Journal of Risk and Financial Studies
Vol. 1, No. 1, 2020, pp. 1-24

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

Application of Logical Analysis of Data (LAD) to Credit Risk Ratings for Banks in Zimbabwe

Isabel Linda Moyo[1], Aditi Kar Gangopadhyay[2], Victor Gumbo[3], Eriyoti Chikodza[4] and Brian Jones[5]

[1]*Lecturer and PhD Student, Dept. of Statistics and Operations Research, National University of Science and Technology, P. O. Box AC 939, Bulawayo, E-mail: ilzulu82@gmail.com; isabel.zulu@nust.ac.zw*
[2]*Dept of Mathematics, IIT Roorkee, India*
[3]*Dept of Mathematics, University of Botswana, Botswana*
[4]*Dept of Mathematics and Computer Science, Great Zimbabwe University, Zimbabwe*
[5]*Dept of Statistics and Operations Research, National University of Science and Technology, Zimbabwe*

Received: 8 January 2020; Revised: 13 February 2020;
Accepted: 16 March 2020; Publication: 16 May 2020

Abstract: *As data is now becoming available and accessible, credit risk management departments in financial institutions are now engaging machine learning techniques to produce more reliable internal credit risk rating systems. In this paper, data on 17 Zimbabwean banks are used to apply and test the Logical Analysis of Data (LAD) - a supervised learning data mining technique, to generate an objective, transparent, consistent, accurate, self- contained and generalisable credit risk rating system that has varying levels of granularity and is Basel compliant. This system gives an understanding of relationships between the uses of credit ratings, the different options for rating system design and the effectiveness of internal credit rating systems. Such a system becomes useful in decision making pertaining to the determination of the amount allocated as regulatory capital in banks, which is a buffer in banks against distress and bank failure.*

Keywords: *Basel compliance, credit risk ratings, Logical Analysis of Data (LAD), regulatory capital.*

1. INTRODUCTION

Hammer (1986) defined the Logical Analysis of Data (LAD), as a combinatorics, optimization and Boolean logic based methodology for analyzing archives of observations. Peter L. Hammer was the first researcher to introduce the concept of LAD, at a conference in 1986. Over the years, many authors (Crama *et al*, 1988; Boros *et al*, 1997; Alexe *et al*, 2007; only to mention a few) have built theoretical formulations of LAD for the purpose of applications on medical, industrial and economics case studies. Alexe *et al*, (2007) points out that the ability of LAD models to generate patterns or rules, makes it a popular methodology in classification, ranked regression, clustering, detection of subclasses, feature selection and other problems.

The advent of artificial intelligence and machine learning has played a big role in the expansion of the research area of LAD in terms of customizing the definition of the LAD main concepts to generate efficient algorithms for pattern generation; as well as providing a basis of simplifying complex analysis of the difficult problems embedded on LAD, cascading down to its application in various areas of specializations.

On the other hand, the financial crisis episode revealed major weaknesses in the management of credit risk in many financial institutions, and this was implicated to weak regulatory and supervisory systems in place. The financial crisis hits the global markets even after the Basel Committee on Banking Supervision (BCBS) had been established in 1974, by the G10 central banks, to set up rules for banking regulation that were to be included in the drafting of national regulation laws. These meetings were carried out at the Bank of International Settlements (BIS) in Basel, Switzerland and that is why the rules were called the Basel Capital Accord. Up to date there are three capital accords, Basel I, Basel II and Basel III. Basel I which was introduced in 1988, had its focus on credit risk and its aim was on setting up regulatory minimum capital requirements so as to ensure that banks could be in a position to pay back depositors' fund, at any given time. According to Baesens *et al*. (2016), Basel I accord faced drawbacks in that

- the solvency of the debtor was not properly taken care of, as the risk weights depended only on the exposure class and not on the obligor or product characteristics,
- there was insufficient recognition of collateral guarantees to mitigate credit risk,
- various opportunities were made available for regulatory arbitrage by making optimal use of loopholes in the regulation to minimize capital,
- it only considered credit risk and not operational or market risk.

Basel II was then introduced in 1997, to address the shortcomings of the Basel I ac- cord. It had three key pillars where Pillar 1 allowed quantitative models to be built by banks, which were to be reviewed by an overseeing supervisor as highlighted by Pillar 2. Sound processes were to be introduced that would help evaluate risk and supervisory monitoring. Once the risk models were approved, they could then be disclosed to the market as described in Pillar 3. This was to inform and convince the investors that the bank was sound and had solid risk management strategies. Basel III was a direct result of the global financial crisis and it took effect in 2013. It built on Basel II to strengthen global capital standards.

Its key focus was on tangible equity capital since it had the greatest loss-absorbing capacity. The Basel III accord reduced the reliance on models that are developed internally by banks and also on ratings that are obtained from the external rating agencies, placing a great emphasis on stress testing.

The economic instability inherent within Zimbabwean banks evidenced by the level of bank distress report by many researchers, prompts more research work to be done in terms of evaluating and improving the credit quality of banks. Banks' distress which has usually led to bank failure in Zimbabwe, is feared to reproduce cyclical recessions which usually leads to severe financial crisis. Thus more accurate early warning systems for banks are a necessary condition to avoid bank failures and could serve the regulators in their efforts to minimise bailout costs.

Kogan and Lejeune (2010) discuss on why central banks now rely on their own internally developed ratings and hence credit risk rating models within their credit risk management framework. They highlight that external credit assessment institutions have been criticised intensely with the collapse of large financial institutions attributed to unanticipated external ratings due to conflict of interest in the credit agencies. Hammer *et al*. (2006) also highlights on criticisms such as lack of comprehensibility, procyclicality, black box, lack of predictive and crisis warning power and regional bias. This is reflected on the Basel III mandate on the reduced reliance on ratings from external rating agencies.

The BCBS of the BIS still favor Internal Ratings Based (IRB) approach where banks use their own internal rating estimates to define and calculate default risk, for as long as robust regulatory standards are met and the internal rating system is validated by the national supervisory authorities. According to Treacy and Carey (2000), the internal rating system provides autonomy to a bank's management in that it defines credit risk in line with the bank's core business and best international practices.

Banks now develop internal risk models, used to make evaluation of credit risk exposure more accurate and transparent. This in turn improves profitability of banks, as credit policies and processes are made to be more efficient and this also improves the data quality and hence an expected translation to substantial savings on capital requirements.

In this work, a credit risk rating system was built using LAD so as to evaluate the creditworthiness of banks in Zimbabwe. The rating system would enable the central bank, the Reserve Bank of Zimbabwe (RBZ), to rate its counterparts as frequently as desired, as the model can be rerun on an ongoing basis to assess the solvency of the financial institutions and possibly anticipate rating modifications of agencies rather than waiting for external credit rating agencies to modify their ratings. Kogan and Lejeune (2010) also highlights that developing a credit rating system using LAD will have an advantage of the following characteristics:

- Self-containment: the method does not use lagged ratings as independent variables so as to accommodate banks that have not been rated before.
- Objectivity: the system relies only on measurable characteristics of the rated entities.
- Transparency: the system has formal explicit specifications.
- Accuracy: the system is in agreement with learned, opaque rating systems.
- Consistency: discrepancies are easily dealt with.
- Generalisability: can be applied for subsequent years on banks who were not used in the system building.
- Basel compliant: satisfies Basel II accord requirements.

The rest of the paper is organised as follows: Section 2 reviews literature on the use of LAD in credit risk modeling, whilst Section 3 discusses the LAD methodology in detail. Section 4 presents results whilst Section 5 discusses the results and Section 6 concludes the study and presents recommendations.

2. LITERATURE REVIEW

The LAD rating system is based on the learning of the ratings disclosed by RBZ, which are the Camels ratings, as well as the explanatory variables affecting each bank. Hammer *et al*. (2012) emphasised that an effective LAD rating system requires a set of reliable ratings that is not plagued by moral hazard, which defines what constitutes a reliable rating. This has been a weakness of external credit rating agents, who have been severely criticised for failing to predict some major cases of creditworthiness deterioration due to moral hazard.

2.1. The Camels Rating

The RBZ uses the Camels rating system, which enables the central bank to carry out its supervisory/regulatory mission more effectively, by categorizing banks into appropriate supervisory groups. The main aim of this uniform rating system is to identify those institutions whose financial, operational, managerial or compliance weaknesses require special supervisory attention and/or warrant a higher than normal degree of supervisory concern (RBZ Banking Circular No. 01-97/BSD). Kogan and Lejuene (2010) emphasise the growing reliance on credit ratings as due to the growth of investment opportunities associated with globalization of the world economies. Due to the riskiness of these opportunities, investors rely on the credit ratings for the assessment of credit risk.

Table 1
Camels rating framework

Rating scale	*Rating analysis*	*Bank characteristics*
1	outperforming	sound banking in all aspect, strong performance and risk management practices, positive trends to key performance measures and in regulatory compliance
2	satisfactory	moderate weakness, can identify and mitigate risks accordingly, weakness managed well, compliance of regulations done regularly
3	less than satisfactory	functional weakness causing supervisory concern, do not comply with regulations management lack ability to overcome weaknesses efficiently
4	distress	expresses moderate to serious operational failure, risk management at a lower efficiency scale, key performance indicators negative, regulations not complied with, failure not yet imminent but highly likely, under supervisory review
5	failure	critical financial weakness leading to failure, unsatisfactory performance, serious financial restructuring, management not able to combat with severity and volume of problems

2.2. LAD model for bank credit ratings

Hammer *et al*. (2012), stated in their study that the main purpose for using the LAD method to develop bank ratings is fourfold.

- It allows for the representation of higher order interaction between variables, when the complexity of financial systems and the interconnections between financial institutions, is considered.
- Every explanatory variable is allowed to have distinct effects on different rating classes.
- The derivation of rating systems with the varying, user-specified number of rating classes, is enabled.
- The generation of the patterns used in the rating system, is not constrained by any limitation on the number of variables that can be used.

Rocha (2018) then describes a good credit risk model as one that allows the bank to identify risks on their balance sheet. This helps to smoothen the volatility of movements on the balance sheet so as to reduce the likelihood of nasty surprises detrimental to the operations of the bank. A good credit risk model was also said to place tighter controls over profit and loss groups thereby excessively reducing the risks of compliances breaches. This could then facilitate data reconciliation and accurate reporting on figures which would lead to more transparent decision making.

Kogan and Lejuene (2010) express the importance of variables in LAD model building, as LAD is on discovering how the interactions between the values of the small group of variables (generated patterns) affects the outcome (bank ratings). And hence the magnitude of importance of a variable is measured by the frequency of its appearance in the model's pattern. According to Mortada *et al*. (2011), the true edge of LAD over other approaches such as neural networks, is its ability to generate interpretable patterns, which strengthen the motivation of applying LAD in the work so as to produce a model that will be useful within the internal banking system as well, to dictate and mitigate risk before distress and failure occurs.

Rocha (2018) also reiterates that complying with regulations to calculate risk- weighted assets as stated in the Basel accords, banks have to understand the internal dynamics that influence the calculation of the capital requirements. An accurate compliance figure comes with an extensive methodology that attributes risks back to their source and how the risk levels change with the changing institutional variables. This is of significance in credit risk management and internal stress testing processes, as stated in Basel III accord, and banks can focus on the business lines and geographies that best suit their risk appetite. Risks within the balance sheets are better understood and dealt with in a way that will sustain the operations of business. Compliance with Basel III requires that the internal risk models (credit rating, probability of default, loss given default, exposure at default) be cross-validated and approved by the legislator; and this only will allow banks to adopt the Internal Rating Based approach to calculate their capital provisions.

In view of the above sentiments and mitigating risk inherent within banks, 23 explanatory variables in terms of financial ratios were extracted from the balance sheets and the profit and loss statement of 17 Zimbabwean banks, which were operational between the years 2010 to 2017.

2.2.1 Explanatory variables

The explanatory variables which were considered for this research work were found to be of significance in the performance of Zimbabwean banks as discussed by Moyo *et al*., (2019).

- Return on Equity (*ROE*) calculated as net profit/total equity, shows how much profit a firm generated with respect to the total shareholder equity amount invested and is found on the balance sheet. It indicates how much returns a firm got from its investments. It is the most important indicator that measures a bank's profitability and its potential to grow. ROE indicates the profitability to shareholders of the bank after all expenses and taxes have been deducted (Van Horne and Wachowicz, 2005).
- Return on Assets (ROA) calculated as net profit/total assets, is a ratio of the banks income to the banks total assets (Khrawish, 2011). It measures the ability of a bank to generate income using assets at its disposal. ROA indicates the profitability on the assets of the bank after all expenses and taxes have been deducted (Van Horne and Wachowicz, 2005).
- Net Interest Margin (*NIM*) calculated as net interest income/total assets, focuses on the profit earned on interest activities. NIM is a ratio that measures the returns of a firm after investing its funds in comparison to its expenses on the same investments.
- Market Share (*MS*) is the logarithm of the value of deposits. MS is a capital indicator which is expected to have a positive impact on profitability. Abreu and Mendes (2002) found out that well-capitalised banks have lower expected bankruptcy costs and better profitability. Hoffman (2011) also argues that if a bank has a great market share, then it has an advantage of controlling prices on products in the market. This is because such banks they have monopoly powers arising from a large market share which suggests higher profits.
- Company Size (*CS*) is measured by the accounting value of a bank's total assets. CS is represented by the natural logarithm of total assets. The effect of CS on profitability is generally expected to be positive (Alper and Anbar, 2011).

$$CS=(total\ assets)/(overall\ total\ assets\ in\ all\ banks)$$

- Credit Risk (*CR*) has been proxied by the ratio of loan loss provision to total loans. This ratio reflects on changes in the performance of bank's loan portfolio that affects the profitability of the bank negatively (Aydogan, 1990 in Alper and Anbar, 2011). If the ratio is high it means the quality of the loan portfolio is poor and therefore there is high risk on the loan portfolio. It is calculated as:

CR1=(loan loss provision)/(total loans) and
CR2=(net loans)/(total assets)

- Taxation (*Tax*) is the ratio of tax to operating profits before tax. This ratio is expected to be negative as it entails the direct cost of a bank and hence it reduces profitability. Tax is a compulsory monetary contribution to the state's revenue, assessed and imposed by a government on the activities, enjoyment, expenditure, income, occupation, privilege, property etc of all organizations.

  It is calculated as:

  Tax = (tax) / (operating profits before tax)

- Solvency Risk (*SR*) is the ratio of shareholder's equity to total assets and has been used as a capital indicator. If the SR ratio is high, then the need for external funding becomes low and this usually result in higher profits to the bank. This proves that a bank is able to absorb losses as well as handling risk exposure of shareholders' investments. It is calculated as:

  SR=(shareholder' s equity)/(total assets)

- Cost Efficiency (*CE*) is used in order to estimate how efficiently banks manage their expenses relative to their size. CE is an indicator on cost management efficiency since the operating expenses are an outcome of bank management. Efficient cost management usually leads to improved profitability of banks. It is calculated as:

  CE1=(operating expense)/(total assets) and
  CE2=(operating expense)/(average assets)

- ,mDiversification (*Divers*) is the product of manager's decisions to reduce risk. The importance of diversification through fee based services of banks, is to increase the non-interest income and so a positive relationship is expected between Divers and profitability. It is calculated as:

  Divers=(non interest income)/(gross revenue)

- Business Mix (BM) is the ratio of net income from fees and commission to average assets. This ratio is expected to be positive since it measures the capability of a bank to generate income through fees and commission from account maintenance. It is calculated as:

  BM1=(net income from fees and commission)/(average assets) and *BM2=(net income from fees and commission)/(total assets)*

- Liquidity 1 (*Liq1*) is expected to have a positive coefficient as high liquidity allows a bank to avoid costly borrowing of funds. Insufficient liquidity is one of the major reasons of bank failures hence the interventions of the central bank in Zimbabwe (RBZ) in enforcing the Basel II instrument amid the volatile economic environment. It is calculated as:

  Liq1=(liquid assets)/(average assets)

- Liquidity 2 (*Liq2*) measures the percentage of total assets comprised by loans and we expect a positive coefficient as more loans generate interest income for the bank unless a bank takes on unacceptable levels of risk. It is calculated as:

  Liq2=(net loans)/(total assets)

- Loan Funding Structure (*LFS)* is the ratio of granted loans to received de- posits. If more deposits are transformed into loans, then higher interest margin and profit is realised. This ratio is thus expected to have a positive effect on profitability. It is calculated as:

  LFS=(granted loans)/(received deposits)

- Foreign Exchange Risk Management Efficiency (*FERME*) is the ratio of income from net exchange rate differences to average assets. With the emergence of the alternative market in Zimbabwe, the banks now have minimal access to foreign currency. This ratio is thus expected to have negative impact to profitability. It is calculated as:

  FERME=(income from net exchange rate difference)/(average assets)

- Impact of Managerial Inefficiency (*IMI*) is the logarithms of overhead costs which is expected to have a negative impact on profitability.

- Impact of the economic crisis (*Crisis*) is a control variable in the banking industry. It is calculated as:

$$Crisis = ROA + \frac{equity / total\,assets}{\sigma(ROA)}$$

  This ratio is expected to be negative because it is often associated with a panic or a bank run, where investors sell off assets or withdraw money from savings accounts from a failing institution, for the fear of assets losing value if kept in that financial institution.

- Efficiency, calculated as:

  Efficiency=(total operating expenses)/(gross revenue) is a measure of a bank's capacity to generate revenue from available resources. Lower ratios are usually anticipated as an higher values in the efficiency ratio are interpreted as either increasing costs or decreasing revenues, where both scenarios negatively affect profitability.

- Bank Orientation (*BO*) is measured as a logarithm of the number of branches of a bank. Banks with many branches are likely to be oriented towards retail banking as reported by Beck *et al*., (2005), and this tends to have an effect on the performance of such banks. However they had an inconclusive viewpoint on whether such an impact would be expected to be positive or negative.
- Ownership, is a dummy variable, which is described by *D1* = 1 if government owned and, *D1* = 0 if privately owned. The ownership structure has effects on the principal-agent relationships which could have a positive influence on the profitability of banks (Westman, 2011).
- Loan to Asset Ratio (*LAR*), demonstrates the ability of banks to meet the demand for loans by using total assets owned by banks. If the ratio is high, then the credit performance levels are favorable as it shows that the loan component given in the total structure of the assets is greater (Rivai *et al*., 2007). It is calculated as:

  $$LAR=(total\ gross\ loans)/(total\ loans)$$

- Cost of reserve requirements (*Reserve*) is the opportunity cost of keeping such reserves. It is a control variables and is expected to have a negative impact. It is calculated as:

  $$Reserve=(non\ interest\ income)/(total\ assets)$$

- Bank concentration (*Conc*) reflects on the banks that have used the process of mergers and acquisitions as part of their growth strategies. It is calculated as:

  $$Conc = (MS)^2$$

  which is a quadratic form of M S. Saona (2016) used Conc to study the non linear relationship between the capital ratio and bank's profitability.
- Relative bank size (*RBS*), calculated as *RBS* $= \frac{total\ assets}{overall\ total\ assets\ in\ all\ banks} \times 100$ measures growth.

3. THE LAD METHOD

When using the LAD classification system to construct a credit rating system, a classification problem associated to the bank rating problem is first spelt out, then a LAD model constructed for it and lastly a banking rating system rooted to the developed LAD model defined. The succeeding LAD methodology outline is as explained by Kogan and Lejeune (2010).

3.1 Binarisation

According to Boros *et al*. (1997), binarisation uses indicator variables to show whether the value of a numerical variable does or does not exceed a specified level, called a cut-point. It is a process of associating an indicator variable to each cut- point. Cut-points selection is done by solving an associated set covering problem. Kogan and Lejene (2010) outline the process of binarisation in the following manner. Data analysed by LAD are represented by n-dimensional real valued vectors that are called *positive or negative,* which go with the value of the additional binary (0, 1) attribute called the outcome or the class of the observations. This can be considered as a collection of M points (a^j, z^j) where the outcome z^j of observation j has value 1 for a positive one and 0 for a negative one, and a^j is an n dimensional vector. If $a[i]$ is used to denote the explanatory factors corresponding to the components then the dataset can then be represented as shown in the example in Table 2.

Table 2
A set of observations: Source- Kogan and Lejeune (2010)

Observation	*Variables*			*Outcome*
j	$a[1]$	$a[2]$	$a[3]$	Z^j
1	3.5	3.8	2.8	1
2	2.6	1.6	5	1
3	1	2.2	3.7	1
4	3.5	1.4	3.9	0
5	2.3	2.1	1	0

LAD discriminates positive and negative observations by constructing a binary- valued function, f, depending on the n input variables in such a way that it closely approximates the unknown actual discriminator. LAD constructs a function, f, which is defined as a weighted sum of combinatorial patterns. This function, f , the original dataset is first transformed into a binarised dataset (variables now have values 0 and 1). For each variable

$a[i]$, a set of $K(i)$ values is defined, $\{c_{i,k} | k = 1, ..., K(i)\}$, called *cut-points* to which binary variables $\{y_{i,k} \mid k = 1, ..., K(i)\}$ are associated. The values of these binary variables for each observation (a^j, z^j) are then defined as:

$$y^j_{i,k} = \begin{cases} 1 \ \textit{if} \ a^j[i] \geq c_{i,k}, \\ 0 \ \textit{otherwise} \end{cases} \tag{1}$$

Binarised dataset corresponding to the data in Table 2 is shown in Table 3, and the values $c_{i,k}$ of the cut-points k for each variable $a[i]$ are shown, as well as those of the binary variables y^j associated with cut-point k of variable i in observation j.

Table 3
The binarised dataset: Source- Kogan and Lejeune (2010)

Variables	*Obs*		*a[1]*			*a[2]*		*a[3]*		*Out-*
Cut points		*c1,1*	*c1,2*	*c1,3*	*c2,1*	*c2,2*	*c3,1*	*c3,2*	*c3,3*	*come*
		3	*2.4*	*1.5*	*3*	*2*	*4*	*3*	*2*	
					$Y^j_{i,k}$					Z^j
Binary Variables	1	1	1	1	1	1	0	0	1	1
	2	0	1	1	0	0	1	1	1	1
	3	0	0	0	0	1	0	1	1	1
	4	1	1	1	0	0	0	1	1	0
	5	0	0	1	0	1	0	0	0	0

From Table 3, $y^1 = 1$ since $a^1[1] = 3.5 > c_{1,1} = 3$.

3.2 Pattern Generation

Positive (negative) patterns generated by LAD are combinatorial rules obtained as conjunctions of binary variables and their negations. When these patterns are translated to the original variables, they constrain a subset of input variables to take values between identified upper and lower bounds, so that

- a sufficiently high proportion of the positive (negative) observations in the dataset satisfy all pattern conditions, and
- at least one of the pattern conditions is violated by a sufficiently high pro- portion of the negative (positive) observations.

Herrera and Subasi (2013) stated in their paper that a observation, $\theta \in D$ satisfying the conditions of a pattern P, that is, $P(\theta) = 1$, is said to be covered by that pattern. The *degree* of the pattern is defined as the number of variables appearing in a pattern, with values that are constrained. The *prevalence* of a pattern is the fraction of positive (negative) observations covered by that positive (negative) pattern. The fraction of positive (negative) observations among all those observations covered by a positive (negative) pattern is called the *homogeneity.*

In Table 3, pattern $y_{1,3} = 0$ and y3,1 = 1 is a positive one of degree 2, covering one positive observation and zero negative observations. Thus its prevalence is equal to $\frac{100}{3}\%$ and its homogeneity is 100%. A pattern imposes a strict upper bound on the value of $a[1]$ and a strict lower bound on the value of $a[3]$, on the original dataset. When performing analysis using LAD, the *pandect* is first generated, that is, the collection of all patterns in a dataset. A pandect consists of a large number of patterns but these are reduced as many of these patterns are either subsumed by other patterns or those similar to them. This poses the need to set a number of limitations on a collection of patterns to be generated, by restricting their degrees (to low values), their prevalence (to high values), and their homogeneity (to high values) (LAD *control parameters*). These have better quality compared to those patterns with high degree or low prevalence, or low homogeneity.

Herrera and Subasi (2013) also highlighted another definition of a pattern, as a conjunction of literals (binary variables or its negation) which does not contain both a variable and its negation: $P = \Lambda_{j \in N_p}$, where $N_p \subseteq \{1, .., n\}$ and x_j is a binary variable. They defined the degree of a pattern as the number of literals (associated with features) involved in the definition of that pattern. In this study, the patterns are ultimately used to build a decision rule for classifying banks by providing financial statement ratios and CAMELS ratings that distinguish healthy banks from those that are distressed. LAD *models* are then defined as the collections of patterns sufficient for classifying the observations in the dataset. A model has to include sufficiently many positive (negative) patterns to accommodate each of the positive (negative) observations in the dataset and be covered by at least one of the positive (negative) patterns in the model. Good models are those that minimise the number of points in the dataset that are covered simultaneously by both positive and negative patterns in the model. An observation satisfying the conditions of a pattern P, that is, $P(\theta) = 1$, is said to be covered by that pattern. A single pattern generation algorithm involves finding the optimal values for a set of decision variables that minimise a certain objective function subject to a set of constraints.

3.3 The LAD Model

A LAD model generates interpretable patterns that are helpful to bank management in understanding the reasons why a their bank is in a particular credit rating. Hererra and Subasi (2013) define an LAD model as a collection of positive and negative patterns with the same characteristics as the pandect, generated in pattern generation step. The LAD model is defined by a standard LAD algorithm which uses greedy heuristics to solve the set-covering problem, to develop the LAD model. The software WEKA which was implemented in this study, uses any of the two Chvatal's greedy heuristics (Greedy Set Covering, GSC or the Iterated Sampling, IS) to produce feasible solutions to set covering problems. The algorithm generates the minimum number of patterns required to cover the training dataset. The basic assumption of LAD model is that a binary point is positive if it is covered by some positive patterns, and not covered by any negative pattern. Similarly, a binary point covered by some negative patterns, but not covered by positive patterns, is negative. This assumption facilitates the construction of the LAD model for a given dataset, where large patterns are generated and then selecting a subset of them that satisfy the above assumption. Each pattern in the model thus, satisfies certain requirements in terms of prevalence and homogeneity.

3.4 Classification of observations

An observation is classified as positive (*negative*), contained in the dataset or not, if it satisfies the conditions of some of the positive (negative) patterns in the model, and do not satisfy the conditions of any of the negative (positive) patterns in the model.

According to Boros *et al*. (2000), LAD uses a discriminant function that assigns specific weights to the patterns in the model when classifying an observation that satisfies both positive and negative patterns in the model. A simple discriminant function that assigns equal weights to all positive (negative) patterns, can be de- fined by letting p and q represent the number of positive and negative patterns in the model and h and k, represent the numbers of those positive and negative pat- terns in the model that cover a new observation θ. Then the discriminant function $\Delta(\theta)$ is

$$\Delta(\theta) = \frac{k}{p} - \frac{k}{q}$$

and the corresponding classification is determined by the *sign* of the discriminant. LAD leaves *unclassified* an observation with $\Delta(\theta) = 0$ since either the model provides contradictory evidence or the evidence provided is not sufficient. The results of classifying observations can be shown in a *classification matrix.*

Table 4
Classification matrix

	Classification of observations		
Observation classes	*Positive*	*Negative*	*Unclassified*
Positive	a	c	e
Negative	b	d	f

The percentage of positive (negative) observations that have been classified correctly is represented by a (resp. d). The percentage of positive (negative) observations that have been misclassified is c (resp. b). The percentage of positive (negative) observations that remain unclassified is e (resp. f).

$$a + c + e = 100\%,$$

and

$$b + d + f = 100\%.$$

The quality of the classification is

$$\frac{1}{2}(a+d)+\frac{1}{4}(e+f).$$

3.5 Multi-class LAD

LAD can also be adapted to a multi-class problem by applying the same algorithm used for 2-classes, to each pair of classes in the multi-class set (Mortada, 2011). Shaban (2014) also used and supported the one-versus-rest approach, that outlines that a single separator between class c_i (for some i) and all other classes is built, and as such, K different two-class classifiers are built, where K is the total number of classes. S_i becomes the i^{th} classifier separating observations in class c_i (considered to be positive) and observations in $\bar{c}_i$, not in c_i (form a set of negative observations). In the theory formulation step, the patterns generated are used to create a decision function called the discriminant. With the multi-class LAD, the discriminant is then used to generate a score for each class. A new observation will be assigned to the class with the highest score. First, a pattern/class relationship, $\mathbf{D}_{ij}$ ($N_{ij} \times 1$) is created for each class pair pattern set $\mathbf{P}_{ij}$, where Nij is the number of patterns in the set $\mathbf{P}_{ij}$. Each element $d_{ij,n}$ where $n \in \{1, 2,$

..., N_{ij}}, of the matrix, is calculated as the coverage rate of the pattern p_n in $\boldsymbol{P}_{ij}$ with respect to the observations in the class c_i (m_i), normalised by the sum of coverage rates of all the patterns in the set $\boldsymbol{P}_{ij}$ (coverage rates for each class):

$$d_{ij,n} = \frac{\frac{\mathrm{cov}(p_n, c_i)}{m_i}}{\Sigma_{n=I}^{N_{ij}} \left(\frac{\mathrm{cov}(p_n, c_i)}{m_i} \right)}$$

Every pattern p_n in a set $\boldsymbol{P}_{ij}$ is associated with an element $d_{ij,n}$ of the matrix $\mathbf{D}_{ij}$. These elements are normalised weights for the patterns.

The score for a certain class $k \in \{1, 2, ..., K\}$ is found by grouping all the class pair pattern sets that separate class k and all the (K – 1) classes. For each class k, (K – 1) classes per set are grouped such that P_{kj}, for all $j \in \{1, 2, ..., K\}$, where $j \neq k$.

For a new observation θ, the score for a class $k \in \{1, 2, ..., K\}$ is calculated by adding the elements $d_{kj,n}$ of matrix $\boldsymbol{D}_{kj}$ whose corresponding pattern p_n in set P_{kj} covers observation θ for all (K – 1) class pair sets. The maximum score obtained for one class is equal to (K – 1), which is realised when all the patterns in the sets P_{kj} for all $j \in \Delta$ {1, 2, ..., K} where $j = k$, cover observation θ.

The resulting discriminant function therefore takes the shape:

$$\Delta(\theta) = ar\ gmax_{k=1,\ldots,K} \sum_{P_n \in P_{kj}} p_n(\theta) \, . \, dj, n.$$

If a pattern covers the observation θ, then $p_n(\theta) = 1$ otherwise $p_n(\theta) = 0$. The output of the decision function for a new unclassified observation ? is the highest scoring class for that observation.

4. RESULTS

The study sample of 17 banks over 8 years, was randomly divided into the training and validation sets, using cross-validation in WEKA. Cross-validation tests the model on data that it has not seen before, and predictions are compared to the actual results. Cross-validation has 10 fold and splits the data into 10 parts, with the first 9 being used to train the algorithm, and the 10^{th} is used to assess the algorithm. Repeating this process, allows each of the 10 parts of the split dataset a chance to be the held-out test set. Generally in machine learning, a training set is used to derive patterns. Descriptive statistics were performed and presented in Table 5.

Table 5
Descriptive Statistics

Microeconomic factor	*Min*	*Max*	*Mean*	*Std Dev*
MS	15.514	21.266	18.86	1.131
Conc	240.677	452.258	356.954	42.213
ROA	-0.29	0.233	0.048	0.056
ROE	-6.164	0.771	0.133	0.634
NIM	-0.016	0.211	0.077	0.033
CS	16.219	21.38	18.942	1.009
CR	-0.207	0.246	0.015	0.036
CE	0.061	0.64	0.218	0.119
Divers	-0.081	0.855	0.466	0.189
Liq1	0.163	1.894	0.961	0.272
Liq2	0.155	0.978	0.8	0.142
CR2	0.144	0.93	0.665	0.159
LFS	0.189	2.349	0.773	0.337
Efficiency	0.111	3.715	0.85	0.461
IMI	15.332	18.719	17.283	0.754
D1	0	1	0.169	0.376
Reserve	-0.009	0.341	0.117	0.076
Crisis	-0.42	42.731	7.583	5.608
BO	1.386	4.174	2.904	0.703
RBS	0.194	30.931	5.882	6.238

ROA, ROE and NIM measure the efficiency of assets, shareholders and in general, management, in generating profits and are critical indicators of a firm's performance (Moyo *et al*., 2019), which regulators and auditors use to evaluate the wealth of a bank. The negative values of these factors (min values), report on the losses that these institutions experienced in their returns during the period of study. The standard deviation for all values is moderate across all factors and they reflect on the near to similar conditions that these institutions operate under, being in the same industry and under the same regulatory and monitoring system. As reported by Moyo *et al*. (2019), banks in Zimbabwe earn about US$0.05 net income per US$1 of total assets (mean ROA), which is unsatisfactory returns and a wakeup call for bank management. US$0.13 is also earned per US$1 of equity capital (mean ROE), and about US$0.10 for every US$1 loaned out (mean NIM).

Patterns were derived for classes 1, 2, 3 and 4 by applying the One-Versus-Rest strategy since the WEKA package can only handle problems with 2 classes at any given time.

Table 6
List of positive patterns generated by WEKA-LAD classification rules

Factor	*Class1*	*Weights*	*Class 2*	*Weights*	*Class 3*	*Weights*	*Class 4*	*Weights*
MS	≥ 20.72	0.0630	≤ 16.71	0.1220	≤ 17.10	0.0156	≤ 17.31	0.0379
			≥ 19.3	0.0081	≥ 19.43	0.0078	≥ 18.41	0.0076
Conc							≤ 299.46	0.0151
							≥ 317.61	0.0606
ROA	≥ 0.084	0.1181	≥ 0.089	0.1626	≤ −0.134	0.0859		
					≥ 0.028	0.0078		
ROE	≤ 0.38	0.0079	≥ 0.16	0.2846	≥ 0.09	0.0156		
	≥ 0.61	0.0787	≤ 0.61	0.0081	≤ 0.104	0.0703		
NIM	≥ 0.12	0.0787	≤ 0.091	0.098	≥ 0.09	0.0859		
					≤ 0.016	0.0859		
CS	≥ 9.48	0.0472	≤ 18.87	0.2114				
CR	= 0	0.0394						
Divers	≤ 0.31	0.0394	≤ 0.41	0.0570	≤ 0.015	0.0312		
	≥ 0.70	0.0157						
Liq1	≤ 0.62	0.0079						
Liq2	≤ 0.63	0.0472			≥ 0.9	0.0703		
CR2	≥ 0.83	0.0394						
Positive patterns	64		116		68		53	
Negative patterns	166		85		108		70	
Max de-gree	3		3		4		5	
Instances	136		136		136		136	
Discr.Scores		0.5826		0.9518		0.4763		0.1212

The One-Versus-Rest strategy was implemented by decomposing the data analysis problem into 4 distinct classification problems, each of which would stack one of the four classes against the collective of the remaining three classes. There was "class 1 versus class 0", in which all observations that originally belonged to classes 2, 3, and 4 were reassigned to an artificial 0-class, and so on, for classes 2, 3, and 4.

The output of the LAD model presents the interactions between the values of selected variables (which form the patterns), and their effect on the outcome (the credit ratings), as well as extracting the important variables as well as their threshold values, from the model. The importance of a variable is known as the ratio of the frequency of patterns containing the variable to the number of all generated patterns. Factor *M S* is the most important factor on credit rating for all classes with the highest importance for all classes. It appeared 48 times in 64 positive patterns in class 1 (0.75), 115 times in 116 positive patterns in class 2 (0.9914), 36 times in 68 positive patterns in class 3 (0.53) and 47 times in 53 positive patterns in class 4 (0.8868). This then goes on to show that capitalisation plays an important role for all banks in all classes but the threshold values differ per class. *M S* is most critical for banks in class 2 and class 4. This could be explained by the fact that class 2 banks needs a good capital base to upgrade to be class 1 banks, or to maintain their and avoid downgrading to class 3 (less than satisfactory banks). Class 4 banks are distressed banks and would also need a stronger capital base for them to upgrade to class 3. For a bank to be in class 1, its M S should be greater than 20.72 though there are banks with M S less than 16.71 and some with *M S* greater than 19.3 that are in class 2.

The normalised weights ($d_{ij,n}$) for these threshold boundaries were also calculated for all threshold values and M S has more weight for class 2 banks, followed by class 1 then class 4 and lastly class 3 (showing order of priority). For class 1, the order of importance of factors on credit rating is *M S, ROA, ROE, CS, NIM, Divers, CR2, Liq2, CR* and lastly *Liq1*. For class 2, *M S, ROA, ROE, CS, NIM*, and Divers in their order of importance, are the only factors that determine if a bank is in that class. For class 3, *ROA* is the most important factor followed by *M S, NIM, ROE*, *Liq2* and Divers. This order differs from the refined order provided by the normalised weights assigned to these factors. The importance measure de- scribes the general importance of a factor, regardless of its optimal threshold value, whilst the weights describe the coverage or importance of the factor associated to the threshold boundary. For class 4, only the *M S* and *Conc* are important factors describing distressed banks. Both these factors are capitalisation factors, and so this means that distressed banks are best defined using their capital levels. Summing up the weights produced the discriminant scores for each class. Class 2 had the highest discriminant score showing that a large pool of banks fall in class 2 followed by class 1, class 3 and lastly class 4. The patterns generated also uncovered the class identifiers, which is a factor that distinguishes a class from other classes by, and only with its value, regardless of other factors. Factors *Liq1* $\leq$ 0.62 and CR2 $\geq$ 0.83 are only in class 1 and not in other classes, hence *Liq1* and *CR2* are class identifiers for class 1. *Liq2* is a class identifier for class 1 and class 3. The most influential factor is M S with the highest frequency of appearing in patterns and the least influential factor is *Liq1*.

All remaining analyses were performed on the validation dataset and the performance measures of the LAD models were summarised in Table 7.

Table 7
Performance Measures of the LAD Model

Rating class	*1*	*2*	*3*	*4*
Accuracy	79.41%	66.18%	75.00%	86.03%
Kappa	0.212	0.302	0.199	-0.075
MAE	0.289	0.443	0.442	0.356
RMSE	0.472	0.633	0.633	0.536
Coverage Rate	93.38%	90.44%	94.12%	97.06%
Sensitivity	0.333	0.652	0.429	0.000
Specificity	0.878	0.682	0.809	0.921
Precision	0.794	0.666	0.722	0.854
Recall	0.794	0.662	0.750	0.860
TP Rate	0.794	0.662	0.750	0.860
FP Rate	0.583	0.367	0.574	0.932
ROC Area	0.509	0.662	0.573	0.372
PRC Area	0.713	0.648	0.654	0.819
Quality	39.71%	33.09%	37.50%	43.02%

The accuracy measures of the LAD models on the 4 classes were reasonably high and acceptable. The LAD model for class 4 was the best decision model as it had the highest classification accuracy, followed by the model for class 1, class 3 and lastly class 2. The classification quality followed the same trend as in the accuracy, for the 4 classes. The classification quality level were pretty low most probably because of the influence of negative patterns that were generated from the combined classes (the artificial 0-class). The Kappa statistics are greater than 0 for classes 1, 2 and 3, which means that the developed classifier (using the ground truth) performs better than a random classifier (measured by the expected accuracy). With classifiers that are in complete agreement, the Kappa statistic will be equal to 1, and if there is no alignment between the classifiers other than what would be expected by chance, Kappa will be less than or equal to 0. Class 4 has a negative Kappa which is close to 0 hence the developed classifier is not in agreement with a random classifier, maybe because of the number of observations in class 4 which was a small fraction as compared to observations in other classes. The confusion matrices were found as shown below.

	a	*b*	
Class 1:	7	14	$\| a = 1$
	14	101	$\| b = 0$

	a	*b*	
Class 2:	60	14	$\| a = 1$
	32	30	$\| b = 0$

	a	*b*	
Class 3:	9	22	$\| a = 1$
	12	93	$\| b = 0$

	a	*b*	
Class 4:	0	10	$\| a = 1$
	9	117	$\| b = 0$

All performance metrics are drawn from the confusion matrices which outline the number of observations that were classified correctly or misclassified, for each class. The ability of a model to classify banks in the positive 1-class (targeted class), was reflected by the TP Rate (true positive rate), which was highest for class 4, followed by class 1, class 3 and lastly class 2. The ROC area (which measures discrimination, that is, the ability of the classifier to correctly classify) was more than 50% for classes 1, 2 and 3 (more accurate classification) but less than 50% for class 4. This measure on class 4 contradicts with the output of the accuracy measure that had class 4 with the best decision model. This could be due to the fact that accuracy is based on one specific cut-point meanwhile ROC tries all the cut-points and plots sensitivity and specificity. Nonetheless, the accuracy measure is superior to the ROC measure and hence class 4 still has the best decision model. The MAE (mean absolute error) was small for all classes reflecting on the prediction efficiency of the decision models, though the RMSE (root mean squared error) has values that are sufficiently larger than the MAE values since the MAE measures the spread of the errors around the mean whilst the RMSE measures the difference between the predicted values and the actual observed values. The errors are within the acceptable range, and hence all the decision models have good predictive power and demonstrate the robustness of the proposed classification system.

5. DISCUSSION

The LAD model built performs well in classifying the banks into the four classes, and it has varying levels on the effect of the predictor variables on each credit rating. Kick and Koetter (2007) had their findings supporting this outcome, as they outlined that the individual impact of each banks' balance sheet item differs across banks' credit risk rating categories. The factor MS was present in classifying banks into all classes, which shows the level of distress within Zimbabwean banks. According to the findings of Mortada *et al* (2011) in their paper and the conclusions they reached, it can also be generalised in this work that when a bank has factor MS as an indicator for classification model, then the bank has some level of distress which initiates further investigations for the internal credit risk department, on how to mitigate the risk at an early stage as reported by the results. Estrella *et al*. (2002) stated that the combinatorial nature of the LAD model makes it possible to capture high order interaction governing complex and volatile systems, which gives it strength to be used as an early warning system. The decision models built can thus be used to detect weak banks, monitor and rate banks on a proactive basis, at a frequency desired, without waiting for external credit rating agencies modifications and avert the risk of systematic bank crisis.

Hammer *et al*. (2011) outlines that an early warning system can reduce financial costs of a bank crisis as it monitors the emergence of a bank failure well before it actually occurs. The risk of spillover across the whole financial system and economy, due to interbank linkages is also monitored. This usually leads to bank crisis as an insolvent bank precipitates financial distress to its counterparts. They further alluded that credit risk rating systems are vital in the banks' operations as they are used in the loan approval, management reporting, pricing, determination of the covenants and collaterals of a credit line, limit setting and loan loss provisioning as well as setting regulatory capital. Again, to take into cognizance is that the credit risk rating affects each and every decision and operation of the financial institution throughout the life cycle of the granted credit, hence the need for a robust credit risk rating model.

6. CONCLUSION

LAD produced high accuracy measures in detecting the threshold values and char- acteristic patterns for banks conditions. The credit ratings provided by the RBZ were determined using the CAMELS rating system methodology and these have been used to group banks in executing the LAD modeling procedure which has since produced discriminant scores per class or grouping of banks. The discriminant scores calculated show the general status of most banks in Zimbabwe, just satisfactory, with moderate weakness being managed well and just being able to identify and mitigate risks accordingly. Nothing is exceptional in the performance of the general populace of the banks. The LAD thus provides measures and targets that can be aimed by banks so as to upgrade to better classes and attain the most in terms of profitability, and depart as much as possible from distress.

REFERENCES

Abreu, M. and Mendes, V. (2002), "Commercial Bank Interest Margins and Prof- itability: Evidence from E.U. countries", *Econometrica*, **26**(3),393-415.

Alexe, G., Alexe, S., Bonate, T.O. and Kogan, A. (2007), "Logical analysis of data - the vision of Peter L. Hammer", *Annals of Mathematical Artificial Intelligence* **49**, 265–312.

Alper, D. and Anbar, A. (2011), "Bank Specific and Macroeconomic determinants of commercial bank profitability: Empirical evidence from Turkey", *Business and Economics Research Journal*, **2**(2),139-152.

Baesens, B., Rosch, D. and Scheule, H. (2016), "Credit risk analytics: Measurement techniques, applications and examples in SAS", *John Wiley and Sons, New Jersey*.

Basel Committee on Banking Supervision (2001), "The internal ratings based approach", *Basel Committee on Banking Supervision*.

Basel Committee on Banking Supervision (2004), "Bank failures in mature economies", *Basel Committee on Banking Supervision, Working Paper 13*.

Basel Committee on Banking Supervision (2001), "International convergence of capital measurements and capital standards: a revised framework (Basel II)", *Basel Committee on Banking Supervision*.

Beck, T., Demirguc-Kunt, A., and Maksimovi, V. (2005), "Financial and Legal Constraints to Growth: Does firm size matter?", *The Journal of Finance* **60**(1),137- 177.

Boros, E., Hammer, P.L., Ibaraki, T. and Kogan, A. (1997), "Logical analysis of numerical data, *Mathematical Programming", **79,*** 163–190.

Boros, E., Hammer, P.L., Ibaraki, T., Kogan, A., Mayoraz, E. and Muchnik, I. (2000), "An implementation of logical analysis of data", *IEEE Transactions on Knowledge and Data Engineering* **12**(2), 292–306.

Crama, Y., Hammer, P.L. and Ibaraki, T. (1988), "Cause-effect relationships and partially defined Boolean functions", *Annals of Operations Research*, **16**, 299-326.

Estrella, A., Park, S. and Perstiani, S. (2002), "Capital ratios and credit ratings as predictors for bank failures", *Federal Reserve Bank New York: Economic Policy Review,* 32-52.

Hammer, P.L. (1986), "Partially defined Boolean functions and cause-effect relationships", *International Conference on Multi-Attribute Decision Making Via OR-Based Expert Systems*, University of Passau, Passau, Germany.

Hammer, P.L., Kogan, A. and Lejeune, M.A. (2006), "Modeling country risk ratings using partial orders", *European Journal of Operational Research* **175**(2), 836–859.

Hammer, P.L., Kogan, A. and Lejeune, M.A. (2011), "Reverse- engineering country risk ratings: a combinatorial non-recursive model", *Annals of Operational Research* **188**, 185–213.

Hammer, P.L., Kogan, A. and Lejeune, M.A. (2012), "A logical analysis of banks' financial strength ratings", *Expert Systems with Applications* **39,** 7808–7821.

Herrera, J.F.A. and Subasi, M.M. (2013), "Logical Analysis of Multiclass data", Rutcor-Rutgers Center for Operations Research, RRR 5-2013.

Hoffman, P.S. (2011), "Determinants of the profitability of the US Banking Industry", *International Journal of business and Social Science* , **2**(22), 255-269.

Khrawish, H.A. (2011), "Determinants of Commercial Banks Performance: Evidence from Jordan", *International Research Journal of Finance and Economics* **81**, 148-159.

Kick, T. and Koetter, M. (2007), "Slippery slopes of stress: Ordered failure events in German banking", *Journal of Financial Stability* **3,** 132-148.

Kogan, A., and Lejeune, M.A. (2010), "Combinatorial methods for constructing credit risk ratings", *Handbook of Quantitative Finance and Risk Management*, Springer Science + Business Media e-ISBN 978-0-387-77117-5.

Mortada, M.A., Yacout, A. and Lakis, A. (2011), "Diagnosis of rotor bearings using logical analysis of data", *Journal of Quality in Maintenance Engineering* **17**(4), 371-397.

Moyo, I.L., Gumbo, V., Chikodza, E. and Jones, B. (2019), An empirical study on the factors influencing the performance of financial institutions in Zimbabwe, *International Journal of Finance and Banking Studies*, **8**(4) https://doi.org/10.20525/ijfbs.v8i4.51