



Journal Homepage: -[www.journalijar.com](http://www.journalijar.com)  
**INTERNATIONAL JOURNAL OF  
 ADVANCED RESEARCH (IJAR)**

Article DOI:10.21474/IJAR01/9008  
 DOI URL: <http://dx.doi.org/10.21474/IJAR01/9008>



### RESEARCH ARTICLE

## THE APPLICATION OF PRINCIPAL COMPONENT ANALYSIS AND FACTOR ANALYSIS TO STOCK MARKETS RETURNS.

Samreen Fatima<sup>1</sup>, Tayyab Raza Fraz<sup>2</sup> and Rfia Shafi<sup>3</sup>.

1. Assitant professor, Ph.D Department of Statistics, Faculty of Science, University of Karachi, Pakistan.
2. Lecturer, M.Phil Department of Statistics, Faculty of Science, University of Karachi, Pakistan.
3. Assitant professor, M.Sc Department of Statistics, Faculty of Science, University of Karachi, Pakistan.

### Manuscript Info

#### Manuscript History

Received: 02 March 2019  
 Final Accepted: 04 April 2019  
 Published: May 2019

**Key words:-**  
 PCA, FA.

### Abstract

It is evident from literature that few Asian markets have strong linkage with US markets. The movement in one market often affects the other market if they have linkage. Current study explores the linkage among the major Asian and US stock markets using PCA and FA techniques. PCA identifies patterns in series on the variability while FA defines the structures using covariance and correlation. In this study, weekly closing returns of fourteen stock markets namely: KSE-100 (Pakistan), Nikkie225 (Japan), S&P 500(US), Nasdaq Composite and DJI (US), KLSE (Malaysia), BSESN (India), HIS (Hong Kong), JKSE (Indonesia) SSE(China), KS11(Korea), TWII (Tiwian), CSE(Sri Lanka) and TASI (Saudi Arabia) spanning from 1<sup>st</sup> January , 2001 to 14<sup>th</sup> January, 2019 are used as case study. Only first nine PCA's are constructed and from them first six PCA's are chosen as they contain almost 79.4% of total variability. First two PCA provide important information, such that PC1 consist of a group of US and few Asian stock markets (BSESN, NIKKIE 225, HIS, KS11 and TWII). Whereas, PC2 contains all US and Asian markets (JKSE, KLSE and SSE). Furthermore, no relationship of KSE-100, CSE and SSE with US markets is found. Moreover, Factor analysis with VARIMAX method gives quite different results as compare to PCA. FA1 comprises of a group of US markets and FA2 represents only two Asian stock markets (KS11 and TWII). Finally, FA method is found more appropriate as it utilizes correlation/covariance as well.

Copy Right, IJAR, 2019.. All rights reserved.

### Introduction:-

Managing High-dimensional data is very difficult task for the researcher. Extensive computer resources and special statistical methodologies are required to handle and model such data. Various methods have been devised for dimension reduction; significant among them are: Principal Component Analysis (PCA) and factor analysis (FA) that are commonly employed for clustering of data in the field of medicine, bio sciences, economics and finance etc. PCA is a well-known statistical technique for reducing the dimensionality and the complexity of data with minimum loss of information. In PCA, a large number of inter-related variables can be structured into new group of uncorrelated variables. Moreover, this new set of variables are ordered sequentially with the first component

**Corresponding Author:- Samreen Fatima<sup>1</sup>**

Address:-Assitant professor, Ph.D, Department of Statistics, Faculty of Science, University of Karachi, Pakistan.

explaining as much of the variation as it can. Each principal component is a linear combination of the original variables in which the coefficients indicate the relative importance of the respective variable.

The PCA is obtained by linear combination of original (manifest) variables by considering sum of diagonal elements of covariance or correlation matrix as their total variance and decompose them so that,  $V(PC_1) > V(PC_2) > \dots > V(PC_k)$ . In contrast, factor analysis not only explains the diagonal terms but also the off-diagonals (covariance/correlation) as well. However, the aim of both techniques is to reduce the dimensionality of a data set. Whereas, dimension of the dataset is defined as 'n' with the relationship  $n > q$  where, q is the number of factors in PCA/FA.

PCA was first introduced by (Pearson, 1901) as a tool in exploratory data analysis and also for constructing predictive models. In the field of finance PCA is specifically used to identify the effectiveness of the components having high variability. Moreover, PCA is considered as a very powerful tool capable of handling big data problems efficiently. In 1931, Turnstone developed multiple factor model based on the psychological theory by the name of Factor Analysis (FA) based on research work of (Pearson, 1901) and (Spearman, 1987). Furthermore, FA and PCA were compared and explained as two different techniques by (Jolliffe & Jolliffe, 1986). In contrast, recent study (Fan, Liao, & Mincheva, 2013) showed that PCA is closely related to factor analysis. However, heavy tailed errors and outliers of PCA and FA were discussed by ((Fan, Wang, Wang, & Zhu, 2017) & (Sun, Zhou, & Fan, 2018)).

According to the researchers, PCA is one of the most growing techniques applied to study the market cross-correlation and systemic risk ((Billio, Getmansky, Lo, & Pelizzon, 2012), (Kritzman, Li, Page, & Rigobon, 2010) and (Zheng, Podobnik, Feng, & Li, 2012)). In addition to this, for summarizing the large data sets in a way that relationship and patterns can be understood easily, factor analysis (FA) is used. In contrast to PCA, FA normally regroup the variables into a limited set of clusters based on shared variance. Hence, it helps to isolate (clusters) and concepts. According to (Bartholomew, Knott, & Moustaki, 2011), FA reduces dimensionality of large sets of measurable and unobservable variables that are represented by fewer latent variables sharing a common variance. Furthermore, (Egloff, Leippold, & Wu, 2010) suggested a two-factor model of volatility by applying PCA to analyze the dimensionality of the volatility dynamics, capturing the long and short-term fluctuations in the volatility term structure. Both optimistic and pessimistic view of investors by using PCA is discussed by (M. Baker & Wurgler, 2006). Whereas, (S. R. Baker, Bloom, & Davis, 2016) described a policy which measures uncertainty index of Economic policy. As a case study, (Chong, Yap, & Mohamad, 2013) applied Factor Analysis on twenty eight financial ratios of trading companies listed with Malaysian Stock Exchange and examined their distributional properties and found seven are positively skewed. Moreover, (Armeanu & Lache, 2008) selected eight variables of insurance companies which were operated in the Romanian market in 2006. Armeanu used PCA method to examine the financial strength of the insurance companies. Further discussion on principal component analysis for analysis of stock portfolio management is done by (Pasini, 2017). Their result showed that PCA was useful to diversify the risk which is helpful for the investor.

The main objectives of the study are manifold: 1) To explore the linkage among the Asian and US stock markets. 2) To make groups based on the variability using two statistical techniques Principal Component Analysis and Factor Analysis. 3) To observe the similarities among the selected Asian Stock markets.

The outline of this paper is as follows: Firstly, section 2 presents a brief introduction of Principal Component Analysis and Factor Analysis. Secondly, section 3 deals with data analysis using Principal Analysis and Factor Analysis methods based on data from 1<sup>st</sup> January 2001 to 16<sup>th</sup> January, 2019 of the selected stock markets data of closing price data. Finally, section 4 summarizes the conclusion.

## **Materials and Methods:-**

### **Statistical Analysis**

Statistical analysis is carried out using the following well-known techniques. In this study, these techniques are applied using Eviews 9 and Minitab 17 software.

### **Component Analysis**

Principal component analysis (PCA) is a technique mostly used to highlight strong patterns present in a dataset. If a number of possibly correlated variables are transformed into a smaller number of uncorrelated variables, they are called principal components. Dealing with large number of variables, PCA is found extremely useful in eliminating or reducing dimensions.

Let  $X_1, X_2, \dots, X_p$ , is a vector of ‘p’ random variables whose variance and covariance matrix exists. Principal components are specific linear combinations such that these linear combinations are independent and obtained by applying Eigen analysis on its covariance matrix, which provides new axes. The new axes represent the directions with maximum variability and provide a simpler description of the covariance structure. Therefore, Principal component mainly depends on the covariance/correlation matrix and hence the assumption of multivariate normality is not required.

The linear combinations of ‘p’ random variables and Eigen vectors is mathematically defined as a system of linear regression equations,

$$\begin{matrix} Y_1 = e_{11}X_1 + e_{12}X_2 + \dots + e_{1p}X_p \\ \vdots \\ Y_p = e_{p1}X_1 + e_{p2}X_2 + \dots + e_{pp}X_p \end{matrix}$$

Which can be represented in matrix notation as follows,

$$y = \varepsilon' X$$

Where  $i=1,2,\dots,p$ , matrix of Eigen vectors  $\varepsilon = (e_{ti})_{p \times p}$ ,  $X = x_{p \times 1}$  and  $Y$  is also  $y_{p \times 1}$  vector.

Furthermore,  $X = (x_i)_{n \times p}$  has population variance covariance matrix represented by

$$\text{Var}(Y_i) = \sum_{k=1}^p \sum_{l=1}^p e_{ik} e_{il} \sigma_{kl} = e_i' \Sigma e_i$$

In general, population covariance matrix or population data are not available therefore MLE is used in place of population covariance matrix ( $\Sigma$ ).

The sample estimate of  $\Sigma$  is obtained by sample variance-covariance matrix, given by

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{x})(X_i - \bar{x})'$$

From this sample variance-covariance matrix estimates of Eigen values,  $\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_p$  the corresponding Eigenvectors  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p$  are computed. Since we know that,

$$\begin{aligned} \text{Trace}(S) &= \sigma_1^2 + \sigma_2^2 + \dots + \sigma_p^2 = \hat{\lambda}_1 + \hat{\lambda}_2 + \dots, \hat{\lambda}_p = \sum \hat{\lambda}_j \\ \hat{\lambda}_1 &> \hat{\lambda}_2 > \dots, > \hat{\lambda}_p \end{aligned}$$

The Eigen values  $\hat{\lambda}_j$  are the variance of the corresponding PC.

So, the estimated principal components using the Eigenvectors as coefficients can be written as

$$\begin{matrix} \hat{Y}_1 = \hat{e}_{11}X_1 + \hat{e}_{12}X_2 + \dots + \hat{e}_{1p}X_p \\ \vdots \\ \hat{Y}_p = \hat{e}_{p1}X_1 + \hat{e}_{p2}X_2 + \dots + \hat{e}_{pp}X_p \end{matrix}$$

In this study returns of daily closing prices of the stock market i.e.,  $p = 14$  ( $x_1, x_2, \dots, x_{14}$ ) have been considered for analysis. Where,  $X_i$ 's are the returns of the  $i^{\text{th}}$  series (see in abstract).

Furthermore, the first ‘k’ principal components out of ‘p=14’ components are selected in such a way that the proportion of variation explained by those first ‘k’ principal components is to be as large as possible.

**Factor Analysis**

In factor analysis the p-observable variables ( $x_1, x_2, \dots, x_p$ ) are grouped according to higher correlations, i.e. groups of variables are formed in such a way that variables within a particular group are highly correlated. Factor analysis explores the latent structure defined by linear combination of manifest variables. FA has few latent variables  $f_1, f_2, \dots, f_m$  which also explain covariance and correlation among the observable variables.

The factor analysis model is written in vector form as,

$$X - \mu = \Lambda f + \epsilon$$

Where,  $\Lambda$  is the matrix of factor loadings, 'f' represents vector of common factors and ' $\epsilon$ ' represents specific factor. FA basically focuses on modeling the covariance among the Y's. So, the variance/covariance matrix ( $\Sigma$ ) can be expressed in terms of  $\Lambda$  and  $\psi$  as  $\Sigma_p = \Lambda\Lambda' + \psi$ . Here the variance of ' $x_i$ ' is partitioned into communality (due to f's) and a specific variance (unique to  $x_i$ ), given by  $\text{var}(x_i) = \lambda^2_{i1} + \lambda^2_{i2} + \dots + \lambda^2_{im} + \psi_i$ .

Where,  $c_i^2 = \lambda^2_{i1} + \lambda^2_{i2} + \dots + \lambda^2_{im} =$  Communality and  $\psi_i =$  Specific variance. The communality,  $c_i^2$  is also referred as the  $i^{\text{th}}$  explained by the specific variance and  $\Psi_i$  is called unique variance or residual variance. The larger communality for each variable is indicative of a successful factor analysis.

One disadvantage of the factor analysis model is that it does not always fit the data. In such cases, two problems are faced regarding the estimates: (1) Number of factors are not known, and (2) it is unclear that what the factors are? The other statistical procedures, however, do not lead to such obvious consequences regarding the estimates or tests in case of failure of assumptions. A transformation of the factor loadings and the resulted transformation of the factors are called factor rotation. It happens in many cases that all factor loadings obtained from initial loadings by an orthogonal transformation are not easily interpretable. It is therefore a common practice to rotate them in some way to achieve a simpler structure. In fact, it is always desirable to have a pattern of loadings such that each variable is loaded high on a single factor and has small loadings on the remaining factors. This transformation rotates the common factors in m-dimensional space. Moreover, there are various methods of carrying out Factor analysis, however in this study Factor Rotation using VARIMAX method is used.

#### Factor Rotation using VARIMAX method

The most common method for factor rotation is the VARIMAX procedure, which maintains axes at right angles. This is a method that also minimizes the number of variables with high loadings on a factor.

Let the rotated matrix of factor loadings be denoted by  $L^* = [l_{ij}^*]$  and  $c_i^2$  represents the  $i^{\text{th}}$  communality then,  $\tilde{l}_{ij} = l_{ij}^* / c_i$  is defined as the rotated coefficients in terms of square root of communalities. If P is an ' $m \times m$ ' orthogonal matrix such that  $L^* = LP$  and  $F^* = P'F$ , then the matrix P is chosen to maximize the following:

$$V = \frac{1}{k} \sum_{j=1}^m \left[ \sum_{i=1}^k (\tilde{l}_{ij}^*)^4 - \frac{1}{k} \left( \sum_{i=1}^k \tilde{l}_{ij}^{*2} \right)^2 \right]$$

Where, maximized value of 'V' means that squares of the loadings are spread out on each variable as much as possible. The interpretations of common factors become simpler by finding groups of very large and very small coefficients in any column of the rotated matrix of factor loadings.

#### Data Analysis and Results:-

Weekly closing returns of fourteen markets data: KSE-100 (Pakistan), Nikkie 225 (Japan), S & P 500, Nasdaq Composite and DJI (US), KLSE (Malaysia), BSESEN (India), HIS (Hong Kong), JKSE (Indonesia) SSE(China), KS11(Korea), TWII (Tiwian), CSE (Sri Lanka) and TASI (Saudi Arabia) markets are obtained from yahoo finance.com. Furthermore, for visual understanding data from 1<sup>st</sup> January, 2001 to 14<sup>th</sup> January, 2019 is plotted in a graph given below, Figure 1.

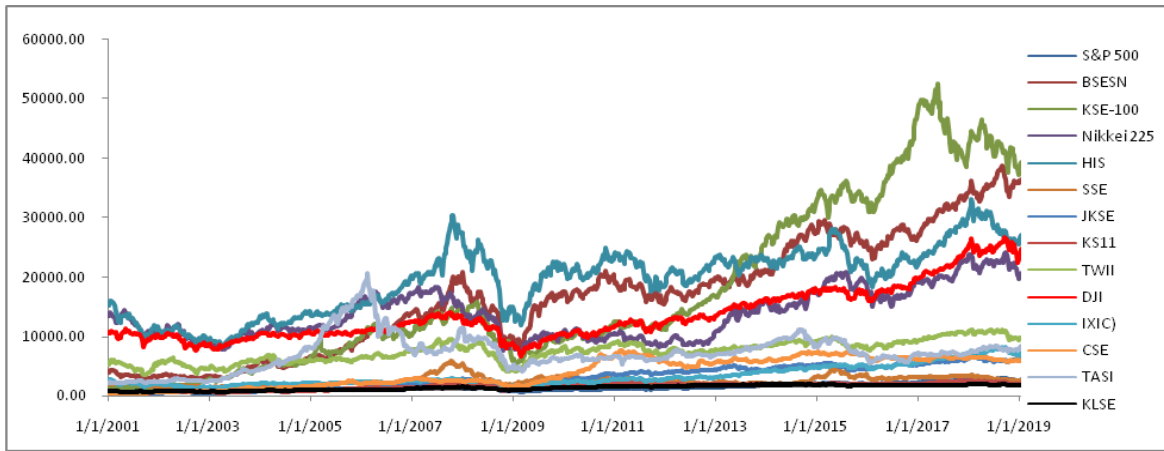


Figure 1:-time series plot of daily share price of considered stock markets.

Returns of above mentioned indices are calculated by taking the logarithmic ratio of  $p_t$  and  $p_{t-1}$ ,  $r_t = \log (p_t / p_{t-1})$ .

Moreover, descriptive statistics of the returns are summarized in Table 1 which shows that all countries have positive mean returns indicating the fact that prices increased during the selected period. Furthermore, stock markets except CSE are negatively skewed and show a high probability of earning returns which is greater than mean. Whereas, kurtosis for each series is much larger than normal distribution value indicating that the tails of each series are thicker than the normal distribution. Nevertheless, DJI, Nikkie225, S&P 500 and TASI have thicker tails among all of them which is also confirmed by Jarque-Bera test statistics at 5% level of significance. Therefore, the null hypothesis of normality needs to be rejected.

Table 1:-Provides general statistics of returns of fourteen stock markets for the period 1<sup>st</sup> January, 2001 to 14<sup>th</sup> January, 2019.

	Mean	Median	Maximum	Minimum	Std. Dev.	Skewness	Kurtosis	Jarque-Bera	Probability
DLOG(BSESN)	0.002	0.004	0.132	-0.174	0.030	-0.495	6.382	486.973	0.000
DLOG(CSE)	0.003	0.001	0.180	-0.113	0.025	0.787	9.533	1770.363	0.000
DLOG(DJI)	0.001	0.003	0.107	-0.200	0.023	-1.072	11.926	3303.779	0.000
DLOG(HIS)	0.001	0.002	0.117	-0.178	0.029	-0.259	5.780	313.484	0.000
DLOG(IXIC)	0.001	0.003	0.131	-0.175	0.030	-0.561	6.448	515.459	0.000
DLOG(JKSE)	0.003	0.004	0.116	-0.233	0.030	-0.952	9.211	1654.738	0.000
DLOG(KLSE)	0.001	0.002	0.067	-0.114	0.018	-0.771	7.937	1048.966	0.000
DLOG(KS11)	0.001	0.003	0.170	-0.229	0.030	-0.695	9.112	1540.763	0.000
DLOG(KSE-100)	0.003	0.006	0.109	-0.201	0.031	-1.157	8.389	1348.776	0.000
DLOG(NIKKEI 225)	0.000	0.002	0.114	-0.279	0.030	-1.091	11.202	2824.341	0.000
DLOG(SP&500)	0.001	0.002	0.114	-0.201	0.024	-0.924	10.737	280.49	0.000
DLOG(SSE)	0.000	0.000	0.139	-0.159	0.033	-0.285	5.808	321.887	0.000
DLOG(TASI)	0.001	0.003	0.138	-0.238	0.034	-1.400	10.954	2787.907	0.000
DLOG(TWII)	0.001	0.002	0.183	-0.131	0.028	-0.197	6.881	596.650	0.000

Notes: Authors calculations using Eviwes 9 software.

The pair wise correlations between the markets are computed to measure the relationship (see Table 2). Correlation matrix shows that all three considered markets of US have strong positive correlation. It is interesting to note that KSE-100, TASI and CSE have weak correlations with all markets under consideration, showing that price movements in these markets are independent to certain extent. However, remaining pairs of Asian and US markets returns series have moderate correlations between 0.5 to 0.7.

**Table 2:-**Correlation of selected returns

	BSESN	CSE	DJI	HIS	IXIC	JKSE	KLSE	KS11	KSE-100	Nikkie 225	S&P-500	SSE	TASI	TWII
BSESN	1.00													
CSE	0.19	1.00												
DJI	0.48	0.13	1.00											
HIS	0.59	0.16	0.56	1.00										
IXIC	0.47	0.16	0.84	0.58	1.00									
JKSE	0.44	0.13	0.32	0.48	0.30	1.00								
KLSE	0.42	0.16	0.31	0.53	0.32	0.52	1.00							
KS11	0.53	0.11	0.50	0.65	0.54	0.45	0.48	1.00						
KSE-100	0.16	0.04	0.13	0.16	0.11	0.12	0.14	0.17	1.00					
Nikkie225	0.49	0.16	0.58	0.63	0.58	0.40	0.39	0.61	0.12	1.00				
S&P500	0.49	0.15	0.97	0.59	0.91	0.34	0.32	0.53	0.11	0.60	1.00			
SSE	0.16	0.06	0.15	0.36	0.15	0.22	0.24	0.22	0.09	0.22	0.15	1.00		
TASI	0.13	0.10	0.13	0.14	0.13	0.15	0.15	0.12	0.13	0.20	0.14	0.11	1.00	
TWII	0.46	0.13	0.42	0.62	0.48	0.42	0.48	0.69	0.19	0.51	0.45	0.23	0.12	1.00

Notes: Authors calculations using Eviwies 9 software.

### PCA and Factor Analysis

Weekly averages of daily return series of selected Asian and US markets are used for the PCA and factor analysis. These methods assume that the variables have no serial correlations. Moreover, PCA combines the selected stock markets (having maximum variability) in terms of the similarities of their co-movements. Furthermore, it is well known that markets with correlated and close co-movement patterns are not good from portfolio diversification perspective. Whereas, PCA technique not only investigates the co-movement patterns of the Asian stock markets but also provides information for portfolio diversification prospects for investors. In short, there is no rule of thumb to determine the number of clusters of returns with similar contemporaneous movement patterns. Therefore, in order to identify similar movement pattern in each PCA, the Eigen values and the associated percentage of variance are considered.

However, the criteria to select appropriate number of the PCS is based on the proportion of percentages associated with each PC. In the current study, 14 stock market returns are considered and nine statistically significant PCS are considered from 14, given in Table 3. First, PCA consists of S&P500, BSESN, Nikkie 225, HIS, KS11, TWII, DJI and NASDAQ using the criteria  $1/\sqrt{\text{total no. of variables}}$ . The first PC contains 42% cumulative variation to

the total variation in the returns of the fourteen stock markets under consideration. This PC shows strong link between stable Asian stock markets and three US markets. Moreover, second PC indicates interesting results that is, JKSE, KLSE and SSE have strong relationship with each other however, they all have opposite movements with selected US markets. Furthermore, PC3 contains only TASI, whereas, PC4 shows that CSE and KSE-100 move in opposite directions. While, PC5 contains only SSE and PC6 shows that SSE and TASI have opposite movements. Moreover, in PC7, JKSE moves opposite to TWII and KLSE and PC8 represents only BSESN. At last, PC9 indicates Nikkie 225 and JKSE have negative influence on KLSE. Therefore, extracting first nine PCS contribute to 91.5% variation in all. The rest of the five PCS account for remaining 8.5% of variation. Historical literature shows that mostly PC1 or only few PCS contain maximum information ((Pérignon, Smith, & Villa, 2007), (Kritzman et al., 2010), (Billio et al., 2012), and (Zheng et al., 2012)). To put it in a nut shell, it is decided that first six PCS are significant among the nine constructed PCS as they contain 79.4% variation and the remaining three PCS have less than 5% variation.

Table 3(a & b) provide the result of PCA including: Eigen values, eigenvectors, and proportion of variability using both covariance and correlation matrixes.

**Table 3a:-**Eigen analysis of the Correlation Matrix

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
Eigenvalue	5.93	1.43	1.036	0.977	0.904	0.834	0.659	0.548	0.478	0.434	0.3	0.279	0.164	0.02
Proportion	0.424	0.102	0.074	0.07	0.065	0.06	0.047	0.039	0.034	0.031	0.021	0.02	0.012	0.001
Cumulative	0.424	0.526	0.6	0.67	0.735	0.794	0.841	0.88	0.915	0.946	0.967	0.987	0.999	1

Notes: Authors calculations using Minitab 17 software.

**Table 3b:-Eigen vectors**

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
S&P 500	0.341*	-0.421*	-0.06	0.036	0.063	-0.039	-0.183	0.102	0.031
BSESN	0.29*	0.09	0.028	-0.102	-0.19	0.081	-0.078	-0.809*	0.33
KSE-100	0.093	0.203	-0.406	0.621*	-0.481	-0.376	-0.135	0.01	-0.063
Nikkei 225	0.317*	-0.039	-0.013	-0.001	0.067	0.077	0.26	-0.121	-0.419
HIS	0.342*	0.119	0.12	-0.001	0.073	-0.078	0.102	-0.138	0.122
SSE	0.136	0.32*	0.06	0.129	0.706*	-0.55	-0.062	-0.07	0.001
JKSE	0.245	0.333*	0.113	-0.119	-0.053	0.249	-0.556*	0.016	-0.599*
KS11	0.324*	0.118	0.169	0.058	-0.117	0.077	0.396*	0.064	-0.118
TWII	0.3*	0.188	0.15	0.061	-0.144	0.028	0.479*	0.276	0.02
DJI	0.329*	-0.419*	-0.064	0.053	0.058	-0.056	-0.205	0.079	0.029
NASDAQ	0.331*	-0.388*	-0.045	0.025	0.046	-0.057	-0.067	0.132	0.064
CSE	0.097	0.082	-0.459	-0.732*	-0.195	-0.418*	0.089	0.079	-0.083
TASI	0.096	0.154	-0.722*	0.104	0.365	0.516*	0.121	0.001	0.083
KLSE	0.252	0.372	0.106	-0.122	-0.081	0.142	-0.308	0.432	0.551*

Notes: \* indicates selected stock markets of each PC. Authors calculations using Minitab 17 software.

In Factor analysis maximum likelihood with unrotating method is adopted. Whereas, in maximum likelihood the first factor contains S&P500, BSESN, NIKKIE 225, HIS, KS11, TWII, DJI and NASDAQ same as the first PC. Furthermore, FA2 indicating interesting result that JKSE, KLSE and SSE have strong relationship to each other while they have opposite effect with all selected US markets. In addition to this, FA3 having TASI, FA4 showing CSE and KSE-100 both have negative influence on each other and FA5 has SSE only. Whereas, FA6 represents SSE and TASI move in opposite direction, in FA7 TWII and KLSE have negative impact on JKSE and FA8 has BSESN. Furthermore, Nikkie 225 and JKSE have negative impact on KLSE in FA9. Overall results obtained from FA are same as PCA.

**Table 4a:-Principal Component Factor Analysis of the Correlation Matrix :Unrotated Factor Loadings**

Variable	FA1	FA 2	FA 3	FA 4	FA 5	FA 6	FA 7	FA 8	FA 9	Communality
S&P 500	0.83	-0.503	-0.061	0.036	0.06	-0.036	-0.148	0.075	0.021	0.98
BSESN	0.706	0.108	0.028	-0.101	-0.18	0.074	-0.064	-0.599	0.228	0.973
KSE-100	0.228	0.242	-0.413	0.613	-0.457	-0.344	-0.11	0.008	-0.044	0.998
Nikkei 225	0.772	-0.047	-0.013	-0.001	0.064	0.071	0.211	-0.089	-0.289	0.743
HIS	0.834	0.142	0.123	-0.001	0.069	-0.072	0.083	-0.102	0.085	0.765
SSE	0.332	0.383	0.061	0.128	0.671	-0.502	-0.051	-0.051	0.001	0.985
JKSE	0.597	0.398	0.115	-0.117	-0.05	0.227	-0.452	0.012	-0.414	0.972
KS11	0.791	0.142	0.172	0.058	-0.111	0.07	0.321	0.047	-0.081	0.808
TWII	0.731	0.225	0.153	0.06	-0.137	0.025	0.389	0.204	0.014	0.825
DJI	0.802	-0.501	-0.065	0.052	0.055	-0.051	-0.167	0.059	0.02	0.939
IXIC)	0.807	-0.464	-0.046	0.025	0.044	-0.052	-0.055	0.097	0.044	0.888
CSE	0.236	0.098	-0.467	-0.723	-0.185	-0.381	0.072	0.058	-0.058	0.999
TASI	0.234	0.184	-0.735	0.103	0.347	0.471	0.098	0.001	0.058	0.994
KLSE	0.615	0.445	0.108	-0.12	-0.077	0.13	-0.25	0.32	0.381	0.935
Variance	5.9391	1.4299	1.0364	0.9766	0.9036	0.8339	0.6589	0.5481	0.4779	12.8044
% Var	0.424	0.102	0.074	0.07	0.065	0.06	0.047	0.039	0.034	0.915

Notes: Authors calculations using Minitab 17 software.

Next Factor loading VARIMAX technique is applied and the results of FA are quite different from PCS. FC1 has all US markets. While the remaining factors that is FC2 to FC9 contain only individual markets i.e, TWII, SSE, BSESN, JKSE, TASI, CSE, KSE-100 and KLSE respectively.

**Table 4b:-Result of FA using VARIMAX method**

Variable	FA1	FA 2	FA 3	FA 4	FA 5	FA 6	FA 7	FA 8	FA 9	Communality
S&P 500	0.944*	0.228	0.038	-0.125	-0.094	0.044	-0.039	-0.028	0.082	0.98
BSESN	0.289	0.294	0.03	-0.863*	-0.16	0.045	-0.082	-0.065	0.137	0.973

KSE-100	0.052	0.096	0.035	-0.047	-0.034	0.058	-0.009	-0.988*	0.034	0.998
Nikkei 225	0.459	0.603	0.113	-0.18	-0.272	0.157	-0.08	0.005	-0.137	0.743
HIS	0.404	0.553	0.29	-0.362	-0.155	0.027	-0.055	-0.036	0.226	0.765
SSE	0.063	0.128	0.974*	-0.032	-0.07	0.044	-0.016	-0.035	0.076	0.985
JKSE	0.154	0.245	0.088	-0.15	-0.892*	0.05	-0.042	-0.042	0.235	0.972
KS11	0.304	0.795*	0.056	-0.183	-0.16	0.019	-0.007	-0.06	0.134	0.808
TWII	0.216	0.829*	0.062	-0.067	-0.056	0.019	-0.048	-0.099	0.259	0.825
DJI	0.929*	0.198	0.043	-0.13	-0.093	0.034	-0.031	-0.048	0.069	0.939
IXIC)	0.881*	0.292	0.036	-0.107	-0.034	0.034	-0.06	-0.022	0.088	0.888
CSE	0.078	0.057	0.017	-0.061	-0.036	0.044	-0.99*	-0.009	0.053	0.999
TASI	0.07	0.058	0.043	-0.033	-0.043	0.987*	-0.043	-0.058	0.051	0.994
KLSE	0.152	0.309	0.106	-0.147	-0.253	0.07	-0.072	-0.044	0.841*	0.935
Variance	3.1918	2.4402	1.081	1.0426	1.0405	1.0202	1.0131	1.0075	0.9675	12.8044
% Var	0.228	0.174	0.077	0.074	0.074	0.073	0.072	0.072	0.069	0.915

Notes: \* indicates selected stock markets in FA. Authors calculations using Minitab 17 software.

### Conclusion:-

PCA is a method for multivariate data analysis and is used in many fields to extract relevant information from confusing data sets. Also, PCA provides means of identifying patterns in data and expressing the data in a way that highlights their similarities and differences among them. An advantage of PCA is its quality of quantifying the importance of each dimension in describing the variability of a data set. This method reduces the number of dimensions, without loss of information to great extent. Moreover, this paper presents the possible use of PCA in the stock volatility domain to reduce the dimension of data. In this case study, we used three US states and eleven Asian countries returns variables describing closing price data. First five PCS contains approximately 79.45% of the information provided by the original nine constructed PCS. Furthermore, PCA and FA (Unrotated Factor Loadings) methods yield same groups. Whereas, VARIMAX factor rotating method reveals a picture of the stock markets different from those of PCS and factor analysis. According to the available empirical analysis VARIMAX factor rotating method has been proved to be a good extracting method so far. In contrast to PCA, FA explains covariances and correlations between the variables and also unveils the hidden structure of the series. The results obtained from First FA shows US market only whereas second FA displays KS11 and TWII. However, remaining seven FA contains individual stock markets returns.

### Acknowledgements:-

The authors are thankful to Prof. Dr. Junaid Sageer Siddiqui for their valuable comments which improved the work considerably.

### References:-

1. Armeanu, Dan, & Lache, Leonard. (2008). Application of the model of principal components analysis on Romanian insurance market. *Theoretical and Applied Economics*, 6(6), 11.
2. Baker, Malcolm, & Wurgler, Jeffrey. (2006). Investor sentiment and the cross-section of stock returns. *The journal of finance*, 61(4), 1645-1680.
3. Baker, Scott R, Bloom, Nicholas, & Davis, Steven J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593-1636.
4. Bartholomew, David J, Knott, Martin, & Moustaki, Irini. (2011). *Latent variable models and factor analysis: A unified approach* (Vol. 904): John Wiley & Sons.
5. Billio, Monica, Getmansky, Mila, Lo, Andrew W, & Pelizzon, Lorian. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of financial economics*, 104(3), 535-559.
6. Chong, K, Yap, Ben, & Mohamad, Zulkifflee. (2013). A Study on the Application of Factor Analysis and the Distributional Properties of Financial Ratios of Malaysian Companies. *International Journal of Academic Research in Management (IJARM)* Vol, 2, 83-95.
7. Egloff, Daniel, Leippold, Markus, & Wu, Liuren. (2010). The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis*, 45(5), 1279-1310.



8. Fan, Jianqing, Liao, Yuan, & Mincheva, Martina. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(4), 603-680.
9. Fan, Jianqing, Wang, Dong, Wang, Kaizheng, & Zhu, Ziwei. (2017). Distributed estimation of principal eigenspaces. arXiv preprint arXiv:1702.06488.
10. Jolliffe, Flavia R, & Jolliffe, FR. (1986). *Survey design and analysis*: Ellis Horwood Chichester.
11. Kritzman, Mark, Li, Yuanzhen, Page, Sebastien, & Rigobon, Roberto. (2010). Principal components as a measure of systemic risk.
12. Pearson, Karl. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.
13. Pérignon, Christophe, Smith, Daniel R, & Villa, Christophe. (2007). Why common factors in international bond returns are not so common. *Journal of International Money and Finance*, 26(2), 284-304.
14. Pasini, G. (2017). Principal Component Analysis for Stock Portfolio Management. *International Journal of Pure and Applied Mathematics*, 115(1), 153-167.
15. Spearman, Charles. (1987). The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4), 441-471.
16. Sun, Qiang, Zhou, Wen-Xin, & Fan, Jianqing. (2018). Adaptive huber regression. *Journal of the American Statistical Association*(just-accepted), 1-35.
17. Zheng, Zeyu, Podobnik, Boris, Feng, Ling, & Li, Baowen. (2012). Changes in cross-correlations as an indicator for systemic risk. *Scientific reports*, 2, 888.