A DISCRIMINANT FUNCTION ANALYSIS APPROACH TO STOCK SELECTION IN THE CAPITAL MARKET OF NIGERIA

Nwakuya Maureen.T. (PhD) &
Ofokansi Arch-a-Michael
Department of Mathematics and Statistics
University of Port Harcourt, NIGERIA

ABSTRACT

In this paper, an approach of selecting stocks in the capital market of Nigeria based on Discriminant Function Analysis is presented. Attempt was made to classify in two groups a sample of 20 companies similar in terms of business profile (manufacturing), that traded stocks on the Nigerian Stock Exchange during the period of January 2010 to December 2015 into high value and low value group (i.e. stocks that can generate more profit and less profit respectively). Initial grouping is made according to the Earnings per share ratio and splits the sample into ‘10’ high value and ‘10’ low value group. The analysis was performed on multivariate data of the companies and their six financial ratios. The findings of the study reveal that the Price per Earnings ratio and Closing price contributes the most to selecting stocks in the capital markets of Nigeria whereas Market capitalization contributes the least. Cross validation technique applied shows 85% prediction accuracy of generalizing the discriminant function model to predicting the group of new stocks of unknown membership. Thus, the Discriminant function derived in this paper can help potential investors in predicting group membership of newly listed company when selecting stocks that can generate more profit compared to others.

Keywords: Discriminate Functions, Financial Ratios, Cross Validation, Nigerian Stock Exchange.

INTRODUCTION

Investing as defined by Investopedia (2016) is “the act of committing money or capital to an endeavor with the expectation of additional income or profit”. In the society today as a worker, there is a limit to how many hours a day you put in at work and you get paid based on the amount of time per day you put in at work. Making an investment can be achieved in several ways which include; putting money into securities (stocks), bonds, mutual funds, real estate, or most at times starting one’s own business. etc. In Stock exchanges around the world, various types of securities are traded and stock belongs among them. Investing in stocks listed in the stock market involves the buying of shares issued by a company or combination of different companies across several sectors in this market. A useful and reliable way of indicating a company’s performance and financial situation is by analyzing its financial ratios and compare with other firms with the same business profile. In the stock market, some common financial ratios include; Earnings per share, Return on Equity, Return on Assets, Payout ratio, Dividend yield, Dividend per share, Market Capitalization, Beta coefficients, Price per earnings ratio, Debt/Equity ratio, etc. When selecting stock that can give more profit which is seen as being of high value, situations arise where an investor will be interested in investigating the difference between all the companies alongside their financial ratios simultaneously. Thereby referring to the group of companies that can give
more profit as “high valued stocks” and the ones that can give less profit as “low valued stocks”.

An important Statistical technique that can achieve this is called Discriminant Function Analysis. Discriminant Analysis as its commonly called, is concerned with separating distinct sets of objects into groups and assigning new objects of unknown origin to one of the two (or more) distinct groups on the basis of measurement of some variables. Hafez et al. (2015). In this research, we will be concerned with the Two-group discriminant analysis. In general, high value stocks would have sound balance sheets that show good financial health. Jo Vu, (2011) observed that “Financial ratios are often examined to supplement an investor’s decision in equity markets”. The big problem for value investing will be in the decision making process, value minded investors are faced with the problem of selecting stocks based on their financial ratios that are of high-value so as to make more profit and diversify their portfolio. It is against this background that this study was conceived.

Aim and Objectives: The main aim of this study is to apply the method of discriminant analysis to classify and then predict stocks in the capital market of Nigeria on the basis of high value and low value.

LITERATURE REVIEW
Nature and Applications of Discriminant Analysis

Theoretically, discriminant analysis involves deriving a random variable (function) which is a linear combination of independent variables that will discriminate best between objects in the groups defined beforehand. It can also assign new objects to one group among a number of groups. Actually, two sets of techniques based on the purpose of analysis exist, i.e., Predictive discriminant analysis and Descriptive discriminant analysis.

Stevens (1996) described the distinction between Predictive and Descriptive discriminant analysis in the following way: “in predictive discriminant analysis, the focus is on classifying subjects into one of several groups (or to predicate group membership) whereas in descriptive discriminant analysis, the focus is on revealing major differences among groups.” In their research, Oghojafor et al. (2012) developed a set of discriminatory functions which helped in predicting the willingness of subscribers to drop their current service provider. Okeke & Amobi (2014) employed discriminant analysis to classify retail bank customers on the basis of users and non users, and then they identified which variables contribute to the classification. Analysis of the data used for their study showed that the seven variables and individual factors variables analyzed significantly predicted group membership. In another study still under discriminant analysis, Hur-Yagba et al. (2015) examined the usefulness of multiple discriminant analysis model in analyzing the financial health of manufacturing companies in Nigeria. The study recommended that manufacturing companies in Nigeria should use the Altman discriminant model to help them detect signs of bankruptcy several years before it occurs.

Financial Ratios its predictive power
A significant number of studies like Jo Vu, 2011; Okicic et al. 2013; Armeanu et al. 2012; Stancu & Stancu, 2014, have shown that financial ratios have predictive power, which implies it can also be used for analyzing stocks.
Some Quantitative models in Stock Selection

Discriminant analysis in Stock Selection
Considering literature on researches embarked on using discriminant analysis to select stocks. Siquera et al. (2012) in their research analyzed the cross-sectional relation between fundamental and financial variables, besides the CAPM (Capital Asset Pricing Model) and the average stock return using discriminant analysis. They examined stocks traded on the Sao Paulo Stock exchange and found that the discriminatory predictive capacity obtained a significant level of success. In a different research, Okicic et al. (2013) proposed an approach of stock selection based on discriminant analysis by investigating some fundamental variables and average stock returns on the underdeveloped capital markets of Bosnia and Herzegovina.

DATA & METHODOLOGY
Data, Population, Sample
The study used secondary data. The secondary data was collected online through NSE contact center in which the data contained published annual reports from December 2010 – December 2015 on financial ratios for firms (companies) that issue stock in NSE. The population of this study comprises of companies similar in terms of business profile (manufacturing industry) that issued stocks traded on the Nigerian Stock Exchange between 2010 and 2015. The annual data on six financial ratios was obtained for 20 companies namely:


The six financial ratios associated with the 20 samples include: Price per Earnings ratio, Market Capitalization, Earnings per Share, Dividend Yield, Payout ratio, Closing price.

Organization of Multivariate Data
Multivariate Data is a collection or measurement of p number of variables and n number objects/cases. That is p variables (j=1,2,…, p) and n objects/items (i=1,2,…,n) arranged in a rectangular array called Y of n rows and p columns. Y is called the data matrix where

\[
Y = \begin{bmatrix}
  y_{i1} & y_{i2} & \cdots & y_{ij} & \cdots & y_{ip} \\
  y_{i1} & y_{i2} & \cdots & y_{ij} & \cdots & y_{ip} \\
  \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  y_{n1} & y_{n2} & \cdots & y_{nj} & \cdots & y_{np}
\end{bmatrix}
\]
Thus the annual report on the six financial ratios for each company is similar to a multivariate data matrix with financial ratios and companies representing \( p \) variables and \( n \) objects respectively which makes the data collected suitable for discriminant analysis.

**Assumptions of Discriminant Analysis**

- The variables \( Y_1, Y_2, \ldots, Y_p \) are independent of each other.
- Groups are mutually exclusive and the group sizes are not too different from each other.
- The number of independent variables is less than the sample size.
- The variance-covariance matrix of the independent variables are similar within each group of the dependent variable.
- The independent variables follow a normal distribution

**Discriminant Analysis method**

Johnson & Wichern (1992) observed that some situations may arise where one may be interested in separating two groups of objects or assigning a new object to one of two groups. It is convenient to label the groups \( G_1 \) and \( G_2 \) respectively.

The objects are separated or classified on the basis of measurements of \( p \) random variables \( Y^T = [Y_1, Y_2, \ldots, Y_p] \). The observed values of \( Y \) differ to some extent from one group to another.

Johnson & Wichern (1992) have suggested replacing the population parameters with their sample counterparts. Suppose we have \( n_1 \) observations of the multivariate random variable \( Y^T \) from \( G_1 \) and \( n_2 \) observations from \( G_2 \) with \( n_1 + n_2 - 2 \geq p \)

The respective data matrices are \( Y_1(n_1 \times p) \) and \( Y_2(n_2 \times p) \) where \( (n_i \times p) \) indicates dimension of matrix i.e. \( n \) rows and \( p \) column, while \( Y_1(n_1 \times p) \) and \( Y_2(n_2 \times p) \) represent group 2 respectively. The sample covariance matrices \( S_1 \) and \( S_2 \) can be combined (pooled) to derive a single unbiased estimate of \( \Sigma \). In particular, the weighted average of the sample covariance matrices is computed as;

\[
S_{pl} = \frac{(n_2 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}
\]

\( S_{pl} \) is an unbiased estimate of \( \Sigma \) if the data matrices \( Y_1 \) and \( Y_2 \) contain random samples from the populations \( G_1 \) and \( G_2 \) respectively. Thus the linear discriminant function which maximally separates two populations is given as

\[
D = [\bar{Y}_1 - \bar{Y}_2]S_{pl}^{-1}Y
\]

*where* \( Y \) is the random variable matrix (independent variables)

\( S_{pl}^{-1} \) is the inverse matrix of \( S_{pl} \)

In the next section we shall cover the descriptive and predictive aspect of discriminant analysis.

**Descriptive Aspect of Discriminant Analysis**

The importance of discriminant analysis in describing group differences can be accessed by considering its eigenvalue. Citing from Rencher (2002),

Let \( H \) represent the matrix in place of \( [\bar{Y}_1 - \bar{Y}_2][\bar{Y}_1 - \bar{Y}_2]^T \)

\( E \) represent \( S_{pl} \)

and let \( a = [\bar{Y}_1 - \bar{Y}_2]S_{pl}^{-1} \)
The solutions of \((E^{-1} H - \lambda I) a = 0\) are the eigenvalues \(\lambda_1, \lambda_2, \ldots, \lambda_s\) of \(E^{-1} H\), so we rank them as \(\lambda_1 > \lambda_2 > \cdots > \lambda_s\). The largest eigenvalue \(\lambda_1\) after ranking is the maximum value of \(\lambda = a^T H a / a^T E a\) that maximally separates the means of each group and therefore describes group differences. An eigenvalue \(\lambda_1\) that is greater than 1 indicates a good model.

We may wish to know how well the variables separate the groups. This can be gotten from the canonical correlation \(r\) which serves as a means of measuring the association between the groups in the dependent variable and the discriminant function. Rencher (2002), derived the canonical correlation as

\[
r = \frac{\lambda_1}{\sqrt{1 + \lambda_1}}
\]

A high level of \(r\) within the range of values \((0 < r < 1)\) indicates a high level of association between the groups in the dependent variable and the discriminant function. Also the squared canonical correlation \(r^2 = \frac{1}{1 + \lambda_1}\) in discriminant analysis is analogous to the \(R^2\) coefficient of determination in regression analysis.

In order to test the significance of the discriminant function, we will apply Wilk's \(\Lambda\) test for two group case. As seen in Rencher (2002) p282,

\[
\Lambda = \frac{1}{1 + \lambda_1},\quad \Lambda \text{ ranges from the values 0 to 1 and the values of } \Lambda \text{ closer to 0 indicates the significance of the discriminant function.}
\]

In the presence of other independent variables, we are interested in assessing the contribution from each of the variables. We standardize the discriminant function coefficients to offset the different scales among the independent variables.

Rencher (2002) p278, has shown that the relative contribution of each independent variables to separation of two groups can be suitably assessed by comparing the coefficients \(a_r, r = 1, 2, \ldots, p\) in the discriminant function;

The standardized coefficient is of the vector form;

\[
a^* = (\text{diag} S_p) ^{1/2} a
\]

where “\text{diag}” denotes “diagonal elements of”

The absolute value of these standardized coefficients is used to rank these variables in order of their contribution to separating groups.

We shall now proceed to consider the Predictive aspect of Discriminant analysis

**Predictive aspect of Discriminant Analysis.**

Suppose we have a new object to be classified into one of the two groups, the predictive aspect of discriminant analysis enables the allocation of new objects to groups. Consider an object whose group membership is unknown, we can assign it to a group on the basis of the \(p\) measured variables, \(Y_0\), associated with this object. Applying Fisher (1936) classification procedure when there are two populations, a simple classification can be based on the discriminant function

\[
D = a^T Y = [\bar{Y}_1 - \bar{Y}_2] S_{pl}^{-1} Y_0
\]

Where \(Y_0\) is the vector of measurement on a new object we wish to classify into one of the two groups (populations).

To determine whether \(Y_0\) is closer to \(\bar{Y}_1\) or \(\bar{Y}_2\), we check if its discriminant score closer to the centroid \(\bar{D}_1\) or \(\bar{D}_2\).
We evaluate each object’s discriminant score \( D_{1i} = a^T Y_{1i} \) for each observation \( Y_{1i} \) from the first group and then \( \overline{D_1} = \frac{1}{n_1} \sum_{i=1}^{n_1} D_{1i} = a^T \overline{Y}_1 \) is centroid for first group. Similarly we obtain \( \overline{D_2} \) as centroid for second group.

Now suppose \( \overline{Y} \) is the overall mean vector matrix for each \( p \) random variables (combining group 1 and 2). We can calculate its discriminant score \( \overline{D} = a^T \overline{Y} \) as the threshold value.

Fisher (1936) linear classification procedure assigns:
- \( Y_0 \) to \( G_1 \) if \( D_0 = a^T Y_0 \) is closer to \( \overline{D}_1 \) than it is to \( \overline{D}_2 \) and
- \( Y_0 \) to \( G_2 \) if \( D_0 \) is closer to \( \overline{D}_2 \) than it is to \( \overline{D}_1 \)

**Cross Validation**

Cross validation which is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. For the purpose of this study, we shall apply the hold-out-one sample method which will be performed on times.

This method involves leaving out one object in the data and then perform discriminant analysis on the remaining \( n-1 \) objects to predict the membership of that sample that was left out. This process will be repeated \( n \) times leaving out a different sample each time.

At the end of the iteration, the prediction will be compared with the original classification to access the level of accuracy of the discriminate function of predicting new objects.

**DATA ANALYSIS**

This section presents the Discriminant Analysis results using IBM SPSS 21 and Minitab 17 software on our sample of 20 companies that are similar in terms of business profile (manufacturing) traded on the Nigerian Stock Exchange from 2010 – 2015.

Recall from the previous section that we will be analyzing the multivariate data of average annual financial ratios of listed companies from 2010 – 2015. For a sustainable characterization, we shall transform the Price per Earnings ratio, Market Capitalization and Closing Price by taking their natural logarithms respectively for each company as shown in Table 1. Also since we are using Earnings per Share (EPS) ratio as the categorical dependent variable, the first ten companies with larger values of EPS ratio will be categorized under Group “1” (high value stocks) while the remaining ten companies will be categorized under Group “0” (low value stocks) as shown in Table 1 below.

For the purpose of this study, let
- \( logMKTCAP \) represents (logarithmic) Market Capitalization
- \( YDVD \) represents Dividend Yield
- \( logCLPR \) represents (logarithmic) Closing Price
- \( PAYR \) represents Payout ratio
- \( logPE \) represents (logarithmic) Price per Earnings ratio
TABLE 1: ANNUAL 2010 to 2015 TRANSFORMED AVERAGE DATA OF FINANCIAL RATIOS FOR SELECTED COMPANIES BASED ON STOCK PRICES

<table>
<thead>
<tr>
<th>s/n</th>
<th>firm</th>
<th>logCLPR</th>
<th>LogPE</th>
<th>YDVD</th>
<th>PAYR</th>
<th>logMKTCAP</th>
<th>GROUP</th>
<th>EPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GUINESS</td>
<td>2.3153</td>
<td>1.35</td>
<td>0.031</td>
<td>0.69</td>
<td>11.4879</td>
<td>1</td>
<td>9.17</td>
</tr>
<tr>
<td>2</td>
<td>ENAMELWA</td>
<td>1.5371</td>
<td>1.18</td>
<td>0.013</td>
<td>0.18</td>
<td>9.3390</td>
<td>1</td>
<td>7.61</td>
</tr>
<tr>
<td>3</td>
<td>UACN</td>
<td>1.5893</td>
<td>0.98</td>
<td>0.045</td>
<td>0.37</td>
<td>10.8357</td>
<td>1</td>
<td>4.84</td>
</tr>
<tr>
<td>4</td>
<td>FLOURMILL</td>
<td>1.7615</td>
<td>1.13</td>
<td>0.044</td>
<td>0.48</td>
<td>11.1080</td>
<td>1</td>
<td>4.76</td>
</tr>
<tr>
<td>5</td>
<td>PRESCO</td>
<td>1.3308</td>
<td>0.73</td>
<td>0.034</td>
<td>0.20</td>
<td>10.3308</td>
<td>1</td>
<td>4.67</td>
</tr>
<tr>
<td>6</td>
<td>7UP</td>
<td>1.9593</td>
<td>1.35</td>
<td>0.033</td>
<td>0.59</td>
<td>10.7658</td>
<td>1</td>
<td>3.82</td>
</tr>
<tr>
<td>7</td>
<td>BETAGLAS</td>
<td>1.3504</td>
<td>0.87</td>
<td>0.023</td>
<td>0.14</td>
<td>10.0494</td>
<td>1</td>
<td>3.22</td>
</tr>
<tr>
<td>8</td>
<td>GLAXOSMITHE</td>
<td>1.6133</td>
<td>1.28</td>
<td>0.026</td>
<td>0.47</td>
<td>10.6089</td>
<td>1</td>
<td>3.30</td>
</tr>
<tr>
<td>9</td>
<td>CAP</td>
<td>1.5231</td>
<td>1.28</td>
<td>0.053</td>
<td>0.79</td>
<td>10.3004</td>
<td>1</td>
<td>1.80</td>
</tr>
<tr>
<td>10</td>
<td>UNILEVER</td>
<td>1.5934</td>
<td>1.47</td>
<td>0.021</td>
<td>0.61</td>
<td>11.1712</td>
<td>1</td>
<td>1.32</td>
</tr>
<tr>
<td>11</td>
<td>PZ</td>
<td>1.4710</td>
<td>1.40</td>
<td>0.019</td>
<td>0.46</td>
<td>11.0452</td>
<td>0</td>
<td>1.26</td>
</tr>
<tr>
<td>12</td>
<td>ASHAKACEM</td>
<td>1.3140</td>
<td>1.36</td>
<td>0.020</td>
<td>0.40</td>
<td>10.6642</td>
<td>0</td>
<td>1.15</td>
</tr>
<tr>
<td>13</td>
<td>BERGER</td>
<td>0.9446</td>
<td>0.92</td>
<td>0.078</td>
<td>0.64</td>
<td>9.3501</td>
<td>0</td>
<td>1.13</td>
</tr>
<tr>
<td>14</td>
<td>DANGSUGAR</td>
<td>0.9275</td>
<td>0.99</td>
<td>0.079</td>
<td>0.73</td>
<td>11.3224</td>
<td>0</td>
<td>0.84</td>
</tr>
<tr>
<td>15</td>
<td>BOCGAS</td>
<td>0.8043</td>
<td>1.00</td>
<td>0.036</td>
<td>0.40</td>
<td>9.4093</td>
<td>0</td>
<td>0.73</td>
</tr>
<tr>
<td>16</td>
<td>VITAFOM</td>
<td>0.6949</td>
<td>0.84</td>
<td>0.060</td>
<td>0.41</td>
<td>9.6237</td>
<td>0</td>
<td>0.71</td>
</tr>
<tr>
<td>17</td>
<td>INTBREW</td>
<td>1.2058</td>
<td>1.55</td>
<td>0.011</td>
<td>0.33</td>
<td>10.7004</td>
<td>0</td>
<td>0.47</td>
</tr>
<tr>
<td>18</td>
<td>SCOAB</td>
<td>0.7452</td>
<td>1.45</td>
<td>0.017</td>
<td>0.47</td>
<td>9.5580</td>
<td>0</td>
<td>0.35</td>
</tr>
<tr>
<td>19</td>
<td>ALEX</td>
<td>1.0338</td>
<td>1.72</td>
<td>0.006</td>
<td>0.33</td>
<td>9.3762</td>
<td>0</td>
<td>0.24</td>
</tr>
<tr>
<td>20</td>
<td>CUTIX</td>
<td>0.2231</td>
<td>1.01</td>
<td>0.074</td>
<td>0.74</td>
<td>9.3944</td>
<td>0</td>
<td>0.17</td>
</tr>
</tbody>
</table>

We first perform the test of equality of population covariance matrix using Box’s M test, which tests the null hypothesis of homogenous covariance matrices. And if they are homogenous, the discriminante analysis provides a meaningful result.

TABLE 2: Test of Covariance matrix

<table>
<thead>
<tr>
<th>Box's M</th>
<th>20.731</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approx.</td>
<td>.957</td>
</tr>
<tr>
<td>df1</td>
<td>15</td>
</tr>
<tr>
<td>df2</td>
<td>1304.526</td>
</tr>
<tr>
<td>Sig.</td>
<td>.500</td>
</tr>
</tbody>
</table>

According to Table 3, Box’s M test is not significant, since the p-value is greater than the level of significance 5% (ie. p-value = 0.50 > 0.05), therefore we accept the homogenous matrices hypothesis.
Testing the independent variables (financial ratios) for normality at $\alpha = 0.05$ using Minitab 17 yields the above graphs and their respective p-values indicate they are normally distributed. Using SPSS 21 software yields Table 3 which indicates the eigenvalue, which measures the difference between the groups in the discriminant function. Observe that the eigenvalue is 2.870 ($>1$) indicating that the function is a good model.
Also appearing in Table 3 above is the canonical correlation, which determines how much in percentage the function explains the discrimination between groups. We can achieve this by increasing the canonical correlation to its squared value. Thus $r^2 = 0.861^2 = 0.741$, i.e. the function explained 74.1% of the discrimination between groups.

Table 4 above shows that the discriminant function is statistically significant since $p$-value = 0.001 < 0.05, with Wilk’s lambda being 0.258 which is closer to 0 than it is to 1. This indicates that the two groups; “High value” and “Low value” stocks seem to differentiate quite well.

Observe from Table 5 below:

| Table 5: Standardized Canonical Discriminant Function Coefficients |
|-------------------------|----------------|----------------|----------------|----------------|
| Function:               | $logCLPR$      | $logPE$        | $YDVD$         | $PAYR$         |
|                         | -1.367         | 1.609          | 1.194          | -0.888         |
|                         |               |               |               | 0.515          |

The absolute value of the standardized function coefficients shows that logPE, logCLPR, are the most important variables in selecting stocks with 1.609 and 1.367. The next important predictors are $YDVD$ and $PAYR$. Notice also that logMKTCAP is the least important since it is the lowest with 0.515. Hence Price per earnings ratio and Payout ratio strongly predicts allocation of stocks to the high value or low value group in NSE, while Market capitalization has least effect in allocation.

From Table 6 above which shows the Canonical Discriminant Function coefficients, the Discriminant function can be arranged as follows:

$D = -8.798 + 0.724\times logMKTCAP - 4.584\times PAYR + 52.658\times YDVD + 5.89\times logPE - 4.195\times logCLPR$

We can determine the threshold score value using the canonical discriminant function, this threshold value will be used for the new classification of companies in the “high value” and “low value” groups. The threshold score is estimated by taking the average values of the financial ratios across companies and computing them in the function, i.e.

$D = -8.798 + 0.724\times (10.3376) - 4.584\times (0.47) + 52.658\times (0.036) + 5.89\times (1.19)$

$-4.195(1.2969) = -0.0038$

\[ \therefore D = -0.0038 \]
The threshold score is – 0.0038 and it will be used to separate companies between the two groups. After finding the discriminant function and the threshold score, one should perform centroid calculation of each group. The centroid which is the arithmetic mean of the discriminant scores of each group is shown below.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.607</td>
</tr>
<tr>
<td>1</td>
<td>-1.607</td>
</tr>
</tbody>
</table>

Thus from table 7, a stock is classified as belonging to the high value group if the calculation result of its discriminant score is lower than the threshold – 0.0038 (ie < – 0.0038). On the other hand, a stock is classified as belonging to the group of low value if the calculation result of its discriminant score is greater than the threshold – 0.0038 (ie. > – 0.0038).

Notice from Table 8 that the initial ordering of companies in Table 1 based on EPS variable is quite different from the new order in which the Discriminant score for each company was compared with the threshold value.

Table 8: Reclassification of 20 Companies

<table>
<thead>
<tr>
<th>New order</th>
<th>Firm</th>
<th>Old Order</th>
<th>Previous state</th>
<th>Current state</th>
<th>Discr. Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GUINESS</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-4.61</td>
</tr>
<tr>
<td>2</td>
<td>7UP</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>-2.237</td>
</tr>
<tr>
<td>3</td>
<td>PRESCO</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>-1.727</td>
</tr>
<tr>
<td>4</td>
<td>ENAMWELA</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>-1.675</td>
</tr>
<tr>
<td>5</td>
<td>BETAGLAS</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>-1.493</td>
</tr>
<tr>
<td>6</td>
<td>FLOURMILL</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>-1.372</td>
</tr>
<tr>
<td>7</td>
<td>GLAXOSMITH</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>-1.304</td>
</tr>
<tr>
<td>8</td>
<td>UACN</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>-1.173</td>
</tr>
<tr>
<td>9</td>
<td>CAP</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>-0.98</td>
</tr>
<tr>
<td>10</td>
<td>UNILEVER</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>-0.426</td>
</tr>
<tr>
<td>11</td>
<td>PZ</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0.166</td>
</tr>
<tr>
<td>12</td>
<td>BOCGAS</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0.593</td>
</tr>
<tr>
<td>13</td>
<td>BERGER</td>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0.603</td>
</tr>
<tr>
<td>14</td>
<td>ASHAKEM</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0.641</td>
</tr>
<tr>
<td>15</td>
<td>VITAFOAM</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>1.484</td>
</tr>
<tr>
<td>16</td>
<td>INTBREW</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>2.087</td>
</tr>
<tr>
<td>17</td>
<td>DANSUGAR</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>2.115</td>
</tr>
<tr>
<td>18</td>
<td>SCOA</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>2.278</td>
</tr>
<tr>
<td>19</td>
<td>ALEX</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>2.588</td>
</tr>
<tr>
<td>20</td>
<td>CUTIX</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>3.523</td>
</tr>
</tbody>
</table>
Cross Validation classification

In this analysis we shall use cross validation from SPSS 21 to test how accurately our predictive model (discriminant function) will perform in practice. Table 9 presents the cross validated results.

<table>
<thead>
<tr>
<th>GROUP</th>
<th>Predicted Membership</th>
<th>Count</th>
<th>%</th>
<th>Group</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Cross-validated</td>
<td></td>
<td>0</td>
<td>80.0</td>
<td>20.0</td>
<td>100.0</td>
</tr>
<tr>
<td>%</td>
<td></td>
<td>1</td>
<td>10.0</td>
<td>90.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

It can be observed that with the Cross Validated for the total sample of 20 cases, 17 (85%) overall are correctly classified. Note that this percentage is similar to the coefficient of determination, R\(^2\), in the regression model. Of the High value group 90% are correctly identified. Also, 80% of the Low value group are correctly classified.

This cross validation prediction of group membership provides a summary of how well the analysis would be at classifying new stocks that have not been included in the original sample of companies. So far, we can deduce that the discriminant analysis validates the initial grouping of stocks according to the categorical dependent variable EPS before we started the analysis.

CONCLUSION AND RECOMMENDATION

So far, we have investigated the possibility of applying Discriminant analysis to Stock selection and analysis on the capital markets of Nigeria. The results of the analysis shows that the six financial ratios applied together with the discriminant function model are significant in selecting stocks. Thus the main aim of this paper was to develop a discriminant function model that can be used to predict the group membership of a new manufacturing company that issue stock, when a potential investor is selecting which stock to buy. Note that the High value group of stocks can generate more profit. Cross validation prediction accuracy of 85% clearly indicates that the model can be reliably generalized to companies of unknown group membership. The following recommendation are made;

- Business managers should appreciate the importance of frequently subjecting their financial statements to ratio analysis to determine the financial health and profitability of their company.
- We recommend the use of financial ratios and Discriminant function analysis in selecting stocks.
- For further study in the capital markets of Nigeria, the method of Discriminant analysis can be applied to selecting stocks in other sectors such as; Oil, Health, Utilities, Banking, etc. Also more financial ratios can be added for the purpose of the study.
REFERENCES


