On Construction of Robust Composite Indices by Linear Aggregation

SK Mishra
Department of Economics
North-Eastern Hill University, Shillong (India)
mishrasknehu@yahoo.com

Abstract

In this paper we construct thirteen different types of composite indices by linear combination of indicator variables (with and without outliers/data corruption). Weights of different indicator variables are obtained by maximization of the sum of squared (and, alternatively, absolute) correlation coefficients of the composite indices with the constituent indicator variables. The Differential Evolution method has been used for global optimization. Seven different types of correlation are used: Karl Pearson, Spearman, Signum, Bradley, Shevlyakov, Campbell and modified Campbell. Composite indices have also been constructed by maximization of the minimal correlation. We find that performance of indices based on robust measures of correlation such as modified Campbell and Spearman, as well as that of the maxi-min based method, is excellent. Using these methods we obtain composite indices that are autochthonously sensitive and allochthonously robust. This paper also justifies simple mean-based composite indices, often used in construction of human development index.

Keywords: Composite index, Linear aggregation, Principal components, Differential Evolution, Global optimization, Robust correlation, Signum, Bradley, Absolute correlation, Shevlyakov, Campbell, Hampel, Outliers, Mutilation of data

JEL Classifications: C13, C43, C61, C87, C88

On Construction of Robust Composite Indices by Linear Aggregation

SK Mishra
Department of Economics
North-Eastern Hill University, Shillong (India)
mishrasknehu@yahoo.com

I. Introduction: A composite index ($I_k : k = 1, n$) is often a (weighted) linear aggregation of numerous indices ($x_{kj} : j = 1, m; k = 1, n$) such that $I_k = \sum_{j=1}^{m} w_j x_{kj}$. As to the assignment of weights to different indices, there are two approaches: the first in which the weights are determined on the basis of some information or considerations exogenous to the data on index variables ($x_{kj} \in X$), and the second in which the weights are endogenously determined such that $w = f(X)$. The most robust composite index is the one that uses exogenously determined weights since in that case w is used as a parameter and, therefore, $I = \varphi(X \mid w)$. However, when weights are endogenously determined, we have $I = \varphi(X, f(X))$. In this latter case, the composite index, I, depends not only on the index variables, X, but also on the specification of the function, $f(.)$, that obtains weights, w, from X.

To make this point clearer, let x_{kj} be perturbed such that $x_{kj} \Leftarrow x_{kj} + \partial_{kj}$ where $\partial_{kj} \neq 0$. If weights are not derived from X, then $I_k \Leftarrow I_k + w_j \partial_{kj}$ and $I_{i \neq k} \Leftarrow I_{i \neq k}$. That is, the perturbation affects I_k only. However, if weights are derived from X, a perturbation of one of the values of x_{kj} would in most cases alter the values of w as well as the values of $I_k \forall k = 1, n$. A perturbation of x_{kj} will pervade throughout even though all of $x_{ip} : i \neq k \wedge p \neq j$ have remained unchanged. The extent of pervasiveness, which is not a desirable property of the composite index, would depend on the specification of $w = f(X)$.

There is an additional point to be noted. When $I = \varphi(X \mid w) = Xw$, the weight, w_j, which may be viewed as $\partial I / \partial x_j$ is constant and hence I is indeed a linear combination of X. However, when $I = \varphi(X, f(X))$, the weight, w_j, in general , is not constant and, therefore, I is not a linear combination of X. In that case, these weights, which may also be viewed as the rate of substitution among different constituent indices, lose interpretability in any simple manner and hence go far off the desirable property of easy comprehensibility.

II. Two Desirable Propertied of a Composite Index: Now we enunciate two desirable properties of a composite index: (i) change in x_{kj} best be reflected into a change in I_k, which we call sensitiveness, and (ii) change in x_{kj} be least reflected into changes in $I_i : i \neq k$, which we will call robustness. Sensitiveness implies stronger correlation between the composite index, I, and the constituent index variables, $x_j \in X$. On the other hand, robustness implies insensitiveness of w to changes in X.

III. The Simplest Method of Construction of a Composite Index: Perhaps the simplest method of constructing a composite index is to obtain $I_k = (1/m)\sum_{j=1}^{m} x_{kj} : k = 1, n.$ It implies $w_j = 1/m : j = 1, m.$ Viewed as such, this method yields a robust composite index. It also follows the law of insufficient reason; that in absence of any indubitable basis of determining the weights assigned to different index variables, they all carry equal weights. In the last few years, after it was used for construction of the 'human development index', this method has won many adherents. In applying this method, on many occasions, the index variables, $x_j s,$ are standardized or normalized in some manner such that $x_j \Leftarrow g(x_j)/norm(g(x_j)),$ where $g(x_j)$ is a monotonic function of x_j. The $norm(g(x_j))$ may be $\max_k(g(x_{kj})),\ \hat{\sigma}(g(x_j),$ $\underset{k}{med} \mid x_{kj} - med(x_j) \mid,$ etc. The choice of a suitable norm is important. Certain types of norm may run against the desirable property of robustness. On the other hand, sensitiveness of the composite index constructed by this method is rather suboptimal.

IV. The Method of the Principal Components Analysis: The well known Principal Components Analysis (PCA) is another method to obtain the composite index. It attributes two properties to $I = Xw$: first, that it maximizes the sum of squared (Karl Pearson's or product moment) coefficients of correlation between I and $x_j; j = 1, m,$ and the second, that it is orthogonal to any other index, $J = Xv : v \neq w,$ that may be derived from X by maximization of the sum of squared correlation coefficients between J and x_j. Stated differently, the PCA-based composite index satisfies two criteria: i) $I = Xw : \sum_{j=1}^{m} r^2(I, x_j)$ is maximum, and ii) if $I = Xw : \max \sum_{j=1}^{m} r^2(I, x_j)$ and $J = Xv : \max \sum_{j=1}^{m} r^2(J, x_j); w \neq v,$ then the coefficient of correlation between the two such composite indices, $r(I, J) = 0.$ Except in an extremely special case when the constituent index variables themselves are pair-wise orthogonal, I and J both cannot attain the global maximum. The global maximizer composite index is unique.

The technique to obtain such a (unique) global maximizer composite index by PCA consists of, first, obtaining the matrix, $R,$ of correlation coefficients, $r_{ij} \in R,$ between each pair of index variables, x_i and x_j, and, then, obtaining the eigenvalues ($\lambda_1, \lambda_2, ..., \lambda_m$) and the eigenvectors ($u_1, u_2, ..., u_m$) of $R.$ The eigenvectors are then normalized to satisfy the condition $\| u_j \| = 1 \ \forall \ j$ (or, sometimes, $\| u_j \| = \lambda_j \ \forall \ j$), where $\|\bullet\|$ denotes the Euclidean norm. These normalized eigenvectors are used as weights to construct the composite indices. The index constructed by using the eigenvector associated with the largest eigenvalue is often used as the first best composite index. This index attains the global maximum mentioned earlier.

The composite index thus obtained has many optimal properties. However, this PCA based index is often elitist (Mishra, 2007-b), with a strong tendency to weight highly correlated subset of X favourably and relegating poorly correlated index variables to the subsequent principal components. In practice, when one has to use only one composite index to represent X, the poorly correlated index variables remain largely unrepresented. Since correlation is no measure of importance, many highly important but poorly correlated index variables may thus be undermined by the PCA-based composite index.

The said elitist property of the PCA based index may possibly be ameliorated by application of multi-criteria analysis. It has been suggested (Mishra, 1984) that multiple PCA-based composite indices ($I_j : j = 1, m$) obtained by using different eigenvectors of R (of X) can be subjected to multi-criteria decision-making/concordance analysis (Hill and Tzamir, 1972; van Delft and Nijkamp, 1976) for establishing outranking relationship among the objects (A_k) represented by $x_k = (x_{k1}, x_{k2}, ..., x_{km)} : k = 1, n.$ Each composite index, I_j, will take on a weight according to its explanatory power measured by the eigenvalue, λ_j (of R), associated with it. Since PCA-based composite indices are much fewer than the number of index variables in X, it is expected that this approach will be sharper than the approach that applies multi-criteria decision-making tools on X itself (Munda and Nardo, 2005-a and 2005-b). It may be noted, however, that the earlier approach derives endogenous weights from X itself, while the latter approach needs exogenous weights.

Another possible approach to abate the elitist tendency of the composite indices is to derive them not by maximization of the sum of squared correlation coefficients between the composite index and the constituent index variables as the PCA does, but by maximization of the sum of absolute (product moment) correlation coefficients between them (Mishra, 2007-a). That is: the composite index, $I = Xw$ maximizes $\sum_{j=1}^{m} \left| r(I, x_j) \right|$. This sort of index is said to be inclusive in nature since it does assign suitable weights to poorly correlated indicator variables. Yet another possible method to obtain a composite index, $I = Xw$, consists of maximization of the minimal absolute or squared (product moment) correlation coefficient: $I = Xw : \max[(\min_j \left| r(I, x_j) \right|]$. This approach assigns the most egalitarian weights to all index variables and hence favours the poorly correlated indicator variables most (Mishra, 2007b).

V. Replacement of Pearson's Correlation Coefficient by Robust Correlation Coefficient: Arithmetic mean, standard deviation and product moment correlation coefficient are the members of the same family, based on minimization of the Euclidean norm. All of them are very much sensitive to perturbation, errors of observation or presence of outliers in the dataset. If the weights, w, in $I = Xw$ are obtained by maximization of the product moment correlation (whether $\max \sum_{j=1}^{m} r^2(I, x_j)$, $\max \sum_{j=1}^{m} | r(I, x_j) |$ or $\max[\min_j(| r(I, x_j) |)]$, errors of observation, effects of perturbation or presence of outliers on weights would surely be substantial and pervasive. Therefore, there is a need to replace product moment correlation coefficient by some more robust measure of correlation.

Since the formula of computing the product moment correlation is fundamental to development of many other measures of correlation, we present it here. The product moment coefficient of correlation is defined as:

$$r(x_1, x_2) = \text{cov}(x_1, x_2) / \sqrt{\text{var}(x_1) \cdot \text{var}(x_2)} \quad \ldots \quad (1)$$

where, $\bar{x}_a = \frac{1}{n} \sum_{i=1}^{n} x_{ia}$; $\text{cov}(x_1, x_2) = \frac{1}{n} \sum_{i=1}^{n} x_{i1} x_{i2} - \bar{x}_1^2 \bar{x}_2^2$ and $\text{var}(x_a) = \text{cov}(x_a, x_a)$. The quarter square identity (Gnanadesikan and Ketttenring, 1972) gives us:

$$\sum_{i=1}^{n} x_{i1} x_{i2} = \frac{1}{4} \left[\sum_{i=1}^{n} (x_{i1} + x_{i2})^2 - \sum_{i=1}^{n} (x_{i1} - x_{i2})^2 \right]$$

$$= \frac{1}{4}\left[\sum_{i=1}^{n} x_{i1}^2 + \sum_{i=1}^{n} x_{i2}^2 + 2\sum_{i=1}^{n} x_{i1}x_{i2} - \sum_{i=1}^{n} x_{i1}^2 - \sum_{i=1}^{n} x_{i2}^2 + 2\sum_{i=1}^{n} x_{i1}x_{i2}\right] = \frac{1}{4}\left[4\sum_{i=1}^{n} x_{i1}x_{i2}\right].$$

Exploiting this identity we may write

$$r(x_1, x_2) = (1/4)\left[\text{var}(x_1 + x_2) - \text{var}(x_1 - x_2)\right] \Big/ \sqrt{\text{var}(x_1) \cdot \text{var}(x_2)} \quad \ldots \quad (2).$$

This formula (2) is of a great relevance for development of some other formulas of correlation.

There is one more identity that may be interesting. This identity is given as:

$$\sum_{i=1}^{n}\sum_{j=1}^{n}(x_{i1} - x_{j1})(x_{i2} - x_{j2}) = \sum_{i=1}^{n}\sum_{j=1}^{n}(x_{i1}x_{i2} - x_{i1}x_{j2} - x_{j1}x_{i2} + x_{j1}x_{j2})$$

$$= \sum_{i=1}^{n}\left[nx_{i1}x_{i2} - x_{i1}\sum_{j=1}^{n} x_{j2} - x_{i2}\sum_{j=1}^{n} x_{j1} + \sum_{j=1}^{n} x_{j1}x_{j2} \right]$$

$$= n\sum_{i=1}^{n} x_{i1}x_{i2} - \sum_{i=1}^{n} x_{i1}\sum_{j=1}^{n} x_{j2} - \sum_{i=1}^{n} x_{i2}\sum_{j=1}^{n} x_{j1} + n\sum_{j=1}^{n} x_{j1}x_{j2} \quad \ldots \quad (3)$$

Now, since $\sum_{i=1}^{n} x_{i1}x_{i2} \equiv \sum_{j=1}^{n} x_{j1}x_{j2}$ and $\sum_{i=1}^{n} x_{ia} \equiv \sum_{j=1}^{n} x_{ja}$ for $a = 1,2,$ we rewrite (3) as

$$2\left[n\sum_{i=1}^{n} x_{i1}x_{i2} - \sum_{i=1}^{n} x_{i1}\sum_{i=1}^{n} x_{i2}\right] = 2n^2\left[\frac{1}{n}\sum_{i=1}^{n} x_{i1}x_{i2} - \bar{x}_1\bar{x}_2\right] = 2n^2\,\text{cov}(x_1, x_2) \quad \ldots \quad (4)$$

Further simplified, $\sum_{i=1}^{n}\sum_{j=i}^{n}(x_{i1} - x_{j1})(x_{i2} - x_{j2}) = n^2\,\text{cov}(x_1, x_2)$. However, for $i = j$ the terms take on zero value and, thus, $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_{i1} - x_{j1})(x_{i2} - x_{j2}) = n^2\,\text{cov}(x_1, x_2)$. This invariance of sum over $i \le j$ and $i < j$ has important bearings when n is not vary large.

VI. Robust Measures of Correlation: Statisticians have proposed a number of formulas, other than the one that obtains Pearson's coefficient of correlation, that are considered to be less affected by errors of observation, perturbation or presence of outliers in the data. Some of them transform the variables, say x_1 and x_2, into $z_1 = \phi_1(x_1)$ and $z_2 = \phi_2(x_2)$, where $\phi_a(x_a)$ is a linear (or nonlinear) monotonic (order-preserving) rule of transformation or mapping of x_a to z_a. Then, $r(z_1, z_2)$ is obtained by the appropriate formula and it is considered as a robust measure of $r(x_1, x_2)$. Some others use different measures of central tendency, dispersion and co-variation, such as median for mean, mean deviation for standard deviation and so on. In what follows, we present a few formulas of obtaining different types of correlation efficient.

VI.1. Spearman's Rank Correlation Coefficient: If x_1 and x_2 are two variables, both in n observations, and $z_1 = \Re(x_1)$ and $z_2 = \Re(x_2)$ are their rank numerals, then the Pearson's formula applied on (z_1, z_2) obtains the Spearman's correlation coefficient (Spearman, 1904). There is a simpler (but less general) formula that obtains rank correlation coefficient, given as:

$$\rho(x_1, x_2) = r(z_1, z_2) = 1 - 6\sum_{i=1}^{n}(z_{i1} - z_{i2})^2 / [n(n^2 - 1)] \quad \ldots \quad (5)$$

VI.2. Signum Correlation Coefficient: Let c_1 and c_2 be the measures of central tendency or location (such as arithmetic mean or median) of x_1 and x_2 respectively. We transform them to $z_{ia} = (x_{ia} - c_a)/|x_{ia} - c_a|$ if $|x_{ia} - c_a| > 0$, $else\ z_{ia} = 1$. Then, $r(z_1, z_2)$ is the signum correlation coefficient (Blomqvist, 1950; Shevlyakov, 1997). Due to the special nature of transformation, we have

$$r(z_1, z_2) \cong \mathrm{cov}(z_1, z_2) = (1/n)\sum_{i=1}^{n} z_{i1} z_{i2} \quad \ldots \quad (6)$$

In this study we will use median as a measure of central tendency to obtain signum correlation coefficients.

VI.3. Bradley's Absolute Correlation Coefficient: Bradley (1985) showed that if $(u_i, v_i);\ i = 1, n$ are n pairs of values such that the variables u and v have the same median = 0 and the same mean deviation (from median) or $(1/n)\sum_{i=1}^{n}|u_i| = (1/n)\sum_{i=1}^{n}|v_i| = d \neq 0$, both of which conditions may be met by any pair of variables when suitably transformed, then the absolute correlation may be defined as

$$\rho(u, v) = \sum_{i=1}^{n}\left(|u_i + v_i| - |u_i - v_i|\right) \Big/ \sum_{i=1}^{n}\left(|u_i| + |v_i|\right). \quad \ldots \quad (7)$$

VI.4. Shevlyakov Correlation Coefficient: Hampel et al. (1986) defined the median of absolute deviations (from median) as a measure of scale, $s_H(x_a) = \underset{i}{median}\,|\, x_{ia} - \underset{i}{median}(x_{ia})\,|$ which is a very robust measure of deviation, and using this measure, Shevlyakov (1997) defined median correlation,

$$r_{med} = \left[med^2\,|\,u\,| - med^2\,|\,v\,|\right] \Big/ \left[med^2\,|\,u\,| + med^2\,|\,v\,|\right] \quad \ldots \quad (8)$$

where u and v are given as $u_i = (x_{i1} - med(x_1))/s_H(x_1) + (x_{i2} - med(x_2))/s_H(x_2)$ and $v_i = (x_{i1} - med(x_1))/s_H(x_1) - (x_{i2} - med(x_2))/s_H(x_2)$.

VI.5. Campbell's Correlation Matrix: Unlike the coefficient of correlation defined by the formulations above that consider correlation between any pair of variables at a time (and thus presuming that other variables do not exist, while indeed they do exist), Campbell (1980) obtained the entire matrix of robust correlation coefficients simultaneously, discounting for the effects of outliers. The main idea behind Campbell's correlation is to obtain $V = Z'\Omega^{-1}Z$ instead of $Z'IZ$ where $\Omega^{-1} \neq I$, but an inverted Aitken-Mahalanobis distance matrix (Aitken, 1935; Mahalanobis, 1936) defined in a specific manner.

Campbell's method is an iterative method that obtains the $m-$element vector of weighted (arithmetic) mean, $\bar{x}$, and weighted variance-covariance matrix, $V(m, m)$, in the following manner. Initially, all weights, $w_i; i = 1, n$ are considered to be equal, $1/n$, and the sum of weights, $\sum_{i=1}^{n} w_i = 1$. Further, we define $d_0 = \sqrt{m} + b_1/\sqrt{2};\ b_1 = 2,\ b_2 = 1.25.$
Then we obtain

$$\begin{aligned} \bar{x} &= \sum_{i=1}^{n} w_i x_i \Big/ \sum_{i=1}^{n} w_i \\ V &= \sum_{i=1}^{n} w_i^2 (x_i - \bar{x})'(x_i - \bar{x}) \Big/ \left[\sum_{i=1}^{n} w_i^2 - 1\right] \quad \ldots \quad (9) \\ d_i &= \left\{(x_i - \bar{x})V^{-1}(x_i - \bar{x})'\right\}^{1/2}; i = 1, n \end{aligned}$$

$$w_i = \omega(d_i)/d_i; i = 1, n : \omega(d_i) = d_i \ \mathit{if}\ d_i \leq d_0 \ \mathit{else}\ \omega(d_i) = d_0 \exp[-(1/2)(d_i - d_0)^2 / b_2^2]$$

It may be noted that execution of the last operation redefines $w_i; i=1,n$ which may be significantly different from the $w_i; i=1,n$ in the first operation. If this process is repeated, $w_i; i=1,n$ stabilizes and so stabilize $\bar{x}$, V, and $d_i; i=1,n.$ We will call it Campbell (type-I) procedure. A few points are worth noting. If V is ill-conditioned for ordinary inversion, a generalized (Moore-Penrose) inverse of V or V^+ may be used for V^{-1} and if $d_i = 0$ or $d_i \approx 0$ then $w_i = 1.$ From V one may obtain R, the correlation matrix, since $r_{ij} = v_{ij} / \sqrt{v_{ii} v_{jj}}$.

It may also be noted that there can be other approaches to specify $\omega(d_i)$. Any scheme that assigns lower weight to larger magnitude of d_i will make V a robust measure of covariance. Assigning weights such as $w_i = 1$ for $d_i - s_H(d) \le d_i < d_i + s_H(d),$ $w_i = (1/2)^2$ for $d_i - 2s_H(d) \le d_i < d_i - s_H(d)$ and $d_i + 2s_H(d) \ge d_i > d_i + s_H(d)$ and so on may also be very effective in robustization of correlation matrix. Although Campbell (1980) has not suggested this procedure to assign weights, we will call it Campbell (type-II) procedure since in all other respects it is similar to his method of obtaining the robust correlation matrix.

VII. Computational Considerations: The procedure of construction of the PCA-based composite index defined as $I = Xw : \max \sum_{j=1}^{m} r^2(I, x_j),$ where r is the coefficient of product moment correlation, has a closed form formula (Kendall and Stuart, 1968). Additionally, there are many software applications available for this purpose. However, when the coefficient of correlation is defined differently (e.g. Blomqvist's signum, Bradley's absolute, Shevlyakov's robust or Campbell's multivariate correlation) or the maximand function is other than $\sum_{j=1}^{m} r^2(I, x_j)$, e.g. $\sum_{j=1}^{m} \left| r(I, x_j) \right|$ or $\max[\min_j(| r(I, x_j) |)]$, there is no closed form formula for obtaining the index, $I = Xw.$ Nor are the software application available for this purpose. Therefore, the optimand function has to be optimized directly to obtain the weight vector, w.

We accomplish this by using the Differential Evolution (DE) method of global optimization. The DE is one of the most recently invented methods of global optimization that has been very successful in optimization of extremely difficult multimodal functions. The DE is a population-based stochastic search method of optimization grown out of the Genetic algorithms. The crucial idea behind DE is a scheme for generating trial parameter vectors. Initially, a population of points (p in d-dimensional space) is generated and evaluated (i.e. f(p) is obtained) for their fitness. Then for each point (p_i) three different points (p_a, p_b and p_c) are randomly chosen from the population. A new point (p_z) is constructed from those three points by adding the weighted difference between two points ($w(p_b$-$p_c)$) to the third point (p_a). Then this new point (p_z) is subjected to a crossover with the current point (p_i) with a probability of crossover (c_r), yielding a candidate point, say p_u. This point, p_u, is evaluated and if found better than p_i then it replaces p_i else p_i remains. Thus we obtain a new vector in which all points are either better than or as good as the current points. This new vector is used for the next iteration. This process makes the differential evaluation scheme completely self-organizing. Operationally, this method consists of three basic steps: (i) generation of (large enough) population with N individuals [x = (x_1, x_2, …, x_m)] in the m-dimensional space, randomly distributed over the entire domain of the function in question and evaluation of the individuals of the so generated by finding f(x); (ii) replacement of this current population by a better fit new population, and (iii) repetition of this replacement until satisfactory results are obtained or certain criteria of termination are met. The crux of the problem lays in replacement of the current population by a new population that is better fit. In this context, the meaning of 'better' is in the Pareto improvement sense. A set S_a is better than another set S_b *iff :* (i) no $x_i \in S_a$ is inferior to the corresponding member of $x_i \in S_b$; *and* (ii) at least one member $x_k \in S_a$ is better than the corresponding member $x_k \in S_b$. Thus, every new population is an improvement over the earlier one. To accomplish this, the DE procedure generates a candidate individual to replace each current individual in the population. The candidate individual is obtained by a crossover of the current individual and three other randomly selected individuals from the current population. The crossover itself is probabilistic in nature. Further, if the candidate individual is better fit than the current individual, it takes the place of the current individual, else the current individual stays and passes into the next iteration (Mishra, 2006).

VIII. A Simulation Experiment and Robustness of Correlation Matrices: Now we propose to compute different measures of correlation coefficient listed above and to compare their performance as to robustness in presence of outliers and mutilating perturbations in the data (indicator variables, X). This exercise is based on simulated data. We generate a single variable, $x_1 : x_{i1}; i = 1, n$ ($n = 30$) randomly and scale the values such that each x_{i1} lies between 10 and 50 with equal probability. With x_1 we generate $x_{ij}; i = 1, n; j = 1, m$ ($m = 6$) such that $x_{ij} = x_{i1} + e_{ij}$, where e_{ij} are random and uniformly distributed between (-10, 10). As a result of this mutilation the correlation between any two variables, $x_j, x_k \in X$ would be appreciably large. These six variables are then used to construct composite index, $I = Xw.$ The generated variables (X) and the correlation matrix (R) obtained from them by using different formulas (Pearson, Spearman, Signum, Bradley, Shevlyakov and Campbell) are presented in Table-1 and Table-2.1 through Table-2.3.

It may be noted that unless we add e_{ij} to x_{ij} , the coefficient of correlation $r(x_i, x_j)$ between any two variables $x_i, x_j \in X$ is unity. Once errors are introduced, correlation decreases. The range and magnitude of e_{ij} determines the reduction in the magnitude of correlation. We have chosen (-10, 10) as the range of e_{ij} so as to keep high correlation among the variables, and all x_{ij} to lie between zero and sixty. With this X we compute thirteen composite indices as detailed out in section VIII. Then we mutilate or introduce an outlier into X and compute thirteen composite indices as spelt out in section VIII and compare them. For mutilation, we add 1000 to x_{11} (the first observation on x_1) which shifts the median of x_1 from 30.46484 to 31.02664 and mean of x_1 from 29.89639 to 63.229724. Now, x_{11} is clearly an outlier observation. Effect of this outlier permeates through all correlation coefficients, presented in Table 3.1 through 3.3.

A perusal of Tables 3.1 through 3.3 reveals that Karl Pearson's, Signum, Bradley's and Campbell's (type-I) correlation matrices have been evidently poor at containing the effects of mutilation. A number of correlation coefficients have changed significantly in magnitude, sign or both. However, Shevlyakov's correlation matrix has been affected only slightly. Campbell (type-II) correlation matrix has been most robust (table-3.4).

IX. Construction of Composite Indices: As mentioned above, from X we construct thirteen indices by using the following procedures:

(i) By averaging over variables: $I_{0i} = (1/m)\sum_{j=1}^{m} x_{ij}$

(ii) By maximizing $\sum_{j=1}^{m} | r(I_1, x_j) | : I_1 = Xw_1$, where $r(I_1, x_j)$ is Pearson's correlation between I_1 and x_j

(iii) By maximizing $\sum_{j=1}^{m} r^2(I_2, x_j) | : I_2 = Xw_2$, where $r(I_2, x_j)$ is Pearson's correlation between I_2 and x_j

(iv) By maximizing $\sum_{j=1}^{m} | \tilde{r}(I_3, x_j) | : I_3 = Xw_3$, where $\tilde{r}(I_3, x_j)$ is Bradley's correlation between I_3 and x_j

(v) By maximizing $\sum_{j=1}^{m} | \rho(I_4, x_j) | : I_4 = Xw_4$, where $\rho(I_4, x_j)$ is Spearman's correlation between I_4 and x_j

(vi) By maximizing $\sum_{j=1}^{m} \rho^2(I_5, x_j) : I_5 = Xw_5$, where $\rho(I_5, x_j)$ is Spearman's correlation between I_5 and x_j

(vii) By maximizing $\sum_{j=1}^{m} | \widehat{r}(I_6, x_j) | : I_6 = Xw_6$, where $\widehat{r}(I_6, x_j)$ is the signum correlation between I_6 and x_j

(viii) By maximizing $\sum_{j=1}^{m} \widehat{r}^2(I_7, x_j) : I_7 = Xw_7$, where $\widehat{r}(I_7, x_j)$ is the signum correlation between I_7 and x_j

(ix) By maximizing $\sum_{j=1}^{m} | \breve{r}(I_8, x_j) | : I_8 = Xw_8$, where $\breve{r}(I_8, x_j)$ is the Shevlyakov correlation between I_8 and x_j

(x) By maximizing $\sum_{j=1}^{m} \breve{r}^2(I_9, x_j) : I_9 = Xw_9$, where $\breve{r}(I_9, x_j)$ is the Shevlyakov correlation between I_9 and x_j

(xi) I_{10} obtained from the first principal component of $\not{R}$, where $\not{R}$ is the Campbell's correlation matrix with the $\omega(d)$ as defined in Campbell (1980) mentioned above.

(xii) I_{11} obtained from the first principal component of $\not{R}$, where $\not{R}$ is the Campbell's correlation matrix with the $\omega(d)$ defined in Campbell (type-II) above.

(xiii) I_M obtained by $\max\left(\min_j (| \hat{r}(I_M, x_j) |\right)$ where $\hat{r}(I_M, x_j)$ is any specific (Pearson's, Spearman's, Signum or Shevlyakov or any other type of) correlation between I_M and x_j. Thus I_M is a class of indices whose members are different according to the type of correlation coefficient they use, but generically they all use the maxi-min criterion. In this paper we will use Spearman's correlation only to obtain I_M.

The thirteen types of composite indices enumerated above have been constructed with and without mutilation of x_{11} of X. The composite indices, the weights used to construct them and the relevant correlation of the composite indices (I) with the constituent indicator variables (X) are presented in Tables 4.1 through 5.2. Except I_0, which is constructed by a simple arithmetic averaging of variables ($I_{i0} = (1/6)\sum_{j=1}^{6} x_{ij}$) all other composite indices (I_1 through I_{11} and I_M) are based on maximization of different types of correlation. Since I_0 is not based on correlation, it is not relevant to compare its correlation with the constituent variables across the Tables 4.2 and 5.2. They are presented only for the completeness of those tables.

X. A Comparison of the Two Properties of Composite Indices: Earlier in section II of this paper we have noted two desirable properties of indicators, viz. sensitiveness to *autochthonous* changes ($alpha,\ \alpha$), and robustness or immunity to *allochthonous* changes ($beta,\ \beta$). We define them as follows:

$Alpha(\alpha) = \left((1/n_p)\sum_p (I_{pu} - I_{pv})^2\right)^{0.5}$; $p \in N = (1,2,...,n)$; where n_p =no. of elements of N that refer to the mutilated row(s) p of X (that contain(s) outliers); in our present case, n_p =1. Higher value of α indicates higher sensitivity and is a desirable property.

$Beta(\beta) = \left((1/n_q)\sum_q (I_{qu} - I_{qv})^2\right)^{0.5}$; $q \in N = (1,2,...,n)$; where n_q =no. of elements of N that refer to un-mutilated rows q of X (that do not contain outliers); presently, n_p =29 =(30 - n_p)=(30-1). A lower value of β is preferable to a higher value. Ideally β should be zero. Further, $n_p + n_q = n,$ and $I_{.u}(Table-4.1)$ and $I_{.v}(Table-5.1)$ are the composite indices constructed from un-mutilated (outlier-free) variables and mutilated (outlier-infested) variables.

A perusal of Table-6 reveals that the beta values of mean-based, Campbell-I, (S- and A-) Spearman, Campbell-II and maxi-min correlation based composite indices are lower. That means that in these composite indices the effects of outliers/mutilation are largely contained only by those observations that are directly affected and their effects do not percolate or pervade through all other observations. On the other hand, the alpha values (direct sensitivity) of S-signum, Campbell-II, Campbell-I, Mean-based and S-Spearman indices are relatively much higher than those of the other indices. Taking both criteria together, Mean-based, Campbell-I and S-Spearman composite indices are better than others. Among the correlation-based indices, Campbell-I is the best one. If S-Spearman weights are used on X to compute composite index, I_5 has a good performance.

XI. Concluding Remarks: When dealing with the real data obtained from the field, one does not know the location, magnitude or sign of outliers/errors of observation. When these (defective) data are used for sophisticated multivariate analysis, the results may be far from the reality. Correlation matrices (or covariance matrices) make a basis for a number of statistical methods. When correlation matrices are affected by outliers/errors/mutilations, the subsequent results become misleading. The composite indices are only a case in the large spectrum.

Our findings suggest that when we suspect the data to contain outliers or errors of a large magnitude, we should use a robust measure of correlation such as Campbell-I. For constructing indices, either the simple mean-based method (with suitable scaling of indicator variables) or the Campbell-I correlation or S-Spearman method should be used. In particular, S-

Spearman weights should be used on X rather than $\Re(X)$. For multivariate analysis such as the principal component analysis (Devlin, et al. 1981), the factor analysis, the discriminant analysis and the canonical correlation analysis including the regression analysis, one should prefer to use robust measures of correlation (covariance) than the Karl Pearson's correlation.

References

- Aitken, A. C. (1935) "On Least Squares and Linear Combinations of Observations", *Proceedings of the Royal Society of Edinburgh*, 55: 42-48.
- Blomqvist, N. (1950) "On a Measure of Dependence between Two Random Variables", *Annals of Mathematical Statistics*, 21(4): 593-600.
- Bradley, C. (1985) "The Absolute Correlation", *The Mathematical Gazette*, 69(447): 12-17.
- Campbell, N. A. (1980) "Robust Procedures in Multivariate Analysis I: Robust Covariance Estimation", *Applied Statistics*, 29 (3): 231-237
- Devlin, S.J., Gnanadesikan, R. and Kettenring, J.R. (1981) "Robust Estimation of Dispersion Matrices and Principal Components", *Journal of the American Statistical Association*, 76 (374): 354-362.
- Gnanadesikan, R. and Kettenring, J.R. (1972) "Robust Estimates, Residuals and Outlier Detection with Multiresponse Data", Biomatrics, 28: 81-124.
- Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J. and W. A. Stahel, W.A. (1986) *Robust Statistics: The Approach Based on Influence Functions,* Wiley, New York.
- Hill, M. and Tzamir, Y. (1972) "Multidimensional Evaluation of Regional Plans serving Multiple Objectives", *Papers in Regional Science*, 29(1): 139-165
- Kendall, M.G. and Stuart, A. (1968): *The Advanced Theory of Statistics*, vol. 3, Charles Griffin & Co. London.
- Mahalanobis, P. C. (1936) "On the Generalized Distance in Statistics", *Proceedings of the National Institute of Science of India,* 12: 49-55.
- Mishra, S.K. (1984) "Taxonomical Analysis of Regional Development by Outranking Relations on Multiple Principal Components". *Hill Geographer*, 3(1): 20-28.
- Mishra, S.K. (2006) "Global Optimization by Differential Evolution and Particle Swarm Methods: Evaluation on Some Benchmark Functions", SSRN: http://ssrn.com/abstract=933827
- Mishra, S. K. (2007a) "Construction of an Index by Maximization of the Sum of its Absolute Correlation Coefficients with the Constituent Variables", Working Papers Series, SSRN: http://ssrn.com/abstract=989088
- Mishra, S. K. (2007b) "A Comparative Study of Various Inclusive Indices and the Index Constructed By the Principal Components Analysis", Working Papers Series, SSRN: http://ssrn.com/abstract=990831
- Munda, G. and Nardo, M. (2005-a) "Constructing Consistent Composite Indicators: the Issue of Weights", Official Publications of the European Communities, European Communities, Luxembourg, http://crell.jrc.ec.europa.eu/Well-being/papers/Munda%20Nardo%20euroreport1.pdf
- Munda, G. and Nardo, M. (2005-b) "Non-Compensatory Composite Indicators for Ranking Countries: A Defensible Setting", Official Publications of the European Communities, Luxembourg, http://crell.jrc.ec.europa.eu/Well-being/papers/Munda%20Nardo%20euroreport2.pdf
- Shevlyakov, G.L. (1997) "On Robust Estimation of a Correlation Coefficient", Journal of Mathematical Sciences, 83(3): 434-438.
- Spearman, C. (1904) "The Proof and Measurement of Association between Two Things", *American. Journal of Psychology,* 15: 88-93.
- Van Delft, A. and Nijkamp, P. (1976) "A Multi-objective Decision Model for Regional Development, Environmental Quality Control and Industrial Land Use," *Papers in Regional Science*, 36(1): 35–57.

Table-1. Generated X(30, 6) Variables to be used as Indictors to construct Composite Indices [Seed for generating random number = 1111]						
Sl No.	X_1	X_2	X_3	X_4	X_5	X_6
1	3.24515	17.11875	18.93120	4.94349	4.70523	9.16500
2	24.84912	18.12915	17.68236	15.48139	26.29670	10.99727
3	50.34351	53.23216	52.77337	44.02273	55.64800	44.15540
4	40.42578	42.36102	36.21973	41.95478	31.63675	38.11307
5	32.62840	44.66287	31.38759	43.42580	35.13244	37.02850
6	31.13495	30.16973	30.22937	19.27427	33.00687	33.99838
7	19.73745	4.94763	15.00810	8.47932	18.76237	19.21965
8	16.90762	13.96999	24.57726	14.65958	19.68803	19.11785
9	4.93962	18.00873	14.51709	15.94525	15.93895	3.04773
10	25.32545	37.60286	26.88260	32.06437	39.63724	38.43864
11	30.01135	48.71366	39.05519	32.78365	42.08059	34.30613
12	29.39361	17.44837	18.65002	33.36702	30.85420	23.32429
13	30.91832	43.30296	40.68762	33.53372	34.14844	42.22184
14	17.41810	22.20521	24.75624	36.97844	27.35229	19.59569
15	41.99813	52.38159	53.82127	49.00823	47.17287	39.47807
16	13.07349	4.42329	19.37725	6.95275	17.82013	14.84487
17	37.90192	30.73150	28.63064	40.55927	31.03877	26.55806
18	28.43454	41.56275	31.12926	30.37443	39.74691	42.77524
19	46.61491	42.26589	51.81282	48.23561	40.84989	48.73060
20	34.93265	46.54018	32.38450	42.98263	33.97258	44.52911
21	13.43902	25.00785	14.55443	27.71442	20.75521	20.55992
22	36.88546	27.41065	37.41027	42.68135	44.31126	26.65092
23	34.88258	30.03870	26.65572	29.95285	20.18235	31.88528
24	29.43575	33.02137	31.16716	27.77126	24.08382	32.52221
25	32.96518	39.92390	47.31983	49.06694	32.98625	40.77812
26	17.42287	36.16056	25.28987	31.42659	26.08310	18.79163
27	24.50782	36.18683	27.17725	28.00860	34.08299	23.04710
28	43.89777	57.91221	49.96161	56.12815	44.68269	40.51109
29	56.54121	41.96319	52.67857	48.17311	44.70189	44.42373
30	46.68000	46.61551	40.08049	41.45077	48.48765	38.37636
Mean	29.89639	33.46730	32.02696	32.58003	32.19488	30.23973
S. Dev.	12.74549	13.91941	12.08567	13.47113	11.2664	12.00505

..

Table-2.1. Correlation Matrix of Indictor Variables, X of Table-1.													
	Karl Pearson's Coefficients of Correlation							Spearman's Coefficients of Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.72248	0.84368	0.80455	0.83642	0.83081		1.00000	0.74772	0.85984	0.82736	0.79088	0.79399
X_2	0.72248	1.00000	0.81841	0.82252	0.79408	0.82192		0.74772	1.00000	0.85806	0.77842	0.83715	0.81491
X_3	0.84368	0.81841	1.00000	0.80499	0.80900	0.82533		0.85984	0.85806	1.00000	0.83582	0.85228	0.85495
X_4	0.80455	0.82252	0.80499	1.00000	0.78561	0.76657		0.82736	0.77842	0.83582	1.00000	0.77486	0.77397
X_5	0.83642	0.79408	0.80900	0.78561	1.00000	0.77099		0.79088	0.83715	0.85228	0.77486	1.00000	0.78776
X_6	0.83081	0.82192	0.82533	0.76657	0.77099	1.00000		0.79399	0.81491	0.85495	0.77397	0.78776	1.00000

..

Table-2.2. Correlation Matrix of Indictor Variables, X of Table-1.													
	Signaum Coefficients of Correlation							Bradley's Coefficients of Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.46667	0.60000	0.73333	0.46667	0.60000		1.00000	0.75579	0.61097	0.97635	0.92616	0.92616
X_2	0.46667	1.00000	0.73333	0.46667	0.73333	0.86667		0.75579	1.00000	0.83998	0.77816	0.68758	0.82650
X_3	0.60000	0.73333	1.00000	0.60000	0.60000	0.73333		0.61097	0.83998	1.00000	0.63123	0.55006	0.67549
X_4	0.73333	0.46667	0.60000	1.00000	0.33333	0.46667		0.97635	0.77816	0.63123	1.00000	0.90268	0.94972
X_5	0.46667	0.73333	0.60000	0.33333	1.00000	0.73333		0.92616	0.68758	0.55006	0.90268	1.00000	0.85312
X_6	0.60000	0.86667	0.73333	0.46667	0.73333	1.00000		0.92616	0.82650	0.67549	0.94972	0.85312	1.00000

..

Table-2.3. Correlation Matrix of Indictor Variables, X of Table-1.													
	Shevlyakov's Coefficients of Correlation							Campbell's Coefficients of Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.72014	0.81308	0.81066	0.86198	0.78165		1.00000	0.72248	0.84368	0.80455	0.83642	0.83081
X_2	0.72014	1.00000	0.85969	0.81083	0.77017	0.82992		0.72248	1.00000	0.81841	0.82252	0.79408	0.82192
X_3	0.81308	0.85969	1.00000	0.59618	0.77754	0.75549		0.84368	0.81841	1.00000	0.80499	0.80900	0.82533
X_4	0.81066	0.81083	0.59618	1.00000	0.59280	0.67165		0.80455	0.82252	0.80499	1.00000	0.78561	0.76657
X_5	0.86198	0.77017	0.77754	0.59280	1.00000	0.73849		0.83642	0.79408	0.80900	0.78561	1.00000	0.77099
X_6	0.78165	0.82992	0.75549	0.67165	0.73849	1.00000		0.83081	0.82192	0.82533	0.76657	0.77099	1.00000

..

Table-3.1. Correlation Matrix of Indictor Variables, X of Table-1 with x_{11} mutilated.													
	Karl Pearson's Coefficients of Correlation							Spearman's Coefficients of Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	-0.17115	-0.14499	-0.33227	-0.40395	-0.27395		1.00000	0.59422	0.73304	0.63382	0.59733	0.61379
X_2	-0.17115	1.00000	0.81841	0.82252	0.79408	0.82192		0.59422	1.00000	0.85806	0.77842	0.83715	0.81491
X_3	-0.14499	0.81841	1.00000	0.80499	0.80900	0.82533		0.73304	0.85806	1.00000	0.83582	0.85228	0.85495
X_4	-0.33227	0.82252	0.80499	1.00000	0.78561	0.76657		0.63382	0.77842	0.83582	1.00000	0.77486	0.77397
X_5	-0.40395	0.79408	0.80900	0.78561	1.00000	0.77099		0.59733	0.83715	0.85228	0.77486	1.00000	0.78776
X_6	-0.27395	0.82192	0.82533	0.76657	0.77099	1.00000		0.61379	0.81491	0.85495	0.77397	0.78776	1.00000

..

Table-3.2. Correlation Matrix of Indictor Variables, X of Table-1 with x_{11} mutilated.													
	Signaum Coefficients of Correlation							Bradley's Coefficients of Correlation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.33333	0.46667	0.60000	0.33333	0.46667		1.00000	-0.13163	-0.09708	-0.19920	-0.23706	-0.18186
X_2	0.33333	1.00000	0.73333	0.46667	0.73333	0.86667		-0.13163	1.00000	0.83998	0.77816	0.68758	0.82650
X_3	0.46667	0.73333	1.00000	0.60000	0.60000	0.73333		-0.09708	0.83998	1.00000	0.63123	0.55006	0.67549
X_4	0.60000	0.46667	0.60000	1.00000	0.33333	0.46667		-0.19920	0.77816	0.63123	1.00000	0.90268	0.94972
X_5	0.33333	0.73333	0.60000	0.33333	1.00000	0.73333		-0.23706	0.68758	0.55006	0.90268	1.00000	0.85312
X_6	0.46667	0.86667	0.73333	0.46667	0.73333	1.00000		-0.18186	0.82650	0.67549	0.94972	0.85312	1.00000

Table-3.3. Correlation Matrix of Indictor Variables, X of Table-1 with x_{11} mutilated.													
Shevlyakov's Coefficients of Correlation								Campbell's Coefficients of Correlation (type-I)					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.67889	0.81969	0.75845	0.76281	0.78429		1.00000	0.70917	0.84808	0.77065	0.80399	0.80835
X_2	0.67889	1.00000	0.85969	0.81083	0.77017	0.82992		0.70917	1.00000	0.81020	0.81946	0.79914	0.81379
X_3	0.81969	0.85969	1.00000	0.59618	0.77754	0.75549		0.84808	0.81020	1.00000	0.80418	0.82204	0.82041
X_4	0.75845	0.81083	0.59618	1.00000	0.59280	0.67165		0.77065	0.81946	0.80418	1.00000	0.74371	0.73492
X_5	0.76281	0.77017	0.77754	0.59280	1.00000	0.73849		0.80399	0.79914	0.82204	0.74371	1.00000	0.73957
X_6	0.78429	0.82992	0.75549	0.67165	0.73849	1.00000		0.80835	0.81379	0.82041	0.73492	0.73957	1.00000

..

Table-3.4. Correlation Matrix of Indictor Variables, X of Table-1 without/with x_{11} mutilated.													
Campbell's Coefficients of Correlation (type-II) without mutilation								Campbell's Coefficients of Correlation (type-II) with mutilation					
	X_1	X_2	X_3	X_4	X_5	X_6		X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.00000	0.69835	0.85152	0.77355	0.83790	0.77019		1.00000	0.70917	0.84808	0.77065	0.80399	0.80835
X_2	0.69835	1.00000	0.82080	0.86480	0.84873	0.79847		0.70917	1.00000	0.81020	0.81946	0.79914	0.81379
X_3	0.85152	0.82080	1.00000	0.84091	0.84685	0.83689		0.84808	0.81020	1.00000	0.80418	0.82204	0.82041
X_4	0.77355	0.86480	0.84091	1.00000	0.72419	0.75929		0.77065	0.81946	0.80418	1.00000	0.74371	0.73492
X_5	0.83790	0.84873	0.84685	0.72419	1.00000	0.75284		0.80399	0.79914	0.82204	0.74371	1.00000	0.73957
X_6	0.77019	0.79847	0.83689	0.75929	0.75284	1.00000		0.80835	0.81379	0.82041	0.73492	0.73957	1.00000

.

Table-4.1. Composite Indices of Variables (X of Table-1) using Different Types of Correlation													
Sl	I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_M
1	9.6848	9.5944	9.6177	10.9405	9.7515	9.2650	11.0274	7.7700	10.7196	10.7352	10.6114	10.5492	8.1250
2	18.9060	19.0150	19.0145	19.2927	18.5125	18.5602	21.0632	17.6713	18.9590	18.9586	20.6704	20.4739	18.3993
3	50.0292	50.1239	50.1306	50.9532	50.0823	49.9333	52.6774	48.1218	50.9465	50.9583	54.7034	54.1938	49.2010
4	38.4519	38.1972	38.1892	38.0034	37.3447	37.2083	36.0914	37.6016	38.2988	38.2865	42.0284	41.6099	37.9471
5	37.3776	37.1184	37.0838	35.3517	37.1659	37.3831	31.0281	35.3749	37.6808	37.6777	40.8249	40.4445	38.3116
6	29.6356	29.8414	29.8517	32.2441	29.9120	29.5057	37.0153	30.4503	31.8634	31.8780	32.4124	32.0803	29.0833
7	14.3591	14.6903	14.7098	16.2656	14.8860	14.8106	21.1928	17.6982	14.5391	14.5414	15.7222	15.5220	14.4692
8	18.1534	18.3421	18.3678	19.1870	18.8608	18.7724	21.2760	19.3216	17.8835	17.8920	19.8752	19.6799	17.7685
9	12.0662	11.9641	11.9523	9.9978	12.5428	12.8656	6.0215	8.5765	11.4477	11.4561	13.1775	13.1123	12.3228
10	33.3252	33.4070	33.3727	32.5735	34.6265	34.8825	31.8010	32.9012	35.2145	35.2319	36.3969	36.0540	35.2301
11	37.8251	37.7862	37.7733	37.9607	38.3274	38.1886	37.2526	34.2784	39.9322	39.9548	41.3382	40.9910	37.6009
12	25.5063	25.5798	25.5590	23.2833	25.7446	26.4176	21.5121	26.9315	23.5040	23.4884	27.8645	27.5644	27.2470
13	37.4688	37.4284	37.4330	38.4283	37.8820	37.5571	38.7233	36.8865	39.0517	39.0665	40.9676	40.5830	37.1165
14	24.7177	24.6495	24.6323	20.8915	25.7334	26.5405	14.9098	23.9056	22.0132	22.0085	27.0038	26.7709	26.5106
15	47.3100	47.2012	47.2095	46.3807	47.4235	47.4571	43.7685	44.4248	46.6412	46.6481	51.7302	51.2760	46.5784
16	12.7486	13.1040	13.1359	14.3481	13.8723	13.8327	18.3287	15.1338	12.4818	12.4931	13.9743	13.8235	12.6015
17	32.5700	32.3962	32.3835	30.7034	31.5801	31.8941	27.8259	31.9606	30.6535	30.6321	35.5940	35.2352	32.7343
18	35.6705	35.7546	35.7307	36.2305	36.7217	36.6596	37.1371	35.3606	38.4259	38.4470	38.9721	38.5996	36.8096
19	46.4183	46.4115	46.4352	46.7313	46.3881	46.2874	46.8038	47.9540	45.2288	45.2232	50.7733	50.2591	45.8981
20	39.2236	38.9691	38.9368	38.1882	38.9258	38.9237	35.2371	38.3915	40.3613	40.3590	42.8457	42.4250	40.0689
21	20.3385	20.1685	20.1324	17.7486	20.7918	21.2274	13.0635	18.9334	20.1940	20.1939	22.1953	22.0001	21.9779
22	35.8917	36.0501	36.0517	33.5710	36.6335	37.3499	31.5743	35.9461	32.9637	32.9570	39.2426	38.8696	36.8800
23	28.9329	28.7301	28.7322	29.6706	27.5020	27.1628	29.9269	29.5770	29.0179	29.0017	31.6332	31.2887	28.0194

24	29.6669	29.5596	29.5664	30.5977	29.1744	28.8313	31.0271	29.5358	30.4181	30.4183	32.4406	32.1179	28.8777
25	40.5067	40.3413	40.3539	38.6864	40.9517	41.1454	34.3036	40.5077	38.3958	38.3919	44.2965	43.8824	40.6689
26	25.8624	25.5995	25.5736	23.4200	25.9414	26.1949	17.9936	21.7052	25.8173	25.8230	28.2443	28.0320	26.2382
27	28.8351	28.7804	28.7605	27.9657	29.0638	29.1856	26.2795	25.7385	29.7640	29.7751	31.5025	31.2385	29.0625
28	48.8489	48.4959	48.4809	46.5420	48.2013	48.3398	40.9129	44.9250	47.9786	47.9740	53.3863	52.9154	48.4806
29	48.0803	48.0955	48.1255	48.9792	47.1065	46.9306	50.6795	49.1912	46.4680	46.4530	52.5981	52.0548	46.6554
30	43.6151	43.6130	43.5997	43.6208	43.1856	43.2062	44.0802	41.9229	44.2925	44.2933	47.6695	47.2113	43.4444

.

Table-4.2. Weights of Indicator Variables and their Correlation with respective Composite Indices Composites												
	Weights assigned to Different Constituent Variables						Correlation of Composite Indices with Constituent Variables					
Index	X_1	X_2	X_3	X_4	X_5	X_6	X_1	X_2	X_3	X_4	X_5	X_6
I_0	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.91593	0.91128	0.92844	0.91192	0.90669	0.91286
I_1	0.16368	0.14988	0.17262	0.15487	0.18517	0.17378	0.91804	0.90738	0.92961	0.90826	0.91042	0.91398
I_2	0.16431	0.14859	0.17551	0.15370	0.18425	0.17363	0.91834	0.90698	0.93015	0.90792	0.91027	0.91402
I_3	0.26368	0.17880	0.23533	-0.04771	0.13128	0.23862	0.78253	0.76542	0.79659	0.69284	0.75378	0.80772
I_4	0.05004	0.10016	0.15874	0.19633	0.28584	0.20889	0.89143	0.92392	0.95729	0.88877	0.90211	0.91413
I_5	0.05995	0.10706	0.14515	0.18814	0.29327	0.20642	0.88921	0.91012	0.95640	0.89989	0.91724	0.90567
I_6	0.27155	0.27253	0.37839	-0.18246	0.18351	0.07648	0.60000	0.73333	0.73333	0.46667	0.73333	0.86667
I_7	0.18934	0.04025	0.21347	0.08052	0.23790	0.23852	0.60000	0.86667	0.86667	0.60000	0.73333	0.86667
I_8	0.15853	0.29999	0.09882	-0.00223	0.19450	0.25040	0.90713	0.95281	0.94576	0.87222	0.89303	0.96780
I_9	0.15733	0.30041	0.09933	-0.00329	0.19545	0.25078	0.90621	0.95338	0.94584	0.87108	0.89398	0.96820
I_{10}	0.91834	0.90698	0.93015	0.90792	0.91027	0.91402	0.91834	0.90698	0.93015	0.90792	0.91027	0.91402
I_{11}	0.91374	0.9281	0.94655	0.91855	0.92468	0.91016	0.91374	0.9281	0.94655	0.91855	0.92468	0.91016
I_M	0.12581	0.12963	0.03704	0.24655	0.24443	0.21653	0.90745	0.90879	0.94438	0.90968	0.90656	0.90656

..

Table-5.1. Composite Indices of Variables (Mutilated X of Table-1) using Different Types of Correlation													
Sl	I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_M
1	176.3515	-3.5179	5.1449	16.7937	25.1186	23.4711	24.8536	-267.4545	4.3735	4.3735	193.8887	193.8887	33.9684
2	18.9060	17.7722	17.8435	17.0849	17.7823	17.7896	18.4352	15.1795	17.5863	17.5863	20.8106	20.8106	17.6480
3	50.0292	50.0771	50.0897	49.8444	49.1205	49.1119	51.9449	48.6928	50.9178	50.9178	55.1043	55.1043	48.7327
4	38.4519	37.7225	37.7448	38.9516	37.2824	37.2737	38.6617	39.7761	37.9240	37.9240	42.3144	42.3144	37.4613
5	37.3776	38.0748	38.0303	38.9486	37.4917	37.5651	38.8839	43.7402	38.3251	38.3251	41.0973	41.0973	37.5581
6	29.6356	29.5658	29.5772	29.0272	29.3682	29.4130	30.4713	26.1988	31.5948	31.5948	32.6458	32.6458	28.8554
7	14.3591	13.6152	13.6605	12.1163	15.1076	15.1144	11.5218	7.8720	13.4789	13.4789	15.8127	15.8127	15.0546
8	18.1534	18.6476	18.6310	18.0630	18.7640	18.7017	18.0923	15.9435	18.4308	18.4308	20.0283	20.0283	18.6740
9	12.0662	13.4604	13.4006	13.7471	12.1933	12.1719	14.9008	17.5244	12.8316	12.8316	13.2930	13.2930	12.1235
10	33.3252	35.1284	35.0433	34.4238	35.1314	35.2847	34.9757	38.4675	36.3480	36.3480	36.6364	36.6364	34.7696
11	37.8251	39.4423	39.3722	39.8868	37.4547	37.5030	42.6276	42.5713	41.4398	41.4398	41.6606	41.6606	36.8415
12	25.5063	24.7674	24.8065	23.4994	26.7495	26.7961	21.5171	24.6584	22.2600	22.2600	28.0058	28.0058	27.2591
13	37.4688	38.8151	38.7446	39.5177	37.6443	37.6502	40.4319	40.4777	40.5836	40.5836	41.2780	41.2780	37.2872
14	24.7177	26.1899	26.1188	25.7780	26.7225	26.6992	24.1029	30.7178	23.3098	23.3098	27.1774	27.1774	27.2174
15	47.3100	48.3096	48.2660	49.0136	46.8059	46.7233	50.1643	50.6563	47.9662	47.9662	52.1239	52.1239	46.7360
16	12.7486	13.1106	13.1070	11.6918	13.8835	13.8297	11.5651	8.1262	12.6771	12.6771	14.0792	14.0792	13.7553
17	32.5700	31.2234	31.2844	31.5183	31.8228	31.8030	30.4215	32.3192	29.3976	29.3976	35.8136	35.8136	32.3210
18	35.6705	37.3139	37.2350	36.9674	36.8582	36.9985	38.1208	39.5555	39.5741	39.5741	39.2423	39.2423	36.3148
19	46.4183	46.3682	46.3640	46.6598	46.6426	46.5509	45.4611	45.1535	45.5666	45.5666	51.1292	51.1292	46.8941
20	39.2236	39.8289	39.7823	40.8865	39.3851	39.4804	40.5768	44.8892	40.9627	40.9627	43.1288	43.1288	39.3745
21	20.3385	21.6020	21.5332	21.8777	21.4835	21.5644	21.3244	27.3529	21.2091	21.2091	22.3338	22.3338	21.6055

22	35.8917	35.8721	35.8890	34.4238	37.0670	37.0134	33.3582	34.9252	32.6171	32.6171	39.4956	39.4956	37.4968
23	28.9329	27.4195	27.4774	28.5684	27.4482	27.4328	27.9209	27.0891	27.9279	27.9279	31.8407	31.8407	27.6446
24	29.6669	29.5860	29.5815	30.4822	28.9193	28.9016	30.7796	29.9417	30.6663	30.6663	32.6771	32.6771	28.8217
25	40.5067	41.9125	41.8322	42.7796	41.5704	41.4593	41.1491	45.3224	40.3103	40.3103	44.6184	44.6184	41.9789
26	25.8624	27.3412	27.2678	28.3796	25.6779	25.6926	29.5215	33.8502	27.3787	27.3787	28.4608	28.4608	25.6050
27	28.8351	29.6906	29.6574	29.8135	28.5322	28.5827	31.6173	32.5931	30.4049	30.4049	31.7347	31.7347	28.2211
28	48.8489	49.4753	49.4382	51.0476	47.8355	47.7915	51.6960	55.2125	49.0206	49.0206	53.7762	53.7762	47.9746
29	48.0803	46.2964	46.3841	46.4158	46.7751	46.6580	45.7727	42.1663	45.1050	45.1050	52.9601	52.9601	47.0867
30	43.6151	42.9597	43.0001	42.7932	42.6287	42.6749	44.1168	42.7824	43.4186	43.4186	47.9945	47.9945	42.4497

...

Table-5.2. Weights of Mutilated Indicator Variables and their Correlation with respective Composite Indices Composites

	Weights assigned to Different Constituent Variables						Correlation of Composite Indices with Constituent Variables					
Index	X_1	X_2	X_3	X_4	X_5	X_6	X_1	X_2	X_3	X_4	X_5	X_6
I_0	0.16667	0.16667	0.16667	0.16667	0.16667	0.16667	0.94438	0.14158	0.1667	-0.03234	-0.11163	0.02685
I_1	-0.01446	0.18181	0.20939	0.18786	0.22462	0.21080	-0.46816	0.89113	0.88603	0.90801	0.91841	0.89730
I_2	-0.00574	0.18075	0.20769	0.18620	0.22310	0.20800	-0.36590	0.91686	0.91472	0.91409	0.91600	0.91000
I_3	0.00459	0.27805	0.22294	0.20009	0.10807	0.18626	0.13231	0.83884	0.79072	0.75836	0.72525	0.75225
I_4	0.01494	0.05166	0.15370	0.24529	0.26840	0.26601	0.77219	0.90834	0.95996	0.89321	0.89321	0.90790
I_5	0.01537	0.05203	0.15862	0.24085	0.26590	0.26723	0.77219	0.90834	0.95996	0.89321	0.89321	0.90790
I_6	-0.13456	0.11201	0.06289	0.33548	0.24834	0.37583	0.46667	0.86667	0.86667	0.60000	0.73333	0.86667
I_7	-0.22132	0.21406	0.07678	0.11939	0.33948	0.47161	0.46667	0.86667	0.86667	0.60000	0.73333	0.86667
I_8	-0.00817	0.32778	0.17590	0.02095	0.20228	0.28126	0.79594	0.96276	0.95954	0.88606	0.93295	0.95633
I_9	-0.00817	0.32778	0.17590	0.02095	0.20228	0.28126	0.79594	0.96276	0.95954	0.88606	0.93295	0.95633
I_{10}	0.90665	0.90851	0.93768	0.89362	0.90059	0.90227	0.90665	0.90851	0.93768	0.89362	0.90059	0.90227
I_{11}	0.90665	0.90851	0.93768	0.89362	0.90059	0.90227	0.90665	0.90851	0.93768	0.89362	0.90059	0.90227
I_M	0.02349	0.00860	0.16803	0.27876	0.23875	0.28237	0.85050	0.85317	0.93059	0.86296	0.85451	0.86607

..

Table-6. Sensitivity and Robustness of Different Composite Indices to Mutilation and Presence of Outliers

Alpha and Beta Values of difference Composite Indices arranged according to value of Beta

Sl No.	Type of Composite Index		Beta	Alpha
1	Mean	I_0	0	166.6667
2	Campbell-I	I_{10}	0.26346	183.2773
3	S-Spearman	I_5	0.43228	14.2061** 158.5300*
4	A-Spearman	I_4	0.52073	15.3671
5	Campbell-II	I_{11}	0.61880	183.3395
6	Maxi-min	I_M	0.65191	25.8434
7	S-Shevlyakov	I_9	1.05643	6.3617
8	A-Shevlyakov	I_8	1.06304	6.3461
9	S-Pearson	I_2	1.09376	4.4728
10	A-Pearson	I_1	1.13635	13.1123
11	Bradley	I_3	2.68379	5.8532
12	A-Signum	I_6	5.80213	13.8262
13	S-Signum	I_7	6.04344	275.2245

* Obtained by using S-Spearman weights to mutilated X;

Alpha and Beta Values of difference Composite Indices arranged according to value of Alpha

Sl No.	Type of Composite Index		Beta	Alpha
1	S-Signum	I_7	6.04344	275.2245
2	Campbell-II	I_{11}	0.6188	183.3395
3	Campbell-I	I_{10}	0.26346	183.2773
4	Mean	I_0	0	166.6667
5	S-Spearman	I_5	0.43228	158.5300* 14.2061**
6	Maxi-min	I_M	0.65191	25.8434
7	A-Spearman	I_4	0.52073	15.3671
8	A-Signum	I_6	5.80213	13.8262
9	A-Pearson	I_1	1.13635	13.1123
10	S-Shevlyakov	I_9	1.05643	6.3617
11	A-Shevlyakov	I_8	1.06304	6.3461
12	Bradley	I_3	2.68379	5.8532
13	S-Pearson	I_2	1.09376	4.4728

** Obtained by using S-Spearman weights to rank of mutilated X (x_{11} is altered; others unaltered)

Note: Computer programs (FORTRAN) for computing correlations and Composite Indices used in this paper are obtainable from the author (contact: mishrasknehu@yahoo.com).