Rating models:
Impact on the Regulatory Capital for Corporate Exposure

BAZZI MEHDI, CHAMLAL HASNA

Laboratory of Computer Science Decision Aiding
Faculty of Science Ain chock, Casablanca ,Morocco

Abstract

Being stimulated by the Basel II Capital Accord, banks adopting the internal rating-base approach (IRBA),have began developing more and more their own internal rating as long as the systems meet specified minimum requirements. In this context, the purpose of this paper is to provide an overview of statistical methods to develop Rating model commonly used in practice as well as analysing the relationship between the number of classes in the master scale and the impact on regularity capital for Corporate Exposure.

Keywords: `Rating models,Risk Management, Basel II, Master scale.`

1. Introduction

After the 2008 financial crisis , banking regulation has developed a reform program meet the lessons of the crisis. This program has caused an increase in banks capital. Therefore, banks need to optimize their return on equity [1] which has doubly penalized by the lower margin of profit and the increased risk of cost.

Despite this, the regulators tolerance has become increasingly stringent with the loopholes in risk measurement and management. Therefore banks are encouraged to establish best practices for risk management, in this case the establishment of the internal rating models under the Advanced Approach (IRBA) of Basel rules [2].

The purpose of this paper is to show how the choice of the modeling method used in the estimation of rating model for corporate exposures [3] can be a determining factor for the optimization of RWA. This will be accomplished by analyzing the relationship between the number of risk class in a rating scale and the impact on the RWA. This analysis will follow several steps:

First, we will give an overview of statistical methods used to build and estimate rating models. The overview leads to a clear understanding of the under lying statistical indicators and algorithms behind each technique. We also highlight the benefits and the drawbacks of the various approaches.

Second, once classification techniques are analyzed, we will ask the question whether the models described are in line with the IRB Approach of Basel II.

Third, an empirical study will be conducted on real corporate portfolio. The observations of the latter are described by relatively large number of mixture of discrete and continuous variables, and where the minority group (Non defaulting clients) represents less (20%). The purpose of the study is to build multiple master scales using the different classification techniques, analyze the difference between the methods, and use the output to identify which technique provides the best result in term of stability, accuracy and robustness. Finally, the relationship between the number of risk grade and the impact on RWA will be analyzed in order to identify potential opportunities for RWA optimization.

2. Statistical Methods Risk Classification:

We define statistical models as the kind of approach which uses econometric methods to classify borrowers according to their risk. Statistical rating results from thorough analysis of public and private information from all relevant sources. The rating process involves a search for explanatory variables which provide as sound and reliable a forecast of the deterioration of borrowers situation as possible. In this section, we describe an overview of parametric and nonparametric models generally considered for statistical risk assessment. Finally, we discuss the benefits and the drawbacks of each approach. Many of the methods are described in more detail in [4].

In general, the establishment of statistical model can be described as follows: Firstly, we use borrowers characteristics indicators like financial information as quantitative variables (balance-sheet variables), behavior variables (account information) or qualitative variables as management quality, competitive position, and growth prospects. Other input may be used like macroeconomic variables which were collected historically and are available for defaulting and non-defaulting borrowers. Let the borrowers characteristics are defined by a vector of n separate variables $(X_1...X_n)$ Observed at time t - L. The variable Y is defined as Y=1 for default and Y=0 for non default. The time lag L between X and Y determines the forecast horizon.

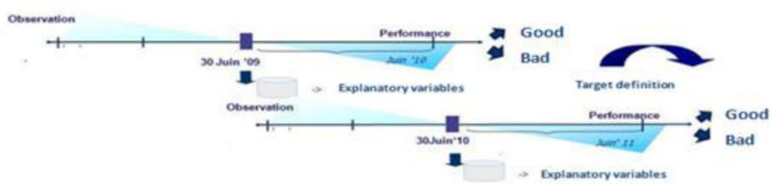

Figure 1: Methodology of construction of dependent variable

2.1. Classification by Bayesian Discriminant Analysis:

Discriminant analysis (DA) is a classification technique applied to corporate bankruptcies by Altman as early as 1968 [1]. In the case of rating models, DA handles the situation in which we have a set of borrowers, each belonging to group (Defaulting and Non defaulting borrowers) and we look for the rules (based on borrowers characteristics) for assigning the borrowers to their groups. This approach is also called Bayesian, because it was developed from bayes theorem:

$$P(G_i|x) = \frac{P(G_i)P(x|G_i)}{\sum_j P(G_i)P(x|G_i)}$$

Where i=1,2,..,n (but in our case we have two groups: defaulter and non defaulter).

- $P(G_i|x)$ is the a posteriori probability of belonging to G_i given x.
- $p_i = P(G_i)$ is the a priori probability of belonging to G_i.
- $f_i(x) = P(x/G_i)$ is the conditional density of the distribution of x, when its group G_i is known.

In addition, the Bayesian approach to discriminant analysis allows cost of incorrect classification C_{ij} to be included. Given i_s a borrower which we want to classify and we look for complete system of Event (A_1, A_2, A_n) in which i_s is classed in the group G_i if he belongs to A_i. According to [5] this partition is that minimize the average global risk, and it is given by following theorem:

Theorem : The optimal classification rule based on the choose of partition $P = \{A_1, A_2, ..A_n\}$,where $A_s = \{x \in P/h_s(x) = min(h_j(x), j = 1, .., n)\}$
with: $h_j(x) = \sum_{i=1}^{n} C_{ij}P(G_i)f_i(x)$.
With the assumption of equal cost and the equiprobability, we have :

$$i_s \textit{ is classed in the group } G_i \Leftrightarrow f(i_s/i) = Max_{l=1...n}f(i_s/l)$$

The problem becomes a comparison of density function within each group which gives an advantage to the group with higher density values. However, in the case of the absence of the homoscedasticity (see figure 2), the accuracy of results is not enough. In order to illustrate this phenomenon we use the Discriminant Analysis of Fisher (1936) Iris data using normal density as example.

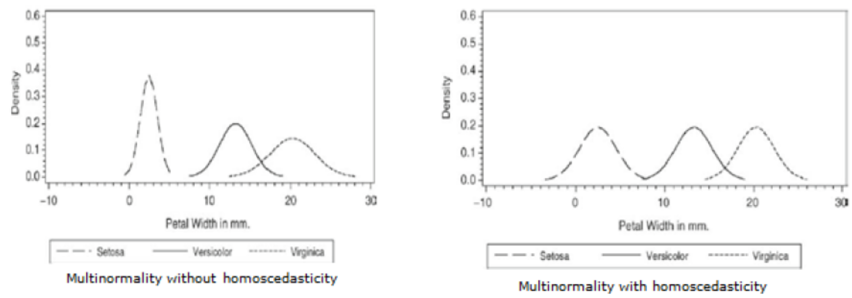

Figure 2: Discriminant Analysis of Fisher (1936) Iris data

In [6] the authors propose a variant of discriminant analysis based on atypicity index and density function.
The atypicity index of group G_l affected to the individual i is given by:

$$ind(i,l) = \sum_{j \in E f(j/l) > f(i/l)} f(j|l)$$

The method proposed consists in the classification of the individual i_s in the group G_{l_0} which verifies:

$$\frac{f(i/l_0)}{ind(i,l_0)} = Max_{l=1\dots n} \frac{f(i/l)}{ind(i,l)}$$

The authors prove that the criterion above gives results better then these given by the Bayesian approach even with homoscedasticity assumption.

2.2. Classification by Logistic regression:

Logistic regression [7] is introduced into software more recently than discriminant analysis, possibly because of its greater complexity of calculation, and has therefore only recently become a regularly used tool for most statisticians. Wiginton(1980) was one of the first to publish credit scoring results using the logistic regression.
When building a credit scoring model-particularly when modeling the probability of default (PD) of customers- the dependent variable Y is binary and takes

two possible values:

$$Y = \begin{cases} 1 \; if\ the\ borrower\ does\ default\ within\ the\ following\ year \\ 0 \; the\ borrower\ does\ not\ default\ within\ the\ following\ year \end{cases}$$

The PD is modeled by using a logistic regression and the score is attributed to each borrower based on explanatory variables that are accurately chosen when building model. Therefore, the probability of occurrence of the default event equals:

$$P(Y = 1|X) \quad = \quad \frac{\exp^{\beta_0 + \sum_j \beta_0 x_j}}{1 + \exp^{\beta_0 + \sum_j \beta_0 x_j}}$$

And

$$score = log(\frac{P(Y = 1|X)}{1 - P(Y = 1|X)}) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$$

With:β_i the parameter of the regression, x_i explanatory variable and $X = \{x_i \; i = 1...p.\}$ The function $log(P|(1-P))$ is called logit function and $\exp^{\beta x_i} \; i = 1...p$ is the odds, i.e. the relation between default probability and the probability of survival. Now it can be easily seen that a variation of a single variable x_i of one unit has an impact of $\exp^{\beta}$ on the odds when β denotes the coefficient of the variable x_k. Hence, the transformed coefficients $\exp^{\beta}$ the odds ratio and they represent the multiplicative impact of a borrowers characteristic on the odds.

In practice, if the borrower observations are highly dispersed, in other words if there are very few observations for given value x of X, it will not be possible to calculate $P(Y = 1|X = 1)$ directly, and we will have to group the value of X in brackets to estimate the probability $P(Y|X)$ by the proportion of the $Y = 1$ given x.

The Strengths of logistic regression can be summarized as:

- The method is theoretically sound
- It directly models a probability
- Many statistical tests, such as tests of significance of coefficients are available. They are asymptotic and even exact.

- However, when the assumption of normality of the distribution is satisfied, the regression logistic is less accuracy than discriminant analysis [8].

2.3. The Classification by the decision tree:

2.3.1. Principal of decision tree

The decision tree technique is to detect criteria for successive divisions of a set of individuals E in two or more segments (called nodes). We start by choosing the variable that by its categories gives the best possible division of the population (the segment down more homogeneous) and then repeat this on each new node until the division is not possible or desirable according a stopping criterion predefined by type of tree. Terminal nodes are called leaves and an individual is assigned to a leaf when it meets all the rules that lead to this leaf.

Main methods of classification decision tree

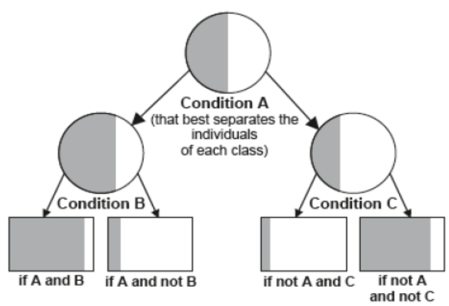

Figure 3: DT example

The main decision tree algorithms are:

- CART (Classification And Rgression Tree) which is suitable for all kinds of variables.
- C5.0 (de J.R.Quinlan) suitable for all kinds of variables.
- Many statistical tests, such as tests of significance of coefficients are available. They are asymptotic and even exact.

- CHAID (Chi-Square Automation Interaction Dtection) initially provided for the consideration of the explanatory and dependent variables, discrete and qualitative.

2.3.2. CART

The CART tree is developed by [9] is a binary tree. The criterion for division of nodes used by the CART tree is the Gini index:

$$GINI(node) = \sum_{r \neq s} C(r|s)p(r|t)p(s|t)$$

$C(r|s)$ is the cost of incorrect assignment of an individual of class j to class i. Equal costs are often considered, for example $C(r|s) = 1$ if $r \neq s$ and $C(r|s) = 0$ if $r = s = 1...k$, in this case:

$$GINI(node) = \sum_{r \neq s} p(r|t)p(s|t) = 1 - \sum_{r=1}^{k} p(r|t)^2$$

More classes are uniformly distributed over the Gini index, the higher the node is most pure, low is its Gini index.

The classification by CART is characterized by its generality and accuracy. Indeed, its generality is due to the fact that the dependent variable may be qualitative or continuous and in this case CART can be used for classification or regression .CART takes into account the cost of incorrect assignment C_{ij} by integrating them into the GINI formula and finally CART handles missing values by replacing them with equally splitting variable or equally reducing variable. Equally splitting variables are those that provide (pretty near) the same purity as the variable nodes without treatment. Equally reducing variables are those which retain the variable distribution of the original.

CART performance is due to its pruning mechanism. Maximum tree is constructed by continuing the process of division nodes as it is possible. Then the algorithm deduces several nested sub-trees by successive pruning, it compares, before choosing the one for which the error rate measured in test or cross-validation is the lowest possible. Another aspect of the performance of CART is its exhaustive search of all possible splits.

2.3.3. C5.0

The C5.0 [10] tree is a development tree C4.5 [11] and TD3 [12] submitted by the same Australian researcher J.Ross Quinlan: it uses the criterion of information gain for splitting nodes . As CART , C5.0 explores all possible splits in the explanatory variable and begins by building up tree ($T_{(max)}$) that seeks to reduce by pruning. However, the pruning process differs from CART .Another difference from CART is that C5.0 is not binary. This is because of its treatment of the qualitative variables which, at the level of a parent node, give rise to a child node for each category.

2.3.4. CHAID

This tree, proposed by Kass GV [13] is an improvement from the first tree AID (1963) Morgan and Sonquist. CHAID uses chi-square test for the variable separation (most significant) for each node, it can only be used with qualitative or discrete variables. Unlike the CART tree, CHAID is not binary, handle missing values as a modality which may be isolated or merged with another categorie. Finally, CHAID does not have a pruning process from a spanning tree (T_{max}) that tries to reduce (post-pruning), but it uses predefined criteria which stop the tree growing (Pre-pruning).
The general strengths and weaknesses of tree are:

- The results are expressed as explicit conditions on the original variables
- Through categorization, non linear relationship between the variables and the score can be easily modeled.
- Interaction present in the data can be identified, parametic methods can model interaction only to limited extend (by introducing dummy variables)
- Probabilities of default have to be calculated in a separate step.

- The definition of the nodes at level $n+1$ is very highly dependent on the definition at level n. consequently, the modification of single variable, if it is located near the top of the tree, may modify the whole tree.

To sum up: trees are particularly used when the data is characterized by a limited number of predictive variables which are known to be interactive.

3. Statistical models and Basel II:

In this section we talk about the classification techniques used in the rating models (described above) and we see even they are in line with IRBA Approach of Basel. But, before this discussion we should define rating systems as done in the in the Basel document. Following $\oint 394$ of the Revised Framework from June 2004 and a rating system *comprises all the methods, processes, controls, and data collection and IT systems that support the assessment of credit risk, the assignment of internal ratings, and the quantification of default and loss estimates*. Therefore the statistical methods described above allow the assignment of internal rating.

The minimum requirements to build an internal rating systems are discussed in part II, section III, H of the Revised Framework. The text concern the assignment of internal rating defines the conditions and constraints that must be taken into account :

- A minimum of 7 rating classes of non-defaulted borrowers ($\oint 404$)

- The number of borrowers in each class must not exceed a certain limit or be lower the certain limit ($\oint\oint 403, 406$). The excessive concentration in single rating class shows that the discrimination power of the master scale is not sufficient enough, and the undue concentration in risk class can introduce instability in the master scale since a small change in the population might have a significant impact on the default rate.

- The level of risk must be different from class to another ($\oint 410$).

- Plausible intuitive and current input data ($\oint\oint 410, 411$).
- All relevant information must be taken into account ($\oint 411$).

The Basel II requirements dont make any preference for a certain method. Therefore the classification techniques discussed her are all possible candidates for the IRB Approach.

The strengths and weakness of the single methods concern some of the minimum requirements. For example, if there are few data the modelers must avoid the decision tree. Methods which allow for statistical tests of significance of coefficients (e.g. the logit model) provide a straightforward way to prove the plausibility of the borrowers input factor (as required by $\oint 410$). When the outcome of the model is continuous variable (e.g. Discriminant Analysis, Logit regression), the master scale can be defined in a more flexible way ($\oint\oint\oint 403, 404, 406$). Finally none of the drawbacks of the models considered here exclude a specific method and bank should rather base their choice on their internal aim and constraints. In the following part, a master scale is built for corporate Exposures using different techniques mentioned above.

4. Classification methods and their impact on the regulatory capital:

4.1. Aim of the analysis:

The study presented in this paper shows how the classification techniques using in building rating models for corporate exposure can impact the regulatory capital in the first hand and in the second hand, if the classification technique has been chosen , how can we optimize the RWA (and thus the regulatory capital) by adjusting the number of rating classes.

Firstly, we apply all classification technique mentioned above in the article. The purpose is to see which technique is best suited for corporate exposure by providing the optimal result in line with the best practice in risk management. Secondly, given the choice of the classification technique, different simulations are realized by taking the number of classes as input. Indeed, we studied the impact of the different techniques on RWA while changing the number of classes.

4.2. Description of the portfolio on which the study was conducted:

As mentioned above, the portfolio on which the study was conducted is corporate exposure.

We describe in the Table below the variables used in the empirical studies

Property	**value**
Source	**Information Risk company , Morocco**
Sample Size	**1663 totaux: 300 Defaults and 1363 Cleans**
Dependent variable	**Binary variable which describe whether the debtor is defaulting or not.**
Default (1)	**Bale II definition default[22]**
Not Default(0)	**Bale II definition default**
Explanatory Variables (40)	**Label**
LogTotalBilan	Logarithm of Total Assets
TotalBilan	Total Assets
LogCA	Logarithm of Turnover
AgeSociete	Age of the company
PassifCirculant	Current Liabilities
LogIMMO	Logarithm of Fixed Assets
TresorerieActifSUM	Cashflow Assets
FraisFinanciersSUM	Interest
ResultatNetSUM	Net income
ResultatNetN1SUM	Net income N-1
DatNaissance	Birthday
ChiffrAffSUM	Turnover
ActifCircuSUM	Current Assets
ActifImmobilisSUM	Fixed Asset
CreancesClientsSUM	Accounts Receivable
StocksSUM	Stock

variables (Next)	**Label**
RotatioStock	Stock*360 / Turnover
RotationCreancesClients	(Accounts Receivable)*360 / Turnover
CAActifCirculantActifImmobilis	(Turnover + Curent Assets)/ Fixed Assets
FraisFinanciersCA	Interest / Turnover
CroissanceRN	Net Income growth
CAFDpropresEndettement	Financin Capacity + Capital / Bank Debt
CAActifImmobilise	Turnover/ Fixed Assets
BFR	working capital needs (WC)
BFRCA	working capital needs/ Turnover
CurrentRatio	Current Ratio
WorkingCapitalTurnoverRatio	Working Capital turnover ratio
RotationBFRCA	WC*360/ Turnover
NetMargin	Net Margin
ROE	Return on Equity
Gearing	Gearing
TresorerieNette	Net Cashflow
EndettementNet	Net Debt
FpDansstructure	percenatge Capital on total Asset
ENFP	Total Debt /Capital

The modeling windows are the dates 12312009 and 12312010 which mean that all performing loans at 12312009 and 12312010 are considered. These loans are analyzed from 01012010 to 31122010 and 01012011 to 12312011 (figure 4).

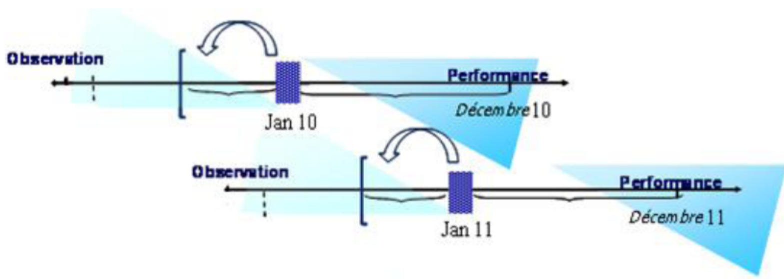

Figure 4: Modelling windows

For the aim of the study, the rating model has been already been done, based on the best practice in the industry [3] and the models shows a good discriminating power. The methodology used to build the models is summarized in the figure 5, but is not the core subject of the study. In fact, the focus is on the master scale and therefore the methodology of the rating model wont be detailed here.

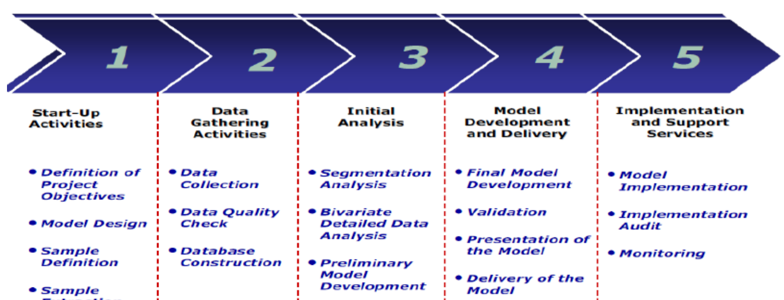

Figure 5: Steps of the model process

The distribution of the number of borrowers in the portfolio is:

Windows	**Non defaults**	**defaults**	**Somme**
2009	843 (83%)	175 (17%)	1018
2010	520 (81%)	125 (19%)	645
Somme	1363 (82%)	300 (18%)	1663

The average default rate on which the models have been performed is (18%).The number of borrowers is acceptable making the portfolio sufficiently granular.

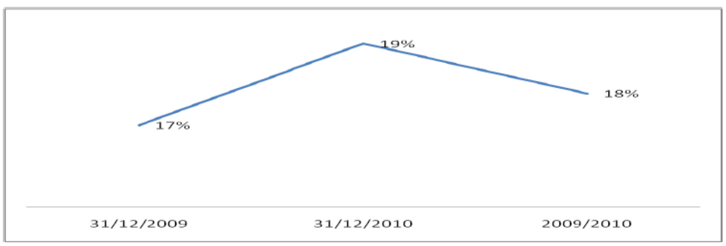

Figure 6: Evolution of the default rate

The default rate series presents a tendency to increase especially after 2008 crisis.

5. Presentation of results:

5.1. Building the Master scale:

As mentioned above, the modeling windows are 31122009 and 31122010. The different classification methods are performed on this sample. The different results are summarized in the following matrix: As first, the form of the master scale changes according the classification technique used. In fact the distribution of default rate differs on the technique used. We also observe this phenomenal when number of classes vary.
Secondly, we observe that for some number of classes, the decision trees (CART, C5.0, CHAID) could not always build a tree with a specified number of classes.

Figure 7: Master Scale with several classification methods

For example CHAID not able to build a tree with five classes and CART could not build tree with seven classes. This could be explained by variety of the splitting criterion. In fact, CART algorithm [9] selects split using towing criterion, C5.0 [10]uses information gain as splitting criterion and it shares with CART its exhaustive search for all possible split which ensures that the optimal split is chosen, and finally CHAID [13] uses the χ^2 test to define the most significant variable for each node.

The third result shows that the numbers of classes are limited. Indeed, with the number of default (300) it is not easy to build a master scale with more than 8 classes without have an over fitting and lack of robustness of the models. Precisely, with decision trees which require a sufficiently large of number of borrowers per node.

As fourth point, we talk about the method used in order to construct a master scale using logistic regression and RAD technique. Unlike decision tree (CART, C5.0, CHAID) which return score in form of several ranges, the two other methods performed above return continuous score which we must subdivide in several rating classes. To do this, we adopted a mixed approach which started with an objective grouping of classes and it continues with more empirical approach. We used an algorithm to define the master scale by linking the probability of default of borrowers to an exponential distribution with a frequency close to the normal distribution (figure 8). These classes have been modified after according to empirical criteria to identify the most satisfactory master scale.

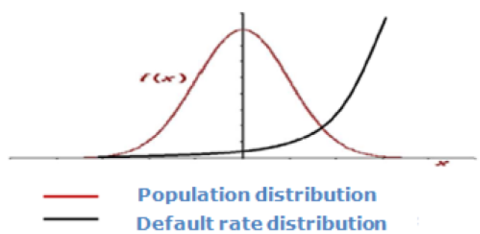

Figure 8: Example of master scale

This technique is frequently used for corporate exposure [14] in accordance with rating scale build by the major rating agencies like StandardδPoors.

Another point which we make sure that is respected when we build the master scale using logit and RAD method is the no inversion of rating classes which means that the rating class A is less risky than B , B is less risky than C and so on. In fact, as the graphic below shows, the discrimination and the progressiveness of default rates are respected.

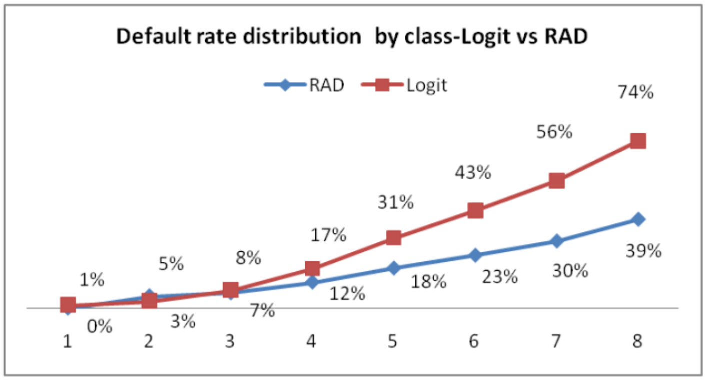

Figure 9: Default rate distribution by classes logit vs RAD

6. Establishing a relationship between the number of classes and the impact on regulatory capital

After showing the impact of the technique chosen on building the rating models, the aim of this section is to establish a relationship between the number of risk classes within the master scale and the impact of the regularity capital. This relationship is quite important in the context in which banks see their margin profit decrease more and more because both the concurrence and their risk cost which get higher with important pressure on banks capital. Thus, to reach an acceptable level of profit (ROE) banks must optimize their RWA. To establish this relationship, a RWA simulation has been conducted. The Exposure at default (EAD) of the portfolio is considered as the same for each loan. This assumption gives a similar weight to each loan and consequently assumes the best granularity of the portfolio.

As consequence, each loan is supposed to have an EAD of 100 kMAD and the simulation results are:

Nb of classes	C5.0	CART	CHAID	LOGIT	RAD
5	831 440	485 636		437 814	443 002
6	830 740	472 170	502 761	436 126	442 245
7	829 848		502 332	431 466	441 777
8			502 332	431 466	441 777

The first analysis of the table above shows that there is a negative relationship between the evolution of the RWA and the number of the rating classes. In fact, the amount of the regularity capital decreases with the number of classes for all the classification techniques.

Logistic regression gives the optimize measure of RWA, RAD and CART methods give goods results also, unlike CHAID and C5.0 which give very high amount of RWA.

Finally, the slope of the curve is close to 0. In other words, more than just getting closer with the increasing number of classes, the curves converge to a certain limit. This shows that RWA do not decrease indefinitely with the number of classes (figure 10).

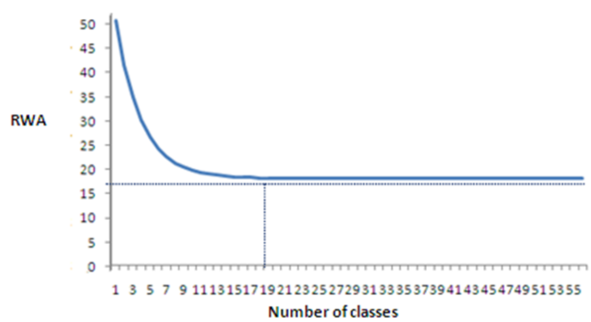

Figure 10: RWA Evolution depending on the number of classes-classical portfolio

7. Conclusion

In this study we described a variety of methods of building and estimating rating models, we showed that all these techniques are in line with IRBA Approach of Basel. In fact, the logit regression might be best suited for Corporate Exposures since it provides better results in term of discriminatory power, stability, and robustness.

Finally, as proved in the empirical results, there is negative relationship between the number of risk classes and the RWA, showed an opportunity for RWA optimization. These opportunities are less significant for our sample but might still have best impact on sample with appropriate number of defaults, a point which warrants attention due to increasing risk costs and pressure on profit margin.

8. Bibliography

[1] **Altman EI (1968)**, Financial Indicators, Discriminant Analysis, and the Prediction of Corporate Bankruptcy, Journal of Finance.

[2] **Basel Committee on Banking Supervision, Jun 2006**,International Convergence of Capital Measurement and Capital Standards.

[3]**Evelyn H. (2006)**, Estimation of Rating model for Corporate Exposures, Springer

[4] **Stephan Tuffery (2013)**,Data Mining and Statistic for Decision Making, Wiley Series in Computational Statistic, Chapter 11.

[5] **Nakach J.P. (1980)**, Methodes de discrimination pour variables de nature quelconque, theorie et pratique :Thse prsente pour obtenir le grade de docteur-es sciences, Pirre et Marie Curie University.

[6] **CHAH S . CHAMLAL H. (2001)**, Proposition dune nouvelle rgle discriminante et compariason avec les rgles lineaire et quadratique, Revue de la statistique applique, tome 49 ,n3(2001), p.61-72.

[7] **Hosmer W, Lemeshow S (2000)**,Applied Logitic Regression, New York, Willey.

[8] **Efron,B(1975)**, the efficiency of logistic regression compared to discriminant analysis. Journal of the American statistician Association, 70,892-898.

[9] **Breiman L., Friedman J., Olshen R., and Stone C.,(1984)**, Classification and Regression Trees, CA: Wadsworth and Brooks.

[10] **R.Quinlan C5.0:An informal Tutorial,(1998)**, http://www.rulequest.com/see5-unix.html.

[11]**R.Quinlan ,(1993),** C4.5 :Programs for machine LearningMorgan Kaufmann,San Mateo,California

[12] **J.R.Quinlan (1986)**, Induction of Decision Trees ,Machine Learning 1 pp 81-106.

[13] **G.V.Kass ,(1980)**, An exploratory technique for investigating large quantities of categorical data ,Applied Statistics, vol.29,pp.119-127.

[14] **Rauhmeier, R. (2006)**, PD-Validation Experience from Banking Practice, Springer.