et al. 2005). Benchmarking involves the comparison of a rating source`s probabilities of default with results from alternative sources. Hence, methods are required to test empirically whether the underlying probabilities of default of two rating buckets stemming from different rating sources are equal.

This paper tackles the two issues of assigning probabilities of default to rating grades and comparing the performance of agencies to other rating sources by means of their probabilities of default from the perspective of a (system of) central bank(s) – the Eurosystem – in the special context of the Eurosystem Credit Assessment Framework. Within this framework, the Governing Council of the European Central Bank (ECB) explicitly specified the ECB´s understanding of high credit standards when deciding that a minimum credit quality of "A"[46] should be required for all sorts of collateral to become eligible for Eurosystem monetary policy operations (European Central Bank 2005).

Hence, we aim at deriving a probability of default equivalent to "A" in this paper in a first step. In the empirical application of our methods which we regard as applicable in the general problem of assignment of probabilities of default to any rating grades we will thus restrict ourselves to a demonstrative single case - the "A" grade. Drawing on the earlier works of Blochwitz and Hohl 2001, Tiomo 2004, and Jafry and Schuermann 2004, we analyze historical default rates published by the two rating agencies Standard&Poor´s and Moody´s and derive the ex-ante benchmark for the Eurosystem Credit Assessment Framework. Technically speaking, we use data of Standard&Poor´s and Moody´s publicly available rating histories (Standard&Poor´s 2005, Moody´s 2005) to construct confidence intervals for the level of probability of default to be associated with "A".

[46] Note that we focus on the coarser category "A" throughout this paper. The "A"-grade comprises three sub-categories (named A+, A, and A- in the case of Standard&Poor´s, and A1, A2, and A3 in the case of Moody´s, respectively). However, we do neither differentiate between them nor treat them separately as the credit threshold of the Eurosystem was also defined using the coarser category.

This results in one of the main outcomes of our work, i.e. the statistical deduction of a benchmark of "A" for the Eurosystem Credit Assessment Framework in terms of a probability of default.

The second aim of this paper is to propose simple mechanisms that allow comparisons between the performance of external credit assessment institutions and any other rating source by means of probabilities of default. In the special context of our paper, this mechanism allows for a performance checking within the Eurosystem Credit Assessment Framework. Our work rests on the studies of Blochwitz and Hohl 2001 and Tiomo 2004, and its basic idea is to compare ex-post data (i.e. the realized default rates of pools of obligors considered eligible) to the ex-ante desired default rate expected by the Eurosystem (i.e. the ex-ante benchmark probability of default derived for the “A”-grade). Blochwitz and Hohl 2001 and Tiomo 2004 conducted Monte Carlo simulations to scrutinize which default rates one might expect for certain rating grades in the worst case. The results of their studies were used to define the thresholds for performance monitoring. In this context, it is important to note that the thresholds were uniformly applied to all rating sources without differentiating according to the number of rated obligors (i.e. the size of a certain portfolio). In our approaches, however, we account for the respective number of objects rated by a certain credit assessment source: We apply the technique of hypothesis testing. Moreover, interpreting p-values as frequencies results in a (simple) validation rule that can – in addition to an approach relying on a determination of fixed upper limit - guarantee a long run convergence to the probability of default of the benchmark.

VII.2. MODELLING DEFAULTS USING A BINOMIAL DISTRIBUTION

It is worthwhile mentioning that the probability of default itself is unobservable because the default event is stochastic. The only quantity observable and hence measurable is the empirical default frequency. In search of the meaning of “A” in terms of a one-year probability of default, we will thus have to make use of a theoretical model that rests on certain assumptions about the rules governing default processes. As it is common practice in credit risk modelling, we follow the so-called cohort method (in contrast to the duration approach, ➔ see Lando and Skoedeberg 2002) throughout this paper and furthermore assume that defaults can be modelled using a binomial distribution (Nickel et al. 2000, Blochwitz and Hohl 2001, Tiomo 2003, Jafry and Schuermann 2004). The quality of each model's results in terms of their empirical significance depends on the adequacy of its underlying assumptions. As such, this section briefly discusses the binomial distribution and analyses the impact of a violation of the assumptions underlying the binomial model. It is argued that postulating a binomial model is taking a risk-averse point of view.

VII.2.1. The cohort method and the binomial model[47]

We decide to follow the cohort method, as the major rating agencies document the evolution of their rated entities over time on the basis of so-called static pools (Standard&Poor´s 2005, Moody´s 2005). A static pool consists of N_Y rated entities with the same rating grade at the beginning of a year Y.

In our case, N_Y denotes the number of entities rated “A” at the beginning of year Y. The so-called cohort method simply records number of entities D_Y that have migrated to the default grade by year-end out of the initial N_Y (Nickel et al. 2000, Jafry and Schuermann 2004).

[47] For a more detailed treatment of the binomial distribution see e.g. Rohatgi (1984), and Moore and McCabe (1999).

Following now the opening remarks to chapter 2, let the (observed) number of defaults in each year Y, D_Y, be binomially distributed with a 'success probability' p and a number of events N_Y - in notational form: $D_Y \approx B(N_Y ; p)$ - then it follows that each individual (“A”-rated) entity has the same (one-year) probability of default (PD) 'p' under the assumptions of the binomial distribution. Moreover, the default of one company has no influence on the (one-year) defaulting of the other companies, i.e. the (one-year) default events are independent.

The number of defaults D_Y can take on any value from the set $\{0,1,2,...N_Y\}$. Each value of this set has a probability of occurrence given by the probability density function of the binomial distribution which, under the assumptions of constant p and independent trials, can be shown to be equal to:

$$b(n_Y ; N_Y ; p) = P(D_Y = n_Y) = \binom{N_Y}{n_Y} p^{n_Y} (1-p)^{N_Y - n_Y} \tag{1}$$

The cumulative binomial distribution is given by

$$B(n_Y ; N_Y ; p) = P(D_Y \leq n_Y) = \sum_{i=0}^{n_Y} b(i; N_Y ; p) \tag{1'}$$

The mean and the variance of the binomial distribution are given by

$$\begin{aligned} \mu_{D_Y} &= N_Y p \\ \sigma_{D_Y}{}^2 &= N_Y p(1-p) \end{aligned} \tag{2}$$

As indicated above, a clear distinction has to be made between the 'probability of default' (i.e. the parameter p in formulae (1) and (1') and the 'default frequency'. While the probability of default is the fixed (and unobservable) parameter 'p' of the binomial distribution, the default frequency is the observed number of defaults in a binomial experiment, divided by the number of trials $\left(df_Y = \frac{n_Y}{N_Y} \right)$ (see column 'default frequency' in table 1).

This observed frequency varies from one experiment to another, even when the parameters p and N_Y stay the same. It can take on values from the set

$$df_Y \in \left\{ \frac{0}{N_Y}, \frac{1}{N_Y}, \frac{2}{N_Y}, ..., 1 \right\}.$$

The default frequency may thus be understood as a random variable. Its mean and variance

can be derived from formulae (1) and (1'):

$$\mu_{df_Y} = p$$
$$\sigma_{df_Y}^{\ 2} = \frac{p(1-p)}{N_Y} \qquad (2')$$

The probability density function can be derived from (1) by setting $f_Y = \frac{n_Y}{N_Y}$:

$$P(df_Y = f_Y) = \binom{N_Y}{f_Y N_Y} p^{f_Y N_Y} (1-p)^{(1-f_Y)N_Y} \qquad (3)$$

As $f_Y \in \left\{ \frac{0}{N_Y}, \frac{1}{N_Y}, \frac{2}{N_Y}, ..., 1 \right\}$, this distribution is discrete.

VII.2.2. The binomial assumptions

It is of crucial importance to note that formulae (1) is derived under two assumptions. Firstly, the (one-year) default probability should be the same for every A-rated company. Secondly, the A-rated companies should be independent with respect to the (one-year) default event. The default of one company in one year should hence not influence the default of another A-rated company within the same year.

VII.2.2.1. The constant p

It may be questioned whether the assumption of a homogeneous default probability for all “A”-rated companies is fulfilled in practice (e.g. Blochwitz and Hohl 2001, Tiomo 2004, Hui et al. 2005, Basel Committee on Banking Supervision 2005b). The distribution of defaults would then not be strictly binomial. Based on assumptions about the distribution of PDs within rating grades, Blochwitz and Hohl 2001 and Tiomo 2004 use Monte Carlo simulations to study the impact of heterogenous PDs on confidence intervals for PDs. The impact of a violation of the assumption of a uniform PD across all entities with the same rating may, however, also be modelled using so called 'mixed binomial distribution' of which the lexian distribution is a special case.

The lexian distribution considers a mixture of 'binomial subsets', each subset having its own PD. The PDs can be different between subsets. The mean and variance of the lexian variable x, being the number of defaults among n companies, are given by[48]

$$\begin{aligned} \mu_x &= n\overline{p}, \\ \sigma_x^2 &= n\overline{p}(1-\overline{p}) + n(n-1)var(p) \end{aligned} \qquad (4)$$

where $\overline{p}$ is the average value of all the (distinct) PD's and var(p) is the variance of these PD's.

Consequently, if a mixed binomial variable is treated as a pure binomial variable, its mean would still be correct, whereas the variance would be under-estimated when the 'binomial estimator np(1-p)' is used (see the additional term in (4)). The mean and the variance will be used to construct confidence intervals (see infra). An under-estimated variance will lead to narrower confidence intervals. Within our context, a narrower confidence interval leads to a lower (upper-)limit and this implies a risk averse approach.

[48] See e.g. Johnson N.I (1969)

VII.2.2.2. Independent trials

Several methods for modelling default correlation have been proposed in literature (e.g. Nagpal/Bahar (2001), Servigny/Renault (2002), Blochwitz/When/Hohl (2003), Blochwitz/When/Hohl (2005). They all point to difficulties of measuring correlation.

Apart from the difficulties involved, there are strong arguments suggesting that accounting for correlation does not seem to be necessary for our purposes: Firstly, looking at Standard and Poor´s historical default experiences in table 1, we see that, except for the year 2001, not more than one company defaulted per year – a fact which indicates that correlation cannot be very high. Secondly, not accounting for correlation leads to confidence intervals that are more conservative[49]. This is to be supported from a risk management perspective since the intention of the Traffic Light Approach is to protect the Eurosystem from losses. Empirical Evidence for these arguments is provided by Nickel et al. 2000.

VII.3. THE (BINOMIAL) DISTRIBUTION OF THE BENCHMARK

Modelling default frequencies using the binomial distribution requires an estimate of its parameters p and N_Y.

VII.3.1. The benchmark`s PD

Table 1 shows data on defaults within Standard&Poor's class of A-rated issuers. (The corresponding results for Moody´s are given in Annex 1.) The first column lists the year, the second column shows the number of A-rated issuers for that year. The column 'Default Frequency' is the observed one-year default frequency among these issuers. The last column gives the average default frequency over the 'available years' (e.g. the average over the period 1981-1984 equals 0.05%).

[49] This is, just as in the case of heterogeneous PD's, due to the increased variance when correlation is positive. As an example, consider the case where the static pool can be divided in two subsets of size N_1 and N_2 ($N_1 + N_2 = N$). Within each subset, issuers are independent, but between subsets they are positively correlated. The number of defaults in the whole pool is then a sum of two (correlated) binomials. The total variance is given by $N_1 p(1-p) + N_1 p(1-p) + 2\sigma_{12}$, which is again higher than the 'binomial variance'.

	Number of Issuers	Default Frequency	Average(1981-YYYY)
1981	494	0.00%	0.00%
1982	487	0.21%	0.11%
1983	466	0.00%	0.07%
1984	471	0.00%	0.05%
1985	510	0.00%	0.04%
1986	559	0.18%	0.07%
1987	514	0.00%	0.06%
1988	507	0.00%	0.05%
1989	561	0.00%	0.04%
1990	571	0.00%	0.04%
1991	583	0.00%	0.04%
1992	651	0.00%	0.03%
1993	719	0.00%	0.03%
1994	775	0.13%	0.04%
1995	933	0.00%	0.03%
1996	1027	0.00%	0.03%
1997	1106	0.00%	0.03%
1998	1116	0.00%	0.03%
1999	1131	0.09%	0.03%
2000	1118	0.09%	0.04%
2001	1145	0.17%	0.04%
2002	1176	0.09%	0.04%
2003	1180	0.00%	0.04%
2004	1209	0.00%	0.04%
average 1981-2004		***0.0400%***	
standard deviation 1981-2004		***0.0671%***	

Source: Standard&Poor's, "Annual Global Corporate Default Study: Corporate defaults poised to rise in 2005"

Table 1: One year default frequency within Standard an Poor`s A rated class

The average one-year default frequency over the whole observation period lasting from 1981 to 2004 is 0,04%, the standard deviation is 0,07%.

The maximum likelihood estimator for the parameter p of a binomial distribution is the observed frequency of success. Table 1 thus gives for each year between 1981 and 2004 a maximum likelihood estimate for the probability of default of S&P's class of A-rated companies, thus 24 estimators.

One way to combine the information in these 24 estimates is to apply the Central Limit Theorem to the arithmetic average of the default frequency over the period 1981-2004 which is 0.04% according to table 1. As such, it is possible to construct confidence intervals for the true mean $\mu_{\bar{x}}$ of the population around this arithmetic average. The Central Limit Theorem states that the arithmetic average $\bar{x}$ of n independent random variables x_i, each having mean μ_i and variance σ_i^2, is approximately normally distributed with parameters

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{n} \mu_i}{n} \text{ and } \sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{n} \sigma_i^2}{n^2}$$

(see e.g. DeGroot (1989), and Billingsley (1995)). Applying this theorem to S&P's default frequencies - random variables with $\mu_i = p$ and $\sigma_i^2 = p(1-p)/N_i$ - this implies that the arithmetic average of S&P's default frequencies is approximately normal with mean

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{n} p}{n} = p \text{ and variance } \sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{n} \frac{p(1-p)}{N_i}}{n^2}.$$

Estimating p from S&P data ($\hat{p} = 0.04\%$ for "A" and $\hat{p} = 0.27\%$ for "BBB"), confidence intervals for the mean - the default probability p - can be constructed. These confidence intervals are given in table 2 for S&P's rating grades "A" and "BBB".

Confidence level	Lower	Upper
S&P A		
95.0%	0.01%	0.07%
99.0%	0.00%	0.08%
99.5%	0.00%	0.09%
99.9%	0.00%	0.10%
S&P BBB		
95.0%	0.17%	0.38%
99.0%	0.13%	0.41%
99.5%	0.12%	0.43%
99.9%	0.09%	0.46%

Table 2: Confidence intervals for the $\mu_{\bar{x}}$ of S&P's 'A' compared to 'BBB'

Please note that in applying the Central Limit Theorem it was assumed that $\mu_i = p$, i.e. that the PD remains constant over time. If this is not the case, then the above confidence intervals are intervals around the average PD, i.e.

$$\mu_{\bar{x}} = \frac{\sum_{i=1}^{n} p_i}{n} = \bar{p}\,.$$

This is, however, consistent with the statement of the rating agencies that they rate 'through the cycle'.

The necessary condition for the application of the central limit theorem is the independence of the annual binomial variables. This is hard to verify. Nevertheless, several arguments in favour the above method can be brought forward.

Firstly, a quick analysis of the data in table 1 shows that there are no visible signs of dependence among the default frequencies. Secondly, and probably most convincingly, the data in table 1 confirm the findings on the confidence intervals that are found in table 2. Indeed, the last column in table 1 shows the average over 2, 3, 24 years. As can be seen, these averages fall within the confidence intervals (see table 2), except for a few of them. For these exceptions it can be argued that, firstly, not all values must be within the limits of the confidence intervals (in fact for a 99% confidence interval one exception is allowed every 100 years, for a 95% interval it is even possible to exceed the limits every 20 years) and, secondly, we did not always compute 24-year averages while the Central Limit Theorem was applied to a 24-year average. When random samples of size 23 are drawn from these 24 years of data, the arithmetic average seems to be within the limits given in table 2. The third reason that supports our findings is a theoretical one. In fact, a violation of the independence assumption would change nothing to the findings about the mean $\mu_{\bar{x}}$. However, the variance would no longer be correct as the covariances should be taken into account. Furthermore, dependence among the variables would no longer guarantee a normal distribution.

The sum of dependent and (right) skewed distributions would no longer be symmetric (like the normal distribution) but also skewed to the right. Assuming positive covariances, this would lead to wider confidence intervals. Furthermore, as the resulting distribution will be skewed to the right and as values lower than zero would not be possible, using the normal distribution as an approximation will lead to smaller confidence intervals. As such, a violation of the independence assumption implies a risk-averse result.

We can thus conclude that there is strong evidence to believe that the **probability of default** for the binomial process that models the observed default frequencies of Standard&Poor's A-grade is somewhere between 0.00% and 0.1%.

An additional argument can be brought forward which supports our findings: Firstly, in the definition of the A-grade we are actually also interested in the minimum credit quality that “A-grade” stands for. We want to know the highest value the probability of default can take to be still accepted as equivalent to “A”. Therefore, we could also apply the Central Limit Theorem on the data for Standard&Poor´s BBB. Table 2 shows that there the PD of BBB is probably higher than 0.1%.

VII.3.2. A rule fort he N-year average & Moody`s default frequencies

The PD of a rating source is unobservable. As a consequence, a performance checking mechanism can not be based on the PD alone. In this section, it is shown that the Central Limit theorem could also be used to design a mechanism that is based on an average observed default frequency.

In fact, in section VII.3.1 it was found that 24-year average of S&P's default frequencies is, according to the Central Limit Theorem, normally distributed:

$$\overline{x}^{S\&P} \approx N(\mu_{\overline{x}^{S\&P}};\sigma_{\overline{x}^{S\&P}}) \tag{6}$$

with $\mu_{\overline{x}^{S\&P}}$ and $\sigma_{\overline{x}^{S\&P}}$ estimated at 0.04% and 0.0155% respectively.

In a similar way the average default frequency of any rating source is normally distributed:

$$\bar{x}^{rs} \approx N(\mu_{\bar{x}^{rs}};\sigma_{\bar{x}^{rs}}) \tag{7}$$

The formulae (6) and (7) can be used to test whether the average default frequency of the rating source is at least as good as the average of the benchmark's by testing the statistical hypothesis

$$H_0 : \mu_{\bar{x}^{rs}} < \mu_{\bar{x}^{SaP}} \text{ against } H_0 : \mu_{\bar{x}^{rs}} \geq \mu_{\bar{x}^{SaP}} \tag{8}$$

Although seemingly simple, such a performance checking mechanism has several disadvantages. Firstly, assuming e.g. 24 years of data for the rating source, the null hypothesis can not be rejected if the annual default frequency is 23 times 0.00% and one time 0.96% ($\bar{x}^{rs} = \frac{23 \times 0.00\% + 1 \times 0.96\%}{24} = 0.04\%$, p-value is 50%). In other words, extreme values for the observed default frequencies are allowed (0.96%). Secondly, the performance rule is independent from the static pool size. A default frequency of 0.96% on a sample size of 10000 represents 96 defaults, while it is only 2 defaults for a sample of 200. Thirdly, a 24-year average seems not workable in practice. Other lengths could be used (e.g. a 10 year average), but even then it remains unworkable as 24 (or 10) years of data must be available before the rating source can be backtested. Taking into account these drawbacks, two alternative performance checking mechanisms will be presented in section VII.4.

This rule can, however, very well be used to test whether the average default frequency of S&P and that of Moody's are significantly different. Under the null-hypothesis that

$$H_0 : \mu_{\bar{x}^{S\&P}} = \mu_{\bar{x}^{Moody's}} \tag{9}$$

the difference of the observed averages is normally distributed, i.e. (assuming independence)

$$\overline{x}^{S\&P} - \overline{x}^{Moody's} \approx N(0; \sqrt{\sigma_{\overline{x}^{S\&P}}^{\;2} + \sigma_{\overline{x}^{Moody's}}^{\;2}}) \tag{10}$$

Using an estimate of the variance, the variable $\frac{\overline{x}^{S\&P} - \overline{x}^{Moody's}}{\sqrt{s_{\overline{x}^{S\&P}}^{\;2} + s_{\overline{x}^{Moody's}}^{\;2}}}$ has a t-distribution with 46 degrees of freedom and can be used to check the hypothesis (9) against the alternative hypothesis $H_1 : \mu_{\overline{x}^{S\&P}} \neq \mu_{\overline{x}^{Moody's}}$.

Using the figures from S&P and Moody's, a value of 0.81 is observed for this t-variable. The value has a p-value (2-sided) of 42% so that the hypothesis of equal PD's for Moody's & S&P's “A” grade cannot be rejected. Hence, these findings support the Governing Council´s decision to fix the eligibility threshold at “A” as outlined in chapter 1.2 without further differentiating between Moody´s and Standard&Poor´s.

In formulae (10) S&P and Moody's “A” class were considered independent. Positive correlation will imply a lower t-value.

VII.3.3. The benchmark`s number of trials (N_Y)

Using the Central Limit Theorem, it was found that the probability of default for A rated companies must be somewhere between 0.0% and 0.1%, with an average estimated at 0.04%.

To allow a performance checking the assignment of PDs to rating grades alone is not enough. In fact, as can be seen from S&P data in table 1, the observed default frequencies often exceed 0.1%. This is because the PD and the (observed) default frequencies are different concepts. A performance checking mechanism should, however, be based on 'observable' quantities, i.e. on the observed default frequencies of the rating source.

In order to construct such a mechanism, it is assumed that the benchmark is a binomial distribution. The mean of this distribution is estimated at 0.04% (but it might be as high as 0.1%). The other binomial parameter is the number of trials N. For the benchmark, N is taken to be the average size of S&P's static pool or N = 792 (see table 1).

This choice seems to be somewhat arbitrary because the average size over the period 2000-2004 is higher (i.e. 1166), but so is the average observed default frequency over that period (0.07%). Should the binomial parameters be based on this period, then the mean and the variance of this binomial benchmark would be higher, confidence limits would then also be higher.

VII.3.3. The benchmark`s number of trials (N_Y)

To summarise the two preceding sections we derived that:

- The PD of the benchmark is at most 0.1%.
- The benchmark can be modelled as a binomial random variable with parameters p = 0.04% and N_Y = 792 issuers.

These findings can now be used to find limits for the observed default frequencies and thus to construct a performance checking mechanism.

VII.4. TWO POSSIBLE PERFORMANCE CHECKING MECHANISMS

The previous section defined the benchmark of “A” in terms of a PD. The PD being unobservable, this section goes one step further and presents two alternative performance checking mechanisms based on the (observable) default frequency:

- A performance checking mechanism that uses a fixed, absolute upper limit for the probability of default as a benchmark.
- Taking the volatility of the “A”-grade into account (see table 1), one could measure its performance by a binomial distribution. The performance checking mechanism would then make use of a stochastic benchmark.

VII.4.1. A performance checking mechanism relying on a fixed benchmark

Using the Central Limit Theorem, we found that the probability of default of the benchmark (p^{bm}) is at most 0.1%. A rating source is thus in line with the benchmark if its default probability is at most 0.1%. Assuming that the rating source's default events are distributed binomial with parameters PD^{rs} and N_Y^{rs}, this means that a performance checking mechanism should check whether

$$PD^{rs} \leq 0.1\% \tag{11}$$

Since PD^{rs} is an unobservable variable, (11) can not be used for validation purposes. A quantity that can be observed is the number of defaults in a rating source's static pool within one particular year i.e. $df *_Y^{rs}$. (the ' indicates that it is the observed value).

The performance checking mechanism should thus check whether observing a value $df\,{*}_Y^{rs}$ for a random variable that is (approximately) normally distributed $df_Y^{rs} \approx N\left(PD^{rs}; \sqrt{\frac{PD^{rs}(1-PD^{rs})}{N_Y^{rs}}}\right)$ is consistent with (11).

This can be done using a statistical hypothesis test. One has to test the null hypothesis that

$$H_0 : p^{rs} \leq 0.1\%$$

against the alternative hypothesis

$$H_1 : p^{rs} > 0.1\%$$

Assuming that H_0 is true, one can compute the probability of observing the value $df\,{*}_Y^{rs}$. This is the so-called p-value of the hypothesis test. This p-value is given by

$$1 - \Phi\left(\frac{df\,{*}^{rs} - 0.1\%}{\sqrt{\frac{0.1\%(1-0.1\%)}{N_y^{rs}}}}\right) \qquad (12)$$

where Φ is the cumulative probability function for the standard normal distribution. Table 3 gives an example for an eligible set of $N_y^{rs} = 10000$ companies.

n	df'(rs)	p-value	probability
0	0.00%		
1	0.01%	99.78%	0.35%
2	0.02%	99.43%	0.77%
3	0.03%	98.66%	1.54%
4	0.04%	97.12%	2.80%
5	0.05%	94.32%	4.60%
6	0.06%	89.72%	6.84%
7	0.07%	82.87%	9.22%
8	0.08%	73.66%	11.24%
9	0.09%	62.41%	12.41%
10	0.10%	50.00%	12.41%
11	0.11%	37.59%	11.24%
12	0.12%	26.34%	9.22%
13	0.13%	17.13%	6.84%
14	0.14%	10.28%	4.60%
15	0.15%	5.68%	2.80%
16	0.16%	2.88%	1.54%
17	0.17%	1.34%	0.77%
18	0.18%	0.57%	0.35%
19	0.19%	0.22%	0.14%
20	0.20%	0.08%	0.05%
21	0.21%	0.03%	0.02%
22	0.22%	0.01%	0.01%
23	0.23%	0.00%	0.00%
24	0.24%	0.00%	0.00%
25	0.25%	0.00%	0.00%

Table 3: Test of credit quality assessment source against the limit of 0.1% for a sample size of 10000

The first column of the table 3 gives different possibilities for the number of defaults observed in year 'Y'. The observed default frequency is derived from it by dividing the number of defaults by the sample size. This is shown in the second column of the table. The third column shows the p-values computed using formulae (12). So the p-value for observing at least 15 defaults out of 10000, assuming that H_0 is true, equals 5.68%. In the same way, it is derived from the table that if H_0 is true, then the probability of observing at least 18 defaults on 10000 is 0.57%, or "almost impossible". This can be stated in another way: if we observe at least 18 defaults then H_0 can almost impossibly be true.

Fixing a confidence level (i.e. a minimum p-value, e.g. 1%), table 3 can be used as a performance checking mechanism:

> If the size of the static pool is 10000, then the rating source is in line with the benchmark only if at most 17 defaults are observed (confidence level of 1%), i.e.
>
> $$df*^{rs}_{Y} \leq 0{,}17\% \,.$$

This technique has a disadvantage of first having to decide on a confidence level. Moreover, fixing only one limit (0.17% in the case above) does not guarantee a convergence to an average of 0.1% or below.

A p-value being a probability, it can be interpreted in terms of 'number of occurrences'. From table 3 it can be seen that, if the null-hypothesis is true, the observed default frequency must be lower than 0.12% in 80% of the cases. In other words, only once every 5 year can a value above 0.12% be observed, else the rating source is not in line with the benchmark.

Transformed into a simplified second performance checking rule, this could mean:

> A rating source, with a static pool of size 10000, is in line with the benchmark if at most once every five year a default frequency above 0.12% is observed. One should never accept a default frequency above 0.17%.

The intervals for other sizes of the static pool are shown in table 4.

	All time	Once in 5y	Never	Average DF
500	*0.00%-0.00%*	*0.20%-0.40%*	*>0.40%*	*0.06%*
1000	*0.00%-0.10%*	*0.20%-0.40%*	*>0.40%*	*0.10%*
2000	*0.00%-0.10%*	*0.15%-0.25%*	*>0.25%*	*0.08%*
3000	*0.00%-0.10%*	*0.13%-0.23%*	*>0.23%*	*0.08%*
4000	*0.00%-0.13%*	*0.15%-0.25%*	*>0.25%*	*0.09%*
5000	*0.00%-0.12%*	*0.14%-0.20%*	*>0.20%*	*0.08%*
6000	*0.00%-0.12%*	*0.13%-0.20%*	*>0.20%*	*0.08%*
7000	*0.00%-0.11%*	*0.13%-0.19%*	*>0.19%*	*0.08%*
8000	*0.00%-0.11%*	*0.13%-0.19%*	*>0.19%*	*0.08%*
9000	*0.00%-0.11%*	*0.12%-0.18%*	*>0.18%*	*0.07%*
10000	*0.00%-0.11%*	*0.12%-0.17%*	*>0.17%*	*0.07%*
50000	*0.00%-0.112%*	*0.114%-0.134%*	*>0.134%*	*0.07%*

Table 4: Performance checking rule based on a fixed benchmark for different static pool sizes

The increase in values in the column 'once in five years' is due to the discrete nature of the variable; for a size of 500, for instance, the value above 0.0% is 0.2%.

The column 'average DF' is an estimated average using 4 out of 5 occurrences at the midpoint of the first interval and 1 out of 5 occurrences at the midpoint of the second. These averages are clearly below the benchmark limit of 0.1%.

VII.4.2. A performance checking mechanism based on a stochastic benchmark

In the preceding section, a performance checking mechanism using a fixed upper limit for the benchmark was derived. That fixed upper limit followed from the Central Limit Theorem and was found to be 0.1%.

Looking at table 1, it seems clear, however, that the benchmark is also stochastic. In this section we develop an alternative checking mechanism, based on a stochastic benchmark. In fact, it was derived in section VII.3 that the benchmark can be seen as an (approximately)

$$df^{bm} \approx N\left(PD^{bm};\sqrt{\frac{PD^{bm}(1-PD^{bm})}{N^{bm}}}\right) \tag{13}$$

where PD^{bm} was estimated at 0.04% and N^{bm} was estimated at 792.

On the other hand, the rating source's default frequency is distributed as

$$df_Y^{rs} \approx N\left(PD^{rs};\sqrt{\frac{PD^{rs}(1-PD^{rs})}{N_Y^{rs}}}\right) \tag{14}$$

Assuming a stochastic benchmark, there is no longer an upper limit for the PD of the rating source. The condition on which to base the performance checking mechanism should be 'the rating source should do at least as good as the benchmark' or, in terms of a PD, this means that the rating source's PD should be lower than or equal to the benchmark's. The hypothesis to be tested is thus:

$$H_0 : PD^{rs} \leq PD^{bm} \text{ against } H_1 : PD^{rs} > PD^{bm}$$

where PD^{bm} was estimated at 0.04% and N^{bm} was estimated at 792.

The test is completely different from the one in the preceding section. Indeed, we can <u>not</u> replace PD^{bm} by 0.04% because this is only an estimate of the benchmark's PD. The true PD of the benchmark is unknown.

The difference of two normal distributed variables also has a normal distribution, thus, assuming that both are independent[50]:

$$df_Y^{rs} - df^{bm} \approx N\left(PD^{rs} - PD^{bm};\sqrt{\frac{PD^{rs}(1-PD^{rs})}{N_Y^{rs}} + \frac{PD^{bm}(1-PD^{bm})}{N^{bm}}}\right) \tag{15}$$

[50] If the rating source's eligible class and the benchmark are dependant then the variance of the combined normal distribution should include the covariance term.

PD^{rs} and PD^{bm} are unknown, but if the null hypothesis is true, then their difference should be $PD^{rs} - PD^{bm} \leq 0$. An estimate for the combined variance

$$\frac{PD^{rs}(1-PD^{rs})}{N_Y^{rs}} + \frac{PD^{bm}(1-PD^{bm})}{N^{bm}}$$

is needed. A standard hypothesis test, testing the equality of two proportions, would use a 'pooled variance' as estimator. This pooled variance itself is derived from a 'pooled proportion' estimator (see e.g. Moore and McCabe (1999), and Cantor and Falkenstein (2001)). The reasoning is that - as we test the hypothesis of equal proportions - all observations can be pooled so that there is a total of $N_Y^{rs} + N^{bm}$ observations among which there are $N^{bm}.df'^{bm} + N_Y^{rs}.df'^{rs}_Y$. The pooled proportion is thus

$$df'^{pooled} = \frac{792 \times 0.04\% + N_Y^{rs}.df*_Y^{rs}}{792 + N_Y^{rs}} \tag{16}$$

and the two variances are then

$$\sigma_{bm}^2 = \frac{df*^{pooled}.(1-df*^{pooled})}{792}, \sigma_{rs}^2 = \frac{df*^{pooled}.(1-df*^{pooled})}{N_Y^{rs}} \tag{17}$$

then (10) becomes

$$df_Y^{rs} - df^{bm} \approx N\left(0; \sqrt{df*^{pooled}(1-df*^{pooled})\left(\frac{1}{N_Y^{rs}} + \frac{1}{792}\right)}\right) \tag{18}$$

However, as we have an estimate of the benchmark that is based on 24 past observations, we decide not to touch the variance estimate of the benchmark. So, again taking a risk-averse position, we leave the variance of the benchmark untouched and the hypothesis test uses the distribution given in (19).

$$df_Y^{rs} - df^{bm} \approx N\left(0; \sqrt{\frac{df^{*pooled}(1 - df^{*pooled})}{N_Y^{rs}} + 0.07\%^2}\right) \quad (19)$$

Using the observed default frequency (df'^{rs}) as an estimate for the rating source and using the estimated benchmark values, the p-values of the test are given by:

$$1 - \Phi\left(\frac{df^{*rs} - 0{,}04\%}{\sqrt{0.07\%^2 + \frac{df^{*pooled}(1 - df^{*pooled})}{N_Y^{rs}}}}\right) \quad (20)$$

The results for an estimated benchmark PD of 0.04% and a static pool of 10000 companies are shown in table 5.

	mean	**stand dev**	**df'(bm)**
BM	*0.0400%*	*0.0711%*	*0.0400%*
Size	10000		mean
n	***df'(rs)***	***p-value***	***probability***
0	0.0000%	71.27%	5.10%
1	0.0100%	66.17%	5.31%
2	0.0200%	60.86%	5.43%
3	0.0300%	55.43%	5.43%
4	0.0400%	50.00%	5.34%
5	0.0500%	44.66%	5.16%
6	0.0600%	39.49%	4.91%
7	0.0700%	34.59%	4.59%
8	0.0800%	30.00%	4.23%
9	0.0900%	25.77%	3.84%
10	0.1000%	21.94%	3.44%
11	0.1100%	18.50%	3.04%
12	0.1200%	15.46%	2.65%
13	0.1300%	12.81%	2.29%
14	0.1400%	10.52%	1.95%
15	0.1500%	8.57%	1.65%
16	0.1600%	6.92%	1.37%
17	0.1700%	5.55%	1.14%
18	0.1800%	4.41%	0.93%
19	0.1900%	3.48%	0.75%
20	0.2000%	2.73%	0.61%
21	0.2100%	2.12%	0.48%
22	0.2200%	1.64%	0.38%
23	0.2300%	1.26%	0.30%
24	0.2400%	0.96%	0.23%
25	0.2500%	0.73%	0.18%

Table 5: Test of credit quality assessment source against the limit of stochastic benchmark for a sample size of 10000, using df(bm) = 0.04%*

Analogous to the reasoning in VII.4.1, this table can be used for a performance checking mechanism.

A rating source, with a static pool of size 10000, is in line with the benchmark if at most once every five year a default frequency above 0.1% is observed. One should never observe a default frequency above 0.23%.

The intervals for other sizes of the static pool are shown in table 6. The average default frequency seems to be lower than 0.1% for all sizes. As argued in 0, a higher average than 0.04% could be justified. Table 6 shows the results when an estimate of 0.07% is used for the benchmarks PD.

		p(bm)=0.04%			**p(bm)=0.07%**	
	All time	**Once in 5y**	**Average DF**	**All time**	**Once in 5y**	**Average DF**
500	*0%-0%*	*0.2%-0.6%*	*0.08%*	*0%-0%*	*0.2%-0.8%*	*0.10%*
1000	*0%-0%*	*0.2%-0.5%*	*0.07%*	*0%-0.1%*	*0.2%-0.5%*	*0.11%*
2000	*0%-0.1%*	*0.15%-0.35%*	*0.09%*	*0%-0.1%*	*0.2%-0.45%*	*0.11%*
3000	*0%-0.1%*	*0.13%-0.3%*	*0.08%*	*0%-0.1%*	*0.13%-0.37%*	*0.09%*
4000	*0%-0.1%*	*0.125%-0.275%*	*0.08%*	*0%-0.125%*	*0.15%-0.37%*	*0.10%*
5000	*0%-0.1%*	*0.12%-0.28%*	*0.08%*	*0%-0.16%*	*0.18%-0.34%*	*0.12%*
6000	*0%-0.1%*	*0.12%-0.25%*	*0.08%*	*0%-0.15%*	*0.16%-0.35%*	*0.11%*
7000	*0%-0.1%*	*0.11%-0.24%*	*0.08%*	*0%-0.15%*	*0.17%-0.34%*	*0.11%*
8000	*0%-0.1%*	*0.11%-0.237%*	*0.07%*	*0%-0.15%*	*0.16%-0.32%*	*0.11%*
9000	*0%-0.1%*	*0.11%-0.23%*	*0.07%*	*0%-0.15%*	*0.16%-0.32%*	*0.11%*
10000	*0%-0.1%*	*0.11%-0.23%*	*0.07%*	*0%-0.15%*	*0.16%-0.32%*	*0.11%*
50000	*0%-0.1%*	*0.11%-0.21%*	*0.07%*	*0%-0.15%*	*0.152%-0.29%*	*0.10%*

Table 6: Performance checking rule based on a stochastic benchmark for different static pool sizes

VII.5. SUMMARY AND CONCLUSION

Using the binomial distribution as a model for default events and arguing that a violation of its underlying assumptions implies a risk averse point of view, this paper has shown that the probability of default of an A rated issuers is at most 0.1%.

The PD being an unobservable quantity, three alternative performance checking mechanisms based on the (observable) default frequency (i.e. the number of defaults in the pool divided by the size of the pool) were proposed. The first rule outlined in section 3.2 defines limits on the average default frequency over N years. From a risk management point of view, however, it might leave too much flexibility. Moreover, it is expected to be more difficult to implement in practice as long time series of historical default experiences are required. The next mechanism proposed defined a fixed upper limit on the PD. The limit was derived from the Central Limit Theorem. The resulting performance checking rule was given in table 4. It yields a long run average default frequency that seems below the fixed benchmark PD of 0.1%. The final and preferred option was based on a stochastic benchmark. The benchmark is again derived applying the Central Limit Theorem. The results are shown in table 6. These results yield a long run average default frequency that is close to the 0.1%.

REFERENCES

A. AGRESTI/ B. COULL: Approximate is better than “exact” for interval estimation of binomial proportions: in The American Statistician. May 1998.

A. AGRESTI/ B. CAFFO: Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures: in The American Statistician. November 2000.

Basel Committee on Banking Supervision: International Convergence of Capital Measurement and Capital Standards. A Revised Framework. 2005a.

Basel Committee on Banking Supervision. Studies on the validation of internal rating systems (revised): A Working Paper. Bank for International Settlements. 2005b.

P. BILLINGSLEY: Probability and Measure. No. 3. John Wiley & Sons. 1995.

S. BLOCHWITZ/ S. HOHL: The worst case or what default rates do we have to expect from the rating agencies?: A Working Paper. Deutsche Bundesbank. Frankfurt. 2001.

S. BLOCHWIZ/ C.S. WHEN/ S. HOHL: Reconsidering Ratings: A Working Paper. Deutsche Bundesbank. Frankfurt. 2003.

L.D. BROWN/ T.T. CAI/ A. DASGUPTA: Interval Estimation for a Binomial Proportion: in Statistical Science, Vol. 16. No. 2. 2001.

T. CAI: One-sided confidence intervals for discrete distributions: in J. Statistical Planning and Inference Vol. 131. 2005.

R. CANTOR: Moody´s Investors Service´s response to the consultative paper issued by the Basel Committee on Bank Supervision “A new capital adequacy framework”: in Journal of Banking and Finance. Vol. 25. 2001.

R. CANTOR/ E. FALKENSTEIN: Testing for Rating Consistency in Annual Default Rates: in Journal of fixed income. 2001.

R. CANTOR/ C. MANN: Are corporate bond ratings procyclical: in Moody´s Special Comment (October). 2003.

R. CANTOR/ F. PACKER/ K. COLE: Split ratings and the pricing of credit risk: in Journal of Fixed Income. December 1997.

J. CHRISTENSEN/ E. HANSEN/ D. LANDO: Confidence Sets for Continuous-Time Rating Transition Probabilities: in Journal of Banking and Finance. Vol. 28. 2004.

M. CROUHY/ D. GALAI/ R. MARK: Prototype risk rating system: in Journal of Banking and Finance. Vol. 25. 2001.

M.H. DEGROOT: Probability and statistics. Second edition. Addison-Wesley Publishing Company. 1989.

E. ELTON/ J. MARTIN/ J. GRUBER/ D. AGRAWAL/ C. MANN: C. Factors affecting the valuation of corporate bonds: in: Cantor, R. (Ed.), Recent Research on Credit Ratings (special issue): in Journal of Banking and Finance Vol. 28. 2004.

European Central Bank: The implementation of monetary policy in the Euro area. General documentation on Eurosystem monetary policy instruments and procedures. 2004.

European Central Bank: Introductory statement to the press conference given by Jean-Claude Trichet, President of the ECB. Frankfurt. November 2004.

J. FONS: Understanding Moody`s corporate bond ratings and rating process. Moody´s Special Comment. May 2002.

Governing Council of the European Central Bank: Eurosystem collateral framework: Inclusion of nonmarketable assets in the single list. Press Release. July 22 2005.

HEITFIELD: Studies on the Validation of Internal Rating Systems. Basel Committee on Banking Supervision: A Working Paper 14. 2005.

C.-H. HUI/ T.-C. WONG/ C.-F. LO/ M.-X. HUANG: Benchmarking Model of Default Probabilities of Listed Companies: in Journal of Fixed Income. September 2005.

J. HULL/ M. PREDESCU/ A. WHITE: A. The relationship between credit default swap spreads, bond yields, and credit rating announcements: in: Cantor, R. (Ed.), Recent Research on Credit Ratings (special issue): In Journal of Banking and Finance Vol 28. 2004.

Y. JAFRY/ T. SCHUERMANN: Measurement, Estimation and Comparison of Credit Migration Matrices: in Journal of Banking & Finance Vol. 2004.

N.L. JOHNSON. Discrete distributions. Houghton Mifflin Company. Boston. 1969.

J.P. KRAHNEN/ M. Weber. Generally accepted rating principles: a primer: in Journal of Banking and Finance. Vol. 25. 2001.

D. LANDO/ T.M. SKØDEBERG: Analyzing rating transitions and rating drift with continuous observations: in Journal of Banking and Finance. Vol. 26. 2002.

Moody´s. Moody´s Default Report 2005: An Annual Default Study. 2005.

D.S. MOORE/ G.P. MCCABE: Introduction to the practice of statistics. W.H. Freemand and Company. New York. 1999.

R.G. NEWCOMBE: Two-sided confidence intervals for the single proportion: comparison of seven methods: Statistics in Medicine Vol. 17. 1998.

P. NICKEL/ W. PERRAUDIN/ S. VAROTTO: Stability of Rating Transitions: in Journal of Banking and Finance. Vol. 24. 2000.

J. REICZIGEL: Confidence intervals for the binomial parameter: some new considerations: A Working Paper. Szent István University. Budapest 2004.

V.K. ROHATGI: Statistical Inference. Wiley Series in Probability and mathematical statistics. John Wiley & Sons. 1984.

A. SPANOS: Statistical foundations of econometric modelling. Cambridge University Press. 1986.

Standard&Poor´s. Annual global corporate default study: corporate defaults poised to rise in 2005. Global fixed income research. 2005.

D. TASCHE. A traffic lights approach to PD validation. Working Paper. Deutsche Bundesbank. Frankfurt. 2003.

A. TIOMO. Credit risk and variability of default rates: an empirical analysis using simulations. Working Paper. Banque de France. Paris. 2004.

S.E. VOLLSET: Confidence intervals for a binomial proportion. Statistics in Medecine. Vol. 12. 1993.

Annex 1: Historical data on Moody´s A-grade

YEAR	Issuers	1Y-Def.-Freq.
1981	376	0.00%
1982	387	0.26%
1983	432	0.00%
1984	472	0.00%
1985	524	0.00%
1986	579	0.00%
1987	555	0.00%
1988	553	0.00%
1989	587	0.00%
1990	614	0.00%
1991	609	0.00%
1992	694	0.00%
1993	740	0.00%
1994	880	0.00%
1995	968	0.00%
1996	1071	0.00%
1997	1133	0.00%
1998	1154	0.00%
1999	1173	0.00%
2000	1237	0.00%
2001	1287	0.16%
2002	1301	0.16%
2003	1279	0.00%
2004	1244	0.00%
	Mean	**0.02%**
Standard Deviation		**0.07%**