Predicting a bank’s credit rating change using financial data and public financial news

Dingilizwe Jacob Nkomo,
Lecturer,
Harare Institute of Technology, P. O. Box BE 277, Harare, Zimbabwe,

Washington Chiwanza,
Lecturer,
Harare Institute of Technology, P. O. Box BE 277, Harare, Zimbabwe,

Walter Gachira,
Lecturer,
Harare Institute of Technology, P. O. Box BE 277, Harare, Zimbabwe,

Runesu Chikore,
Lecturer,
Harare Institute of Technology, P. O. Box BE 277, Harare, Zimbabwe,

ABSTRACT
This research integrated published financial news with accounting ratios to measure the viability of increasing the discriminatory power of a credit rating model. The aim was the utilization of information in news articles to create additional predictors of credit rating changes. Modal sentiment in news articles were integrated with calculated financial ratios from the published financial statements. The integrated data points were then analyzed in Eviews resulting in a model which was then used in determining credit rating. The model was found to be stable and having a good discriminatory power. Credit ratings of sample banks were then determined as at 31 December 2013. The output from the model was found to be almost the same as ratings assigned by GCR and RBZ. Data points from outside the sample were used to validate capability of the model in measuring ratings of banks outside the sample.

Key words: Credit Rating, News Sentiments, Discriminatory Power

Contribution/Originality
The credit ratings play an important role in today’s financial markets, where high-quality ratings are a basic requirement for a financial market to function properly. The credit ratings are of high significance to both rated banks, depositors and the regulators who will be in need of assessing the financial strength of the bank. Motivated by deficiency of existing credit rating research, this study aims at utilizing the rich information in news articles to create additional predictors of credit rating changes.

I. INTRODUCTION
The major cause of the global financial crisis was the use of structured financial products whose credit ratings were biased due to the use of flawed computer models by the credit rating agencies (Blinder, 2007). This necessitated the improvement in the rating criteria that is used by the credit rating agencies (CRAs) in an effort to help manage credit risks in the banking sector.
example of evidence of changes in rating criteria was the downgrading of 73% of formally triple–A rated mortgage-backed securities in 2006 to junk in 2010 (Financial Crisis Inquiry Commission, 2011). Credit Rating Agents (CRAs), themselves, are forced by market conditions to produce ratings that are not volatile (European Central Bank, 2004). In generating stable ratings, the credit rating agents’ rating models will be based on the through-the-cycle rating philosophy which avoids sentiments like news sentiments since the sentiments increases the volatility of the ratings. The through-the-cycle rating philosophy is most appropriate when the ratings will be used for long term investment decision purposes. This differs for the Zimbabwean banking system in which the balance sheet of banks comprises mainly of short term deposits and short term loans. This study aims at utilizing the rich information in news articles to come up with a quantitative credit rating model which incorporates financial data and public financial news. The research is designed to implement a firm-level news analysis system to study whether news coverage, topics and sentiment can be used to model the changes in rating agencies’ future credit rating assignments. The following null hypothesis will be tested which is that the combination of public news with financial accounting data can improve the predictive power of credit rating models against an alternative hypothesis that the combination of public news with financial accounting data cannot improve the predictive power of credit rating models.

Credit Rating of Banks in Zimbabwe

It is mandatory for all the banks in Zimbabwe to be rated by an independent credit rating agency. These banks are obligated to publish their ratings in line with the Banking Act (Chapter 24:20) as read with Banking Regulations Statutory Instrument 205 of 2000, so as to promote transparency, accountability, and effective market discipline. The credit ratings assigned by the credit rating agency to an individual bank could affect the banks’ cost of funds after they are published. Credit ratings are also a valuable tool for depositors, debtors and regulators in assessing the financial strength of banks. The results of the banks’ respective credit rating are usually published on the semi-annual basis when the banks’ semi-annual results are announced. The external CRA that rates most banks in Zimbabwe is Global Credit Rating, (GCR). The rating results that are published by GCR usually expire after 12 months from the issue date, thereby, mandating the rated banks to seek for another rating from the external agent. The Reserve Bank of Zimbabwe, (RBZ) also rates banks in Zimbabwe. The Banking Act [Chapter 24:20] empowers the RBZ to monitor, supervise and investigate efficiency of banking institutions in Zimbabwe.

II. LITERATURE REVIEW

The Determinants of Credit Rating

According to Adams (2003), the two major determinants of rating grade are financial ratios and corporate governance mechanism. Ashbaugh-Skaife (2006) states that most rating processes concentrate on the financial ratios because they are relatively easy to analyze compared to corporate governance which calls for great analytical skills and is also bound to manipulation by the researchers.

Financial Ratios

Financial ratios can be used to depict the performance of any business. These ratios are derived from the items in the financial statement (Adams, 2003). The financial ratio variables proxy firm specific factors such as leverage, liquidity and firm size (Ederington (1993), Kamstra et al.
Key ratios considered in coming up with ratings are categorized into five categories which are: size (Ohlson, 1999), financial leverage (Altman, 1968), profitability (Altman, 1968), interest leverage (Huang et al., 2004) and liquidity ratios (Ohlson 1999).

**Corporate Governance**

Unlike the financial ratios which measure the firm performance, the corporate governance determines the firm performance (Shleifer and Vishny, 1997). Corporate governance refers to the system of rules, practices and processes by which a company is directed and controlled. The governance structure specifies the distribution of rights and responsibilities among different participants in the corporation (such as the board of directors, managers, shareholders, creditors, auditors, regulators, and other stakeholders) and specifies the rules and procedures for making decisions in corporate affairs. The factors that are usually considered under corporate governance are: ownership structure and board independence (Bhojraj, 2003 and Ashbaugh-Skaife et al. 2006). According to Bhojraj, (2003), firms with strong governance should receive a higher rating. Similarly mechanisms that induce firms to disclose information in a timely and transparent manner should reduce information risks and therefore improve a firm's rating.

**Economic Impact of Text Data in Mass Media**

Fang and Peress (2009) documented that firms with no news coverage earn higher returns than the firms with high news coverage. They documented this after their study on the cross-sectional relation between media coverage and expected stock returns on all companies listed on NYSE. Another similar study was conducted by Keremati et al (2011), who documented the existence of strong price drift after the bad news is published. It was found that the news coverage effects last for months. Lau et al. (2011) realized the importance of text mining modules in extracting relevant information from the full text of news articles. Their research was based on the text articles that are extracted from websites. Two techniques were adopted to achieve their goal of extracting information from the websites. The first technique adopts dictionaries that contain lists of positive and negative words. This dictionary based approach is straight forward to implement and can often produce good results (Tetlock et al. 2008). The second approach used by Lau et al. (2011) adopted machine learning techniques to construct classifiers that can be used to predict sentiments of the texts. It was found that the Opinion-Finder system provides a synthetic approach that refines the sentiment recognition process by considering the text around known sentiment words. This approach delivers better performance by combining the benefits of a keyword-based approach and machine learning techniques (Wilson et al. 2005).

**III. DATA AND METHODOLOGY**

The researchers used annual data for the banks in Zimbabwe from 2009 to 2013. Data of eight banks (five commercial banks, one building society, one merchant bank and one savings bank) which are all operational in Zimbabwe was used.

<table>
<thead>
<tr>
<th>Type of bank</th>
<th>Names of chosen banks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Banks</td>
<td>CBZ; Met Bank; Barclays Bank, Standard Chartered Bank and NMB</td>
</tr>
<tr>
<td>Merchant Bank</td>
<td>Tetrab Investment Bank</td>
</tr>
<tr>
<td>Building Society</td>
<td>FBC Building Society</td>
</tr>
<tr>
<td>Savings Banks</td>
<td>POSB</td>
</tr>
</tbody>
</table>

Table 3.1: The selected banks to be used as a sample
The Ordered Logistic Model
Since the logistic regression model has the least requirements regarding certain statistical assumptions compared to the other alternative methods models (discriminant analysis model, probit model, neuron networks), the researchers used the ordered logistic model to predict the bank’s likely credit grade. The ordered logistic regression is an extension of the binomial logistic regression model under which the dependent variables have more than two possible outcomes (Brown, 2001; Sohn et al, 2007). The logistic regression constitute of a linear regression equation which can be expressed as shown below:

$$Z_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_n X_{i,n} + \epsilon_i \quad \text{equation 3.1}$$

Where:

- $X_{i,j}$ is the $j^{th}$ independent variable coming from bank $i$. These explanatory variables are a combination of both the quantitative and qualitative factors.
- $Z_i$ is the output from the linear regression component of the ordered logistic regression model. This variable will be transformed (as shown in equation 3.2) to attain the rating of a bank.
- $\beta_i$ is the coefficient of the $i^{th}$ predictor variable. The coefficients measure the effect on the odds of bank failure of a unit change in the corresponding independent variables.
- $\beta_0$ is the $y$-intercept.
- $n$ refers to the number of financial indicators included in the scoring function.
- $\epsilon_i$ is the error term for the $i^{th}$ predictor variable. Because $P_i$ can take on only the values of 0 and 1, the error $\epsilon_i$ is dichotomous as well.

The left hand-side-value from equation 3.1 above can take any value from $-\infty$ to $+\infty$ which make it difficult to give a rating scale in that range. To overcome this problem of linear probability model, the logistic function is used such that the output values are in the interval between 0 and 1 instead of the interval $-\infty$ to $+\infty$ as is in the linear regression model. The outcome logistic regression model could be shown as in the equation 3.2 below:

$$P_i = \left[ \frac{1}{(1 - e^{-Z_i})} \right] \quad \text{equation 3.2}$$

The graphical representation of the logit model outcome is shown in Graph A in the Appendix. Hence, by fitting equation 3.1 into equation 3.2, the full logistic regression equation would appear as below:

$$P_i = \left[ \frac{1}{1 - e^{-(\beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \ldots + \beta_n X_{i,n} + \epsilon_i)}} \right] \quad \text{equation 3.3}$$

Where log is the natural logarithm and $P_i$ is the probability that bank $i$ will fail in the next period. Since the ordered logistic model is non-linear, its parameters cannot be estimated using the ordinary-least-square method. Hence, the researchers made use the maximum likelihood method to estimate the parameters of the model. After the estimation of the parameters, it was possible to
calculate the probability of occurrence of each possible outcome, both within the sample and out-of-sample.

The Maximum Likelihood
The estimation of parameters using the maximum likelihood initially involved the assumption of the distribution of the data. For each training data-point, there exist a vector of features, $\mathbf{x}_i$, and an observed class $y_i$. The probability of that class is either $p_i$, if $y_i = 1$, or $1 - p_i$, if $y_i = 0$. The likelihood is then:

$$L(\beta_0, \ldots, \beta_k) = \prod_{i=1}^{n} P_i^{y_i} (1 - P_i)^{1-y_i} \quad \text{equation 3.4}$$

After fitting the logistic regression model, an assessment of the significance of the overall model with p coefficients for the predictors included in the model was done. This test was done through the use of the Likelihood Ratio Test which tests the joint null hypothesis that all slope coefficients except the constant are zero i.e. $H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$. A small p-value will lead to the rejection of $H_0$ and the concluding that at least one (or more) of the p coefficients are different from zero. This statistic is computed as:

$$LR\text{ Statistic} = -2[l(\hat{\beta}) - l(\beta)] \quad \text{equation 3.5}$$

Where $l(\beta)$ is the restricted log likelihood (the model constant only) and $l(\hat{\beta})$ is the chosen model with p predictors. This statistic helps in testing the overall significance of the model.

Dependent Variable and Independent variables
The dependent variable is the creditworthiness indicator assigned by RBZ. This dependent variable consists of 5 rating grades (from 1 to 5) of the scores previously assigned to a bank by RBZ. The independent variables that were used in this research were deduced from the reviewed literature. They include both qualitative and quantitative information and were categorized into two categories which are (i) the category of financial ratios and (ii) the category of public news.

Public News Analysis
The first step in this research is to use the text mining modules to extract relevant information from the full text of news articles. The researchers used the machine learning techniques to construct classifiers that can be used to predict sentiments of text in the analyzed news article. The researchers specifically used the Opinion-Finder, a machine based text mining software, due to the fact that its use delivers better performance by combining the benefits of a keyword-based approach and machine learning techniques.

The use of text mining modules to extract relevant information from the full text of news articles can be divided into three modules which are: (i) firm name extraction, (ii) topic clustering and (iii) sentiment analysis. Through combining the outputs from the modules, the research obtained the counts of positive ($pos_{sent}$) and negative sentence ($neg_{sent}$) in a news article, its topic, and the public firm IDs associated with the news articles. The positive (negative) sentiment score of an article $z$ is the fraction of positive (negative) sentences:

$$pos_z = \frac{pos_{sent}_z}{total_{sent}_z}, \quad \text{equation 3.6}$$
$$neg_z = \frac{neg_{sent}_z}{total_{sent}_z}, \quad \text{equation 3.6}$$
Combining Accounting Data and Financial News
The public news extracted from the Opinion-Finder is integrated with the financial ratios from the financial statements using the logit model to predict the rating change of a financial institution. The major reason for converting the qualitative news into a quantitative number is to enable the news to be put in the logit model as inputs.

Calibrating the Rating Model
The researchers calibrated the rating model with the objective of assigning a default probability to each possible score from the designed model. The default probabilities were classified into 5 rating classes, which are identical to the 5 tier rating class that is used by RBZ. In order to assign the model’s output in accordance to the 5 tier rating scale, the researchers used the ordinal scale which is capable of assigning the ratings in ranks. Instead of dividing the range of \( P_i \) into two regions to produce a dichotomous response, the range of \( m \) is dissected by \( m - 1 \) boundaries or thresholds into \( m \) regions. Denoting the thresholds by \( \mu_0 < \mu_1 < \mu_2 < \ldots < \mu_{m-1} \), and the resulting response from mapping the ratings were done as shown on equation 3.17 below:

\[
P_i = \begin{cases} 
0 & \text{if } P_i^* \leq \mu_0 \\
1 & \text{if } \mu_0 < P_i^* \leq \mu_1 \\
2 & \text{if } \mu_1 < P_i^* \leq \mu_2 \\
\vdots & \text{......equation 3.7} \\
m - 1 & \text{if } \mu_{m-2} < P_i^* \leq \mu_{m-1} \\
m & \text{if } P_i^* > \mu_{m-1} 
\end{cases}
\]

Where \( P_i^* \) is an unobserved continuous random variable representing the rater’s risk evaluation of issuer i, \( P_i \) is the observed rating category by a rating agency for issuer i, \( Z_i \) is the linear-regression component of a logistic regression model, \( \mu_i \) denotes threshold parameters (cut-off points). Based on the calculated thresholds the researchers assigned rating grades to the banks depending on the calibrated scale.

IV. DATA PRESENTATION AND ANALYSIS

News Sentiment Analysis
The qualitative outputs of negative sentiment, positive sentiment or neutral sentiment were translated into numbers of -1, +1 or 0 respectively (i.e. where 1 was for the overall positive sentiment, 0 was for the overall neutral sentiment and -1 was for the negative sentiment level in the news articles). The major reason for converting the sentiment level for each article into numbers was to allow for an easy to calculation of the modal sentiment level using excel. The modal sentiment results for each bank appeared as is shown Table A in the Appendix.

Dependent Variable Analysis
The ratings assigned by RBZ were chosen as best ratings to use as the dependent variable in the model construction process. The RBZ 5 tier rating scale had 4 data points that are described out of the 5 possible data points for the Zimbabwean banking sector case. This means that the 5 tier rating scale explains 80% of possible data points from the 5 tier rating scale.
Correlation results

The ratios that were used in this research are: (i) Capital Adequacy Ratio (CAR), (ii) non-performing loan ratio (NPLR), (iii) Earnings Per Share (EPS), (iv) liquidity ratio (LR), (v) return on equity (ROE), (vi) cash-and-cash equivalent to total assets ratio (CCTA), (vii) loans to deposit ratio (LDR), (viii) EBIT to total assets ratio (EBITTA) (ix) Return on Assets (ROA), (x) Total Loans to Total assets (TLTA). The correlation coefficient matrix of the data of the banks used in the sample from 2009 to 2012 is as shown in Table B in the Appendix. Using the guidelines for the analysis of correlation coefficients by Sharma (2007), the following interpretations were made on the correlation of the variables:

- There is a strong correlation of 84% between ROA and ROE. This indicated that the two ratios are highly related and they almost explain the same information in the model. The high correlation could be based on the fact that the two ratios are all profitability ratios and hence they almost explain the same thing. Therefore, using the two ratios that are highly correlated will result in stability problems since the model will become too sensitive to small changes in any of the two ratios (Brooks, 2008).

- There is a strong correlation of 82% between the liquidity ratio (LR) and the cash & cash equivalent to total asset ratio (CCTA). The strong correlation between these two variables implies that they cannot be jointly used in the same model.

- There is a moderately high correlation between LDR and CAR; LDR and EBITTA; LOTA and CCTA; CAR and ROA. Jointly using these variables in the same model will lead to the model being moderately unstable.

- The correlation between other correlation coefficients are either moderately low correlated or have a low correlation.

The researchers circumvented jointly using the variables that had a strong correlation. Therefore, the variables that were used in the model building process were having a low correlation. This was done with the intention of creating a stable model, based on the guidance from Brooks (2008).

Model of Accounting Ratios Only

A forward stepwise procedure increased the number of covariates until all remaining five coefficients were above the 5% significance level. The optimal model that was designed using the accounting ratios only is:

$$\log \left( \frac{P}{1-P} \right) = 1.9 \text{CAR} + 6.7\text{ROE} - 7.9 \text{TLTA} + 11.5 \text{LOTA} + 9.2 \text{LR} - 1.1 \text{EPS} - 0.7 \text{LDR} \quad \text{…equation 4.1}$$

The statistics of the model were as shown in the table below:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo R-squared</td>
<td>0.408</td>
</tr>
<tr>
<td>Akaike info criterion</td>
<td>2.178</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-29.200</td>
</tr>
<tr>
<td>Prob(LR statistic)</td>
<td>0.000001</td>
</tr>
</tbody>
</table>

Table 4.3: The statistics for the model without sentiment
Model of Financial News and Accounting Ratios
Using the forward stepwise model selection procedure and encompassing the sentiment level in each model, the optimal model appeared as:

\[
\log \left( \frac{P}{1-P} \right) = 7.2 \text{CAR} - 1.2 \text{EPS} - 1.0 \text{LDR} + 15.9 \text{LOTA} + 13.5 \text{LR} + 30.9 \text{NPLR} + 2.3 \text{ROE} - 1.1 \text{SL} - 9.3 \text{TLTA}
\]

…equation 4.2
The statistics of the model were as shown in the table below:

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pseudo R-squared</td>
<td>0.556</td>
</tr>
<tr>
<td>Akaike info criterion</td>
<td>1.885</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-21.930</td>
</tr>
<tr>
<td>Prob(LR statistic)</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Table 4.4: The statistics for the model with sentiments

Comparison of the Models
Therefore, the model with news sentiments has a smaller Akaike information criterion than the model without news sentiments. This implies that more information would have been left out when the model without news sentiments is used to describe the real creditworthiness of a bank compared to when the model with news sentiments is used. Hence the model with news sentiments has a better goodness of fit than the model without news sentiments. Moreover, the model that encompasses the news sentiments has a pseudo $R^2$ of 56% whereas the model with no news sentiments has a pseudo $R^2$ of 41%. Therefore the model with news sentiments has a high predictive power compared to the model with no news sentiments. This implies that the news sentiments can help in improving the predictive power of the model. Since the model with news sentiments has favorable statistics of a higher pseudo $R^2$ and a lower Akaike information criterion than the model without news sentiments, it can be concluded that the model with news sentiments is a better model to use when rating banks than the model without the news sentiments. Therefore, the researchers selected the null hypothesis which states that the news sentiments can improve the predictive power of the model.

Analysis of the Selected Model
The overall significance test of the model and selected individual variables were conducted to ensure robustness of the model. T tests were carried out on the independent variables to determine the significance of individual selected determinant variables in the model. All the dependent variables in the ultimate variables (containing both news sentiments and accounting ratios) were tested at a 95% confidence level. Results for the analysis are encompassed in Table 4.3 and Table 4.4. The hypothesis that was being tested is:

$H_0$: the model is significant

$H_1$: the model is not significant

The output results with a p-value of 0.00 suggested that the model is statistically significant and the coefficients are not jointly equal to zero. Hence $H_0$ has been accepted.
Model Design
Since the statistical tests has indicated that the news sentiments increase the discriminatory power of the model, the rating model that will be used in rating the banks will encompass the news sentiments. The output logistic regression model can be represented in the form:

\[ P_i = \frac{1}{1 + e^{-(7.2\text{ CAR} - 1.2\text{ EPS} - 1.0\text{ LDR} + 15.9\text{ LOTA} + 13.5\text{ LR} + 30.9\text{ NPLR} + 2.3\text{ ROE} - 1.1\text{ ISL} - 9.3\text{ TLTA})}} \]  

.. equation 4.3

Pi indicates the forecasted risk level of the bank and exists in a rate from the best anticipated level to the worst anticipated level. The ranges of Pi is shown on table 4.5. The component equations existed in ordered form. These were the critical models in the design of a rating system. The component output rating equations could be shown as below:

\[ \text{I\_RBZ} = 7.19327062482\text{ CAR} - 1.17908508941\text{ EPS} - 1.04242292561\text{ LDR} + 15.9300431259\text{ LOTA} + 13.4807115941\text{ LR} + 30.9338123881\text{ NPLR} + 2.32628819152\text{ ROE} - 1.0736222802\text{ SL} - 9.2772574973\text{ TLTA} \]  

.. equation 1

\[ \text{RBZ\_1} = \text{@CLOGISTIC}(5.32218555264 - \text{I\_RBZ}) \]  

.. equation 2

\[ \text{RBZ\_2} = \text{@CLOGISTIC}(8.27926492234 - \text{I\_RBZ}) - \text{@CLOGISTIC}(5.32218555264 - \text{I\_RBZ}) \]  

.. equation 3

\[ \text{RBZ\_3} = \text{@CLOGISTIC}(11.4551081093 - \text{I\_RBZ}) - \text{@CLOGISTIC}(8.27926492234 - \text{I\_RBZ}) \]  

.. equation 4

\[ \text{RBZ\_4} = 1 - \text{@CLOGISTIC}(11.4551081093 - \text{I\_RBZ}) \]  

.. equation 5

It can be observed that the output model comprise of 5 equations with dependent variables which are I\_RBZ, RBZ\_1, RBZ\_2, RBZ\_3 and RBZ\_4. The model comprises of no simultaneous block such that the number of independent block is 1.

Calibration of the Model
Since there was no bank with a rating 5 in the sample used, the results of the banks used in the sample were made up of 4 rating categories as is shown below. The limits had been assigned based to the 5 equations generated from the ordered equations from the EViews. The ratings have been limited into four possible rating bounds. These bounds are through the output of the regression component of the logistic regression. The possible infinite values will be captured through the use of the logarithm function. The range limits for the model designed is shown below:

\[ \begin{align*}
P_i &= \begin{cases} 
1 & \text{if } P_i^* \leq 5.32218555264 \\
2 & \text{if } 5.32218555264 < P_i^* \leq 8.27926492234 \\
3 & \text{if } 8.27926492234 < P_i^* \leq 11.4551081093 \\
4 & \text{if } P_i^* > 11.4551081093 
\end{cases}
\]  

..equation 4.4

Determination of the bank rating
The outcome of the ratings of the banks in the sample appeared as shown in Table C in the Appendix. The estimates using the model show that CBZ has a rating of 1 which implies that the bank has the highest credit quality in the sample. The rating of 1 assigned to CBZ implies that the bank has extremely low risk factors which mean that the bank has the lowest chance of defaulting on its creditors, (depositors included). The model has estimated the credit rating grades of Barclays, FBC Building Society and Tetrad Investment bank as having a risk grade of “3” (as at 31 December 2013). This implies that these banks are fairly strong. These banks have an adequate protection against risk factors and are considered to be sufficient for prudent investment. However, there is considerable variability in risk during economic cycles. The rating
results are indicating that four banks in the sample are categorised under “class 4” of the 5 tier grading scale (as at 31 December 2013). This implies that these banks (ZB Bank, NMB, MetBank and POSB) are weak and unsafe to invest with.

Level Validation Results
The level validation process involved benchmarking the ratings assigned to each bank against the rating that has been assigned to the bank by any other rating entity. The results of the of the level validation of the banks that are in the sample are shown in the Table D in the Appendix. The level validation showed that the model is capable of predicting the level of credit worthiness of CBZ, Standard Chartered Bank, Barclays bank, NMB MetBank, FBC Building Society and Tetrad Investments banks. However, the model has failed to match the results for ZB Bank. The failure of the model to depict the same credit rating with that of RBZ could have been caused by the slight differences in the state of the two models. The differences in the rating models can also be shown on the different results that have been assigned by RBZ and GCR to CBZ and to Standard Chartered Bank as well as the different ratings to Tetrad Investment Bank and to MetBank. Based on the ratings as of 2013, GCR presented that the Tetrad Investment Bank and MetBank have the same level of risk (of BB+) whereas RBZ argues that the two bank have different credit rating, and it assumes that TIB has a higher credit rating of 3 compared to MetBank which it assigned a credit rating of 4. Therefore, credit rating models are always different although highly related. This could be due to the different rating methodologies that are applied by different rating agencies different rating entities (Altman, 2006).

Backtesting using Data-points outside the Sample
The model was also validated through the use of the data points which are outside sample. The randomly selected bank outside the sample was Stanbic and it had a rating of 1 which is consistent with RBZ’s rating of 1 and GCR’s rating of AA-. Thus the designed model is capable of reproducing the ratings that will be assigned by RBZ. Hence, it can be concluded that the model is capable for predicting the creditworthiness of any banking institution operating within Zimbabwe.

Model Quality Assurance
The model has all the qualities for good rating models. In qualitative terms, it can be argued that the model has the following properties which make it to be in line with the regulatory enforcements; completeness, acceptance and consistence to regulation.

Prediction of Credit Rating
Using the determined coefficients for CBZ in 2013, the model predicted that CBZ will be have a credit score of 1 in 2014. This means that the bank will be capable of withstanding the vagaries of business conditions and is highly resistant to the outside influences such as economic instability. The rating of 1 assigned to CBZ also indicates that the bank has substantial compliance with laws and regulations and exhibits the strongest performance and risk management practices relative to its size, complexity and risk profile and give no cause for supervisory concerns.

V. CONCLUSIONS
This research has presented the importance of the published news in credit rating. Through the use of machine learning techniques to analyze the published news, this research found the
capability of integrating the sentimental level of published news with the annual financial results of a banking institution in an effort to predict the creditworthiness of a banking institution. The news sentiments improved the discriminant power as well as the goodness of fit of the model. This necessitated the null hypothesis to be accepted. The improvement of the model’s discriminatory power after integrating news with financial ratios could be based on the fact that the published news affects the day to day operation of the bank’s stakeholders. The level of sentiments circulating relative a specific bank will affect the decision of the depositors as far as banking with that bank is concerned. This is further supported by the research that was done by Peress (2009) which documented that the effects of public news stays for longer periods and in turn affects the share price of a firm. Based on the results that have been deduced from this research project, it can be concluded that it is possible to design an offsite risk surveillance system that is capable of predicting early warning systems. This surveillance system will help in recognizing the banks that will be exhibiting extremely unsafe conditions leading them to be of greatest supervisory concern. Immediate action will then be taken so as to reduce significant risk to the deposit funds.
REFERENCES

- European Central Bank (2004), Market dynamics associated with credit ratings a literature review, occasional paper series.
APPENDIX

Table A

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>Modal sentimental level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009</td>
</tr>
<tr>
<td>Commercial Bank in Zimbabwe</td>
<td>+1</td>
</tr>
<tr>
<td>(CBZ)</td>
<td></td>
</tr>
<tr>
<td>Standard Chartered Bank</td>
<td>+1</td>
</tr>
<tr>
<td>NMB</td>
<td>-1</td>
</tr>
<tr>
<td>Stanbic</td>
<td>+1</td>
</tr>
<tr>
<td>Met Bank</td>
<td>0</td>
</tr>
<tr>
<td>Barclays Bank</td>
<td>+1</td>
</tr>
<tr>
<td>Tetrad Investment Bank</td>
<td>+1</td>
</tr>
<tr>
<td>FBC Building Society</td>
<td>0</td>
</tr>
<tr>
<td>POSB</td>
<td>0</td>
</tr>
</tbody>
</table>

The modal sentimental levels in numerical terms
Source: Primary data analysis using OpinionFinder

Table B

<table>
<thead>
<tr>
<th></th>
<th>CAR</th>
<th>CCTA</th>
<th>EBITTA</th>
<th>EPS</th>
<th>LDR</th>
<th>LOTA</th>
<th>LR</th>
<th>NPLR</th>
<th>ROA</th>
<th>ROE</th>
<th>SL</th>
<th>TLTA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAR</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCTA</td>
<td>0.08</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBITTA</td>
<td>0.41</td>
<td>-0.13</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPS</td>
<td>-0.02</td>
<td>0.10</td>
<td>0.21</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDR</td>
<td>0.47</td>
<td>-0.16</td>
<td>0.47</td>
<td>-0.18</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOTA</td>
<td>-0.15</td>
<td>-0.24</td>
<td>0.01</td>
<td>-0.31</td>
<td>0.31</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LR</td>
<td>0.16</td>
<td>0.82*</td>
<td>0.00</td>
<td>0.08</td>
<td>-0.07</td>
<td>-0.19</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NPLR</td>
<td>-0.09</td>
<td>-0.06</td>
<td>0.00</td>
<td>-0.29</td>
<td>0.09</td>
<td>0.24</td>
<td>-0.15</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROA</td>
<td>0.43</td>
<td>0.05</td>
<td>0.60*</td>
<td>0.35</td>
<td>0.14</td>
<td>0.13</td>
<td>0.10</td>
<td>0.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROE</td>
<td>0.05</td>
<td>0.13</td>
<td>0.46</td>
<td>0.52</td>
<td>0.00</td>
<td>0.18</td>
<td>0.08</td>
<td>0.04</td>
<td>0.84*</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>0.29</td>
<td>-0.29</td>
<td>0.00</td>
<td>0.05</td>
<td>0.19</td>
<td>-0.10</td>
<td>-0.19</td>
<td>-0.26</td>
<td>-0.05</td>
<td>-0.06</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>TLTA</td>
<td>-0.31</td>
<td>0.34</td>
<td>-0.36</td>
<td>0.17</td>
<td>-0.25</td>
<td>0.19</td>
<td>0.13</td>
<td>-0.14</td>
<td>0.04</td>
<td>0.25</td>
<td>0.01</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The correlation matrix of the data points

COPY RIGHT © 2014 Institute of Interdisciplinary Business Research
Table C

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Bank in Zimbabwe (CBZ)</td>
<td>1</td>
</tr>
<tr>
<td>Barclays Bank</td>
<td>3</td>
</tr>
<tr>
<td>ZB Bank</td>
<td>4</td>
</tr>
<tr>
<td>NMB Bank</td>
<td>4</td>
</tr>
<tr>
<td>Met Bank</td>
<td>4</td>
</tr>
<tr>
<td>FBC Building Society</td>
<td>3</td>
</tr>
<tr>
<td>People’s Own Savings Bank (POSB)</td>
<td>4</td>
</tr>
<tr>
<td>Tetrad Investment bank (TIB)</td>
<td>3</td>
</tr>
</tbody>
</table>

The rating results from the model

Graph A

Source: All possible outcomes in the designed model
### Table D

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>Year</th>
<th>Rating</th>
<th>RBZ</th>
<th>GCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial Bank in Zimbabwe (CBZ)</td>
<td>2013</td>
<td>1</td>
<td>1</td>
<td>A+</td>
</tr>
<tr>
<td>Standard Chartered Bank</td>
<td>2013</td>
<td>1</td>
<td>1</td>
<td>AA-</td>
</tr>
<tr>
<td>Barclays Bank</td>
<td>2013</td>
<td>3</td>
<td>3</td>
<td>AA-</td>
</tr>
<tr>
<td>ZB Bank</td>
<td>2013</td>
<td>4*</td>
<td>2</td>
<td>BBB+</td>
</tr>
<tr>
<td>NMB</td>
<td>2013</td>
<td>4</td>
<td>4</td>
<td>BBB-</td>
</tr>
<tr>
<td>Met Bank</td>
<td>2013</td>
<td>4</td>
<td>4</td>
<td>BB+</td>
</tr>
<tr>
<td>FBC Building Society</td>
<td>2013</td>
<td>3</td>
<td>3</td>
<td>BBB-</td>
</tr>
<tr>
<td>Tetrad Investment Bank (TIB)</td>
<td>2013</td>
<td>3</td>
<td>3</td>
<td>BB+</td>
</tr>
</tbody>
</table>

*Rating different to the RBZ rating
Level validation of the banks historical ratings GCR and RBZ

### Table E

<table>
<thead>
<tr>
<th>Bank Name</th>
<th>Number of analyzed articles</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2009</td>
</tr>
<tr>
<td>CBZ</td>
<td>140</td>
</tr>
<tr>
<td>Met Bank</td>
<td>108</td>
</tr>
<tr>
<td>NMB</td>
<td>43</td>
</tr>
<tr>
<td>Stanbic</td>
<td>98</td>
</tr>
<tr>
<td>Standard Chartered Bank</td>
<td>107</td>
</tr>
<tr>
<td>Barclays Bank</td>
<td>129</td>
</tr>
<tr>
<td>Tetrad Investment Bank</td>
<td>111</td>
</tr>
<tr>
<td>FBC Building Society</td>
<td>74</td>
</tr>
<tr>
<td>POSB</td>
<td>98</td>
</tr>
</tbody>
</table>

*Source: Herald, Newsday, DailyNews, Financial Gazette and the Sunday Mail*