International Journal of Theoretical and Applied Mathematics
2016; 2(2): 100-109
http://www.sciencepublishinggroup.com/j/ijtam
doi: 10.11648/j.ijtam.20160202.21

Selection of Stocks on the Ghana Stock Exchange Using Principal Component Analysis

Abonongo John*, Oduro F. T., Ackora-Prah J.

College of Science, Department of Mathematics, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Email address:
abonongojohn@gmail.com (A. John)
*Corresponding author

To cite this article:
Abonongo John, Oduro F. T., Ackora-Prah J. Selection of Stocks on the Ghana Stock Exchange Using Principal Component Analysis. *International Journal of Theoretical and Applied Mathematics.* Vol. 2, No. 2, 2016, pp. 100-109. doi: 10.11648/j.ijtam.20160202.21

Received: July 19, 2016; **Accepted**: September 12, 2016; **Published**: December 10, 2016

Abstract: A major problem in stock selection is the use of the right procedure(s) in identifying the best stock(s). The principal component analysis was employed as a data reduction technique in selecting stock(s) that characterize each sector on the Ghana Stock Exchange. The results indicated that, among the 9 stocks in the Finance sector, only 3 stocks (CAL, ETI, and GCB) were able to characterize the sector. The Distribution sector had 2 stocks (PBC and TOTAL) among the 4 stocks characterizing the sector. The Food and Beverage sector had only FML characterizing the sector out of the 3 stocks. Also, the information Technology had CLYD characterizing the sector out of the 2 stocks. The Insurance sector had EGL characterizing the sector out of the 2 stocks. The Manufacturing sector had only 2 stocks (PZC and UNIL) characterizing the sector out of the 10 stocks and for the Mining sector, 2 stocks (TLW and AGA) among the 4 stocks were the best. In effect, the 34 stocks considered from the Ghana Stock Exchange were reduced to 12 stocks (CAL, ETI, GCB, PBC, TOTAL, FML, CLYD, EGL, PZC, UNIL, TLW and AGA). The results also indicated that the selected stocks were able to explain much of the variance in their respective sectors compared to the rest of the stocks in that same sector and thus could be considered for further analysis and probably investment.

Keywords: Principal Component Analysis, Stock Selection, Screen Plot, Uncertainty

1. Introduction

Investing on the stock market is poised with high risks and high gains, hence, it attracts a great number of investors. Also, as far as information regarding stocks is concerned, it is often complex and has a lot of uncertainty, making it difficult to select attractive stocks. Even though the selection of attractive stocks is not easy for investors, Principle Component Analysis (PCA) can guide an investor in telling attractive stocks from unattractive ones. The PCA is more suitable in studying the covariance structure of a vector time series. It is appropriate when one have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables that will account for most of the variance in the observed variables; a variable reduction procedure.

Principal Component Analysis technique has been extensively used in many studies in (e.g., [8]) described the joint structure with a model that can potentially be used for scenario estimation and analysis of the risk of interest rate-sensitive portfolios. Three variations of the principal component analysis technique to decompose global interest rate and yield curve implied volatility structure were examined, highlighting that global yield curve structure can be explained with 15 to 20 factors, whereas implied volatility structure needs at least 20 global factors, furthermore in (e.g., [8]) also used principal component analysis in the granting of loan. The result showed that the utility of principal component analysis in the banking sector to decrease the size of data, without much loss of information in (e.g., [2]) performed a selection of optimal SNP sets that capture intragenic genetic variation. Their results revealed that principal component analysis may be a strong tool for establishing an optimal SNP set that maximizes the amount of genetic variation captured for a candidate gene using a minimal number of SNP set in (e.g., [4]) used the principal component analysis in investigating the structure of light curves of RRabstar. They concluded that the principal component analysis was an effective way to account for

many aspects of RRab.

Again, the decomposition of interrelated variables into uncorrelated components makes it convenient to use in analyzing the complex structure of financial markets. It has been applied to the study of market cross-correlation and systemic risk measurement (e.g., [5]) and to produce market indices in (e.g., [1]. [7]), also used principal component analysis technique reducing from the 19 stocks to 9 stocks for Nigerian stock exchange. The main task of feature extraction is to select or combine the features that preserve most of the information and remove the redundant components in order to improve the efficiency of the subsequent classifiers without degrading their performances. The result exhibited principal component analysis merit of quantifying the essentials of each dimension for describing the variability of a data set. In (e.g., [7]) further supported the use of principle component analysis for the identification of the most essential factors and in the process, considerably reducing the number of input variables to an efficient and sufficient sets (e.g., [6]), also applied principle component analysis on daily frequency observations on stock market indexes, long term and short term rate and interest rate spot exchange for nine countries (e.g., [6]) also showed that principal components analysis may be used to reduce the effective dimensionality of the scenario specification problem in several cases in (e.g., [10]) applied principle component analysis to the Korean composite stock price index (KOSPI) and the Hangseng Index (HIS) to reduce the data points into two components and observed that the co movement stocks clusters.

Moreover, the eigenvalue one criterion (e.g., [3]) is an approach for retaining and interpreting any component with an eigenvalue greater than one (1). That is, each observed variable contributes one unit of variance to the total variance in the data set. Hence any component that shows an eigenvalue greater than one (1) is accounting for a greater amount of variance than the rest of variables and components with eigenvalue less than one (1) is accounting for less variance than had been contributed by one variable. This criterion is very useful for its ability to always retain the correct number of components especially when a small number of variables are being analyzed and the variables communalities are high (e.g., [9]) investigated the accuracy of the eigenvalue one criterion and recommended its use when less than 30 variables are being considered and communalities are greater than 70 or when the analysis is based on over 250 observations and the mean communality is greater than or equal to 60. Again, the components can be selected using the screen test, with the screen test, the eigenvalues are plotted with their associated components. The breaks between the components that appear before the break are assumed to be meaningful and are retained. Also those appearing after the break are assumed to be trivial.

The purpose of this paper is to apply the principal component analysis in selecting attractive stocks from seven sectors on the Ghana Stock Exchange. This is to provide investors with a simple technique in selecting winning stocks for investments.

2. Materials and Methods

2.1. Source of Data and Methods of Data Analysis

This paper used secondary data of 34 stocks from the Ghana Stock Exchange (GSE) and Annual Report Ghana databases comprising the daily closing prices from the period 02/01/2004 to 16/01/2015.

The daily index series were converted into compound returns given by;

$$X_t = \log\left(\frac{p_t}{p_{t-1}}\right) \tag{1}$$

Where X_t is the continuous compound return at time t, p_t is the current closing stock price index at time t and p_{t-1} is the previous closing stock price index.

2.2. Stationarity Test: PP and KPSS Tests

This paper employed two quantitative unit root tests namely; the Phillip-Perron (PP) unit root test and the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test in other to establish the existence or non-existence of unit root in the time series under study so as to be able to ascertain the nature of the process that produces the time series.

The KPSS test was used to test the null hypothesis that the data generating process is stationary, H_o: $I(0)$ against the alternative that it is non-stationary, H_1: $I(1)$. It assumes that there is no linear trend term and is given by;

$$Y_t = X_t + Z_t, Z_t\ I(0) \tag{2}$$

Where X_t is a random walk, $X_t = X_{t-1} + v_t$; $v_t \sim N(0, \sigma_v^2)$ and Z_t is a white noise series. The previous pair of hypothesis is equivalent to;

$$H_o: \sigma_v^2 = 0$$

$$H_1: \sigma_v^2 > 0$$

If H_o is true, the model becomes $Y_t = constant + Z_t$; $Z_t I(0)$ hence Y_t is stationary. The test statistic is given by;

$$KPSS = \frac{1}{T^2}\sum_{t=1}^{T}\frac{s_t^2}{\hat{\sigma}_\infty^2} \tag{3}$$

Where T is the number of observations, $\hat{\sigma}_\infty^2$ is an estimator of the long-run variance of the process Z_t.

The PP statistic test of the hypothesis:

H_o: unit root against

H_1: stationary about deterministic trend

Under the H_o of $p = 0$, the PP test Z_p and Z_τ statistics have the same asymptotic distributions as the ADF t-statistic and normalized bias statistics. The PP test is categorized into two statistics known as Phillips Z_p and Z_τ tests given by;

$$Z_p = n(\hat{p}_n - 1) - \frac{1}{2}\frac{n^2\hat{\sigma}^2}{s_n^2}\left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}\right) \tag{4}$$

$$Z_\tau = \sqrt{\frac{\hat{\gamma}_{0,n}}{\hat{\lambda}_n^2}}.\frac{\hat{p}_n - 1}{\hat{\sigma}} - \frac{1}{2}\left(\hat{\lambda}_n^2 - \hat{\gamma}_{0,n}\right)\frac{1}{\hat{\lambda}}.\frac{n\hat{\sigma}}{s_n} \tag{5}$$

$\hat{\gamma}_{j,n} = \frac{1}{n}\sum_{i=j+1}^{n} \hat{\varepsilon}_i . \hat{\varepsilon}_{i-j}$, for $j = 0$, then $\hat{\gamma}_{j,n}$ is a maximum likelihood estimate of the error terms while $\hat{\gamma}_{j,n}$ is the covariance between the error terms *j*-periods apart *for* $j > 0$.

$\hat{\lambda}_n^2 = \hat{\gamma}_{0,n} + 2\sum_{j=1}^{q}\left(1 - \frac{j}{q+1}\right)\hat{\gamma}_{j,n}$, when there exists no autocorrelation between the error terms, $\hat{\gamma}_{j,n} = 0$ for $j > 0$, then $\hat{\lambda}^2{}_n = \hat{\gamma}_{0,n}$.

2.3. Principal Component Analysis

This study employed this method in selecting stock(s) that characterized each sector. It involves a mathematical method that changes a number of correlated variables into a smaller number of uncorrelated variables known as principal components. The first principal component accounts for as much of the variance in the series (data) whereas each succeeding component accounts for as much variance in the series. Also, it is an eigenvector/value based approach employed in dimensionality reduction of multivariate data. It assists in finding patterns in data and expressing the data in a manner that highlights their differences and similarities.

Given a *n*-dimensional variable $x = (x_1, \dots, x_n)'$ with covariance matrix Σ_x, a few linear combinations of x_i can explain the Σ_x structure. If x is the monthly lag return of n assets, then the Principal Component Analysis (PCA) can be used to study the origin of variation of these *n* asset returns. Also, PCA can be applied to either the covariance matrix $\Sigma_x$orto the correlation matrix $\rho_x$of x. The correlation matrix is the covariance matrix of the standardized random vector $x^* = S^{-1}x$, where S is the diagonal matrix of the standard deviation of the components of x. Using covariance matrix, if $w_i = (w_{i1}, \dots, w_{in})'$ where $i = 1, \dots, n$, then

$$X_i = w_i'x = \sum_{j=1}^{n} w_{ij}x_j \tag{6}$$

is a linear combination of the random vector x. If x consists of the returns of n stocks, then X_i is the return of a portfolio that assigns weight w_{ij} to thej^{th}stock. By standardizing the vector w_i, we get $w'w_i = \sum_{j=1}^{n} w_{ij}^2 = 1$. From properties of linear combination, so, random variables:

$$Var(X_i) = w_i' \sum_x w_i \,, i = 1, \dots, n \tag{7}$$

$$Cov\left(X_i, X_j\right) = w_i' \sum_x w_j \,, j = 1, \dots, n \tag{8}$$

PCA assists in determining linear combination $w_i$such that X_i and X_j are uncorrelated for $i \neq j$ and the variances of X_i are as large as possible.

The first principal component of x is the linear combinations w_i such that $X_1 = w_1'x$ that maximizes $Var(X_1)$ subject to the constraint $w_1'w_1 = 1$. The second principal component of x is the linear combination $X_2 = w_2'x$ that maximizes $Var(X_2)$ subject to the constraints $w_2'w_2 = 1$ and $Cov\,(X_1, X_2) = 0$. The ith principal component of x is the linear combination $X_i = w_i'x$ that maximizes $Var(X_i)$ subject to the constraints $w_i'w_i = 1$ and $Cov\left(X_i, X_j\right) = 0$. Since the covariance matrix Σ_x is non-negative definite, it has a spectral decomposition.

Also, if $(\lambda_1\,, e_1), \dots, (\lambda_k\,, e_k\,)$ are the eigenvalues and eigenvectors pairs of Σ_x where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$. Then, the ith principal component of x is given by

$$X_i = e_i'x \sum_{j=1}^{k} e_{ij}\, x_j, i = 1, \dots, k \tag{9}$$

Also,

$$Var(X_i) = e_i' \sum_x e_i = \lambda_i, i = 1, \dots, k \tag{10}$$

$$Cov\left(X_i, X_j\right) = e_i' \sum_x e_j = 0, i \neq j \tag{11}$$

If some eigenvalues λ_i are equal, the choice of the corresponding eigenvectors e_i and X_i is not unique. In additionally we have

$$\sum_{i=1}^{k} Var(x_i) = t.x\,(\Sigma_x) = \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} Var(X_i). \tag{12}$$

Also,

$$\frac{Var(X_i)}{\sum_{1=1}^{k} Var(X_i)} = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k} \tag{13}$$

Thus, the proportion of the total variance in x explained by the ith principal component is simply the ratio between the ith eigenvalue and the sum of all eigenvalues of Σ_x. Since $t.x\,(\rho_x) = k$, the proportion of variance explained by i^{th} principal component becomes $\frac{\lambda_i}{k}$ when the correlation matrix is used to perform the PCA. The results of the PCA is that a zero eigenvalue of Σ_x or ρ_x, indicates the existence of an exact linear relationship between the components of x. If the smallest eigenvalue $\lambda_k = 0$, then $Var(X_k) = 0$. Hence, $X_k = \sum_{j=1}^{k} e_{kj}x_j$ is a constant and there are only $k - 1$ random quantities in x, therefore the dimension of x can be reduced.

3. Results and Discussion

3.1. Descriptive Statistics

From Table 1, it is evident that, the Finance sector had seven of the mean returns found to be positive, ranging from 0.0006 to 0.0022 and two of the mean returns were found to be negative (-0.0006 to -0.0003). Volatility (standard deviation) was high in ETI (0.0646) with the least found in HCF (0.0124). The highest and least mean returns were found in ETI and TBL respectively. The variability between risk and returns as a measure of the coefficient of variation (CV%) ranges from -7144.1700 (SOGEGH) to 7749.5900 (ETI). Also five mean returns were positively skewed (4.6600 to 28.3400) and the rest four negatively skewed (-20.8100 to -0.1700) and the kurtosis was high ranging 108.5460 to 850.2200. The Distribution sector had three of its mean returns strictly positive (0.0001 to 0.0017) with the exception of PBC (-0.0019). MLC and PBC had the highest and least mean returns respectively. The sector had high volatility in MLC (0.0582) with the least found in GOIL (0.0210). Also the sector exhibited variability ranging 230.6400 (PBC) to 16906.9400 (GOIL). Two mean returns were positively skewed (1.9100 to 9.4500) and the other two negatively skewed (-13.9600 to -1.0800). The kurtosis was

high ranging from 132.8100 to 363.0600.

The Food and Beverage sector has two positive mean returns, ranging from 0.0008 to 0.0012 with the exception of CPC (-0.0005). FML and CPC had the highest and least mean returns respectively. The sector exhibited high volatility in CPC (0.0458) whereas GGBL (0.0155) exhibited low volatility. The CV% ranged from -14476.9400 (CPC) to 1953.0700 (GGBL). Also two out of the three mean returns were negatively skewed (-3.5600 to -0.0300) and the kurtosis was high ranging from 11.0700 to 71.0900. The Information Communication Technology sector has the two negative mean returns, ranging from -0.0002 to -0.0001. The sector recorded a higher volatility in TRANSOL (0.0352) and low volatility in CLYD (0.0260). The sector had CV% ranging from -95087.0400 (TRANSOL) to -24856.2300 (CLYD). Also this sector has all the two mean returns positively skewed. The kurtosis ranged from 32.8200 to 79.8900. Also, the Insurance sector has its two mean returns positive (0.0002 to 0.0010). Volatility was high in EGL (0.0380) than SIC (0.0304). The sector had CV% ranging 3159.3300 (SIC) to 16299.1900 (EGL). The sector exhibited negative skewness in EGL (-16.8800) and positive skewness in SIC (24.3700). Also the sector had kurtosis ranging from 347.1800 to 692.5800.

The ten stocks in the Manufacturing sector had five positive mean returns, ranging from 0.0001 to 0.0009 and five negative mean returns, ranging from -0.0014 to -0.0001 with the highest mean returns found in UNIL and least mean returns found in ALW. Volatility was high in ALW (0.0445) compared to PKL (0.0038). The sector was found to have CV% ranging from -51723.4200 (SPL) to 65846.8300 (PZC). Out of the ten stocks, six were positively skewed ranging 0.2800 to 6.9500 whereas the remaining four were negatively skewed ranging from -15.2300 to -0.5800. The kurtosis was ranging from 35.9900 to 390.0100. The Mining sector had all the stocks recording positive mean returns, ranging 0.0011 to 0.0018. Volatility was high in GRS (0.0609) and low in ADDs (0.0263). The sector had coefficient of variation (CV) ranging from 2441.7100 (AADs) to 3341.0500 (GSR). The skewness was all positive ranging from 28.7400 to 29.800. This sector had kurtosis ranging from 866.000 to 909.04000.

Furthermore, the highest mean returns for the period under study was found in EBG (0.0022) and the least mean returns found in PBC (-0.0019). Also 24 of the stocks exhibited positive mean returns whereas 11 exhibited negative mean returns over the sample period. It is also evident that, over the sample period, volatility was high in ETI (0.0646) from the Finance sector and lower in PKL (0.0038) from the Manufacturing sector. The coefficient of variation for the entire sample period was high in PZC (65846.8300) and low in TRANSOL (-95087.0400), i.e. from the Manufacturing sector and Information Communication Technology sector respectively. The Manufacturing sector have six of the mean returns positively skewed (0.2800 to 6.9500) and four negatively skewed (-15.2300 to 0.5800).

Out of the 35 stocks, 22 had their mean returns positively skewed as against 13 stocks having their mean returns negatively skewed. The excess kurtosis for all the sectors and stocks for that matter were all positive indicating that all the mean returns were more peaked. Also the excess kurtosis for the entire sample period had the mean returns of GSR (909.0400) in the Mining sector more peaked than CPC (11.0700) in Food and Beverage sector.

The results revealed that, investors in the Finance sector saw gains in CAL, EBG, ETI, GCB, HFC, SCB and UTB since their mean returns were positive whereas investors of SOGEGH and TBL recorded losses (negative mean returns). Volatility (standard deviation) was high in ETI, CAL and EBG as indication of their risk levels. There was high probability of gains for investors of CAL, EBG, ETI, GCB and UTB whereas there was high probability of loss for investors of HCF, SCB, SOGEGH and TBL because the two groups recorded positive and negative skewness respectively. The sector was seen to be volatile since all the excess kurtosis were greater than three. The Distribution sector recorded more gains than losses. That is the mean returns of GOIL, MLC and TOTAL were positive whereas that of PBC was negative, an indication of loss for investors. The mean returns of MLC was commensurate with the risk taken by investors since it recorded the highest mean returns and standard deviation in the distribution sector. The skewness of GOIL and TOTAL was negative posing investors of these two stocks to high probability of loss whereas investors of MLC and PBC had high chances of gains (positive skewness). There existed high volatility trends in these stocks. Also, the Food and Beverage sector saw investors of FML and GGBL achieving gains compared to CPC investors who experienced losses during the same period. Investors of CPC were not compensated for assuming risk since they made losses but recorded the highest volatility (standard deviation) in the sector. It was also indicative that investors of CPC had high chances of making losses. Investors GGBL also had high chances of making losses than gains. Investors of FML had high chances of gains than losses it recorded a positive skewness. Investing in this sector was also volatile. Investors in the Information Communication Technology sector saw the two stocks (EGL and SIC) making losses even though the two had high chances of making gains than losses once the skewness were all positive and that investors were compensated for the risk they assumed. The sector was also seen to be volatile. Again, investors in the Insurance sector saw gains but there was high probability for investors of EGL making losses compared with investors of SIC who had high chances of making gains. This sector was also seen to be volatile since all the excess kurtosis was greater than three. The Manufacturing sector had investors of AYRTN, CLMT, PZC, UNIL and SWL making gains as compared to investors of ALW, SPL, PKL, GWEB and ACI who recorded losses in the same period. It is also evident that investors who made losses in this sector were not compensated as their mean returns recorded high standard deviations. Also, there was high probability of gains for investors of AYRTN, CMLT, SPL, UNIL, SWL and ACI even though investors of SPL and ACI recorded losses. The sector even though recorded same

losses as gains but there was high chances of making gains than losses as it is indicative of the skewness signs. Lastly, the Mining investors making gains and that the two sectors also saw investors having high chances of gains. The two sectors were all volatile.

Moreover, it was clear that most of the sectors and stocks for that matter recorded much gains than losses for investors since most of them recorded positive of their mean returns. For the entire sample period, most of the stocks had their skewness positive or asymmetric in nature indicating that the upper tail of the distribution of the returns was ticker than the lower tail and that there were more chance of gains than losses. The excess kurtosis for all the stocks were greater than three (3) meaning the underlying distribution of the returns were leptokurtic in nature and heavy tailed and that there was more frequently extremely large deviations from the mean returns than a Gaussian distribution. This confirms that investors have been experiencing high levels of volatility on the GSE.

Table 1. *Descriptive Statistics of the Returns Series.*

Sector	Mean	St. Dev	CV	Min	Max	Skewness	Kurtosis
Finance							
CAL	0.0013	0.0425	338.3000	-0.6652	0.7360	4.6600	220.3900
EBG	0.0022	0.0436	1951.6000	-0.1280	1.3075	28.3400	850.2200
ETI	0.0008	0.0646	7749.5900	-0.6934	1.7029	17.3300	542.7100
GCB	0.0009	0.0193	2258.1400	-0.1214	0.4202	11.4400	246.0000
HFC	0.0006	0.0124	2052.6200	-0.1903	0.1597	-0.1700	108.5400
SCB	0.0006	0.0287	5221.8800	-0.7851	0.2331	-20.8100	597.2100
SOGEGH	-0.0003	0.0209	-7144.1700	-0.4314	0.0926	-12.7100	238.7600
TBL	-0.0006	0.0235	-3682.0600	-0.5218	0.1675	-13.0600	288.6200
UTB	0.0011	0.0388	3587.1800	-0.1139	1.1493	27.3600	811.8400
Distribution							
GOIL	0.0001	0.0210	16906.9400	-0.5039	0.1871	-13.9600	363.0600
MLC	0.0017	0.0582	3458.9500	-0.7640	1.0512	9.4500	246.4900
PBC	-0.0019	0.0223	2309.6400	-0.1513	0.2319	1.9100	132.8100
TOTAL	0.0009	0.0433	4638.8700	-0.9029	0.8354	-1.0800	353.8000
Food and Beverage							
CPC	-0.0005	0.0758	-14476.9400	-0.3010	0.3010	-0.0300	11.0700
FML	0.0012	0.0216	1800.6600	-0.1878	0.1886	0.7200	41.0800
GGBL	0.0008	0.0155	1953.0700	-0.2218	0.1160	-3.5600	71.0900
Info. Technology							
CLYD	-0.0002	0.0460	-24856.2300	-0.4260	0.4260	0.5900	32.8200
TRANSOL	-0.0001	0.0352	-95087.0400	-0.4771	0.7782	2.4900	79.8900
Insurance							
EGL	0.0002	0.0380	16299.1900	-0.8248	0.1552	-16.8800	347.1800
SIC	0.0010	0.0304	3159.3300	-0.1139	0.8653	24.3700	692.5800
Manufacturing							
ALW	-0.0014	0.0445	-3255.4600	-0.5136	0.4467	-0.5800	347.1800
AYRTN	0.0005	0.0146	3233.0300	-0.1681	0.2865	6.9500	186.5500
CMLT	0.0003	0.0185	3796.4500	-0.1249	0.1249	0.3900	35.9900
PZC	0.0001	0.0317	65846.8300	-0.7721	0.2956	-15.2300	390.0100
SPL	-0.0001	0.0330	65846.8300	-0.2219	0.5133	3.4900	72.2200
UNIL	0.0009	0.0187	2031.5300	-0.2333	0.2333	2.2300	92.5900
PKL	-0.0001	0.0038	-7560.0700	-0.0670	0.0770	-0.6900	332.8000
GWEB	-0.0003	0.0187	-6560.9500	-0.2218	0.1249	-1.7300	56.0600
SLW	0.0002	0.0192	12812.0800	-0.1761	0.1761	0.2800	45.6600
ACI	-0.0002	0.0261	-12157.5100	-0.3010	0.3980	4.1600	121.3700
Mining							
TLW	0.0017	0.0527	3058.0600	-0.0792	1.5883	28.7400	866.0000
AGA	0.0012	0.0335	2867.3500	-0.0911	1.0200	29.7400	906.2200
GSR	0.0018	0.0609	3341.0500	-0.0748	1.8579	29.8000	909.0400
AADs	0.0011	0.0263	2441.5100	-0.0258	0.7959	29.1690	873.0800

3.2. Further Analysis

In testing, for stationarty in the return series using the PP and KPSS tests. All these tests as shown in Table 2 revealed that, for the PP tests, $p - values$ were very significant at 5% significance level and therefore the null hypothesis of non-stationary or unit root was rejected. In the case of the KPSS test, we failed to reject the null hypothesis of stationary since the test was significant at the 5% significance level. Therefore, the returns series were all stationary at the 5% level of significance for all the three tests.

Table 2. *PP Test and KPSS Test of the Return Series.*

	PP Test		KPSS Test	
Sector	**Test Statistic**	**P-value**	**Test Statistic**	**Critical value (5%)**
Finance				
CAL	-40.2780	0.0000**	0.0423	0.1480
EBG	-31.0930	0.0000**	0.1170	0.1480
ETI	-31.0870	0.0000**	0.0268	0.1480
GCB	-31.3140	0.0000**	0.0404	0.1480
HFC	-32.9370	0.0000**	0.0504	0.1480
SCB	-30.1270	0.0000**	0.0243	0.1480
SOGEGH	-28.2020	0.0000**	0.0370	0.1480
TBL	-32.4470	0.0000**	0.0542	0.1480
UTB	-31.5000	0.0000**	0.0370	0.1480
Distribution				
GOIL	-37.3130	0.0000**	0.1119	0.1480
MLC	-51.0150	0.0000**	0.0973	0.1480
PBC	-41.6420	0.0000**	0.0615	0.1480
TOTAL	-31.2340	0.0000**	0.1007	0.1480
Food and Beverage				
CPC	-48.2120	0.0000**	0.0103	0.1480
FML	-38.4000	0.0000**	0.0167	0.1480
GGBL	-31.4690	0.0000**	0.0517	0.1480
Info. Technology				
CLYD	-54.0670	0.0000**	0.0230	0.1480
TRANSOL	-52.9740	0.0000**	0.0175	0.1480
Insurance				
EGL	-30.3310	0.0000**	0.0497	0.1480
SIC	-31.2470	0.0000**	0.0695	0.1480
Manufacturing				
ALW	-14.9730	0.0000**	0.0510	0.1480
AYRTN	-35.1650	0.0000**	0.0603	0.1480
CMLT	-46.2060	0.0000**	0.1335	0.1480
PZC	-35.6130	0.0000**	0.1436	0.1480
SPL	-35.3890	0.0000**	0.0220	0.1480
UNIL	-39.8050	0.0000**	0.0734	0.1480
PKL	-30.7990	0.0000**	0.4480	0.1480
GWEB	-56.0410	0.0000**	0.0390	0.1480
SWL	-51.4340	0.0000**	0.0208	0.1480
ACI	-46.4080	0.0000**	0.0270	0.1480
Mining				
TLW	-31.5120	0.0000**	0.0511	0.1480
AGA	-31.2520	0.0000**	0.1278	0.1480
GSR	-31.0840	0.0000**	0.0563	0.1480
AADs	-30.0030	0.0000**	0.1477	0.1480

** Significance level: 5%

Figure 1, 2, 3, 4, 5, 6 and 7 show the screen plots of Finance, Distribution, Food and Beverage, Information Communication Technology, Insurance, Manufacturing and Mining sectors respectively. The results show that, for the Finance sector between component 1 and component 2 there exists a large break in eigenvalues whereas small breaks in eigenvalues start from component 3. Therefore the components before the small breaks are retained. This indicates that components 1 and 2 have large eigenvalues compared to the rest of the components. For the Distribution sector, the breaks are all equal but the last break where the eigenvalue levels off is at component 3 hence the eigenvalues before component 3 are retained. Therefore, component 1 and component 2 are retained. The screen plot for the Food and Beverage has a large break between component 1 and component 2 hence they were retained. Again, for the Information Communication Technology and Insurance sectors only component 1 and 2 are retained since the large break is between the two components. Also, for the Manufacturing sector, the large breaks are between component 1 and component 2 and from component 2 to component 3 but the small break in eigenvalues starts at component 3 hence components 1 and 2 are retained. The Mining sector had component 1 and component 2 retained since the large break existed between component 1 and component 2 and also from component 3 the eigenvalue is levelling off.

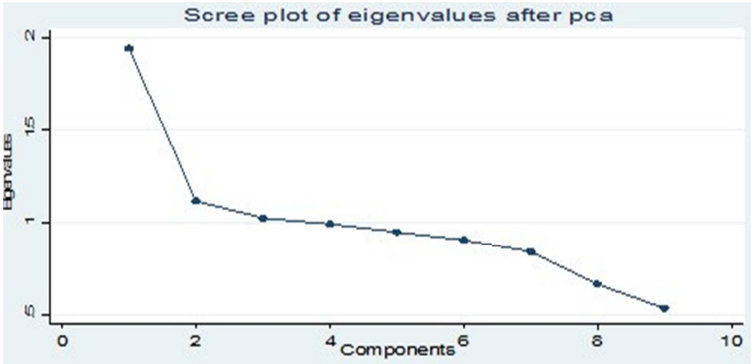

Figure 1. *Screen plot of the Finance Sector.*

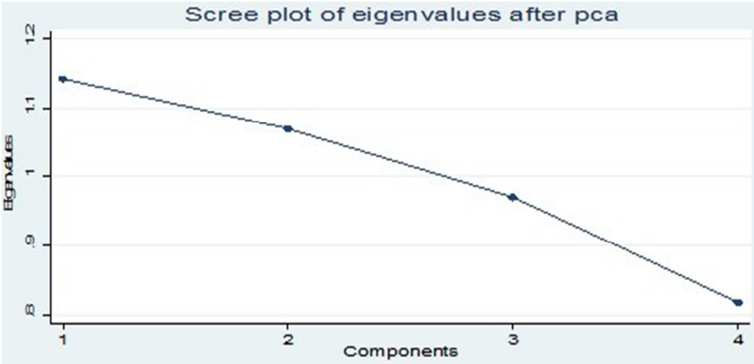

Figure 2. *Screen plot of the Distribution Sector.*

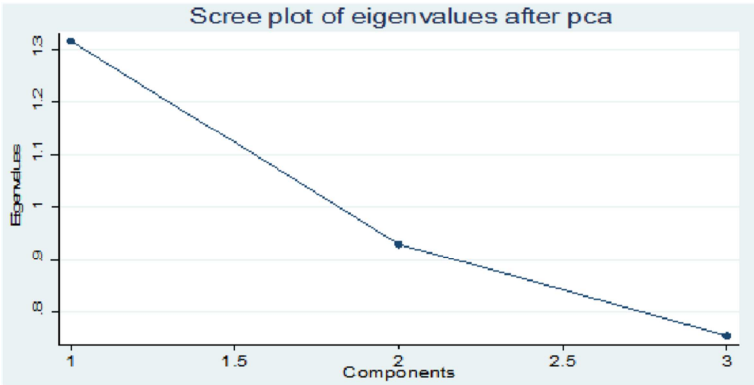

Figure 3. *Screen plot of the Food and Beverage Sector.*

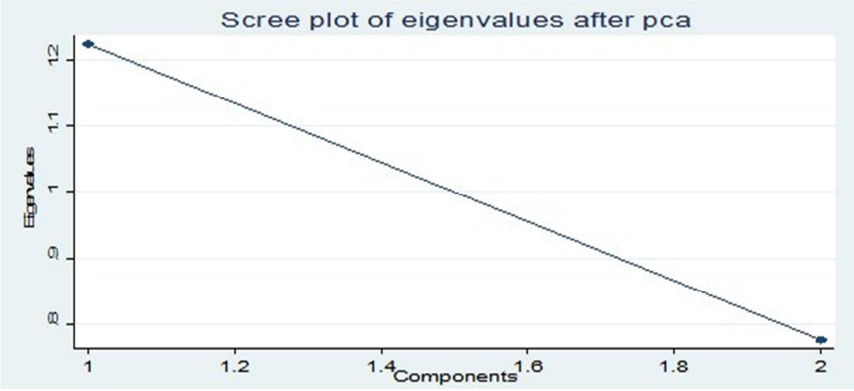

Figure 4. *Screen plot of the Info. Technology Sector.*

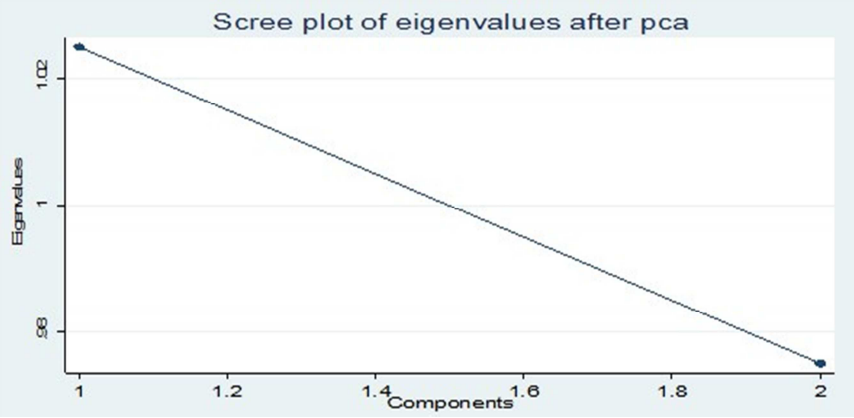

Figure 5. *Screen plot of the Insurance Sector.*

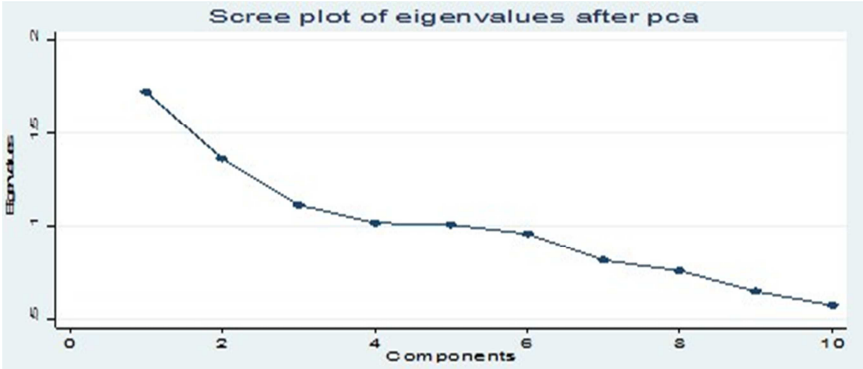

Figure 6. *Screen plot of the Manufacturing Sector.*

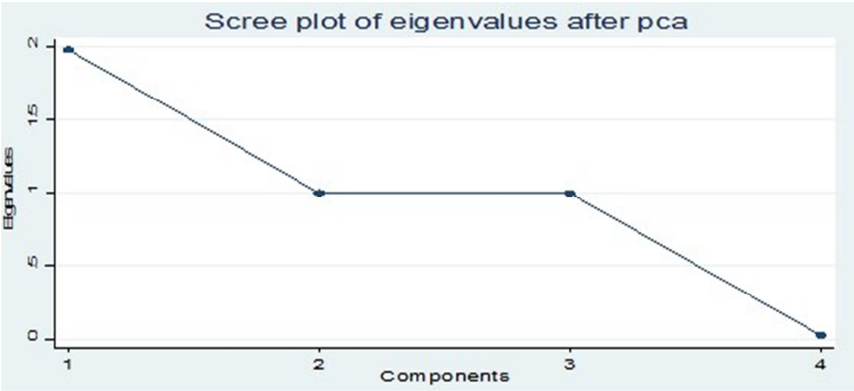

Figure 7. *Screen plot of the Mining Sector.*

The principal component analysis was employed in selecting the stocks that characterize each sector. For each sector, the PCA was employed in selecting the components that explains much of the variance in that sector. Also using the Eigen-value-one criterion, component(s) with eigenvalue greater than one (1) were retained. Therefore it is evident from Table 3 that, component 1 and 2 were retained by most of the sectors. The component loadings were set at 0.5 and that variable(s) with loadings greater than 0.5 was/were selected. The Finance sector had ETI, GCB and CAL selected with loadings 0.670, 0.7576 and -0.6696 respectively. The Distribution sector had PBC and TOTAL selected with loadings 0.7391 and 0.6022 respectively. The Food and Beverage sector had FML selected from comp1 with loadings 0.8835. The information communication technology sector had CLYD selected in comp1 with loadings -0.7071. The insurance sector had EGL selected with loadings (0.7071). The manufacturing sector had PZC and UNIL selected in comp1 and comp2 with loadings 0.5932 and 0.6121 respectively. Also, the mining sector had TLW and AGA selected with loadings -0.7076 and 0.7071 respectively. The results also indicates that the selected stocks are able to explain much of the variance in their respective sectors and hence could be considered for further analysis and probably investment.

Table 3. *Principal Component Analysis of the Returns Series.*

Eigenanalysis		**Eigenvectors**		
Component	**Eigenvalue**	**Variable**	**Comp1**	**Comp2**
Finance				
Comp1	1.9427	CAL	0.3410	-0.6696*
Comp2	1.1120	EBG	0.1203	0.0989
Comp3	1.0284	ETI	0.6760*	0.0672
Comp4	0.9948	GCB	-0.4761	0.7576*
Comp5	0.9527	HFC	0.4084	0.4604
Comp6	0.9085	SCB	-0.2928	0.2650
Comp7	0.8445	SOGEGH	0.3021	0.1046
Comp8	0.6696	TBL	-0.0010	0.2696
Comp9	0.5388	UTB	0.0252	-0.1060
Distribution				
Comp1	1.1419	GOIL	0.5069	-0.4923
Comp2	1.0699	MLC	-0.4223	0.3927
Comp3	0.9704	PBC	0.7391*	-0.0617
Comp4	0.8178	TOTAL	0.5291	-0.6022*
Food and Beverage				
Comp1	1.3153	CPC	0.4404	-0.4270
Comp2	0.92970	FML	0.8835*	0.6168
Comp3	0.7551	GGBL	-0.1927	0.6524*
Info. Technology				
Comp1	1.2246	CLYD	-0.7071	0.7071*
Comp2	0.7754	TRANSOL	0.7071	0.7071
Insurance				
Comp1	1.0249	EGL	0.7071	0.7071*
Comp2	0.9752	SIC	0.7071	-0.7071
Manufacturing				
Comp1	1.7249	ALW	0.1948	0.3752
Comp2	1.3642	AYRTN	0.0382	0.0736
Comp3	1.1156	CMLT	0.4806	0.2934
Comp4	1.0187	PZC	0.5932*	0.1777
Comp5	1.0114	SPL	0.3371	0.2577
Comp6	0.9610	UNIL	-0.4500	0.6121*
Comp7	0.8175	PKL	0.0276	-0.0162
Comp8	0.7604	GWEB	0.1550	0.5424
Comp9	0.6501	SWL	0.4932	-0.2705
Comp10	0.5761	ACI	-0.2103	0.2690
Mining				
Comp1	1.9774	TLW	-0.7076*	0.0007
Comp2	1.0019	AGA	0.0020	0.7071*
Comp3	0.9981	GSR	-0.0023	0.7066
Comp4	0.0226	AADs	0.7071	0.0010

* Selected stock under each component.

4. Conclusion

This paper employed the principal component analysis in selecting attractive stocks on the Ghana Stock Exchange. The results showed that, all the stocks on the exchange were highly volatile but there was higher probability of making gains than losses. The results also indicated that, among the 9 stocks in the Finance sector, only 3 stocks (CAL, ETI and GCB) were able to characterize the sector. The Distribution sector had 2 stocks (PBC and TOTAL) among the 4 stocks characterizing the sector. The Food and Beverage had only FML characterizing the sector. Also, the information Technology CLYD characterizing the sector. The Insurance sector had EGL characterizing the sector out of the 2 stocks. The Manufacturing sector had only 2 stocks (PZC and UNIL) characterizing the sector out of the 10 stocks and for the Mining sector, 2 stocks (TLW and AGA) among of the 4 stocks were the best ones. In effect, the 34 stocks were reduced to 12 stocks. The selected stocks are much better to be considered by investors in the various sectors on the Ghana Stock Exchange for productive investment since they explain much of the variance their respective sectors compared to stocks from the same sector.

References

[1] Feeney, G. and Hester, D. *"Stock Market indices: a principal Component Analysis»*, in D Hester and J Tobin (eds), Risk aversion and portfolio choice, Wiley, (1967). New York.

[2] Horne, B. and Camp, N. Principal component analysis for selection of optimal snp sets that capture intraganic, genetic variation. *Genetic Epidemiology*, (2004). 26: 11–21.

[3] Kaiser, H. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, (1960). 20: 141–15.

[4] Kanbur, S. and Marian, H. Principal component analysis of RR lyre light curves. (2004). http//:www.astro.umass.edu/shashi/paper 7.pdf.

[5] Kritzman, M., Yaunzhen, L., Sebastien, P., and Roberto, R Principal component as a measure of systematic risk. *MIT Sloan School Working Paper 4785*., (2011).

[6] Loretan, "Generating market risk scenarios using principal component analysis: Methodological and practical considerations". (1997). *Federal Reserves Board*, htt//:www.bis.org/publ/ecsc07.pdf.

[7] Mbeledegu, N., Odoh, M., and Umeh, M Stock feature extraction using principal component analysis. *International Conference on Computer Technology and Science. IACSIT Press Singapore*, (2012). DOI: 10.7763/IPCSIT V47.44.

[8] Novosyolov, A. and Satchkev, D. Factor term structure modelling using principal component analysis. *Journal of Asset Management*, (2008). 9 (1): 49–60.

[9] Sterns, J. *Applied Multivariate Statistics for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates. (1986).

[10] Wang, Y. and In-Chan, C." Market index and stock price direction prediction using machine learning techniques: An empirical study on the kospi and hsi. *Science Directs*, (2013). 1: 1–13.