An Application of Factor Analysis on Gross Domestic Product Data of Bangladesh

Yesmin Akhter*, Md. Mahsin and Mohammad Zakaria Mohaimin*****

Abstract: *We analyze Bangladesh's gross domestic product (GDP) data using a factor analysis model to find out the contributing factors that affect GDP. The Factors are calculated in two different methods and the adequacy of the factor model is tested. We use Principal Component and Maximum Likelihood factor analysis approaches and apply them to GDP data for the year 1999-2000. Data has been collected by the Bangladesh Bureau of Statistics on 17 contributing sectors. The analysis has revealed that seventeen sectors have been classified into three factors that are contributing to Bangladesh's GDP. These three factors for principal component analysis are renamed as service factor, agriculture & infrastructure factor, and fishing & mining factor. In maximum likelihood method the factors are renamed as service factor, agriculture & infrastructure factor, and education factor. Lastly the factor scores as district-wise for the three factors are compared. Since the availability of gross domestic product data is very scarce, the data for the year 1999-2000 is used for the analysis.*

Introduction

Factor analysis is a statistical approach that can be used to analyze interrelationships among a large number of variables and to explain these variables in terms of their common underlying dimensions (factors). The approach involves finding a way of condensing the information contained in a number of original variables into a smaller set of dimensions (factors) with a minimum loss of information (Hair, J.F. et al., 1992). Factor analysis is applied as a data reduction or structure detection (structural simplification) method. In this study, we apply this to the Gross Domestic Product (GDP) (market value of final goods and services produced in an economy in a year) data of Bangladesh for the period 1999-2000. To calculate the GDP, data from 64 districts were collected on 17 production sectors by the Bangladesh Bureau of Statistics (BBS 1995 to 2000). It is reasonable to assume that many of these sectors will be related to each other and hence the variables can be grouped by their correlations. This means, all variables within a particular group are highly correlated among themselves but have relatively small correlations with variables in a different group. It is conceivable that each group of variables represents a single underlying construct, or factor, that is responsible for the observed correlations. In this study, we analyze the GDP data with an aim to find the underlying factors that are influencing Bangladesh's GDP.

* Yesmin Akhter, Lecturer, Institute of Statistical Research and Training (ISRT), University of Dhaka. Email: yakhter@isrt.ac.bd

** Md. Mahsin, Lecturer, Institute of Statistical Research and Training (ISRT), University of Dhaka. Email: Mahsin@isrt.ac.bd

*** Mohammad Zakaria Mohaimin, MS in applied Statistics, Institute of Statistical Research and Training (ISRT), University of Dhaka. Email: joti_ms@ist.ac.bd

Description of the method

A. The Model: Let the observable random vector $\mathbf{X}$, with *p* components, has mean μ and covariance matrix Σ. The factor model postulates that $\mathbf{X}$ is linearly dependent upon a few unobservable random variables $F_1, F_2, \cdots\cdots, F_m$, called *common factors*, and *p* additional sources of variation $\varepsilon_1, \varepsilon_2, \cdots\cdots, \varepsilon_p$, called *error* or, sometimes, *specific factors*. Then the factor analysis model is

$$\begin{aligned} X_1 - \mu_1 &= l_{11}F_1 + l_{12}F_2 + \cdots\cdots + l_{1m}F_m + \varepsilon_1 \\ X_2 - \mu_2 &= l_{21}F_1 + l_{22}F_2 + \cdots\cdots + l_{2m}F_m + \varepsilon_2 \\ &\vdots \\ X_p - \mu_p &= l_{p1}F_1 + l_{p2}F_2 + \cdots\cdots + l_{pm}F_m + \varepsilon_p \end{aligned}$$

In matrix notation,

$$\underset{(p\times 1)}{\mathrm{X} - \mu} = \underset{(p\times m)}{\mathrm{L}}\ \underset{(m\times 1)}{\mathrm{F}} + \underset{(p\times 1)}{\varepsilon} \tag{1}$$

Here, μ_i is the *mean* of i^{th} variable, F_j is the j^{th} *common factor*, $L = l_{ij}$, $i = 1,2,.., p,\ j = 1,2,.., m,$ is the matrix of factor loadings where l_{ij} is the loading of i^{th} variable on j^{th} common factor and ε_i is the i^{th} *specific factor.* The unobservable random factors **F** and the unique factors ε satisfy the following assumptions: $E(\mathbf{F}) = \underset{(m\times 1)}{\mathbf{0}}$, $Cov(\mathrm{F}) = \underset{(m\times m)}{\mathrm{I}}$, $E(\boldsymbol{\varepsilon}) = \underset{(p\times 1)}{\mathbf{0}}$ $Cov(\varepsilon) = \underset{(p\times p)}{\psi}$, Where $\boldsymbol{\psi}$ is diagonal matrix. The covariance matrix of $\mathbf{X}$ can be easily obtained as

$$\begin{aligned} \Sigma &= Cov(\mathrm{X}) = E(\mathrm{X} - \mu)(\mathrm{X} - \mu)' \\ &= \mathrm{LL}' + \psi \end{aligned}$$

Also it can be shown that $Cov(\mathrm{X}, \mathrm{F}) = \mathrm{L}$. (See Johnson, R.A. and D.W. Wichern, 1982 for more information.) Therefore,

$$\begin{aligned} Var(X_i) &= l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2 + \psi_i \\ \mathrm{Cov}(X_i, X_k) &= l_{i1}l_{k1} + \cdots + l_{im}l_{km} \\ \mathrm{Cov}(X_i, F_j) &= l_{ij} \end{aligned} \tag{2}$$

B. Communality: Communalities measure how strongly the variables are explained by the factors. A factor analysis model contains both common factors and specific factors. Therefore, the variability in the observation are also split up by common factors and specific factors. The proportion of variability of the i^{th} variable $\left(V\left(X_i\right)=\sigma_{ii}\right)$ contributed by the common factors is called the i^{th} communality. The proportion of variance due to the specific factors is often called the uniqueness or specific variance. Let us denote the i^{th} communality by h_i^2. Then from the covariance structure in (2), we can write $\underbrace{\sigma_{ii}}_{Var(X_i)} = \underbrace{l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2}_{communality} + \underbrace{\psi_i}_{specific\ variance} = h_i^2 + \psi_i$,

where $h_i^2 = l_{i1}^2 + l_{i2}^2 + \cdots + l_{im}^2,\ i = 1,2,...,p.$ i.e. the i^{th} communality is the sum of squares of loadings of the i^{th} variable on m common factors.

3. Estimation of parameters: In factor analysis, we try to describe the covariance relationships among many variables in terms of a few underlying common factors. Because the sample covariance matrix S is an unbiased estimator of the population covariance matrix Σ ,we usually do factor analysis on the sample covariance matrix S or the sample correlation matrix R . For a factor analysis to be meaningful, the variables have to be highly correlated. That means, the off-diagonal elements of the covariance matrix (or equivalently of the correlation matrix) have to be very different from zero. Thus, if Σ is significantly different from a diagonal matrix, factor analysis model is then entertained and the initial problem is to estimate the factor loadings l_{ij} and the specific variances $\psi_{i.}$

There are two methods of estimating the parameter of a factor analysis model. These are principal component method and the maximum likelihood method. In the present study, we consider both methods for estimating parameters.

A. Method of Principal Component: The principal component factor analysis of the sample covariance matrix **S** is specified in terms of its eigenvalue-eigenvector pairs $\left(\hat{\lambda}_1,\hat{e}_1\right),\left(\hat{\lambda}_2,\hat{e}_2\right)$,........, $\left(\hat{\lambda}_p,\hat{e}_p\right)$ where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p \geq 0$

Let m<p be the number of common factors.

The matrix of estimated loading $\left\{\tilde{l}_{ij}\right\}$ is given by

$$\tilde{\mathbf{L}} = \left[\sqrt{\hat{\lambda}_1}\,\hat{\mathbf{e}}_1 \quad \sqrt{\hat{\lambda}_2}\,\hat{\mathbf{e}}_2 \quad \cdots \quad \sqrt{\hat{\lambda}_m}\,\hat{\mathbf{e}}_m\right] \tag{3}$$

The estimated specific variances are provided by the diagonal elements of the matrix $\mathbf{S}-\tilde{\mathbf{L}}\tilde{\mathbf{L}}'$, so

$$\widetilde{\boldsymbol{\psi}} = \begin{bmatrix} \hat{\psi}_1 & 0 & \cdots & 0 \\ 0 & \hat{\psi}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\psi}_p \end{bmatrix} \text{ With } \widetilde{\psi}_i = s_{ii} - \sum_{j=1}^{m} \widetilde{l}_{ij}^{\,2} \tag{4}$$

Communalities are estimated as

$$\widetilde{h}_i^2 = \widetilde{l}_{i1}^2 + \widetilde{l}_{i2}^2 + \cdots + \widetilde{l}_{im}^2 \tag{5}$$

The principal component factor analysis of sample correlation matrix is obtained by starting with **R** (sample correlation matrix) in place of **S** (sample covariance matrix).

The proportion of total sample variance due to the j^{th} factor is

$$\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance due} \\ \text{to } j \text{ th factor} \end{pmatrix} = \begin{cases} \dfrac{\hat{\lambda}_j}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for factor analysis of } \mathbf{S} \\ \dfrac{\hat{\lambda}_j}{p} & \text{for factor analysis of } \mathbf{R} \end{cases} \tag{6}$$

B. Method of Maximum Likelihood

If the common factor **F** and the specific factors $\boldsymbol{\varepsilon}$ can be assumed normally distributed then maximum likelihood estimate of the factor loadings and specific variances may be obtained. When $\mathbf{F}_j$ and $\boldsymbol{\varepsilon}_j$ are jointly normal, the observations $\mathbf{X}_j - \boldsymbol{\mu} = \mathbf{L}\mathbf{F}_j + \boldsymbol{\varepsilon}_j$ are then normal, and the likelihood is

$$\begin{aligned} L(\mu, \Sigma) &= (2\pi)^{-\frac{np}{2}} |\Sigma|^{-\frac{n}{2}} e^{-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n} (\mathrm{x}_j - \bar{\mathrm{x}})(\mathrm{x}_j - \bar{\mathrm{x}})' + n(\bar{\mathrm{x}} - \mu)(\bar{\mathrm{x}} - \mu)'\right)\right]} \\ &= (2\pi)^{-\frac{(n-1)p}{2}} |\Sigma|^{-\frac{(n-1)}{2}} e^{-\left(\frac{1}{2}\right) tr\left[\Sigma^{-1}\left(\sum_{j=1}^{n} (\mathrm{x}_j - \bar{\mathrm{x}})(\mathrm{x}_j - \bar{\mathrm{x}})'\right)\right]} \\ &\quad \times (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\left(\frac{n}{2}\right)(\bar{\mathrm{x}} - \mu)' \Sigma^{-1} (\bar{\mathrm{x}} - \mu)} \end{aligned} \tag{7}$$

which depends on **L** and $\boldsymbol{\psi}$ through $\Sigma = \mathbf{L}\mathbf{L}' + \boldsymbol{\psi}$.

It is desirable to make **L** well defined by imposing the computationally convenient *uniqueness condition*

$$\mathbf{L}'\boldsymbol{\psi}^{-1}\mathbf{L} = \boldsymbol{\Delta}, \quad \text{a diagonal matrix} \tag{8}$$

The maximum likelihood estimates of the communalities are

$$\hat{h}_i^2 = \hat{l}_{i1}^2 + \hat{l}_{i2}^2 + \cdots + \hat{l}_{im}^2 \quad \text{for i = 1,2, …….,p.}$$

$$\begin{pmatrix} \text{Proportion of total} \\ \text{sample variance due} \\ \text{to } j \text{ th factor} \end{pmatrix} = \begin{cases} \dfrac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{s_{11} + s_{22} + \cdots + s_{pp}} & \text{for factor aanalysis of S} \\ \dfrac{\hat{l}_{1j}^2 + \hat{l}_{2j}^2 + \cdots + \hat{l}_{pj}^2}{p} & \text{for factor analysis of R} \end{cases} \tag{9}$$

C. A Large – Sample Test for the Number of Common Factors:

The assumption of a normal population leads directly to a test of model adequacy. Suppose the *m* common factor model holds. In this case $\Sigma = \mathbf{LL}' + \boldsymbol{\psi}$, and testing the adequacy of the *m* common factor model is equivalent to testing

$$H_0 : \underset{(p\times p)}{\Sigma} = \underset{(p\times m)}{\mathbf{L}} \underset{(m\times p)}{\mathbf{L}'} + \underset{(p\times p)}{\boldsymbol{\psi}}$$

vs. $H_1 : \Sigma$ any other positive definite matrix.

When Σ does not have any special form, the maximized likelihood function is proportional to

$$|\mathbf{S_n}|^{-n/2} e^{-np/2} \tag{10}$$

Under H_0 the maximum of likelihood function is proportional to

$$|\hat{\Sigma}|^{-\frac{n}{2}} \exp\left(-\left(\frac{1}{2}\right) \text{tr}\left[\hat{\Sigma}^{-1}\left(\sum_{j=1}^{n}(\mathrm{x}_j - \bar{\mathrm{x}})(\mathrm{x}_j - \bar{\mathrm{x}})'\right)\right]\right)$$

$$= |\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}|^{-n/2} \mathbf{exp}\left(-\frac{1}{2} n \,\text{tr}\left[(\hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}})^{-1} \mathbf{S}_n\right]\right) \tag{11}$$

Now we define the likelihood ratio statistic for testing H_0 as

$$-2 \ln \boldsymbol{\Lambda} = -2 \ln \left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$
$$= -2 \ln \left(\frac{|\hat{\Sigma}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\text{tr}(\hat{\Sigma}^{-1} - \mathbf{S}_n) - p\right] \tag{12}$$

with

$$\nu - \nu_0 = \frac{1}{2} p(p+1) - \left[p(m+1) - \frac{1}{2} m(m-1)\right] = \frac{1}{2}\left[(p-m)^2 - p - m\right] \quad \text{d.f.} \tag{13}$$

Since $\text{tr}(\hat{\Sigma}^{-1} - \mathbf{S}_n) - p = 0$ provided $\hat{\Sigma} = \hat{\mathbf{L}}\hat{\mathbf{L}}' + \hat{\boldsymbol{\psi}}$ is the maximum likelihood estimate of $\Sigma = \mathbf{LL}' + \boldsymbol{\psi}$. Thus we have

$$-2\ln\Lambda = n\ln\left(\frac{|\hat{\Sigma}|}{|S_n|}\right) \tag{14}$$

Bartlett (Bartlett, M.S.1954) has shown that the chi-square approximation to the sampling distribution of $-2\ln\mathbf{\Lambda}$ can be improved by replacing *n* in (14) with the multiplicative factor $(n-1-(2p+4m+5)/6)$.Using Bartlett's correction, we reject H_0 at the α level of significance if

$$(n-1-(2p+4m+5)/6)\ln\frac{|\hat{\mathbf{L}}\hat{\mathbf{L}}'+\hat{\boldsymbol{\psi}}|}{|\mathbf{S_n}|} > \chi^2_{[(p-m)^2-p-m]/2,(\alpha)} \tag{15}$$

provided *n* and $n-p$ are large. Since the number of degrees of freedom, $\frac{1}{2}\left[(p-m)^2-p-m\right]$, must be positive, it follows that

$$m<\frac{1}{2}\left(2p+1-\sqrt{8p+1}\right) \tag{16}$$

in order to apply the test (15).

D. Rotation of Factors: Using the estimation method described above, we estimate the initial loadings by an orthogonal transformation. An orthogonal transformation of the factor loadings, and the implied orthogonal transformation of the factors, is called factor rotation. The estimated loadings have the same ability to reproduce the covariance or correlation matrix. But the initial loadings may not be easily interpretable and hence it is a usual practice to rotate the factors until a simple structure is obtained. Ideally, we look for the set of loadings such that each variable loads highly on a single factor and has small-to-moderate loadings on the remaining factors.

If $\hat{L}$ is the $p\times m$ matrix of estimated factor loadings obtained by any method (principal component, maximum likelihood, and so forth) then $\hat{\mathrm{L}}^* = \hat{\mathrm{L}}\mathrm{T}$, where $\mathrm{TT}' = \mathrm{T}'\mathrm{T} = \mathrm{I}$

is a $p\times m$ matrix of "rotated" loadings. For this rotation, the estimated covariance or correlation matrix also remains unchanged, since

$$\hat{\mathrm{L}}\hat{\mathrm{L}}' + \hat{\psi} = \hat{\mathrm{L}}\mathrm{TT}'\hat{\mathrm{L}} + \hat{\psi} = \hat{\mathrm{L}}^*\hat{\mathrm{L}}^{*'} + \hat{\psi}$$

From the above equation we see that the communalities also remain unchanged. There are several types of rotations to choose form. The most popular one is the *varimax criterion* suggested by Kaiser (Kaiser, H.F. 1958). For this we define $\tilde{l}_{ij}^* = \hat{l}_{ij}^* / \hat{h}_i$ to be the final rotated coefficient scaled by the square root of the communalities. Then the varimax procedure selects the orthogonal transformation **T** that makes

$$V = \frac{1}{p}\sum_{j=1}^{m}\left[\sum_{i=1}^{p}\tilde{l}_{ij}^{*4} - \left(\sum_{i=1}^{p}\tilde{l}_{ij}^{*2}\right)\Big/p\right] \qquad (17)$$

as large as possible. By scaling the rotated coefficients $\hat{l}_{ij}^{*}$, variables with small communality gets more weight in the determination of simple structure. After the transformation **T** is determined, the loadings $\tilde{l}_{ij}^{*}$ are multiplied by $\hat{h}_i$ to keep the original communalities unchanged.

Analysis

There are 64 observations, one from each district, on 17 sectors (variables), which are contributing to Bangladesh's GDP. From a factor analysis perspective the first three eigen values, 9.525, 2.561, 0.977 of the correlation matrix suggest a factor model with $m = 3$ common factors. Because the first two eigen values are greater than unity and the third one is closer to unity and together they explain almost 77% of the total variance. Again a scree plot also shows that the first three eigen values load very highly on the observed variables. Therefore, we consider a factor model with $m = 3$ common factors. Also the adequacy of the three factors model is tested by applying the large sample test for the number of common factors. The estimated value of the test statistics using equation [15] is 78.71 and the 5% critical value is $\chi^2_{88,.05} = 115.84$. Since the critical value is not exceeded by the calculated value 78.471, hence we fail to reject H_0 .

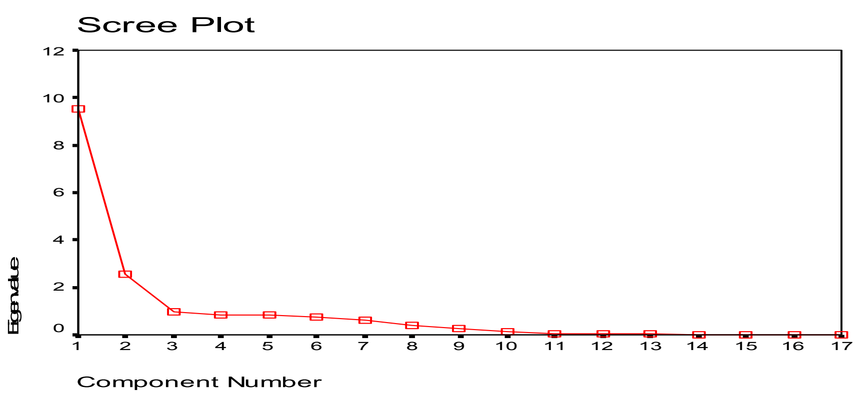

Figure 1: Scree plot

We conclude the data do not contradict a three-factor model. The principle component and maximum likelihood factor analysis approaches were then applied to the correlation matrix of the GDP data. The resulting estimated factor loadings, estimated rotated factor loadings and communalities are presented in the table for both methods.

Table 1: Results of principal component method with varimax rotation

Variables	Three - factor solution						
	Estimated factor loadings			Estimated rotated factor loadings			Communality
	F_1	F_2	F_3	F_1^*	F_2^*	F_3^*	$\hat{h}_i^2$
Crops	.033	**.478**	-.551	-.051	**.714**	-.142	.533
Forest	.195	**.872**	-.133	.026	**.786**	.445	.816
Animal	.064	**.545**	-.041	-.041	**.457**	.303	.303
Fishing	.239	**.647**	.311	.106	.344	**.666**	.572
Mining	.126	.430	**.669**	.031	-.061	**.803**	.649
Manufacturing	**.855**	-.318	.063	**.899**	-.164	-.031	.836
Electricity	**.943**	-.190	-.034	**.962**	.008	-.019	.926
Construction	**.515**	.259	-.313	.461	**.466**	-.024	.430
Wholesale	**.876**	-.114	.008	**.881**	.031	.051	.781
Hotel	**.963**	-.190	.056	**.981**	-.044	.055	.967
Transportation	**.976**	-.140	.015	**.984**	.021	.054	.973
Finance	**.963**	-.151	-.031	**.974**	.038	.010	.951
Real Estate	**.983**	-.009	-.020	**.967**	.145	.106	.967
Public Ad	**.940**	-.259	-.022	**.973**	-.053	-.052	.952
Education	**.707**	.218	.026	**.651**	.254	.243	.548
Health	**.972**	.166	-.006	**.922**	.271	.221	.973
Community	**.808**	.483	.012	**.700**	.482	.405	.886
Cumulative proportion of total sample variance explained	56.031	71.095	76.841	54.513	66.876	76.841	

From the above Table 1 we can see that, the proportion of total variance explained by the three-factor solution (76.841) is appreciably larger than that for the one factor solution (56.03). In varimax rotation three-factor solution expresses the same proportion of total variance as one factor solution. According to unrotated solution of principal factor analysis we can see that the last 12 variables of the gross domestic product loads highly on the first factor F_1 and might be called as *Service Factor* and first 4 variables loads highly on the second factor F_2 and might be called as *Agriculture Factor*. Only one variable "mining and quarrying" loads highly on the third factor F_3 and can be renaming as *Mining Factor*. After varimax rotation we have found that the "construction" variable which had high loading on the first factor (F_1) of unrotated solution have high loading on the second factor (F_2^*) of the rotated solution. Similarly the "fishing" variable which had high loading on the F_2 before rotation have high loading on F_3^* after rotation.

Table 2: Results of maximum likelihood method with varimax rotation

Variables	Three - factor solution						
	Estimated factor loadings			Estimated rotated factor loadings			Communality
	F_1	F_2	F_3	F_1^*	F_2^*	F_3^*	$\hat{h}_i^2$
Crops	.037	**.310**	.170	-.021	**.352**	-.049	.125
Forest	.200	**.685**	.493	.022	**.867**	.015	.745
Animal	.059	**.278**	.207	-.014	**.351**	.005	.122
Fishing	.237	**.448**	.360	.100	**.610**	.069	.383
Mining	.131	**.216**	.212	.049	**.319**	.066	.110
Manufacturing	**.872**	-.487	.027	**.797**	-.198	.571	.999
Electricity	**.926**	-.041	-.319	**.978**	-.025	.062	.939
Construction	**.480**	.225	.044	**.436**	.302	.040	.283
Wholesale	**.862**	-.039	-.087	**.835**	.094	.214	.751
Hotel	**.960**	-.138	-.151	**.947**	-.001	.255	.962
Transportation	**.969**	-.011	-.205	**.978**	.072	.141	.960
Finance	**.943**	-.028	-.286	**.982**	.008	.083	.972
Real Estate	**.980**	.095	-.111	**.956**	.212	.150	.982
Public Ad	**.921**	-.093	-.364	**.988**	-.094	.058	.988
Education	**.733**	-.197	.453	.519	.247	**.672**	.769
Health	**.975**	.221	.016	**.908**	.383	.168	.999
Community	**.809**	.451	.233	**.679**	.657	.142	.911
Cumulative proportion of total sample variance explained	55.339	64.222	70.999	51.493	65.083	70.999	

So according to the rotated solution of principal factor analysis we can rename the factors as *Service, Agriculture & Infrastructure, and Fishing & Mining Factor* respectively.

From the above Table 2 we can see that, the proportion of total variance explained by the three-factor solution (70.99%) is appreciably larger than that for the one factor solution (55.34%). In varimax rotation three-factor solution expresses more proportion of total variance than one factor solution. According to unrotated solution of maximum likelihood analysis we can see that the last 12 variables of the gross domestic product loads highly on the first factor F_1 and might be called as *Service Factor* and first 5 variables loads highly on the second factor F_2 and might be called as *Agriculture & Infrastructure factor*. And no variable has high loading on third factor. After varimax rotation we have found that the "education" variable which had high loading on the first factor (F_1) of unrotated solution have high loading on the third factor (F_3^*) of the rotated solution.

So according to the rotated solution of maximum likelihood method we can rename the factors as *Service, Agriculture & Infrastructure, and Education Factor* respectively.

Comparison of results obtained by two methods

According to the rotated solution of principal component method (PCM) we renamed three factors as *Service, Agriculture & Infrastructure,* and *Fishing & Mining Factor* where as in maximum likelihood method (MLM) we renamed three factors as *Service, Agriculture & Infrastructure,* and *Education Factor* respectively. The common variables of service factor both in MLM and PCM are same except education variable and construction variable. Education, which was in service factor in PCM, is in third factor (*Education Factor*) in MLM. Construction, which was in Agriculture & Infrastructure Factor in PCM, is in the Service Factor in MLM. In PCM third factor (*Fishing & Mining Factor*) is constructed by fishing variable and mining variable only whereas in MLM both these variables are included in the second factor (*Agriculture & Infrastructure Factor*). In MLM the third factor (*Education Factor*) consists of only a single variable “education” of gross district product.

Sorted factor scores by Districts

The estimated values of the unobserved random factors are called factor scores. The sorted factor scores by districts for the principle component approach using regression method are given in the following table:

Table 3: Sorted Factor Scores Table by Districts

Factor 1		Factor 2		Factor 3	
District	**Factor Score**	**District**	**Factor Score**	**District**	**Factor Score**
Dhaka	7.17612	Mymensingh	2.8142	Brahamanbaria	2.88663
Chittagong	2.49538	Narail	2.65717	Sylhet	2.47533
Khulna	0.52531	Bogra	2.39893	Habiganj	2.23617
Narayanganj	0.49076	Comilla	2.09287	Chittagong	2.2207
Bandarban	0.46775	Faridpur	2.01663	Cox's Bazar	2.22058
Comilla	0.36773	Tangail	1.65246	Bogra	1.58574
Gazipur	0.32167	Naogaon	1.5098	Mymensingh	1.47769
Mymensingh	0.25718	Jessore	1.1575	Dinajpur	1.38043
Rajshahi	0.23588	Chittagong	1.02814	Khulna	1.12377
Natore	0.1626	Rangpur	0.97971	Comilla	1.08484
Tangail	0.1194	Dinajpur	0.97467	Chandpur	0.79669
Rangpur	0.09109	Kishoreganj	0.67029	Bhola	0.68717
Faridpur	0.06538	Chandpur	0.6674	Netrakona	0.66669
Pabna	0.04615	Gaibandha	0.63937	Bagerhat	0.59465
Sirajganj	0.03536	Noakhali	0.45517	Patuakhali	0.58555
Jessore	0.02575	Jamalpur	0.39354	Barisal	0.43122
Barisal	0.0237	Sirajganj	0.33135	Noakhali	0.38458

Narsingdi	-0.03693	Kurigram	0.30428	Lakshmipur	0.34513
Sylhet	-0.06606	Sunamganj	0.30117	Kishoreganj	0.14973
Noakhali	-0.06943	Barisal	0.29681	Pirojpur	0.03695
Kishoreganj	-0.08844	Netrakona	0.22071	Barguna	-0.02618
Dinajpur	-0.09739	Jenaidah	0.21759	Rajshahi	-0.0777
Jamalpur	-0.09867	Pabna	0.19445	Rangamati	-0.18032
Kushtia	-0.13818	Rajshahi	0.11858	Naogaon	-0.19237
Naogaon	-0.17417	Sathkira	0.10526	Pabna	-0.20066
Gaibandha	-0.18	Khulna	0.08088	Jessore	-0.22712
Bogra	-0.19088	Thakurgaon	0.05781	Bandarban	-0.24326
Kurigram	-0.19965	Nilphamari	0.02831	Narsingdi	-0.2801
Brahamanbaria	-0.20539	Kushtia	-0.00911	Sunamganj	-0.2871
Sathkira	-0.21703	Bhola	-0.13278	Feni	-0.29041
Nilphamari	-0.22979	Natore	-0.13674	Gazipur	-0.30827
Bagerhat	-0.24018	Moulvibazar	-0.14277	Sathkira	-0.31612
Moulvibazar	-0.24253	Narsingdi	-0.24363	Kurigram	-0.31835
Sunamganj	-0.24384	Bagerhat	-0.3011	Gaibandha	-0.32437
Nawabganj	-0.24706	Sherpur	-0.36407	Tangail	-0.35983
Munshiganj	-0.25397	Nawabganj	-0.3727	Thakurgaon	-0.37529
Chandpur	-0.25626	Manikganj	-0.37682	Sirajganj	-0.42194
Jenaidah	-0.25838	Lakshmipur	-0.38033	Madaripur	-0.4261
Lakshmipur	-0.27109	Gazipur	-0.40632	Narayanganj	-0.43665
Manikganj	-0.27324	Patuakhali	-0.48154	Gopalganj	-0.44832
Netrakona	-0.27936	Lalmonirhat	-0.48569	Natore	-0.45574
Bhola	-0.28435	Joypurhat	-0.52823	Jhalokati	-0.45732
Feni	-0.28516	Pirojpur	-0.58184	Jamalpur	-0.46792
Thakurgaon	-0.29291	Barguna	-0.62627	Munshiganj	-0.47908
Rangamati	-0.29711	Chuadanga	-0.62892	Moulvibazar	-0.48242
Sherpur	-0.31127	Gopalganj	-0.63965	Nawabganj	-0.48472
Chuadanga	-0.31721	Madaripur	-0.64243	Rangpur	-0.48824
Cox's Bazar	-0.32338	Panchagarh	-0.66674	Rajbari	-0.4906
Madaripur	-0.32392	Sylhet	-0.6737	Lalmonirhat	-0.50125
Gopalganj	-0.32499	Magura	-0.67539	Sherpur	-0.56088
Pirojpur	-0.32952	Feni	-0.70032	Jenaidah	-0.57145
Habiganj	-0.35424	Shariatpur	-0.73415	Panchagarh	-0.57349
Lalmonirhat	-0.35594	Munshiganj	-0.75623	Shariatpur	-0.58526
Rajbari	-0.36415	Rajbari	-0.78737	Manikganj	-0.5879
Jhalokati	-0.37892	Dhaka	-0.9146	Kushtia	-0.58866
Joypurhat	-0.37993	Cox's Bazar	-0.92138	Nilphamari	-0.60151
Shariatpur	-0.38492	Narayanganj	-0.96977	Joypurhat	-0.63406
Patuakhali	-0.38929	Meherpur	-1.01905	Magura	-0.64515
Magura	-0.40067	Jhalokati	-1.04143	Khagrachhari	-0.69028
Panchagarh	-0.40152	Brahamanbaria	-1.06787	Chuadanga	-0.71068
Meherpur	-0.41885	Habiganj	-1.19565	Meherpur	-0.74324
Khagrachhari	-0.42479	Khagrachhari	-1.23365	Dhaka	-1.16945
Barguna	-0.42957	Rangamati	-1.6537	Faridpur	-1.62186
Narail	-0.57667	Bandarban	-1.87309	Narail	-3.03866

The above Table 3 represents the 64 districts according to their scores on each factor in descending order. Each district has individual scores on each factor. We can see that the district Dhaka has highest score on the first factor where as the districts Mymenshing and Brahmanbaria have highest scores on second factor and third factor respectively.

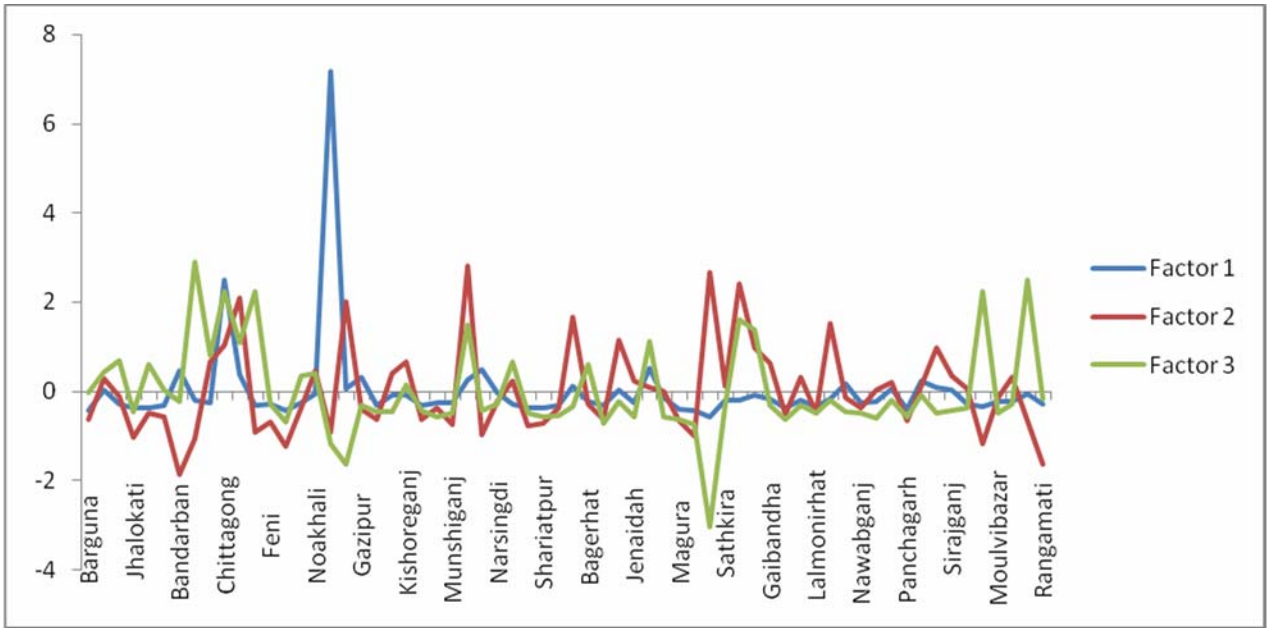

Figure 2: Factor scores of 64 districts

The above graph is plotted by districts against their scores on each factor. Scores of districts on each factor can be compared from this figure such as district Dhaka have highest score on factor 1, but have low score on 2nd factor and negative score on 3rd factor.

Conclusion

From the above analysis it is evident that three major factors are influencing Bangladesh's GDP. For the principal component analysis the factors are Service *Factor*, *Agriculture & Infrastructure Factor*, and *Fishing & Mining Factor*. The first factor consist of manufacturing sector, electric, gas and water supply, wholesale & retail trade, hotel & restaurants, transport, storage & communication, financial intermediation, real estate renting & business service, public administration & defense sector, education sector, health and social work sector, community, social & personal service sector. Second factor consists of crops & horticulture, forestry & related services, animal farming, construction and the third factor consists of fishing sector, mining & quarrying. In maximum likelihood method the three factors are *Service Factor*, *Agriculture & Infrastructure Factor, and Education Factor*. The first factor consists of manufacturing sector, electric, gas and water supply, construction, wholesale & retail trade, hotel & restaurants, transport, storage & communication, financial intermediation, real estate renting & business service, public administration & defense sector. Second factor consists of crops & horticulture, forestry & related services, animal farming, fishing sector, mining & quarrying; and the third factor consists of education sector. Both MLM and PCM give the almost equal but not the same result. The common variables of service factor both in MLM and PCM are the same except education variable and construction variable. Education, which was in service factor in PCM, is in third factor (*Education Factor*) in MLM. Construction, which was in agriculture & infrastructure factor in PCM, is in the service factor in MLM. In PCM third factor (*Fishing & Mining Factor*) is constructed by fishing variable and mining variable only whereas in MLM both these variables are included in the second factor (*Agriculture & Infrastructure Factor*). In MLM third factor (*Education Factor*) consists of only a single variable "education" of gross district product. According to the test of hypothesis we found that three-factor model is adequate for the data. Finally we calculate the scores of districts on each factor and found that the district Dhaka have the highest score on the first factor whereas the districts Mymenshing and Brahmanbaria have the highest scores on second factor and third factor respectively.

References

Bangladesh Bureau of Statistics (BBS). 1995 -2000. *Provisional Estimates of Gross Domestic Products* 1995-1996 to 1999-2000.

Bartlett, M.S.1954. *A Note on Multiplying Factors for Various Chi-Squared Approximation,* Journal of the Royal Statistical Society (B), 16. 296-298.

Johnson, R.A. 1982 and Wichern, D.W. *Applied Multivariate Statistical Analysis,* Prentice-Hall, New Jersey.

Hair, J.F. et al. 1992: *Multivariate data analysis*, 3rd ed. Macmillan, New York.

Kaiser, H.F. 1958. *The Varimax Criterion for Analytic Rotation in Factor Analysis*, Psychometrika, **23** . 187-200.