Journal of Economics and Behavioral Studies (ISSN: 2220-6140)
Vol. 8, No. 6, pp. 40-51, December 2016

The Performance of Maximum Likelihood Factor Analysis on South African Stock Price Performance

Andile Khula, Ntebogang Dinah Moroke
North West University, South Africa
Ntebo.Moroke@nwu.ac.za, aikhula@gmail.com

Abstract: The purpose of this paper is to explore the effectiveness and applicability of Maximum Likelihood Factor Analysis (MLFA) method on stock price performance. This method identifies the variables according to their co-movement and variability and builds a model that can be useful for prediction and ranking or classification. The results of factor analysis in this study provide a guide as far as investment decision is concerned. Stock price performance of the seven well-known and biggest companies listed in the Johannesburg stock exchange (JSE) was used as an experimental unit. Monthly data was available for the period 2010 to 2014.Details of a trivariate factor model is: Factor 1 comprises of Absa and Standard Bank (Financial sectors), Factor 2 has Shoprite and Pick 'n Pay (Retail sectors) while Factor 3 collected Vodacom MTN and Sasol (Industrial sectors). The companies contribute 46.9%, 12.7% and 10.8% respectively to the three sectors and these findings are confirmed by a Chi-square and the Akaike information criterion to be valid. The three factors are also diverse and reliable according to Tucker and Lewis and Cronbach's coefficients. The findings of this study give economic significance and the study is relevant as it gives investors and portfolio manager's sensible investment reference.

Keywords: *Maximum Likelihood Factor Analysis, stock prices*

1. Introduction

Factor analysis is well-known as a statistical method applied in social sciences. In most cases, scholars apply this technique in conducting researches related to psychology. This has led to people falsely believing that the technique is a psychological theory. The development of this technique was essentially to lend a hand in the field of psychology by mathematically explaining psychological theories as far as social aptitudes and conduct is concerned. The methodological procedure for factor analyses is complex and has few guidelines. As noted by Costello and Osborne (2005), there are so many options that this technique offer, but different software packages provide different terminologies which in most cases remain unexplained. The goal of factor analysis is to reduce a vast number of variables according to their relationships and combining those variables into few factors (Muca et al., 2013). In particular, factor analysis is a data reduction technique pioneered by Spearman et al. (1904).The technique can also be used in exploratory or confirmatory studies. Alternative technique to factor analysis is Principal component analysis (PCA). PCA is variance based while factor analysis is covariance based. Cao (2010) however noted that PCA is used by some authors as an *ad hoc* procedure to collect the factors. The author is of the idea that even though the two techniques are closely related, they should be used autonomously. Factor analysis has been reported to be effective under the assumption that the data does not suffer from serial correlations. It should be noted that financial time series data collected over short time intervals such as a week or less is usually associated with serially correlated errors. To protect this assumption, parametric models are recommended to also correct the linearity of the data. On the contrary, Tsay (2005) recommended the use of factor analysis to the residual series. This study explores the effectiveness and applicability of the maximum likelihood factor analysis (MLFA) method in modelling stock price performance in the context of South Africa. Wang (2003) warned that the data from this sector is one of the extremely hard to model with in all fairness due to lack of linearity and parametric nature of the data.

Stock marketers are expected to employ a suitable model(s) to predict stock prices as well as buying and selling of appropriate stock at the right time. This, however, becomes almost impossible if incorrect information on such stocks is relied on due to the use of ineffective model(s). Authors such as Lev and Thiagarajan (1993), Wu et al. (2006) and Al-Debie & Walker (1999) mentioned some of the studies which used the fundamental analysis methods, a well-known method for stock analysis. These authors explained that such studies developed trading rules on the basis of the information linked to macroeconomics, industry, and companies. Tsang et al. (2007) and Ritchie(1996) emphasised that the fundamental analysis assumes stock price is dependent on its intrinsic value and expected investment returns. This method however has its

own short comings depending on the desired predictions to be produced. As suggested by a number of authors, the fundamental analysis is suited when the desire is to produce long-term estimations. Al-Radaideh et al. (2013) recommends the use of this method in both short-term and medium term speculations.

Based on this reason, the current study proposes the application of maximum likelihood factor framework to stock price of South Africa. Studies on the analysis of stock performance in this country have not been fully exhausted especially where the application of the proposed method is concerned. Accordingly, none of the work published ever lent the help of MLFA method to analyse stock price performance of the proposed companies. This method unlike other factor analyses types, affords the researcher to make further diagnosis of the model to confirm its effectiveness before it could be recommended for further analysis. It has also been used by scholars in developing countries and the world at large. The findings may help stock marketers in South Africa to make informed decisions regards to this sector and this may also help boost its performance in the future. The study is beneficial to scholars in respect of the methodology and filling a gap in literature. The findings of the current study could also be appealing for portfolio management in that a factor model vividly condenses complex data for variable estimate. Factor analysis results further gives a perfect representation of the foundation of portfolio risk (Cheng, 2005). The author further highlighted that a factor model may be found useful in relating securities returns to a set of factors.

2. Literature Review

Exploratory Factor Analysis method has in the hitherto been used successfully and has been found to be a powerful tool in data analysis in many areas of study. Several studies on the analysis of stock market have been conducted worldwide but most of them applied other nonlinear models such as the Autoregressive Conditional Heteroscedasticity (ARCH) and a family of the Generalized Autoregressive Conditional Heteroscedasticity (GARCH). Majority of studies around the world have investigated the stock performances using factor analysis technique save for South Africa. Studies which include that by Kumar (2013) employed factor analysis method in determining the factors associated with the performance of stock markets in India. The data used was collected from the national stock exchange (NSE) website of India covering the period from January 2001 to May 2013. The findings revealed industrial performance as one of most dominant factors of stock markets in this country since it accounted for a high percentage of variance compared to other factors. Cheng (2005) used factor analysis technique to study fundamental patterns of the relationship between eight stocks in New York. The data used was collected from the New York Stock Exchange and NASDAC spanning the period from January 1998 to December 2004. Estimation methods such as principal factor analysis and MLFA were used to achieve the study objective. The findings revealed three primary factors, and MLFA proved to outperform its counterpart as it provided more accurate estimations for weekly rates of returns.

Hui (2005), applied factor analysis framework for do portfolio diversification. The study aimed to explore possibilities for diversifying into the United States and Asian Pacific markets in the perspective of a Singaporean investor. The ten weekly stock market indices were collected from the Data Stream of the National University of Singapore covering the period from 1 January 1990 through 30 June 2001. Empirical results revealed two extracted factors from the indexes with the first factor having most significant loading on Hong Kong, the Philippian, Korean, Singaporean as well as Thailand. Australia and New Zealand were dominated by the second factor with Japan, the United States and Taiwan favoured by a different factor. The findings identified United States, Australia and Japan as the three most appropriate and well-developed markets for risk diversification. Taiwan was suggested as the second market for diversification. Xin (2007) used factor analysis in Chinese stock market to reduce the number of the listed companies. The data used was the 2008 medium-term financial targets of 30 companies listed in the mainland China. The author chose eight financial targets to analyse, interpreting the special meaning of each factor. The study found that in a large number of the financial targets in annual report of listed companies, eight original targets can be replaced with the three new factors. A conclusion was reached after observing the factor scores that three common factors were irrelevant, meaning that the information which was included in the factors was non-repetition. Furthermore it was found that for each stock, the high or low scores of its one special factor do not affect other factor scores. It is only the three factors that have reflected most difference in stocks.

Another study which soughtinternational portfolio diversification was done by Valadkhani et al. (2008).The study further sought toexamine relations among the stock market returns of 13 economies.Monthly datacovering the period December 1987 to April 2007 were analysed. The study followed the frameworks of PCA and MLFA in examining discernable patterns of stock market relationships. The results revealedhighly correlated stock returns in a number of Asian countries. Factor loadings were also reported to be highly significant on the first factor. Rotated factor loadings confirmed a significant and linear association between Singaporean, Thailand, Phillipinean, Malaysian, Hong Kong, Indonesian, Taiwan, and Korean stock returns.The results further indicated due to the co-movement in stock returns of the United Kingdom, Germany, United States, Australia, and Japan, these markets could be represented by the second factor loading. Făt andDezsi (2012) also used factor analysis methods to investigate possibilities for portfolio diversification internationally. The study sought to specifically answer the question, "does it pay off?" This study further investigated the underlying structure between the markets. The data used consisted of the 12 daily stock index closing prices covering the period 19 September 1997 to 5 May 2012. Daily data consisted of 3848 observations for each market. The study findings reported significant correlations among the countries with similar economic performance and those situated in the same regions. Furthermore, the results of the two methods had similar implication. What was also revealed by the study were the three main integrated regions such as developed markets from Asia, emerging markets from Europe and two outliers being Romania and Slovakia. Other studies that employed factor analysis on stock markets were done by Gu and Zeng (2014), Bastos and Caiado (2010), Illueca and Lafuente (2002), Shadkam (2014), Tuluca and Zwick (2001) among others. None of the studies published was undertaken in Africa, let alone South Africa. This proves that there is still a gap to fill and awareness has to be made to African stock marketers.

3. Methodology

The initial step in data analysis is ensuring the readiness of data for analyses. Since the data is collected over a period of time, it is essential to impose relevant transformations and check if it does not violate most pertinent assumptions of factors analyses. Once these issues have been looked at, proposed primary data analysis methods are used. Sections 3.2.1 and 3.2.2 give brief description of the preliminary and primary data analyses methods respectively.

Preliminary data analysis: In order to continue safely with the application of factor analysis, the assumptions of linearity must hold. Stock prices are non-linear in nature, so the data should be transformed to linearity. Therefore, this study proposes the application of the Regression Specification Error Test (RESET) in order to uphold this assumption. The test is derived from linear regression model expressed as univariate $AR\ (p)$:

$$X_t = \beta_0 + \sum_{j=1}^{p} \beta_j X_{t-p} + \varepsilon_t, \qquad (1)$$

where according to Gujarati (2003)$\beta_0, \beta_1, \beta_2, \ldots \beta_p$are model parameters and $\varepsilon_t$is an independently and identically distributed (iid) random variable having a mean equal to0 and a variance $\delta^2{}_\varepsilon$. The purpose of AR (p) is to ensure that the error ε_t is minimised (Xaba et al., 2016). For selection of value p that minimizes a certain information criterion,the test statistic according to Rencher (2003)is defined as:

$$F^* = \frac{SSR_0 - SSR_1\ )/\mathrm{r}}{SSR_1/(\mathrm{n-p-r})} \sim F_\alpha(r, n-p-r), \qquad (2)$$

where $r = s + p + 1$. At the α level, the null hypothesis of linearity is rejected in favour of the alternative hypothesis if $F^* > F\alpha(r, n-p-r) or prob\ (F^*) < \alpha$ implying that the true specification is linear.

Due to the nature of the data used in this study, the stationarity tests were performed and reported accordingly. According to Tsay (2010), the price series of an asset tend to be non-stationary as this series have been collected over certain time period. This is as a result of no fixed stock price level. If the series follows a non-stationary process, persistence of shocks will be infinite. Furthermore, in order for the assumptions of asymptotic analysis to be valid, the variables in the regression model should be stationary (Brooks, 2002) also to avoid spurious regression results. Several unit root tests are available. This study followed the methodology of the Augmented Dickey Fuller (ADF) unit root test. Though not really

appropriate, it is advisable to examine the autocorrelation function (ACF) of a series whilst testing for unit root (Brooks, 2002).The ACF may mislead the authors as they have a tendency of revealing extremely persistent but static series. The author therefore discourages the use of the ACF or partial ACF (PACF) in determining the stationarity of the series. Instead of using the ACFs and PACFs, Brooks (2002) and Greene (2003) advice on conducting some formal hypothesis testing procedure that answer the question about unit roots. Hence this study employed the most recommended unit root test**.**

The ADF test is calculated to assess $H_0: \phi = 1$ verus $H_1: \phi < 1$ in the equation:

$$y_t = \phi y_{t-1} + \mu_t. \tag{3}$$

On the contrary, Brooks (2002) suggests the following regression rather than (3):

$$\Delta y_t = \psi y_{t-1} + \mu_{t,} \tag{4}$$

so that a test of $\phi = 1$ is equivalent to a test of $\psi = 0$ (since $\phi - 1 = \psi$). The ADF tests the three kinds of equations such as $\tau, \tau_\mu, \tau_\tau$ which are used in the presence of a constant, trend and when both the constant and a deterministic trend are present respectively. The observed statistics for the Dickey-Fuller (DF) test is:

$$DF = \frac{\hat{\psi}}{S\hat{E}(\hat{\psi})}. \tag{5}$$

The test (5) follows a non-standard distribution. If $p$number of lags of the response variable is 'augmented' then the alternative model becomes:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_1 \Delta y_{t-1} + \mu_{t,} \tag{6}$$

giving birth to the ADF and still conducted on ψ.The augmented test statistic (5) becomes:

$$ADF = \frac{1-\hat{\psi}}{S\hat{E}(\hat{\psi})}, \tag{7}$$

$\hat{\psi}$denoting the least squares estimate of ψ. If the observed absolute value is greater than the critical value, no differencing is required since the series has been rendered stationary. Reject $H_0: y_t \sim I(1)$ in favour of the alternative hypothesis $H_1: y_t \sim I(0)$ if the observed statistic (7) is in excess of the critical value.

Primary Data analysis: This section presents a review of MLFA method. This factor analysis method is recommended by Hair et al. (2010), Tabachnick and Fidel (2001), and Tsay (2005), some of the well-known authors in the field of multivariate analysis. According to Richard and Dean (2002),factor analysis may be considered a descriptive technique to determine the covariance relationship of a considerable number of variables. The correlated variables may be collected into few unobservable and uncorrelated factors. Richard and Dean (2002) notes that X could be treated as an observable random vector. The associated stochastic properties to this variable having$p$factors are its mean denoted asμ, and a covariance matrixknown as Σ. In a factor model,Xhas a linear relationship with latent factors symbolised as $f_1, f_2, \ldots, f_m$. These are alsocalled common factors. Additionally, the model has p sources of variation denoted as $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$,known as specific factors. Given a multiple set of variables, represented as k, Richard and Dean (2002), Rencher (2003), and Wichern and Johnson (2007) suggest the following equation to represent a factor model:

$$\begin{aligned} X_1 - \mu_1 &= l_{11} f_1 + l_{12} f_2 + \cdots + l_{1m} f_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21} f_1 + l_{22} f_2 + \cdots + l_{2m} f_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1} f_1 + l_{p2} f_2 + \cdots + l_{pm} f_m + \varepsilon_p, \end{aligned} \tag{8}$$

where$x = x_1, x_2, \cdots, x_p$, $\mu = \mu_1, \mu_2, \cdots, \mu_p$, $f = f_1, f_2, \cdots, f_m$,$\varepsilon = \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p$.The coefficient $l_{ij}$is known as the loading of the i_{th} variable on the j_{th} factor. It must also be noted that the i_{th} specific factor ε_t is associated only with the i_{th} responsex_i. The p −deviations $x_1 - \mu_1, x_2 - \mu_2, \ldots, x_p - \mu_p$are expressed in terms of $p + m$ unobservablerandom variables$f_1, f_2, \ldots, f_m$, $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p$. Given these many unobservable variables, it may be hopeless to directly verify a factor model from observations on $x_1, x_2, \ldots, x_p$. One may wish to test additional assumptionsabout the F and ε random vectors,implying that

certain covariance associations can be checked in (10). It is therefore assumed that $E(F) = 0$, $Cov(F) = E(FF') = I$, $E(\varepsilon) = 0$, $Cov(\varepsilon) = E(\varepsilon\varepsilon') = \Psi$, $Cov(X) = LL' + \Psi$, or $Cov(X, F) = L$ and $\delta_{ii} = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \Psi_1$ with:

$$Var(X_i) = \tilde{h}_i^2 + \Psi_1. \quad (9)$$

Tsay (2010) defines the i_{th} communality as the sum of squares of the loadings of the i_{th} variable on the m common factors.Prior to factor extraction, it is advisable to determine the possible number of factors to retain. The study used a proportion of variance for this purpose. The corresponding proportion from R is $\sum_{i=1}^{p} \frac{\hat{\lambda}_{ij}^2}{p}$. The contribution of all m factors to $tr(S)$ or p is therefore $\sum_{i=1}^{p}\sum_{j=1}^{m} \hat{\lambda}_{ij}^2$ with the sum of squares of all elements is$\hat{\lambda}$:

$$\sum_{i=1}^{p}\sum_{j=1}^{m} \hat{\lambda}_{ij}^2 = \sum_{i=1}^{p} \hat{h}_i^2 = \sum_{j=1}^{m} \theta_j. \quad (10)$$

The value of m should be relatively large such that the sum of variance accounted contributes sufficiently to $tr(S)$ or p. According to Richard and Dean (2002), the same process can benefit the principal factor method, in which prior estimates of communalities could form $S - \hat{\Psi}$ or $R - \hat{\Psi}$ even though$S - \hat{\Psi}$ or $R - \hat{\Psi}$ will often have some negative eigenvalues. As values of m range from 1 to p, the cumulative proportion of eigen values, $\frac{\sum_{j=1}^{m} \theta j}{\sum_{j=1}^{p} \theta j}$, will exceed 1.0 and then reduce to 1.0 as the negative eigenvalues are added (Rencher, 2003). Note that 80% is reached for a lower value of m, something that is impossible for$S$ or R.A strategy is to choose m equal to the value for which the percentage first exceeds 100%.Hair et al. (2010) suggested a thumb rule of 60% for the sum of the variance accounted for to retain the minimum number of factors. This study adopts Hair et al. (2010) thumb rule as a guiding principle.

Maximum Likelihood Factor Analysis: The assumption of multivariate normality (MVN) needs to be met before this method can be used (Marcoulides, 2008). If the common factors F and the specific ε factors are assumed to follow a normal distribution, then maximum likelihood estimates (MLE) of the loadings and observed variables should be obtained. When $\mathbf{F_j}$ and ε_j are both normal, the observations $\mathbf{X_j} - \boldsymbol{\mu} = \mathbf{LF_j} + \varepsilon_j$ are also normal resulting in the likelihood: L (μ, Σ) = $(\boldsymbol{2\pi})^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} \boldsymbol{e}^{-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})' + n(\bar{x}-\mu)(\bar{x}-\mu)'\right)\right]}$

$$=(\boldsymbol{2\pi})^{-\frac{(n-1)p}{2}} \left|\Sigma\right|^{-\frac{(n-1)}{2}} \boldsymbol{e}^{-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n}(x_j-\bar{x})(x_j-\bar{x})'\right)\right]} \times (\boldsymbol{2\pi})^{-\frac{p}{2}} \left|\Sigma\right|^{-\frac{1}{2}} \boldsymbol{e}^{-\left(\frac{n}{2}\right)\left[\left((\bar{x}-\mu)'\Sigma^{-1}\right)(\bar{x}-\mu)\right]} \quad (11)$$

Equation (11) is dependent on $\boldsymbol{L}$ and Ψ through $\Sigma = \boldsymbol{LL}' + \Psi$. Let $\boldsymbol{X_1}, \boldsymbol{X_2}, \cdots, \boldsymbol{X_n}$which is a random sample from$\boldsymbol{N_p}(\boldsymbol{\mu}, \Sigma)$, where $\Sigma = \boldsymbol{LL}' + \Psi$is the covariance matrix for the $\boldsymbol{m}$ common factor model. The MLE$\hat{\boldsymbol{L}}$, $\hat{\Psi}$, and $\hat{\boldsymbol{\mu}} = \bar{\boldsymbol{x}}$ maximize $\boldsymbol{L}'\Psi^{-1}\boldsymbol{L} = \Delta$ subject to $\hat{\boldsymbol{L}}'\hat{\Psi}^{-1}\hat{\boldsymbol{L}}$, being diagonal.The MLE of the communalities are defined as$\tilde{\boldsymbol{h}}_i^2 = \tilde{\boldsymbol{l}}_i^2 + \tilde{\boldsymbol{l}}_{i2}^2 + \cdots + \tilde{\boldsymbol{l}}_{im}^2for\boldsymbol{i} = \boldsymbol{1}, \boldsymbol{2}, \ldots, \boldsymbol{p}$.

Factor rotation: This study used a transformation technique that helps make the interpretation of the factor structure easy. Factor loadings calculated from an orthogonal transformation have the same ability to reproduce the covariance (or correlation) matrix. This transformation corresponds to a rigid rotation of the coordinate hence it is called factor rotation. If $\hat{L}$ is the $p \times m$ matrix of estimated factor loadings in MLFA, then:

$$\hat{L} *= \hat{L}T, \text{ where } TT' = T'T = 1 \quad (12)$$

is a $p\times m$ matrix of "rotated" loadings. The estimated covariance or correlation matrix remains stable since:

$$\hat{L}\hat{L}' + \hat{\Psi} = \hat{L}TT'\hat{L} + \hat{\Psi} = \hat{L} * \hat{L} * ' + \hat{\Psi}. \quad (13)$$

Equation (13) indicates that the residual matrix:

$$S_n - \hat{L}\hat{L}' - \hat{\Psi} = S_n - \hat{L} * \hat{L} * ' - \hat{\Psi}, \quad (14)$$

remain unchanged. As noted by Richard and Dean (2002), the specific variances $\hat{\Psi}_i$ and the communalities h_i^2 alsoremain unchanged.There are variety of methods proposed when $m > 2$,most common technique being the varimax. This technique uses rotated factor loadings that maximise the variance of the squared loadings

in each column of$\hat{\lambda}*$. The size of the loadings in a column has a bearing on the variance. Similar loadings in a column bring down the variance to be nearly equal to 0. Likewise, the variance approaches maximum as a result of square loadings approaching 0 and 1.Varimaxrotation method attempts to either improve the loadings by making them small or large so as to improve their interpretation (Rencher, 2003). This procedure however fails to give an assurance that all variables will favour only one factor. In all honesty, none of the available procedures can lead to such finding irrespective of the data used.

Factor loadings and the variance are validated for convergence as suggested by (Napper et al., 2008). Factor loading is not a correlation, but rather a measure of the unique association between a factor and a variable. Pattern matrix is interpreted rather than the structure matrix and the reason is pragmatic- it is easier (Tabachnick and Fidell, 2007). The two matrices show different results, such as high and low loadings which are more evident in a pattern matrix. Factors with loadings in excess of 0.32 are considered significant (Tabachnick and Fidell, 2007). Significant loadings suggest that the associated variable is a good measure of the factor.

Goodness of Fit of the factor model: This section discusses the goodness of fit test of factor model. This will affirm the adequacy of the m-factor model for generating the observed covariance's or correlations. The null hypothesis is $H_0: \sum = \Psi + \Lambda\Lambda'$ where Λ has dimensions $p \times m$, and the alternative hypothesis is that $\sum$ is any $p \times p$ symmetric positive definite matrix. The test statistic becomes:

$$\chi^2 = \left[N - 1 - \frac{1}{6}(2p + 5) - \frac{2}{3}m\right] ln \frac{\left|\Psi + \hat{\Lambda}\hat{\Lambda}'\right|}{|S|}, \qquad (15)$$

where $\hat{\Psi}$and $\hat{\Lambda}$are the solutions of the maximum likelihood equation, S is the sample covariance matrix. If the null hypothesis is in fact true, as N becomes large, the statistic tends to be distributed as chi-squared variate with$v = \frac{1}{2}[(p - m)^2 - p - m]$degrees of freedom and the null hypothesis of exactly m common factors is rejected at the α level if $\chi^2 \geq \chi^2_{\alpha;v}$ and accepted otherwise.

Construct validity: The final process is to evaluate the rotated factors for each variable in order to determine importance of the variables to factor structure. The evaluation process start by re-specifying the factor model owing to (1) deletion of a variable(s) from the analysis, (2) the desire to employ a different rotational method for interpretation, (3) extraction of different number of factors if needs be, or (4) it can go to the extent of changing from one extraction method to the other (Hair et al., 2010). Cronbach's alpha and Tucker and Lewis reliability coefficient were used to asses the reliability of the constructs. These coefficients measure how well set of items measure a uni-dimensional latent construct. The two measures range between 0 and 1 with the latter being preferred the most (Cronbach and Shavelson, 2004). Tucker and Lewis coefficient should on the other hand be greater than or equal 0.95 for a factor solution to be considered acceptable (Floyd and Widaman, 1995) with coefficients closer to unity being more desirable (Rossi et al., 2013).

4. Empirical Evidence

The data and assumptions: This study used daily returns for seven stocks listed on the Johannesburg Stock Exchange (JSE) which are ABSA, Standard bank, Shoprite, Pick n Pay, Sasol, MTN and Vodacom. Daily stock returns were collected from January 2010 through December 2014 each consisting of 1248 observations. This data was available at the time of request.The Statistical Analysis Software (SAS) version 9.3 registered to the SAS Institute Inc. Cary, NC, USA and IBM software packages for social scientists (SPSS) version 23 was used for data analysis. The data was subjected to the Box-Cox transformation due to its power to normalise the data and capability to stabilise the variance and improve effect size in the analysis. Furthermore, these data transformation tools constitute important elements of data cleaning and preparation for statistical analyses(Osborne, 2010). A transformed data favours the MLFA method as it performs well when the data is normally distributed. A modified Turkey idea by Box and Cox to take the form of the Box-Cox transformation is presented as:

$$y_t^{\lambda} = \frac{(y_t^{\lambda} - 1)}{\lambda}, \text{ where } \lambda \neq 0; \qquad (16)$$

$$y_t^{\lambda} = log_e(y_i), \text{ where } \lambda=0, \quad (17)$$

wherey_t is defined as the return is for any time, t.

Presented in Figure 4.1 are the original time series plots for the seven stock prices from January 2010 to December 2014.

Figure 1: Original time plots of the seven stock prices

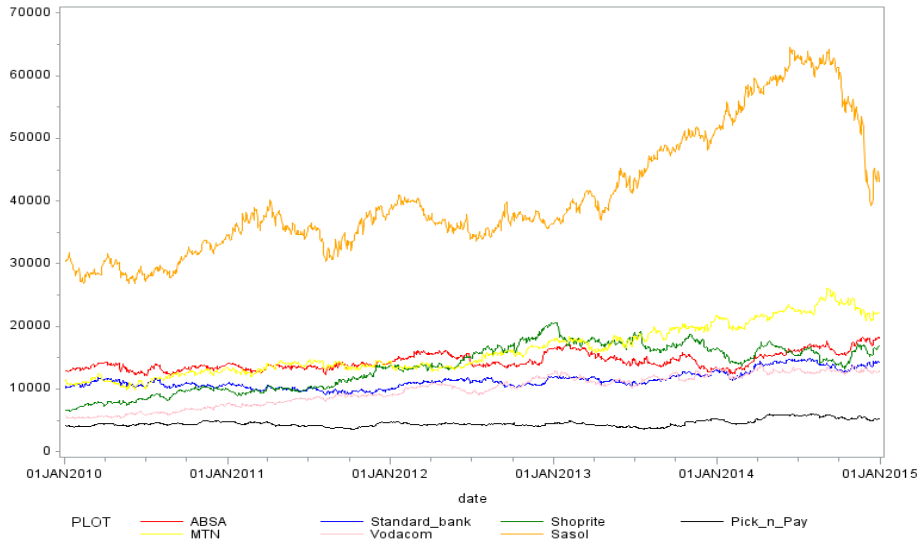

Figure 1 shows results for investigation into properties of time series data used in the study. Stock prices exhibit positive linear trend with seasonal fluctuations during some of the periods. Sasol had a sudden sharp decrease in stock prices between the years 2012 and 2015, with MTN showing a steady increase from 2010. The stock prices of these companies suddenly decelerated towards the end of 2014. Pick 'n Pay appears to be the only company with a constant line from 2010 through 2015. This implies that this company's stock prices performance have been constant with very insignificant changes throughout the years according to the findings. Standard bank, ABSA and Shoprite show mild fluctuations in stock prices between 2010 and 2014 and a ditch towards the end of 2014 and early 2015. This could mean a challenging period for these companies implying a loss in their part during that period. Sasol seem to be the only company that have been surviving the storm all these years except for the beginning of 2015.The stock price performance was at its peak between 2014 and 2015. Though MTN prices were high during these years, the sales of this company were far less than those of Sasol. Most of the stock prices seem to be weakly stationary at levels since none of the stochastic error properties, such as constant mean and variance are revealed. There are time epochs revealed by the seven stock prices implying the possibility of nonlinearity. Upon transformation of the data, all the seven series became stationary at once (see Figure 2) and also conformed to normality and linearity standards (see Table 1).

Figure 2: Log transformed plots of the seven stock prices

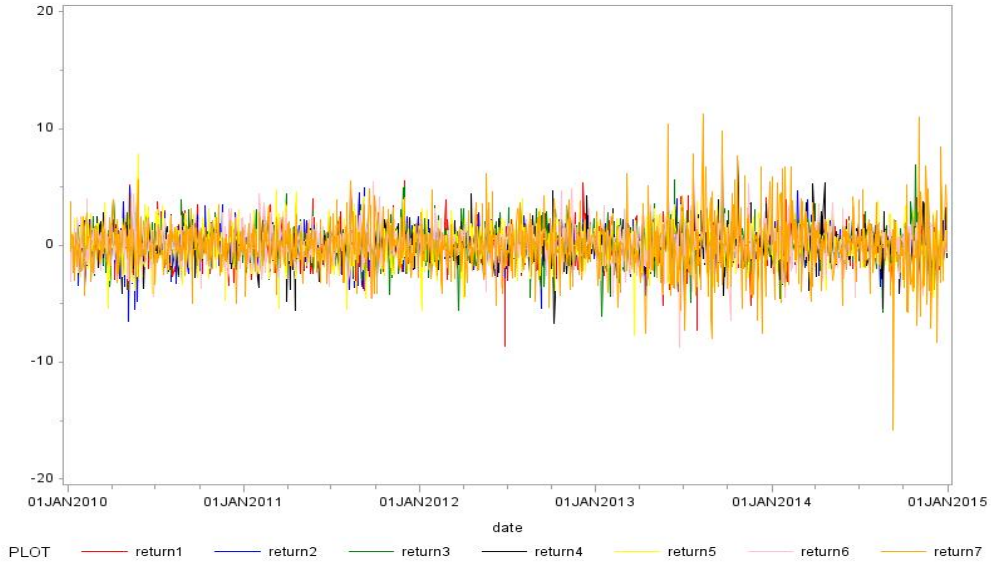

Table 1: Linearity and normality test results

Variables	Linearity RESET p-values	Decision	Normality Kolmogorov-Smirnov p-values	Decision	Stationarity ADF p-values	Decision
ABSA	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0
Standard bank	<.0001	Do not reject H_0	0.09160	Do not reject H_0	<0.0001	Reject H_0
Shoprite	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0
Pick 'n Pay	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0
MTN	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0
Vodacom	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0
Sasol	<.0001	Do not reject H_0	<0.0100	Do not reject H_0	<0.0001	Reject H_0

Both the RESET and Kolmogorov-Smirnov$p$-values fail to reject the null hypotheses of linearity and normality at 10% significance level respectively. On the other hand, the ADF test rejects the null hypothesis of unit root complimenting the results in Figure 2. These findings confirm the applicability of MLFA to the seven stock prices with success.

Findings from MLFA: This section presents the results of MLFA method. Prior to the extraction of relevant factors, a thumb rule was applied as a guiding principle to decide on the appropriate number of factors to retain. The current study adopts Hair et al. (2010) proportion of variance criterion to help determine the number of factors and the results are summarised in Table 2.

Table 2: Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	3.282	2.393	0.469	0.469
2	0.890	0.137	0.127	0.596
3	0.752	0.125	0.108	0.703
4	0.628	0.054	0.090	0.793
5	0.573	0.039	0.082	0.875
6	0.534	0.191	0.076	0.951
7	0.342		0.049	1.000

The rule confirms that only 3 factors are significant and should be retained. The first three factors account for about 70.3% variance and this is in excess of the benchmark of 60% as suggested by Hair et al. (2010).The

MLFA method is then used to extract the three factors and the results are summarised in Table 3.The factor structure is intended to determine relationships between variables and the factors.

Table 3: MLFA solutions

Variables	Unrotated factor loadings			Varimax rotated factor loadings		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor3
ABSA	0.751	.	.	0.726	.	.
Standard bank	0.809	.	.	0.642		.
Shoprite	0.605	.	.	.	.	0.559
Pick'n Pay	0.441	.	-0.324	.	.	0.551
MTN	0.639	.	.	.	0.581	.
Vodacom	0.519	.	.	.	0.575	.
Sasol	0.566	.	.		0.418	.

According to the results in Table 3, the seven variables are all in support of the first factor (see unrotated factor solutions). All these variables have high loadings on this factor as opposed to others. No significant loading is evident as far as the second factor is concerned, however factor 3 has only one significant loading from Pick'n Pay. As a result of various factors' model representative variables are not prominent. This makes interpretation of these results somehow cumbersome. Hence orthogonal rotation was implemented as discussed in previous sections to transform the results so as to be able to clarify the exact meaning of each variable on a factor and also to decide on the naming of common factors. The results of the rotated factors as seen in Table 3 shows ABSA and Standard bank loading high on the first factor and MTN, Vodacom and Sasol on the second factor. Shoprite and Pick 'n Pay is definitely significant members of the third factor. The three factor coefficients in Table 3 are evaluated with the least squares technique according to the standardised scoring coefficients and presented as linear combination of the seven stock prices following (8) as:

$$f_1 = 0.567x_1 + 0.392x_2 - 0.059x_3 - 0.079x_4 - 0.089x_5 - 0.097x_6 + 0.048x_7$$
$$f_2 = -0.151x_1 + 0.118x_2 + 0.003x_3 - 0.066x_4 + 0.378x_5 + 0.369x_6 + 0.168x_7$$
$$f_3 = -0.025x_1 + 0.003x_2 + 0.419x_3 + 0.396x_4 + 0.042x_5 - 0.049x_6 - 0.026x_7 \quad (18)$$

This is a final MLFA model for the seven stock prices where x_1 to x_7 represent Absa, Standard bank, Shoprite, pick 'n Pay, MTN, Vodacom and Sasol and $f_1$to $f_3$are the latent factors. It is evident that Absa and Standard Bank have positive significant contribution to the financial sector while others have proven to have little contribution on this sector. This evidence is proven by the standardised coefficients of each variable. The findings are consistent with the interpretation of Table 3. MTN, Vodacom and Sasol still remain significant positive contributors to the Factor 3 (Industrial sector) so does Shoprite and Pick 'n Pay to Factor 2 (Retail sector). The MLFA model (18) was next checked for adequacy using the chi-square test and the AIC. Factor solutions were also assessed for consistency and convergence using Cronbach's alpha and the Tucker-Lewis coefficients. The latter also checks if the fitted model has reasonable number of factors or not by testing the null hypothesis that certain number of factors are sufficient against the alternative that at least one common factor is needed. The results are summarised in Table 4.

Table 4: Model evaluation test results

Null hypothesis	Prob>ChiSq	Akaike's Information Criterion	Tucker and Lewis's Reliability Coefficient	Cronbach's alpha coefficient
H_0: 1 common factor is sufficient($m = 1$)	142.9337(<.0001)	115.369	0.9188	0.902
H_0: 2 commonfactorsare sufficient($m = 2$)	61.396 (<.0001)	45.590	0.94116	0.895
H_0: 3 common factors are sufficient($m = 3$)	5.528 (>0.1369)**[1]	-0.448**	0.99257**	0.911**

[1] *Indicate significant results*

The chi square results in Table 4 confirm that three factors are needed to explain the seven stock prices. The observed probability on the parenthesis associated with the third suggested hypothesis is significant leading to the acceptance of this hypothesis. Moreover, the AIC attain its minimum value at three common factors implying that three factor solutions are appropriate for these data. Therefore, based on these findings, the seven stock prices can be explained by three common factors. The solution on the third iteration was so close to the optimum to such an extent that PROC FACTOR could not find a better solution. Consequently, SAS displayed the message that convergence criterion is satisfied for three factors as confirmed by Tucker and Lewis's reliability coefficient corresponding to the third hypothesis.As suggested by SAS institute Inc., it is prudent to repeat PROC FACTOR analyses with different prior communality estimates on the appearance of similar message. This ensures the correctness of the factor solution as different prior estimates could lead to similar solution or most probably to even worse local optima, as indicated by the information criteria or the chi-square values. While Cronbach's alpha confirms that the three factors solution is diverse and reliable, the coefficient for the third solution is more significant emphasising this model to be favourable.

5. Conclusion

This paper delivers a first-hand enquiry into communal factor behind day-to-day stock price activities of some of the South African largest and well-known companies listed on the Johannesburg stock exchange. Using the daily data set for ABSA, Standard bank, Shoprite, Pick n Pay, MTN, Vodacom and Sasol covering the period January 2010 to December 2014, a trivariate common factor model for the seven companies was estimated. By modelling the seven stock prices, we implicitly assumed update on monetary rudiments is aggregated in either equity market. As a result there is a common trend of cumulated random information arrivals interconnecting these stock indices. It should be noted that studies on stock performance worldwide and in South Africa have been conducted, but none of these studies analysed stock performance of the seven companies. This makes it difficult for the researcher to compare the current study findings with those already published around the area due to this difference. The findings revealed that among the seven stock indexes, Absa and Standard Bank are the predominant sources of price (attributes a weight of 46.9%) and both are in favour of the first factor (Financial sector). While Shoprite and Pick 'n Pay contributes up to 12.7% to the total variation of the common factor (Retail sectors); Sasol, MTN and Vodacom attribute a weight of 10.8% to the third factor (Industrial sectors). Using a Chi-square test and the AIC, it was confirmed that the three factor models are suitable implying that instead of treating seven companies as individuals, MLFA suggests they be treated as collectives since they share common stochastic trend. The findings suggest that banking institutions in South Africa is the most contributing factor to the well-being of the country. One could infer that, possibly the banking institution contributes more towards tax revenue of the country. Conversely, the contribution that industrial companies make to the economic growth of South Africa is a little bit lower than that of retail companies.

The three factors were also found to be diverse and consistent with the seven variables according to Cronbach's alpha, and Tucker and Lewis reliability coefficients. The findings confirmed that the information included in these factors is non-repetition. Treating these companies as groups could help save time during the period when annual financial reports are delivered and also when policy makers score or rank the companies listed on stock exchange according to their stock price or sales performance. A model of Financial sector (Factor 1) can be used as a classifier for all companies dealing with finances and model 2 (Factor 2) is recommended to use for identification or classification of all retail outlets. The last model (Factor 3) is suitable for any industrial company classification. The findings differ from those by Kumar (2013) which identified industrial performance as one of most dominant factors of stock markets in India. Maximum likelihood factor analysis method has been proven to be effective in reducing multitudinous financial stock prices to few latent factors. The findings of this study are in support of other studies that used MLFA. Factor analysis provides a clear reflection of overall target financial significance. The findings of this study are relevant as it gives investors more reasonable investment reference. One would make an informed decision when selecting a company to make an investment. Stock marketers may also refer to the findings of this study when reworking policies around stocks. They may further use the three MLFA models to make projections of the stock prices for the three sectors into the future. The findings may lessen investors and financial advisors' tasks during the time when they are buying or selling the stocks. The study makes a valid methodological contribution which scholars and financial analysts may refer to. Though the results of this study are not

generalizable to the entire world or other sectors such as macroeconomics, etc. the framework may be followed by scholars analysing similar data sets. Future studies may take into consideration many other companies which were not included in the data. Furthermore, MLFA method may be used in conjunction with other statistical techniques such as logistic regression and discriminant analyses where the three sectors will be used as independent variables to a dichotomous or polychotomous variable of interest.

References

Al-Debie, M. & Walker, M. (1999). Fundamental information analysis: An extension and UK evidence. *The British Accounting Review*, 31(3), 261-280.

Al-Radaideh, Q.A., Assaf, A. & Alnagi, E. (2013).Predicting Stock Prices Using Data Mining Techniques. The International Arab Conference on Information Technology (ACIT'2013).

Bastos, J. A. & Caiado, J. (2010). The structure of international stock market returns. *Rua do Quelhas*, 6, 1200-781.

Brooks, C. (2002). Introductory econometrics for finance: Cambridge University Press. Cambridge

Cao, Y. (2010). A Bayesian Approach to Factor Analysis via Comparing Posterior and Prior Concentration.Thesis submitted to the University of Toronto.

Cheng, W. (2005). Factor Analysis for Stock Performance. *P*rofessional Masters thesis submitted to Worcester Polytechinic Institute.

Costello, A. B. & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation,* 10(7).

Cronbach, L. J. & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and psychological measurement*, 64(3), 391-418.

Făt, C. M. & Dezsi, E. (2012). A Factor Analysis Approach of International Portfolio Diversification: Does it Pay Off? *Procedia Economics and Finance*, 3, 648-653.

Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological assessment*, 7(3),286.

Greene, W. H. (2003). Econometric Analysis. Pearson Education, Upper Saddle.

Gu, A. & Zeng, P. (2014). Sector-Based Factor Models for Asset Returns.*arXiv :1408.2794vi*.

Gujarati, D. N. (2003). *Basic Econometrics* (4e). McGraw-Hill Inc. New York.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson R. E. (2010). Multivariate Data Analysis: A Global Perspective, Prentice-Hall: Upper Saddle River, New Jersey.

Hui, T. K. (2005). Portfolio diversification: a factor analysis approach.*Applied Financial Economics*, 15(12), 821-834.

Illueca, M. & Lafuente, J. A. (2002). International stock market linkages: A factor analysis approach. *Journal of Asset Management*, 3(3), 253-265.

Kumar, R. (2013). The Effect of Macroeconomic Factors on Indian Stock Market Performance: A Factor Analysis Approach. *IOSR Journal of economics and finance*, 8(1),14-21.

Lev, B. & Thiagarajan, S. R. (1993). Fundamental information analysis. *Journal of Accounting research*,31, 190-215.

Muca, M., Puka, L., Bani, K. & Shahu, E. (2013). Principal Component and the Maximum Likelihood methods as tools to analyze large data with a psychological testing example. *European Scientific Journal*, 9(20), 176-183.

Napper, L. E., Wood, M. M. & Klahn, J. A. (2008). Convergent and discriminant validity of three measures of stage of change. J*ournal of the Society of Psychologists in addictive behaviours*, 22(23), 362-371.

Osborne, J. W. (2005). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation*, 15(12), 1-9.

Raykov, T. & Marcoulides, T. R. G. A. (2008). An Introduction to applied multivariate analysis and structural equation modelling. Routledge, New York.

Rencher, A. C. (2003). Methods of multivariate analysis. John Wiley and Sons, New York.

Richard, A. J. & Dean, W. W. (2002). Applied multivariate statistical analysis.Prenticee Hall, London.

Ritchie, J. C. (1996). Fundamental analysis: a back-to-the-basics investment guide to selecting quality stocks: Irwin Professional Publishers.

Rossi, P. H., Wright, J. D. & Anderson, A. B. (2013). Handbook of survey research: Academic Press, San Diego.

Shadkam, E. (2014). FC approach in portfolio selection of Tehran's stock market. *Journal of Asian Finance Economics and Business*, 1(2), 31-37.

Spearman, C. (1904). General Intelligence, objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-292.

Tabachnick, B. G. & Fidell, L. S. (2007). Using Multivariate Statistics. Upper Saddle River, New Jersey, Pearson Allyn & Bacon.

Tsang, P. M., Kwok, P., Choy, S. O., Ng, S. C., Tsang, M. J., Koong, K. & Wong, T. L. (2007). Design and implementation o NN5 for Hong Kong stock price forecasting. *Engineering Applications and Artificial Intelligence*, 20, 453-461.

Tsay, R. S. (2005). *Analysis of Financial Time Series*, 3rd ed. Wiley and Sons, New York.

Tsay, R. S. (2010). *Analysis of financial time series*, 3rd ed., New York, John Wiley & Sons.

Tuluca, S. A. & Zwick, B. (2001). The effects of the Asian crisis on global equity markets.*Financial Review*, 36(1), 125-142.

Valadkhani, A., Chancharat, S. & Harvie, C. (2008). A factor analysis of international portfolio diversification. *Studies in Economics and Finance*, 25(3), 165-174.

Wang, Y. F. (2003). Mining stock price using fuzzy rough set system. *Expert Systems with applications*, 24(1), 13-23.

Wichern, D. W. & Johnson R. A. (2007). Applied Multivariate Statistical Analysis. Upper Saddle River. Prentice Hall, Inc.

Wu, M. C., Lin, S. Y. & Lin, C. H. (2006). An effective application of decision tree to stock trading. *Expert Systems with applications*, 31(2), 270-274.

Xaba, D. L., Moroke, N. D., Arkaah,Y. J. & Pooe, C. (2016). Modeling South African major banks' closing stock prices: A Markov-Switching Approach. *Journal of Economic and Behavioral Studies,* 8(1), 36-40.

Xin, S. (2007). The application of Factor Analysis in Chinese Stock Market. Thesis submitted to the College of science Tianjin Polytechnic University Tianjin (300160).