

## Modified Factor Analysis to Construct Composite Indices: Illustration on Urbanization Index

M.A.C.S.S. Fernando, S. Samita<sup>1</sup> and R. Abeynayake<sup>2</sup>

Postgraduate Institute of Agriculture  
University of Peradeniya  
Sri Lanka

**ABSTRACT.** *This study introduces a modified factor analysis approach to develop a composite index. The methodology is illustrated using an index representing the magnitude of urbanization of a Divisional Secretariat. The new method defines a specific weight to each individual indicator variable and the index assigns a specific numerical value to the level of urbanization of an area. Densities of population, students, houses and common residences, non-resident buildings, business establishments and vehicles were the indicator variables considered in the index building process. Cronbach's alpha was used to verify the internal consistency of these indicator variables. Initially, the grouping patterns in the data have been identified through a Preliminary Factor Analysis. This resulted in a single factor explaining a substantial amount of total variability. The weight corresponding to a particular indicator variable was defined as a function of the correlation coefficient between the indicator variable and the first Principle Component. Then the scaled variables were weighted and used in the final Factor Analysis. A single factor explaining 94% of total variance was selected as the composite index. First, the Index of Urbanization was defined as a linear function of the composite index. Then it was converted into a function of original indicator variables to make it easy to update. The methodology would be applicable to any country for deriving a similar Index of Urbanization. Logical classification of local government authorities, to assist the government in various policy making and development activities is also possible with the new Index.*

**Keywords:** *Composite Index, indicator variables, magnitude of urbanization, Urbanization Index, weighted factor analysis*

### INTRODUCTION

Weighted Factor Analysis (WFA), a modified multivariate technique has extensively been used in composite indicator building procedures for combining sets of sub-indicators. Many of these

studies have applied Principal Component Analysis (PCA) to define weights while Factor Analysis (FA) to analyze the structure of indicator variables. These techniques group together sub-indicators that are collinear to form a composite indicator capable of capturing as much of common information of those sub-indicators as possible (Nardo *et al.*, 2005). Apart from that, the weight deriving methodologies introduced in many composite index

---

<sup>1</sup> Department of Crop Science, Faculty of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka

<sup>2</sup> Department of Agribusiness & Plantation Management, Wayamba University of Sri Lanka, Sri Lanka

\* Author for correspondence: sampathf@sab.ac.lk

building procedures have the weakness of assigning common weight for all the indicator variables belonged to a single subset. Therefore, a new method of defining to the Principle Component (PC) based WFA was introduced in this study because of following reasons: (i) to account for the highest possible variation in the indicators set using the smallest possible number of factors and (ii) to overcome the weakness of defining a common weight for all the indicator variables belonged to a single subset.

The new method defines weights which are specific for each indicator variable of the WFA. Considering importance of the existence of an Urbanization Index (UI), which quantifies the degree of urbanization of an area, the WFA technique was applied to derive a composite index for urbanization.

Urban-rural classification constitutes an important framework for the collection and compilation of population data, in many countries. For local governance, the urban-rural statistical classification is highly significant, mainly for their development and policy making purposes (Bhagat, 2005). Although, many definitions of urban-rural classification have been derived in different research, none of them was based on all types of contexts that have to be considered in such a phenomenon. The magnitude of agglomeration of people and establishments is one of the prominent definitions for urbanization used in many countries. The urban-rural indices that have been constructed so far in those countries are basically based on population density, the predominance of non-agricultural activities and provision of social amenities. In some of the studies, industrialization and infrastructure indicator variables have also been included when building urban indices. However, a clear meaningful mathematical approach has not been developed to evaluate the degree of urbanization of a particular area.

The FA has been applied in developing composite indicators mainly to analyze the structure of the sets of sub-indicators (e.g. Indicator of Relative Intensity of Regional Problems in the Community, General Indicator of Science & Technology). However, in order to come up a meaningful clustering, practical knowledge on the interrelationship among the indicator variables is rather useful. The PCA has been used in composite indicator building processes basically to identify dimensionality of the phenomenon (e.g. Environmental Sustainability Index), to cluster the indicators (e.g. General Indicator of Science & Technology and Index of Success of Software process Improvement) and define the weights (e.g. Internal Market Index, Business Climate Indicator, Environmental Sustainability Index, Human Development Index (De Silva *et al.*, 2000) and General Indicator of Science & Technology. In some indices the PCA has not been successful as expected and thus simpler weighting techniques have been suggested (e.g. Internal Market Index and Environmental Sustainability Index).

In PC and FA, weighting only intervenes to correct for the overlapping information of two or more correlated indicators, and it is not a measure of importance of the associated indicator. Also the information must be comparable for this approach to be used (sub-indicators must have the same unit of measurement). Using equal weighting by combining variables with high degree of correlation, the problem of double counting is possible (*i.e.* if two collinear indicators are included in the composite index with different weights than the unique dimension that the two indicators measure would have a total of both weights in the composite). In many composite indicators, all variables are given the same weight when there are no statistical or empirical grounds for choosing a different scheme (Nardo *et al.*, 2005).

The required number of PCs to be selected to describe a phenomenon is not fixed and based only on explaining sufficient amount of variation in the data. For example in the Internal Market Index (Tarantola *et al.*, 2002), PCA has identified eight main PCs which describe 90 % of the variance and this result has confirmed the expectation of the researchers, that the Internal Market is a multidimensional phenomenon. Conversely, in the Business Climate Indicator, PCA has indicated that a single factor explains a sufficient amount of variance (92%). This has given a statistical justification to the authors for the choice of summarizing a priori the information by means of a single composite indicator. In this case, the phenomenon that the composite indicator aims to measure has one statistical dimension.

The related studies on urbanization indices revealed that the population density is a key determinant of urbanization. Some studies have also considered other infrastructure and industrialization indicators as well. For instance, Liu *et al* (2003) considered some specific infrastructure, industrialization and other families of indicator variables which might have direct relationships with urbanization. Although, some have suggested the importance of using the built-up area rather than the administrative area for calculating average urban density (Angel *et al.*, 2005; Kawamura *et al.*, 1997), it is only possible when reliable data is available. The right indicator variables have been included in many of these research but they have not been combined considering their right commitment to the level of urbanization.

This study constructs a Composite Index of Urbanization using six indicator variables namely, population density, student population density, density of houses and common residences, density of non-residence buildings, density of business establishments and vehicles. A new weight defining technique for the PC-based WFA is introduced. The index quantifies the level of urbanization of any Divisional Secretariat for given values of above six indicator variables. Some of indicator variables, which may affect the level of urbanization, have not been included in this study, since they are not recorded according to Divisional Secretariats. Examples for such variables are, availability of health services, agricultural land use, distribution of education facilities, distribution of tap water facility, road density etc.

## MATERIALS AND METHODS

Literature on existing measures of urbanization and availability of data were considered when selecting the indicator variables for this study. Data were gathered from district offices of the Department of Census and Statistics, local government authorities, Divisional Secretariat offices and provincial education offices. Projected data to 2006 based on the Population and Housing Census survey conducted in 2001 have been included in these handbooks. All the variables except area were obtained as counts.

All variables were converted into respective densities by dividing the counts by total area of the Divisional Secretariat. This minimizes the effect of the area on the counts. Divisional Secretariat data were subjected to few limitations regarding data availability and recording inconsistencies. Data were not available for eight districts: Ampara, Batticaloa, Jaffna, Killinochchi, Mannar, Mullativu, Trincomalee, and Vavunia. Therefore, the study covered only 247 Divisional Secretariats. The indicator variables used were, population density (PD), student population density (SPD), density of houses & common residences (H&CRD), density of non-residence buildings (NRBD) in the density of business establishments (BED), and vehicle density (VD). All the densities were obtained per square kilometer.

Initially, a preliminary FA was performed on the original indicator variables in order to identify the grouping patterns in the indicator variables. After identifying the subsets of indicator variables, a PCA was carried out for each subset. Since the first PC accounts for the largest amount of variability in the data only the first PC was retained. Then, the correlation coefficient between each indicator variable belonged to the particular subset and the first PC were calculated. The following formula was used to determine the weights.

$$w_{ij} = \frac{r_{ij}^2}{\sum_{j=1}^{n_i} r_{ij}^2}$$

$w_{ij}$  = weight correspond to  $j^{\text{th}}$  variable in  $i^{\text{th}}$  subset  
 $r_{ij}$  = correlation coefficient between the first PC of  $i^{\text{th}}$  subset and  $j^{\text{th}}$  variable  
 $i = 1, 2, \dots, m$  (number of subsets)  
 $j = 1, 2, \dots, n_i$  (number of variables in  $i^{\text{th}}$  subset)

Since the formula includes the correlation coefficient between the first PC and each of the variables, the weight will be particular for that variable. This overcomes the weakness of assigning a common weight to all the variables in a single subset. In addition, the weight includes the squared coefficient of correlation which actually represents the coefficient of determination between the variable and the first PC. Therefore, it actually represents the amount of variability, which each of the variable represents out of the total variability in the PC. The two points explained above formulates the rationale of the selected methodology.

After deriving weights, the original variables were divided by their own standard deviation to make the variables with unit variance (scaled). Then, the scaled variables were weighted according to the following formula and the transformed variable was denoted by  $X_i^*$ .

$$X_i^* = \left( \frac{X_i}{\sigma_i} \right) w_i$$

$w_i$  = weight corresponds to  $i^{\text{th}}$  variable

$\sigma_i$  = standard deviation of  $i^{\text{th}}$  variable

$X_i$  = original  $i^{\text{th}}$  variable

There is a limitation in this weight-deriving methodology. It can be statistically proved that, if a subset consists only one indicator variable, the weight corresponding to the variable become one and when a subset consists only two indicator variables the weights of both variables become half. However, since more than two variables are often considered, above limitation will not be serious. Finally, in order to identify the underlying factors determining urbanization a PCs-based FA with covariance matrix option was performed on the transformed variables. Several hypothetical situations described below, would illustrate the improvement of the new weight deriving methodology, under different correlation structures of the indicator variables.

Let the variable matrix as  $\mathbf{X} = (X_i)_{p \times 1}$  and the Correlation matrix of  $\mathbf{X}$  as  $\mathbf{P} = (\rho_{ij})_{p \times p}$ .

**Case I**

When all the correlation coefficients are close to  $\pm 1$ , all the variables may load into one factor (subset). In this kind of a situation, even the Preliminary FA (PFA) yields a highly improved result. The correlation coefficients between the variables and the first PC of the set are close to  $\pm 1$  and almost equal to each other. Therefore, the weights of all the variables are approximately equal and the result is not notably improved with a WFA. In addition, the weight of each variable would be approximately equal to the reciprocal of the number of variables ( $1/p$ ) and the Eigen value of the first factor (or PC) is equal to the number of variables ( $p$ ). The following example illustrates this.

Sample correlation matrix	Correlation coefficients with PC1	Weights
$\begin{pmatrix} 1.000 & 0.918 & 0.996 & 0.953 & 0.916 & 0.916 \\ 0.918 & 1.000 & 0.898 & 0.944 & 0.908 & 0.909 \\ 0.996 & 0.898 & 1.000 & 0.928 & 0.884 & 0.885 \\ 0.953 & 0.944 & 0.928 & 1.000 & 0.986 & 0.986 \\ 0.916 & 0.908 & 0.884 & 0.986 & 1.000 & 0.999 \\ 0.916 & 0.909 & 0.885 & 0.986 & 0.999 & 1.000 \end{pmatrix}$	$\begin{pmatrix} -0.976 \\ -0.955 \\ -0.958 \\ -0.994 \\ -0.976 \\ -0.976 \end{pmatrix}$	$\begin{pmatrix} 0.168 \\ 0.161 \\ 0.162 \\ 0.174 \\ 0.168 \\ 0.168 \end{pmatrix}$

Percentage variance explained in Preliminary Factor Analysis = 94.6%

Percentage variance explained in Weighted Factor Analysis = 94.7%

**Case II**

If all the correlation coefficients are not close to  $\pm 1$  but high or at least one or two variables are not highly correlated with others, then still all the variables may load in to a single factor (subset). In this situation, the result of the PFA is not as improved as the above. More variables are highly correlated with the first Principal Component (PC1) but few (one or two) variables are moderately correlated and hence, the weights of these few variables may differ from others. Consequently, WFA yields an improved result compared to the PFA. The following example illustrates this.

Sample correlation matrix	Correlation coefficients with PC1	Weights
$\begin{pmatrix} 1.000 & 0.708 & 0.876 & 0.822 & 0.783 & 0.788 \\ 0.708 & 1.000 & 0.615 & 0.944 & 0.881 & 0.909 \\ 0.876 & 0.615 & 1.000 & 0.602 & 0.519 & 0.528 \\ 0.822 & 0.944 & 0.602 & 1.000 & 0.961 & 0.986 \\ 0.783 & 0.881 & 0.519 & 0.961 & 1.000 & 0.973 \\ 0.788 & 0.909 & 0.528 & 0.986 & 0.973 & 1.000 \end{pmatrix}$	$\begin{pmatrix} -0.902 \\ -0.931 \\ -0.738 \\ -0.980 \\ -0.946 \\ -0.959 \end{pmatrix}$	$\begin{pmatrix} 0.163 \\ 0.173 \\ 0.109 \\ 0.192 \\ 0.179 \\ 0.184 \end{pmatrix}$

Percentage variance explained in Preliminary Factor Analysis = 83.3%

Percentage variance explained in Weighted Factor Analysis = 87.9%

**Case III**

With a moderately strong correlation structure of the original variables, one factor (subset) situation can still be expected. In this case, a moderate result can be obtained from the PFA. Some variables are moderately correlated with the first Principal Component even in this

situation. Therefore, the weights of few variables may substantially differ from others. Hence, the WFA yields an improved result compared to the Preliminary Factor Analysis. The following example is an illustration for this.

Sample correlation matrix						Correlation coefficients with PC1			Weights	
1.000	0.575	0.979	0.571	0.471	0.478	-0.739		0.116		
0.575	1.000	0.615	0.944	0.881	0.909	-0.941		0.188		
0.979	0.615	1.000	0.602	0.519	0.528	-0.773		0.127		
0.571	0.944	0.602	1.000	0.961	0.986	-0.969		0.200		
0.471	0.881	0.519	0.961	1.000	0.973	-0.924		0.182		
0.478	0.909	0.528	0.986	0.973	1.000	-0.938		0.187		

Percentage variance explained in Preliminary Factor Analysis = 78.4%

Percentage variance explained in Weighted Factor Analysis = 85.7%

When there is more than one factor derived in WFA, they were combined according to the following formula.

$$\text{Urbanization Index} = \frac{\sum_{i=1}^k f_i v_i}{\sum_{i=1}^k v_i}; \quad v_i = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i} \quad \text{and} \quad \sum_{i=1}^k \left( \frac{v_i}{\sum_{i=1}^k v_i} \right) = 1$$

$k$  = the number of factors selected

$f_i$  = factor score of the  $i^{\text{th}}$  factor

$v_i$  = the proportion of variance explained by  $i^{\text{th}}$  factor

$\lambda_i$  = eigen value of the  $i^{\text{th}}$  factor

## RESULTS AND DISCUSSION

Although, literature evidenced use of a single indicator as a composite index in most of the situations, a composite index is a combination of two or more sub-indicators. Therefore, construction of the urbanization index was attempted using both methods. The best method for deriving the index was selected through a thorough analysis on the weights derived under each method.

The PCs based FA with the covariance matrix option was carried out on the transformed indicator variables, in order to identify the underlying sub indicators (factors) of the composite index. The indicator variables were scaled by dividing by the standard deviation. Consequently, each indicator variable was multiplied by a constant (reciprocal of the standard deviation of the original variable), then, the standard deviations of scaled indicator variables become one. This makes, the correlation and covariance matrices of scaled indicator variables equivalent. When a set of variables were multiplied by constants, their correlation structure remains unchanged, and correlation matrices of original and scaled indicator variables become equal. Because of this similarity, the results of PC-based FA on scaled indicator variables (FA without weights) with covariance matrix and the results of PC-based FA on original indicator variables with correlation matrix option become equivalent.

Coefficients of correlations between each pair of indicator variables were above 0.85 and this implied high positive correlation among all variables. PD was identified as the key determinant of level of urbanization, from literature. The high positive correlation structure of remaining five indicator variables suggests their positive association with urbanization. First, PCs based FA on original indicator variables (with correlation matrix option) was performed in order to identify the grouping pattern in the original indicator variables and it was called “PFA”. As justified above, a comparison between FA and WFA is similar to a comparison between PFA and WFA.

**Derivation of the composite index**

First, the results of both PC based FA with correlation matrix option and PCs based WFA with covariance matrix option were compared to identify the effect of weighting. In the PFA, only the first factor which had high negative loadings on all the indicator variables and explained 93.8 % variation was selected.

WFA was based on the results of PFA. Then a PCA was performed on original indicator variables with correlation matrix option to identify the weight of indicator variables. Then the correlation coefficient of each indicator variable with the first PC was used to calculate the weights. After that, the scaled indicator variables were multiplied by the corresponding weights. The weights and standard deviations of original indicator variables are given in Table 1. According to this, all the weights are approximately equal because, the indicator variables are strongly correlated with the first PC.

**Table 1. Weights of indicator variables**

<b>Indicator variable</b>	<b>Correlation coefficient</b>	<b>Weight</b>	<b>Standard deviation</b>
PD	-0.979	0.170321	1884.0
SPD	-0.955	0.162072	337.8
H&CRD	-0.952	0.161056	12458.0
NRBD	-0.992	0.174874	2773.0
BED	-0.958	0.163092	1809.0
VD	-0.974	0.168585	15957.0

Finally, the scaled weighted indicator variables were used in PCs based FA (WFA) with covariance matrix option in order to identify the factor(s) to be used in construction of the composite index. Since all the indicator variables have loaded in to the first factor with high negative loadings which is also meaningful in practical sense, only that factor was selected. Given that all the indicator variables were loaded in to the first factor in both analyses, the results of WFA and PFA became comparable. However, the total percentage variance explained by the first factor has only increased from 93.8% to 94.0%. In fact, in this particular case a significant improvement in percentage of explained variance cannot be expected mainly due to two reasons. Firstly, the result of the PFA was an already improved result, which extracted a single factor explaining 93.8% of total variation. Secondly, the weights derived using almost equal correlation coefficients are obvious to be approximately equal.

**Reliability of the variables**

The value of Cronbach’s alpha of original indicator variables of the study is 0.706 (>0.70), which implies that consistency of the indicator variables is at a satisfactory level. The value of Cronbach’s alpha for scaled indicator variables was equal to 0.987 which implies an enhancement of internal consistency of the original indicator variables by scaling. In the PFA, when all the indicator variables loaded in to first factor, the scaled variables weighted within the factor. The value of Cronbach’s alpha for scaled and weighted indicator variables remains unchanged at 0.987. It implies that the internal consistency of the indicator variables is improved after scaling or scaling and weighting.

Construction of composite index with a particular set of variables has made certain technical problems to the values of weights when there were subsets with one and two indicator variables. Therefore, the weighted factor analysis without Varimax rotation was selected for obtaining the final composite index.

**Deriving the urbanization index from the composite index**

Table 2 provides the details on the parameters, the weights and the factor score coefficients used to formulate the final composite index of urbanization.

**Table 2. Notations of the parameters of the composite index**

Original indicator variables			Scaled and weighted indicator variables			
Notation	Standard deviation	Weight	Notation	Mean	Standard deviation	Factor score coefficient
PD	1883.82	0.170321	PD*	0.0823	0.1703	-0.181
SPD	337.75	0.162072	SPD*	0.0776	0.1621	-0.160
H&CRD	386.93	0.161056	H&CRD*	0.0877	0.1611	-0.157
NRBD	80.23	0.174874	NRBD*	0.0688	0.1749	-0.194
BED	52.86	0.163092	BED*	0.0421	0.1631	-0.163
VD	449.61	0.168585	VD*	0.0452	0.1686	-0.177

The scores of Composite Index (CI) can be represented as a function of transformed (scaled and weighted) indicator variables as follows.

$$CI = (-1) \times \left[ \begin{aligned} &0.181 \left( \frac{PD^* - 0.0823}{0.1703} \right) + 0.160 \left( \frac{SPD^* - 0.0776}{0.1621} \right) + 0.157 \left( \frac{H \& CRD^* - 0.0877}{0.1611} \right) \\ &+ 0.194 \left( \frac{NRBD^* - 0.0688}{0.1749} \right) + 0.163 \left( \frac{BED^* - 0.0421}{0.1631} \right) + 0.177 \left( \frac{VD^* - 0.0452}{0.1686} \right) \end{aligned} \right]$$

The above equation can be further simplified as follows.

$$CI = (-1) \times \left( \begin{aligned} &1.06283 \times PD^* + 0.98705 \times SPD^* + 0.97455 \times H \& CRD^* \\ &+ 1.10921 \times NRBD^* + 0.99939 \times BED^* + 1.04982 \times RV^* \end{aligned} \right) + 0.41537$$



The general formula used in transformation of variables is:

$$\text{Transformed Indicator Variable} = \left( \frac{\text{Original Indicator Variable}}{\text{Standard Deviation of the Indicator Variable}} \right) \times \text{Weight}$$

Therefore, the scores of CI can be represented as following function of original indicator variables.

$$\text{CI} = (-1) \times \left( \begin{array}{l} 0.096093 \times \text{PD} + 0.473641 \times \text{SPD} + 0.405647 \times \text{H \& CRD} \\ + 2.417689 \times \text{NRBD} + 3.083466 \times \text{BED} + 0.393639 \times \text{RV} \end{array} \right) \times 10^{-3} + 0.41537$$

The minimum and maximum values of the of CI score, were -11.1713 to 0.4044 respectively. Colombo Municipal Council area score was the minimum. This implies that the score and the degree of urbanization are negatively related. Therefore, the scores were transformed by multiplying it by minus one (-1) in order to assign the largest value to the most urbanized Divisional Secretariat and the lowest to the least urbanized. [Note that this multiplication is in addition to the -1, which is already in the CI equation]. Then the transformed scores were varied from -0.4044 to 11.1713. In order to obtain a non negative urbanization index, a constant value (0.5) was added to the transformed scores.

The above multiplication by -1 and addition of 0.5 are particular to this study and depend upon numeric values, which were originally obtained for the composite index. Finally, the relationship between the composite index and the Urbanization Index (UI) is in the linear form of the following formula.

$$\text{Urbanization Index} = (-1) \times \text{Composite Index} + 0.5$$

The score of UI can also be represented as follows as a function of original indicator variables with positive coefficients. This formulation will help to update the index according to the changes of the determinants.

$$\text{UI} = \left( \begin{array}{l} 0.096093 \times \text{PD} + 0.473641 \times \text{SPD} + 0.405647 \times \text{H \& CRD} \\ + 2.417689 \times \text{NRBD} + 3.083466 \times \text{BED} + 0.393639 \times \text{RV} \end{array} \right) \times 10^{-3} + 0.91537$$

This uncomplicated linear transformation makes it simple to obtain statistical properties of one index, using statistical properties of the other index. Based on the scores of composite index, the scores of UI were calculated for each Divisional Secretariat. The Urbanization Index, which quantifies the level of urbanization of a Divisional Secretariat, can be effectively used in defining local government authorities: municipal councils, urban councils and *Pradeshiya Sabhas* more logically. It will provide a useful guideline to relevant bodies regarding such classification and also assist in policy making and development activities of the government. Table 3 gives the top 12 Divisional Secretariats according to the level of urbanization obtained under the new method as a comparison with the population density criterion.

**Table 3. Urbanization Index versus population density**

District	Divisional Secretariat	Urbanization index score		Population density (Pop. per sq.km)	
		Magnitude	Rank	Magnitude	Rank
Colombo	Colombo MC	11.6713	1	17889.0	1
Colombo	Dehiwala - Mount Lavinia	6.1628	2	12630.6	2
Colombo	Moratuwa	4.5997	3	10637.9	3
Colombo	Jayawardanapura Kotte	4.5065	4	8227.4	5
Colombo	Rathmalana	3.6888	5	9184.0	4
Galle	Galle Four Gravets	3.1644	6	5797.2	8
Gampaha	Kelaniya	2.6938	7	6317.6	7
Colombo	Kolonnawa	2.6233	8	6389.2	6
Colombo	Maharagama	2.4549	9	5127.6	9
Gampaha	Negombo	2.2972	10	4823.4	10
Colombo	Kesbawa	1.8107	11	4147.6	11
Kandy	Kandy Four Gravets & Gangawata Korale	1.2868	12	2243.1	17

### CONCLUSIONS

PC-based WFA method introduced in this study defines separate weight for each indicator variable in WFA. The explained amount of variability was not improved significantly as a result of weighting. The illustrations with varying correlation structures implied that, this little improvement is caused by the strong correlation structure of the variables and the result is significantly improved when the correlation structure among the variables is not much strong.

The composite indicator developed was based on six indicator variables representing human and building densities of an area. This indicator was used to construct an index representing the magnitude of urbanization of a Divisional Secretariat. Finally, the UI was defined as a linear combination of the composite index and hence as a function of the sub-indicators. This caused to increase the ability to update the index according to the changes of the indicators included. The new index provides a useful guideline for more logical classification of Local Government Authorities and hence assists relevant bodies in policy making and development activities of the government.

Although the degree of internal consistency of the variables included in the WFA was satisfactory according to the Cronbach's Alpha, there might be some other indicator variables which will improve the index as a measure of urbanization. Agricultural land use, road density, distribution of telecommunication facilities, availability of health facilities, *etc*, are some of such indicator variables to be included. Since this study has developed the methodology, it will be interesting if another study use this methodology, to develop an UI including right indicator variables selected through ideas of experts. However, in order to overcome the barriers against gathering data on above indicator variables such study should be facilitated by the government because an extra effort might be needed to obtain data with reference to Divisional Secretariats.

## REFERENCES

- Angel S., Sheppard S.C. and Civco D.L. (2005) *The Dynamics of Global Urban Expansion, Report-Transport and Urban Development Department*, The World Bank Washington D.C.
- Bhagat R.B. (2005). Rural–Urban Classification and Municipal Governance in India. *Singapore J. of Tropi. Geogra.* 26(1), 61-73.
- De Silva A.P.G.S., Thattil R.O., Samitha S. (2000) Weighted Principal Component Analysis Approach to Construct an Index for Human Development, *Sri Lanka J. of Appl. Statis.* 1(1), 13-24.
- Kawamura M., Jayamana S., and Tsujiko Y. (1997). Comparison of Urbanization and Environmental Condition in Asian Cities using Satellite Remote Sensing Data, Poster Session 1, 18th Asian Conference on Remote Sensing.
- Liu G.G., Wu X., Peng C. and Fu A.Z. (2003) Urbanization and Health Care in Rural China, *Contemporary Econ.Poli.* 21(1), 11-24.
- Nardo M., Saisana M., Saltelli A. and Tarantola S. (2005). *Tools for Composite Indicators Building*, Joint Research Centre - European Commission-Directorate General.
- Tarantola S., Saisana M. and Saltelli A. (2002). *Internal Market Index 2002: Technical details of the methodology*