Applied Economics & Business, 2019 3(2) 58 - 75

Construction of a Composite Vegetable Price Index Using Modified Factor Analysis

Siriwardhana, S.M.C.P. *, Thattil, R.O. and Abeynayake R.

Postgraduate Institute of Agriculture, University of Peradeniya, Sri Lanka

**Corresponding Author:*
Email: smcpsirwardhana@gmail.com

ABSTRACT

This study introduces a modified factor analysis approach to develop a composite vegetable price index. The new method uses scaling by dividing the original variables with its mean, a specific weight for each individual indicator variable and the index assigns a specific numerical value to prices of vegetables for a given month. Initially monthly wholesale prices of nineteen vegetable were considered. As some vegetable prices were highly correlated, ten representative variables for highly correlated variables were retained based on variable-cluster analysis and correlation analysis. Green Beans, Leeks, Cabbage, Tomatoes, Brinjals, Pumpkin, Cucumber, Luffa, Ash Plantains, and Green Chili were the indicator variables considered in the index building process.

Initially, the grouping pattern in the data was identified through a Preliminary Factor Analysis. This resulted in a single factor explaining a substantial amount of the total variance. The original variables were divided by their means to scale the variables. The weight corresponding to a particular indicator variable was defined by squaring the Eigen vector coefficient of the given variable of the first Principal Component. Then the weights were assigned to the scaled variables and used in the final Factor Analysis. A single factor explaining 69.8% of total variance was selected as the composite index. First, the Vegetable Price Index was defined as a linear function of the composite index. Then it was converted into a function of original indicator variables by summarizing constant terms to make it easy to update. Cronbach's alpha was used to verify the internal consistency of the indicator variables. This method is not sensitive to variables having comparatively higher variances because of their means. Scaling in mean and weighting improved explaining variance and internal consistency of the variables.

KEYWORDS: Composite index, vegetable price index, weighted factor analysis

Introduction

Sri Lanka being an agricultural country the culture of the country is based on agriculture. The vegetable sub-sector is the second most prominent sub-sector within the Non-perennial agricultural sector, as vegetables are grown throughout the country.

In the process of transformation of agriculture from subsistence into commercial level, the farmers need market information to decide on the choice of crops to be cultivated at the correct time and information on the prices and awareness of marketing. On the other hand, traders need market information to take similar decisions on purchasing, selling, and storage followed by policy makers who require such information to formulate policies on food supply, demand, and consumption and trading, and to impose market regulations. The limitations of univariate analysis is that such analysis is not sensitive to sudden changes such as pest and disease outbreaks and weather fluctuations and it does not provide comprehensive information on vegetable market. Hence, instead of univariate analysis, multivariate analysis can be used to overcome such limitations.

When there are no statistical methods used, in many composite indicators, all variables are given the same weight (Nardo *et al.,* 2005). When using multivariate techniques to construct a composite index, the weights have to be derived for different sets of variables based on PCs within each set (Garcia and Puetra, 1997).Out of different techniques used in constructing indices this research study constructed a composite vegetable price index by using weighted factor analysis and introduced new method of scaling of variables and weights deriving method.

As vegetable price index gives comprehensive information on vegetable market, it is a major requirement to vegetable farmers, buyers, and input suppliers to plan their business while minimizing post-harvest losses.

This research is significant in two aspects. It improves the vegetable market in Sri Lanka by introducing vegetable price index and contribution to the improvement of statistical theory of Weighted Factor Analysis. Following paragraph further, explain these.

- **New scaling method:** In this research study new scaling method was introduced an in that original variable were divided by their mean in order to remove the effect on mean on original variances of variables. Internal consistency and explained variance were increased by the new scaling method.
- **New weight deriving equation:** After identifying factor structure PCA was performed on variables in the retained factor. Then weights were derived for each variable by squiring their Eigen vector coefficients.
- **New index for wholesale vegetable prices:** Vegetable price index is timely requirement to improve vegetable market in Sri Lanka and thereby reduce postharvest losses of vegetables while increasing the profit margin.

Methodology

Data Collection

Monthly wholesale prices of 19 vegetables at production area DECs were collected from Hector Kobbekaduwa Agrarian Research Institute (HARTI) for 11 years from 2005 to 2015. Selected vegetables were Green Beans, Carrot, Leeks, Beetroot, Knol-Khol, Radish, Cabbage, Tomatoes, Ladies Fingers, Brinjals, Capsicum, Pumpkin, Cucumber, Bitter Gourd, Snake Gourd, Luffa, Long Beans, Ash Plantains, and Green Chilies.

Variable Reduction

Initially correlation analysis was done on the prices of vegetables and it showed that some of the prices of vegetables were highly correlated and it is useless to take all variables to index construction. Hence, variable-cluster analysis was performed on originally selected vegetables. As some clusters have more than one variable based on correlation coefficients between each pair of vegetables within a given cluster and availability of vegetable throughout the country, ten representative variables were selected.

Construction of the Index

Weighted Principle Component based factor analysis was used to construct the composite Vegetable price index. After PCA, a preliminary FA was performed on the original indicator variables (vegetable prices) in order to identify the grouping patterns in the indicator variables that were retained as representative variables.

Scaling

The selected variables were divided by their means to make the variables with unit mean (scaled).

$$\mathbf{X}_{\mathbf{i}}^{*}=\left(\frac{\mathbf{X_i}}{\mu_\mathbf{i}}\right) \qquad [1]$$

Where, X_i - Original i^{th} variable; X_i^* - Scaled i^{th} variable; μ_i - Mean of i^{th} variable

Weights

Principle component analysis with correlation matrix option of scaled variables was carried out for the selected subgroups separately. As the Eigen vector coefficient represents the proportion of variation contributed by a given variable to a PC, the weights were calculated by squaring Eigen vector coefficient of a given variable of the PC.

$$\mathbf{w_{pi}} = \mathbf{a_{p1i}^2} \qquad [2]$$

Where, w_{pi} - weight correspond to p^{th} variable in i^{th} subset; a_{p1i} - Eigen vector coefficient of p^{th} variable of the first PC of i^{th} subset; i - 1, 2, …, m (number of subsets); p - 1, 2, …,ni (number of variables in i^{th} subset)

Then, the scaled variables were weighted according to the Equation *3* and the transformed variables were denoted by X_{iw}.

$$\boldsymbol{X_{iw}} = \left(\frac{\boldsymbol{X_i}}{\mu_{\boldsymbol{i}}}\right)\boldsymbol{w_i} \qquad [3]$$

Where, w_i - weight corresponds to i^{th} variable; μ_i- mean of i^{th} variable; X_i - original i^{th} variable

PCs-based FA with covariance matrix option was performed on the scaled and weighted variables in order to get factor score coefficients. Then composite vegetable index was computed as weighted average vegetable price using factor score coefficient as weights.

$$\boldsymbol{CVPI} = \sum\left[\left(\frac{\boldsymbol{Xi}}{\mu_{\boldsymbol{i}}}\right) \times \boldsymbol{w_i} \times \boldsymbol{FSC_i}\right] \qquad [4]$$

Where, CVPI - Composite Vegetable Price Index; X_i - Price of i^{th} variable; μ_i - mean of i^{th} variable; w_i - weight of i^{th} variable; FSC_I - Factor Score Coefficients of i^{th} variable

Since weight, mean and factor score coefficient are constant for a given variable, C_i summarizes the constant terms of i^{th} variable.

$$C_i = \frac{w_i \times FSC_i}{\mu_i} \quad [5]$$

As C_i values are very small, C_i values were reweighted to equal the summation of W_{Ci} to one. W_{Ci} , defines the final weight for i^{th} variable.

$$W_{Ci} = \frac{C_i}{\sum_{i=n} C_i} \quad [6]$$

As final composite vegetable price index, Equation *7* was used.

$$CVPI = \sum_{i=1}^{n} [X_i \times W_{Ci}] \quad [7]$$

Cronbach's alpha was used to verify the internal consistency of these indicator variables.

Results and Discussion

Construction of the Vegetable Price Index

Wholesale prices of 19 vegetables were used in the analysis and their basic statistics are given in Table 1.

Table 1: Descriptive statistics of monthly wholesale prices of the selected 19 vegetables from 2005 to 2015

Vegetable	Average	SD	CV	Maximum	Minimum	Range
Green Beans	80.90	37.36	0.462	237.16	28.64	208.52
Carrot	73.36	38.10	0.519	224.79	25.62	199.17
Leeks	56.78	33.36	0.588	238.53	19.16	219.37
Beetroot	53.93	26.58	0.493	169.78	13.35	156.43
Knol-Khol	38.22	17.64	0.462	114.52	15.45	99.07
Radish	20.30	10.80	0.532	59.18	8.32	50.86
Cabbage	36.58	20.09	0.549	120.36	11.13	109.23
Tomatoes	54.55	32.41	0.594	196.46	12.48	183.98
Ladies Fingers	36.13	18.25	0.505	128.25	10.83	117.42
Brinjals	37.61	21.85	0.581	155.00	11.19	143.81
Capsicum	100.82	55.26	0.548	327.75	32.88	294.86
Pumpkin	26.90	13.25	0.493	103.53	9.35	94.18
Cucumber	20.98	10.54	0.503	83.67	9.27	74.40
Bitter Gourd	57.78	27.41	0.474	153.74	20.53	133.21
Snake Gourd	33.30	18.21	0.547	97.59	12.65	84.94
Luffa	44.56	22.44	0.504	140.10	18.31	121.79
Long Beans	45.82	23.70	0.517	155.25	16.62	138.64
Ash Plantains	42.40	13.85	0.327	107.49	22.22	85.27
Green Chilli	**119.61**	**107.71**	**0.901**	**705.61**	**35.53**	**670.08**

A correlation analysis was performed in order to identify the relationships among the vegetable prices and the correlation coefficients are given in Table 2, 3, 4 and 5.

Table 2: correlation coefficients between monthly wholesale prices of the selected variables

Vegetable	Green Beans	Carrot	Leeks	Beetroot	Knol-Khol	Radish
Carrot	0.787 (0.000)					
Leeks	0.646 (0.000)	0.696 (0.000)				
Beetroot	0.669 (0.000)	0.823 (0.000)	0.702 (0.000)			
Knol-Khol	0.813 (0.000)	0.848 (0.000)	0.655 (0.000)	0.839 (0.000)		
Radish	0.844 (0.000)	0.844 (0.000)	0.692 (0.000)	0.800 (0.000)	0.885 (0.000)	
Cabbage	0.744 (0.000)	0.871 (0.000)	0.569 (0.000)	0.758 (0.000)	0.752 (0.000)	0.742 (0.000)

Table 3: Correlation coefficients between monthly wholesale prices of the selected variables

Vegetable	Green Beans	Carrot	Leeks	Beetroot	Knol-Khol	Radish	Cabbage
Tomatoes	0.601 (0.000)	0.639 (0.000)	0.523 (0.000)	0.603 (0.000)	0.697 (0.000)	0.658 (0.000)	0.667 (0.000)
Ladies Fingers	0.857 (0.000)	0.742 (0.000)	0.680 (0.000)	0.726 (0.000)	0.856 (0.000)	0.844 (0.000)	0.648 (0.000)
Brinjals	0.730 (0.000)	0.708 (0.000)	0.567 (0.000)	0.750 (0.000)	0.760 (0.000)	0.766 (0.000)	0.688 (0.000)
Capsicum	0.788 (0.000)	0.868 (0.000)	0.687 (0.000)	0.893 (0.000)	0.908 (0.000)	0.845 (0.000)	0.807 (0.000)
Pumpkin	0.634 (0.000)	0.570 (0.000)	0.436 (0.000)	0.411 (0.000)	0.571 (0.000)	0.546 (0.000)	0.594 (0.000)
Cucumber	0.770 (0.000)	0.644 (0.000)	0.549 (0.000)	0.675 (0.000)	0.766 (0.000)	0.776 (0.000)	0.562 (0.000)
Bitter Gourd	0.862 (0.000)	0.820 (0.000)	0.620 (0.000)	0.767 (0.000)	0.905 (0.000)	0.922 (0.000)	0.750 (0.000)
Snake Gourd	0.846 (0.000)	0.804 (0.000)	0.659 (0.000)	0.762 (0.000)	0.876 (0.000)	0.928 (0.000)	0.718 (0.000)
Luffa	0.797 (0.000)	0.632 (0.000)	0.634 (0.000)	0.618 (0.000)	0.746 (0.000)	0.755 (0.000)	0.584 (0.000)
Long Beans	0.896 (0.000)	0.776 (0.000)	0.656 (0.000)	0.721 (0.000)	0.811 (0.000)	0.895 (0.000)	0.740 (0.000)
Ash Plantains	0.665 (0.000)	0.622 (0.000)	0.592 (0.000)	0.617 (0.000)	0.647 (0.000)	0.607 (0.000)	0.586 (0.000)

Table 4: Correlation coefficients between monthly wholesale prices of the selected variables

Vegetable	Tomato	Ladies Fingers	Brinjals	Capsicum	Pumpkin	Cucumber	Bitter Gourd
Ladies Fingers	0.657 (0.000)						
Brinjals	0.491 (0.000)	0.790 (0.000)					
Capsicum	0.696 (0.000)	0.791 (0.000)	0.766 (0.000)				
Pumpkin	0.633 (0.000)	0.586 (0.000)	0.409 (0.000)	0.555 (0.000)			
Cucumber	0.501 (0.000)	0.862 (0.000)	0.801 (0.000)	0.728 (0.000)	0.444 (0.000)		
Bitter Gourd	0.683 (0.000)	0.841 (0.000)	0.764 (0.000)	0.866 (0.000)	0.656 (0.000)	0.787 (0.000)	
Snake Gourd	0.716 (0.000)	0.898 (0.000)	0.831 (0.000)	0.827 (0.000)	0.640 (0.000)	0.836 (0.000)	0.925 (0.000)
Luffa	0.601 (0.000)	0.851 (0.000)	0.732 (0.000)	0.682 (0.000)	0.563 (0.000)	0.770 (0.000)	0.800 (0.000)
Long Beans	0.602 (0.000)	0.832 (0.000)	0.830 (0.000)	0.818 (0.000)	0.572 (0.000)	0.825 (0.000)	0.897 (0.000)
Ash Plantains	0.515 (0.000)	0.646 (0.000)	0.625 (0.000)	0.678 (0.000)	0.498 (0.000)	0.596 (0.000)	0.665 (0.000)
Green Chilies	0.633 (0.000)	0.701 (0.000)	0.705 (0.000)	0.855 (0.000)	0.482 (0.000)	0.689 (0.000)	0.798 (0.000)

Table 5: Correlation coefficients between monthly wholesale prices of the selected variables

	Snake Gourd	Luffa	Long Beans	Ash Plantain
Luffa	0.833 (0.000)			
Long Beans	0.889 (0.000)	0.773 (0.000)		
Ash Plantains	0.632 (0.000)	0.638 (0.000)	0.722 (0.000)	
Green Chilies	0.748 (0.000)	0.606 (0.000)	0.776 (0.000)	0.569 (0.000)

All the coefficients of correlations between each pair of indicator variables are positive. Some variables such as Snake Gourd and Bitter Gourd are highly correlated($r = 0.925$ $p = 0.000$). When correlation between prices of two vegetables is high, it indicates that vegetables are substitute to each other. Hence, taking all highly correlated variables into the index is useless. Because of that, representative variables were retained after analyzing variable-cluster and correlation analysis. Figure 1 gives the Dendogram resulting from cluster analysis.

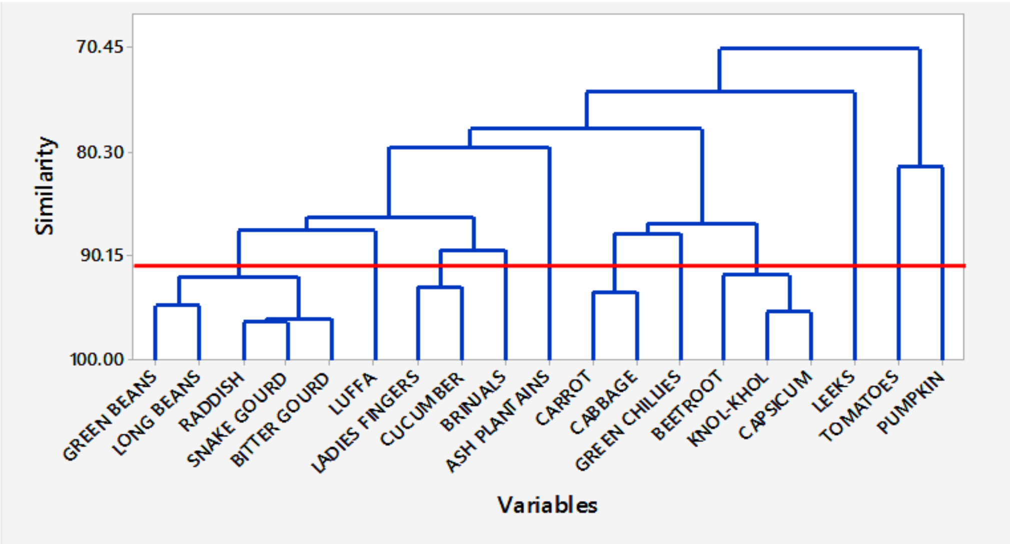

Figure 1: Dendogram of cluster analysis

Eleven clusters at 91.95 Similarity level were retained and representative variables in each cluster was selected according to their availability throughout the country. The cluster analysis was verified by analyzing correlation coefficients among variables within each cluster. The first cluster includes Green Bean, Long Bean, Radish, Snake gourd and Bitter gourd. Table 5 gives the correlation coefficients between the variables in the first cluster.

Table 5: Correlation coefficients of the variables in the first cluster

	Green Bean	Long Bean	Radish	Snake gourd
Long Beans	0.896 (0.000)			
Radish	0.844 (0.000)	0.895 (0.000)		
Snake gourd	0.846 (0.000)	0.889 (0.000)	0.928 (0.000)	
Bitter gourd	0.862 (0.000)	0.897 (0.000)	0.922 (0.000)	0.925 (0.000)

Green Bean was selected as the representative variable for the first cluster, because of its availability throughout the country. Luffa was the only variable in the second cluster. Third cluster consists of Cucumber and Ladies Fingers and their correlation coefficient was 0.862. Cucumber was selected as the representative variable for the third cluster according to their availability. Brinjals was the only variable in the fourth cluster while Ash Plantains was the only variable in the fifth cluster. Carrot and Cabbage existed in sixth cluster and their correlation coefficient was 0.871 (p=0.000). Cabbage was selected as the representative variable for the sixth cluster.

Green chili was the only variable in the cluster seven. Variables belong to eighth cluster are beetroot, Knol-Khol and Capsicum. Capsicum was selected as the representative variable for the eighth cluster. Table 6 shows the correlation coefficients for the variables in the eighth cluster.

Table 6: Correlation coefficients for the variables in the eighth cluster

	Beetroot	Knol-Khol
Knol-Khol	0.839 (0.000)	
Capsicum	0.893 (0.000)	0.908 (0.000)

Leeks, tomato, and pumpkin were selected from ninth, tenth and eleventh clusters respectively. The correlation between Capsicum and Green Chili were 0.855 (p=0.000). Therefore, Capsicum was removed from the index. The correlation coefficients among ten variables retained are given in Table 7.

Table 7: Correlation coefficients of the retained variables

	Green Beans	Leeks	Cabbage	Tomatoes	Brinjals	Pumpkin	Cucumber	Luffa	Ash Plantains
Leeks	0.65 (0.00)								
Cabbage	0.74 (0.00)	0.57 (0.00)							
Tomatoes	0.60 (0.00)	0.52 (0.00)	0.67 (0.00)						
Brinjals	0.73 (0.00)	0.57 (0.00)	0.69 (0.00)	0.49 (0.00)					
Pumpkin	0.63 (0.00)	0.44 (0.00)	0.59 (0.00)	0.63 (0.00)	0.41 (0.00)				
Cucumber	0.77 (0.00)	0.55 (0.00)	0.56 (0.00)	0.50 (0.00)	0.80 (0.00)	0.44 (0.00)			
Luffa	0.80 (0.00)	0.63 (0.00)	0.58 (0.00)	0.60 (0.00)	0.73 (0.00)	0.56 (0.00)	0.77 (0.00)		
Ash Plantains	0.67 (0.00)	0.59 (0.00)	0.59 (0.00)	0.52 (0.00)	0.63 (0.00)	0.50 (0.00)	0.60 (0.00)	0.64 (0.00)	
Green Chili	0.73 (0.00)	0.49 (0.00)	0.76 (0.00)	0.63 (0.00)	0.71 (0.00)	0.48 (0.00)	0.69 (0.00)	0.61 (0.00)	0.57 (0.00)

Maximum correlation coefficient was 0.801(p=0.000) which was between Brinjal and Cucumber.

Scaling Variable

The original variables were divided by their means to make the new variables having unit mean (scaled) as given in *Equation 1*. Here, X_i^* is a new variable. Hence, variance of X_i can be partitioned as in *Equation 8*.

$$\mathbf{Var}(\boldsymbol{X_i}) = \mathbf{Var}\ (\boldsymbol{X_i^*}) \times \mathbf{Var}\ (\boldsymbol{\mu_i}) \qquad [8]$$

As, μ_i, is a constant contribution of μ_i to variance of i^{th} original variable is μ_i^2

$$\mathbf{Var}(\boldsymbol{X_i}) = \mathbf{Var}\ (\boldsymbol{X_i^*}) \times \boldsymbol{\mu_i^2} \qquad [9]$$

Table 8 gives the partitioned variances of the original variables.

Table 8: Partitioned variance of original variables

Variable	Var (X_i^*)	Mean (μ_i)	μ_i^2	$\mathrm{Var}(X_i) = \mathrm{Var}\ (X_i^*) \times \mu_i^2$
Green		80.90		1395.587
Beans	0.213		6545.15	
Leeks	0.345	56.78	3224.21	1113.172
Cabbage	0.302	36.58	1338.21	403.555
Tomatoes	0.353	54.55	2976.12	1050.452
Brinjals	0.337	37.61	1414.48	477.252
Pumpkin	0.243	26.90	723.58	175.639
Cucumber	0.253	20.98	440.17	111.154
Luffa	0.254	44.56	1985.47	503.336
Ash		42.40		191.912
Plantains	0.107		1798.12	
Green		119.61		11602.340
Chilli	0.811		14305.77	

Variance of Variables having large means is obviously high because of effect of variance of mean. Hence, original variable should be divided by its mean to remove variance of mean. Before scaling Green Chilli has the highest variance while cucumber has the lowest variance. However, after scaling green Chilli has the highest variance while Ash Plantains has the lowest variance. Table 9 shows the covariance matrix of the scaled variables.

Table 9: Covariance matrix of the scaled variables

	Green Beans	Leeks	Cabbage	Tomatoes	Brinjals	Pumpkin	Cucumber	Luffa	Ash Plantains	Green Chili
Green Beans	0.213									
Leeks	0.175	0.345								
Cabbage	0.189	0.184	0.302							
Tomatoes	0.165	0.182	0.218	0.353						
Brinjals	0.196	0.194	0.220	0.169	0.337					
Pumpkin	0.144	0.126	0.161	0.185	0.117	0.243				
Cucumber	0.179	0.162	0.155	0.150	0.234	0.110	0.253			
Luffa	0.185	0.188	0.161	0.180	0.214	0.140	0.195	0.254		
Ash Plantains	0.100	0.114	0.105	0.100	0.119	0.080	0.098	0.105	0.107	
Green Chili	0.305	0.261	0.376	0.339	0.369	0.214	0.312	0.275	0.167	0.811

Due to the removal of the effects of the mean, all variables have unit mean and their variances are comparable. Nevertheless, scaling in Standard Deviation makes all variables with unit variance and variances of variables are not comparable. Standard deviations of Mean-Scaled variables are equal to coefficients of variance (CV) of the original variables. Correlation matrix of prices of selected vegetables before and after scaling remains the same.

Weights

Principal Component based Factor Analysis was carried out on the indicator variables (vegetables) in order to identify the underlying sub indicators (factors) of the composite index. Figure 2 shows the Scree plot of the PCA.

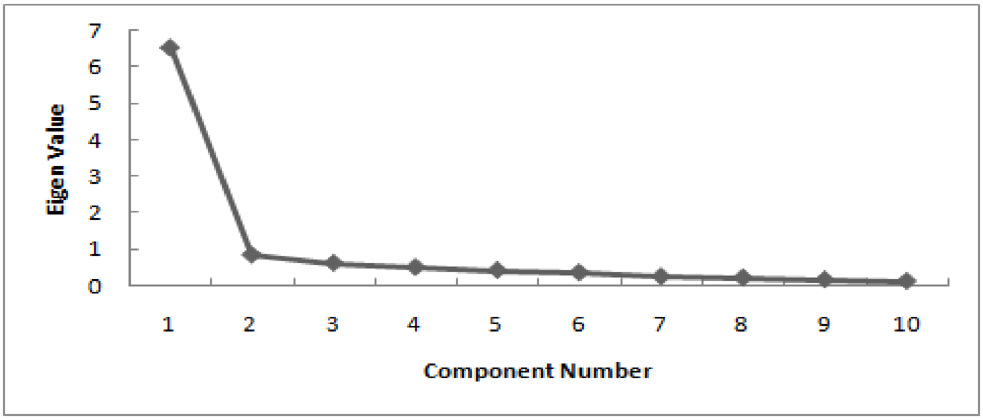

Figure 2: Scree plot of PCA

Only the first PC having Eigen value over one was retained and all variables in the first PC had positive scores. Principal Component based Factor Analysis with correlation option explained 65.6% of total variance. Table 10 gives Principal Component Analysis of the Prices of Selected Vegetables. Since the first PC accounts for the largest amount of variability in the data only the first PC is retained and based on first PC, weights can be derived (Fernando *et al.,* 2012, De Silva *et al.,* 2000).

Table 10: Principal component analysis of the prices of selected vegetables

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
Green Beans	0.36	0.05	-0.01	0.16	-0.01	0.31	-0.33	-0.44	-0.39	-0.55
Leeks	0.29	0.06	-0.59	-0.53	-0.4	0.23	-0.13	0.24	0.08	0.01
Cabbage	0.33	-0.21	0.34	-0.37	0.10	0.37	0.33	-0.26	-0.25	0.47
Tomatoes	0.30	-0.49	0.09	-0.11	-0.32	-0.67	0.14	0.01	-0.22	-0.19
Brinjals	0.33	0.39	0.20	0.01	0.02	0.07	0.63	0.26	0.20	-0.44
Pumpkin	0.27	-0.60	-0.16	0.48	0.15	0.34	0.03	0.37	0.18	0.01
Cucumber	0.33	0.40	0.09	0.34	-0.14	-0.12	-0.18	0.41	-0.47	0.39
Luffa	0.34	0.16	-0.24	0.37	-0.23	-0.16	0.13	-0.54	0.43	0.31
Ash Plantains	0.30	0.08	-0.35	-0.18	0.80	-0.33	-0.05	-0.02	-0.04	0.03
Green Chili	0.32	0.02	0.54	-0.20	0.03	-0.05	-0.55	0.11	0.51	0.00
Eigen value	**6.56**	**0.84**	**0.61**	**0.50**	**0.42**	**0.35**	**0.24**	**0.21**	**0.16**	**0.11**

The equation to calculate weights was derived from PCA general equation. PC is linear combination of all variables taken to analysis (Chatfield and Collins, 1980).

$$Y_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \qquad [10]$$

$$= a_1^T X \qquad [11]$$

As Eigen vector coefficients are orthogonal,

$$a_1^T a_1 = 1$$

The amount of variance of X_i explained by the first PC = a_{i1}^2

As the Eigen vector coefficient represents the proportion of variation contributed by a given variable to a PC, the weights were calculated by squaring Eigen vector coefficient of a given variable of the PC as given in Equation 12.

$$w_{pi} = \frac{a_{p1i}^2}{\sum_{p=1}^{ni} a_{p1i}^2} = a_{p1i}^2, \quad \text{Since} \sum_{p=1}^{ni} a_{p1i}^2 = 1, \qquad [12]$$

Where,

w_{pi} - weight correspond to p^{th} variable in i^{th} subset
a_{p1i} - Eigen vector coefficient of p^{th} variable of the first PC of i^{th} subset
i - 1, 2, …, m (number of subsets)
p - 1, 2, …, ni (number of variables in i^{th} subset)

The weights derived from the Principle Component Coefficients are given in Table 11.

Table 11: The weights derived from the principal component coefficients

Variable	Principal Components Coefficient (a_{p1})	Weight ($w_{pi} = a_{p1i}^2$)
Green Beans	0.355	0.126
Leeks	0.289	0.084
Cabbage	0.327	0.107
Tomatoes	0.296	0.088
Brinjals	0.328	0.108
Pumpkin	0.273	0.075
Cucumber	0.326	0.106
Luffa	0.336	0.113
Ash Plantains	0.304	0.092
Green Chili	0.321	0.103

Then, the scaled variables were weighted according to the Equation 13 and the transformed variables were denoted by X_{iw}.

$$\boldsymbol{X_{iw}} = X_i^* w_i \qquad [13]$$

Where,

$\boldsymbol{X_{iw}}$= weighted i^{th} variable
w_i= weight corresponds to i^{th} variable
$\boldsymbol{X_i^*}$= Scaled i^{th} variable

Factor Analysis

PCs-based Factor Analysis with covariance matrix option was performed on the scaled and weighted variables in order to get factor score coefficients (Table 12). Here, all the variables have the same unit, which is Rs/kg. Hence, PFA with covariance option can be used. However, their variances are very different because of the effects of different means. If variance of a variable is very large, its covariance becomes high though their correlations are low. Hence, factor analysis collects high proportion of variances from such variables while collecting low amount of variances form other variables. Because of that, a variable having comparatively high variance gets considerably high Factor Score Coefficients.

Preliminary Factor Analysis with covariance option extracts 82.9% of the total variance. As Green Chili has the highest variance it gets such a big Factor Score Coefficient like 0.814 while the others get low Factor Score Coefficients. Hence, the contribution of Green Chili to the final index is 69.28%. It means that the index is very sensitive to Green Chili, which has highest variance. Hence preliminary factor analysis cannot be taken to construct the index.

Table 12: Factor analysis with different options performed on the same monthly wholesale prices of selected vegetables

Method	PFA¹		PFA¹		PFA ²		WFA²		WFA²	
	Var. %	FSC	Var. %	FSC	Var. %	FSC	Var. %	FSC	Var. %	FSC
Green Beans	65.63	0.08	88.81	0.14	78.79	0.09	84.72	0.208	79.86	0.13
Leeks	33.86	0.05	54.61	0.11	50.39	0.11	51.61	0.07	48.00	0.07
Cabbage	65.48	0.02	69.89	0.12	72.95	0.12	69.28	0.14	72.51	0.13
Tomatoes	48.39	0.05	57.15	0.12	58.36	0.13	52.41	0.08	53.92	0.09
Brinjals	57.49	0.03	70.90	0.13	70.58	0.13	73.82	0.14	71.93	0.15
Pumpkin	30.12	0.01	48.16	0.11	43.51	0.07	42.71	0.05	41.47	0.04
Cucumber	55.05	0.01	69.39	0.13	67.32	0.09	72.6	0.14	71.00	0.10
Luffa	47.60	0.03	73.96	0.13	67.45	0.09	75.35	0.16	68.37	0.12
Ash Plantains	39.98	0.09	60.06	0.12	52.18	0.04	58.13	0.09	53.10	0.03
Green Chili	98.73	0.81	68.89	0.13	80.90	0.34	58.08	0.13	81.94	0.34
% Variance	82.9		65.6		67.3		68.8		69.8	

Note: ¹ Covariance Matrix; ² Mean-Scaled with Covariance Matrix

PFA with correlation option extracts 65.6% of the total variance. Though the total variance explained by PFA with correlation option is low, this option can overcome the limitation of sensitivity to comparatively higher variance of a variable. However, it creates another limitation that as using correlation matrix for analysis of variance of each variable becomes one and therefore variances are not comparable and factor structure is determined by correlations among variables. Hence, analysis is not able to identify the true factor structure. Irrespective of original variances of variables, only correlations determine factor score coefficients. Because of that the all variables get more or less similar factor score coefficients.

According to the Weighted Factor Analysis method introduced by Fernando *et al.,* (2012) to improve Preliminary Factor Analysis with correlation option total variance explained by model increased from 65.6% to 68.8%. Nevertheless, the limitation of Preliminary Factor Analysis with correlation option still exists.

Scaling in mean can remove the effect of means and variances of all variables become comparable. Scaling can improve factor analysis and after scaling the original variables in mean explained 67.3% of total variance, which is greater than Preliminary

Factor Analysis with correlation option. Hence, Scaling in mean is a good option for different scaled-variables and variables in the same unit but having comparatively different variances.

After scaling the original variables in mean the weights were assigned using with *Equation 3*. Weighted Factor Analysis improved and explained 69.8% of total variances which is the second highest variance explained by a factor analysis for the data. This method takes true variances of variables and their relations with other variables into account and is not sensitive to variables having comparatively higher variances because of their means. Hence, this method is the ideal method for constructing the index. Variables having high true variance and high correlations with other variables get high factor score coefficients and vice versa irrespective of their original variances affected by mean.

Reliability of the Variables

The value of Cronbach's alpha of original indicator variables of the study is 0.8320 and recommended values are 0.7 or higher (>0.70) (Nunnally, 1978). This Implies that consistency of the indicator variables is at a high level. The value of Cronbach's alpha for scaled indicator variables was equal to 0.9310, which implies an enhancement of internal consistency of the original indicator variables by scaling. The value of Cronbach's alpha for scaled and weighted indicator variables was 0.9317. It implies that the internal consistency of the indicator variables was further improved after weighting.

Composite Vegetable Price Index

Composite vegetable price index can be summarized with Equation.14.

$$\boldsymbol{CVPI} = \sum\left[\left(\frac{\boldsymbol{Xi}}{\mu_{\boldsymbol{i}}}\right) \times \boldsymbol{w_i} \times \boldsymbol{FSC_i}\right] \qquad [14]$$

Where,

CVPI - Composite Vegetable Price Index
X_i - Price of i^{th} variable
μ_i - Mean of i^{th} variable
w_i - Weight of i^{th} variable
FSC_i - Factor Score Coefficients of i^{th} variable

CVPI = *(((Green-Beans/80.90)*0.126*0.132) + ((Leeks/56.78)*0.084*0.072) + ((Cabbage/36.58)*0.107*0.127) + ((Tomato/54.55)*0.088*0.091) + ((Brinjals/37.61)*0.108*0.145) + ((Pumpkin/26.90)*0.075*0.038) +((Cucumber/20.98)*0.106*0.103) + ((Luffa/44.56)*0.113*0.116) + ((Ash-Plantains/42.40)*0.092*0.028) + ((Green-Chilies/119.61)*0.0.103*0.338))*

Since weight, mean and factor score coefficient are constant for a given variable, C_i summarizes the constant terms of i^{th} variable.

$$C_i = \frac{w_i \times FSC_i}{\mu_i} \qquad [15]$$

CVPI = *((Green-Beans*0.000206) + (Leeks*0.000106) + (Cabbage*0.000371) + (Tomato*0.000146) + (Brinjals*0.000415) + (Pumpkin*0.000105) + (Cucumber*0.000522) + (Luffa*0.000294 + (Ash-Plantains*0.00006) + (Green-Chilies*0.0002912))*

As C_i values are very small, C_i values were reweighted to equal the summation of W_{Ci} to one. W_{Ci} defines the final weight for i^{th} variable.

$$W_{Ci} = \frac{C_i}{\sum_{i=n} C_i} \qquad [16]$$

Final weights (W_{Ci}) for each vegetable are given in Table 13

Table 13: Final weights (W_{Ci}) of vegetables

Variable	Weight(W_{Ci})
Green Beans	0.0817
Leeks	0.0421
Cabbage	0.1475
Tomatoes	0.0581
Brinjals	0.1648
Pumpkin	0.0418
Cucumber	0.2073
Luffa	0.1168
Ash Plantains	0.0242
Green Chili	0.1157

Then the composite vegetable price index can be presented as Equation 17.

$$CVPI = \sum_{i=1}^{n} (X_i \times W_{Ci}) \qquad [17]$$

CVPI = *((Green-Beans*0.0817) + (Leeks*0.0421) + (Cabbage*0.1475) + (Tomato*0.0581) + (Brinjals*0.1648) + (Pumpkin*0.0418) + (Cucumber*0.2073) + (Luffa*0.1168) + (Ash-Plantains*0.0242) + (Green-Chilies*0.1157))*

The Composite Vegetable Price Index values for 11 years are given in Table 14.

Table 14: The composite vegetable price indexes for 11 years from 2005 to 2015

Month	Index	Month	Index	Month	Index	Month	Index
2005/01	30.40	2007/10	29.40	2010/07	50.79	2013/04	40.59
2005/02	32.64	2007/11	32.81	2010/08	47.09	2013/05	51.24
2005/03	23.71	2007/12	28.74	2010/09	43.18	2013/06	55.35
2005/04	23.64	2008/01	35.86	2010/10	56.51	2013/07	45.53
2005/05	26.17	2008/02	34.16	2010/11	40.53	2013/08	46.34
2005/06	28.53	2008/03	43.71	2010/12	57.63	2013/09	38.29
2005/07	29.62	2008/04	55.80	2011/01	93.24	2013/10	38.58
2005/08	27.05	2008/05	74.07	2011/02	119.29	2013/11	39.03
2005/09	27.70	2008/06	67.14	2011/03	79.18	2013/12	36.61
2005/10	30.38	2008/07	33.74	2011/04	62.23	2014/01	37.08
2005/11	33.27	2008/08	29.46	2011/05	48.89	2014/02	41.08
2005/12	33.74	2008/09	30.14	2011/06	42.39	2014/03	36.68
2006/01	27.19	2008/10	32.37	2011/07	33.05	2014/04	45.54
2006/02	23.67	2008/11	45.66	2011/08	33.12	2014/05	70.22
2006/03	19.78	2008/12	56.33	2011/09	40.49	2014/06	96.57
2006/04	21.04	2009/01	46.00	2011/10	40.74	2014/07	101.91
2006/05	28.87	2009/02	38.82	2011/11	74.19	2014/08	57.99
2006/06	34.66	2009/03	33.11	2011/12	72.79	2014/09	46.52
2006/07	28.81	2009/04	34.12	2012/01	54.01	2014/10	61.03
2006/08	25.09	2009/05	35.83	2012/02	29.11	2014/11	64.47
2006/09	26.66	2009/06	51.62	2012/03	26.16	2014/12	116.13
2006/10	27.42	2009/07	40.90	2012/04	37.58	2015/01	181.36
2006/11	46.71	2009/08	32.38	2012/05	55.62	2015/02	107.40
2006/12	53.13	2009/09	32.14	2012/06	76.56	2015/03	49.50
2007/01	46.26	2009/10	38.21	2012/07	56.31	2015/04	37.40
2007/02	31.81	2009/11	50.09	2012/08	45.70	2015/05	50.75
2007/03	24.71	2009/12	64.49	2012/09	41.22	2015/06	72.28
2007/04	23.20	2010/01	54.45	2012/10	52.80	2015/07	81.65
2007/05	26.58	2010/02	44.30	2012/11	82.43	2015/08	55.47
2007/06	34.51	2010/03	41.55	2012/12	63.55	2015/09	53.18
2007/07	32.86	2010/04	32.27	2013/01	88.57	2015/10	67.55
2007/08	27.86	2010/05	48.69	2013/02	72.51	2015/11	134.36
2007/09	29.41	2010/06	64.63	2013/03	59.46	2015/12	173.75

Conclusions

In this study, all the variables had the same unit, which is Rs/kg. Hence, PFA with covariance option could be used. However, when their variances are very different because of the effects of different means and variance of a variable is very large, its covariance becomes high though their correlations are low. Hence, factor analysis collects high proportion of variances from such variables while collecting low amount of variances form other variables. Because of that, a variable having comparatively high variance gets considerably high Factor Score Coefficients.

Although, the total variance explained by PFA with correlation option is low, this option can overcome the limitation of sensitivity to comparatively higher variance of a variable. Nevertheless, it creates another limitation that as using correlation matrix for analysis variance of each variable becomes one and therefore variances are not comparable and factor structure is determined by correlations among variables. Hence, analysis is not able to identify the true factor structure. Irrespective of original variances of variables, only correlations determine factor score coefficients. Because of that the all variables get more or less similar factor score coefficients.

A new scaling method introduced in this study consists of scaling by dividing by the mean and using weighted values. The new scaling method removes the effect of mean on variance and makes variances of all variables comparable. Scaling can improve factor analysis and after scaling the original variables in mean explained 67.3% of total variance, which is greater than Preliminary Factor Analysis with correlation option. Hence, Scaling in mean is a good option for different scaled-variables and variables in the same unit but having comparatively different variances.

The weights were derived by using squared Eigen vector coefficient of a given variable of the first Principal Component of the subset. Hence, the weights were proportionate to the variation contributed by a given variable to the Principle Component. Weighted Factor Analysis improved and explained 69.8% of total variances which is the second highest variance explained by a factor analysis for the data. This method takes true variances of variables and their relations with other variables into account and is not sensitive to variables having comparatively higher variances because of their means. Variables having high true variance and high correlations with other variables get high factor score coefficients and vice versa irrespective of their original variances affected by mean. Scaling in mean and weighting improved internal consistency of the variables.

References

Chatfield, C. and Collins, A.J. (1980). Introduction to Multivariate Analysis. Chapman and Hall, London, UK.

De Silva A.P.G.S., Thattil R.O., and Samitha S. (2000) Weighted Principal Component Analysis Approach to Construct an Index for Human Development, *Sri Lanka Journal of Applied Statistics* 1(1), 13-24

Fernando, M.A.S.S., Samita, S. Abeynayake R. (2012). Modified Factor Analysis to Construct Composite Indices: Illustration on Urbanization Index, *Tropical Agricultural Research*, 23 (4): 327– 337.

Garcia del V. I. T. and Puetra G. C. (1997). Human development: weighted principal components analysis applied to the data of 1997 report. Joint IASS/IAOS Conference Proceedings, Mexico.

Nardo M., Saisana M., Saltelli A. and Tarantola S. (2005). Tools for Composite Indicators Building, Joint Research Centre - European Commission-Directorate General.

Nunnally, J.C. (1978). Psychometric theory (2nd Ed.). New York: McGraw-Hill.