



OPEN ACCESS

EDITED BY
Bo Wang,
University of Leicester, United Kingdom

REVIEWED BY
Kurdistan Omar,
University of Zakho, Iraq
Heba M. Ezzat,
Cairo University, Egypt

*CORRESPONDENCE
Raju Kumar
✉ raju.kumar@icar.gov.in

RECEIVED 08 August 2023
ACCEPTED 29 September 2023
PUBLISHED 17 November 2023

CITATION

Singh D, Basak P, Kumar R and Ahmad T (2023)
On the methodological framework of
composite index under complex surveys and its
application in development of food
consumption index for India.
Front. Appl. Math. Stat. 9:1274530.
doi: 10.3389/fams.2023.1274530

COPYRIGHT

© 2023 Singh, Basak, Kumar and Ahmad. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

On the methodological framework of composite index under complex surveys and its application in development of food consumption index for India

Deepak Singh¹, Pradip Basak², Raju Kumar^{1*} and
Tauqueer Ahmad¹

¹Division of Sample Surveys, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India,

²Department of Agricultural Statistics, Uttar Banga Krishi Viswavidyalaya, Cooch Behar, West Bengal, India

Indices are created by consolidating multidimensional data into a single representative measure known as an index, using a fundamental mathematical model. Most present indices are essentially the averages or weighted averages of the variables under study, ignoring multicollinearity among the variables, with the exception of the existing Ordinary Least Squares (OLS) estimator based OLS-PCA index methodology. Many existing surveys adopt survey designs that incorporate survey weights, aiming to obtain a representative sample of the population while minimizing costs. Survey weights play a crucial role in addressing the unequal probabilities of selection inherent in complex survey designs, ensuring accurate and representative estimates of population parameters. However, the existing OLS-PCA based index methodology is designed for simple random sampling and is incapable of incorporating survey weights, leading to biased estimates and erroneous rankings that can result in flawed inferences and conclusions for survey data. To address this limitation, we propose a novel Survey Weighted PCA (SW-PCA) based Index methodology, tailored for survey-weighted data. SW-PCA incorporates survey weights, facilitating the development of unbiased and efficient composite indices, improving the quality and validity of survey-based research. Simulation studies demonstrate that the SW-PCA based index outperforms the OLS-PCA based index that neglects survey weights, indicating its higher efficiency. To validate the methodology, we applied it to a Household Consumer Expenditure Survey (HCES), NSS 68th Round survey data to construct a Food Consumption Index for different states of India. The result was significant improvements in state rankings when survey weights were considered. In conclusion, this study highlights the crucial importance of incorporating survey weights in index construction from complex survey data. The SW-PCA based Index provides a valuable solution, enhancing the accuracy and reliability of survey-based research, ultimately contributing to more informed decision-making.

KEYWORDS

multicollinearity, survey weight, food consumption index, principal component, index, household consumer expenditure survey (NSS 68th Round)

1 Introduction

With the rapid growth of large-scale multivariate data generation, which is both complex and expansive, extracting meaningful insights becomes a challenging task due to the voluminous respondents and intricate relationships among variables. Therefore, it is imperative to develop statistical methodologies for summarizing such data to uncover underlying patterns, trends, and contextual factors that facilitate informed decision-making. One widely adopted approach for summarizing and comparing performance across subjects is the use of composite indicators or indices. Composite indicators or indices are mathematical or computational measures designed to effectively quantify multidimensional concepts that cannot be adequately captured by a single indicator alone e.g., industrialization, poverty, Human development, etc. A composite indicator is created by aggregating individual indicators into a single index measure based on a fundamental model and is used for ranking subjects. Composite indicators or indices have proven valuable in various domains of economic, social, and environmental statistics for summarizing and presenting information.

Most of the existing indices reported are either weighted or equally weighted averages of the variables. Wiesmann [1] reported the Global Hunger Index series using the average of three variables, the share of the population with insufficient access to food (provided by FAO); the fraction of the population of children under 5 who are underweight (provided by WHO); and the mortality rates of children under-5 (provided by UNICEF). Athreya et al. [2] reported the food insecurity in rural India by developing the Food Insecurity Index using the weighted average of sub-indices. Gentilini and Webb [3] proposed a multidimensional index of poverty and hunger. The Poverty and Hunger Index uses the five indices which are scaled against the maximum values of each indicator and added up using equal weights as in the construction of the human development index. Hastings [4] developed the Human Security Index (HIS) which is the average of the Social Fabric Index and the basic Human Development Index. Athreya et al. [5] reported the food insecurity in urban India using the same methodology as used by Athreya et al. [2]. Krishnamurthy et al. [6] developed the Vulnerability Index using the three indices which are scaled against the maximum values of each indicator and added up using equal weights. The Baseline Climate and Food Insecurity Index uses three indices which are scaled against the maximum values of each indicator and added up using equal weights [7]. Global Food Security Index [8] used the five indices by using two sets of weightings. One equal weight and the second available option, known as peer panel recommendation, averages the weightings suggested by five members of an expert panel. Every year, the United Nations Development Programme publishes the Human Development Index (HDI) in annually published Human Development Report. The HDI is the geometric mean of the three dimension indices i.e., Health, Education and Income [9]. Since 2019, the United Nations Development Programme is annually publishing the Global Multidimensional Poverty Index (MPI) in the Multidimensional Poverty Index Report [2023 Global Multidimensional Poverty Index (MPI)]. Alkire et al. [10] MPI uses information from ten indicators that are organized into three dimensions: health, education, and living standards. The index is

constructed based on a weighted average using a nested weight structure: equal weight across dimensions and equal weight for each indicator within dimensions. Niti Ayog (Govt. of India) annually reported the Health Index in four rounds from 2014–15 to 2019–20 based on the weighted average of normalized variables in the report “Healthy States, Progressive India: Report on the Ranks of States and Union territories”¹. The Global Hunger index, [11] report published by Concern Worldwide and Welthungerhilfe (Publishes annual publication on the Global Hunger Index) used the standardized scores of undernourishment, child mortality, child stunting, and child wasting to calculate the GHI score for each country. Undernourishment and child mortality each contribute one-third of the GHI score, while child stunting and child wasting each contribute one-sixth of the score in the Global Hunger 9.

Narain et al. [12] developed a new statistical methodology for the construction of index based on normalized indicators. Narain et al. [13, 14] have discussed regional dimensions of disparities in crop productivity in Uttar Pradesh and Madhya Pradesh states of India in the years 2001 and 2002, respectively using the Narain et al. [12] method of construction of the composite index. Narain et al. [15, 16] evaluated economic development in Karnataka and hilly states of India respectively using Narain et al. [12] method of construction of composite index. Narain et al. [17, 18] have estimated socioeconomic development of different districts in Kerala and different states of India respectively using the Narain et al. [12] method of construction of composite index. Rai et al. [19] developed the Livelihood Index for different agro-climatic zones of India based on the statistical background suggested by Narain et al. [12].

However, the challenge of multicollinearity persists in constructing indices from multivariate data, where the combination of weights for correlated variables often leads to the development of poorly constructed composite indices. To address this challenge, the Principal Component Analysis (PCA) based index method is effective. The unique property of PCA in removing multicollinearity has prompted numerous authors to utilize it for the development of indices such as Dahal [20] who developed a Soil Quality Index using principal components analysis. Kumar [21] created an Agricultural Development Index using the principal component technique. Raja and Tawheed [22] evaluated ten districts of Kashmir Valley concerning nine development indicators based on principal component-based Index. Chao and Wu [23] utilized the medical expenditure panel survey to develop a principal component -based index. Senna et al. [24] developed a Water Poverty Index using principal component analysis. Zhou et al. [25] used principal component analysis to develop a Green Finance Development Index. Lieberman-Cribbin et al. [26] developed SES index using principal component analysis on the variables; household income, gross rent, poverty, education, working class status, unemployment, and occupants per room. Mohammed et al. [27] developed the Resilience Index for municipal infrastructure using principal components analysis (PCA) for decision-makers regarding pavement network maintenance planning.

1 Healthy States, Progressive India: Report on the ranks of states and union Territories. Report (2019–20) by Niti Ayog, Government of India.

2 Some existing indices and their limitations

2.1 Multiple factor analysis

It does not serve the purpose to arrive at a meaningful and comparable composite index of development when the indicators are presented in different scale of measurements.

2.2 Aggregation method

The method is not suitable as the composite index of development obtained by use of the method depends on the unit in which the data are recorded.

2.3 Monetary index

Monetary values of developmental indicators may change from place to place and from time to time. All the indicators cannot be converted into monetary values like “death rate,” “birth rate,” “sex ratio,” “literacy rate” etc. cannot be converted into monetary values.

2.4 Ratio index

Developmental Indicators are transformed as ratio in the following manner

$$I_{ij} = (X_{ij} - X_{i \min}) / (X_{i \max} - X_{i \min})$$

The method uses range value in the denominator, which is based on only two observations. Other information is not utilized in this method.

2.5 Narain's index method

Narain et al. [12] method does not consider the effect of multicollinearity which is present among correlated variables and does not incorporate survey weight.

2.6 Principal component analysis

This method uses the estimator $\hat{\Sigma}_{yy}$ which is the function of the superpopulation covariance matrix Σ_{yy} and is not suitable if the data is collected through complex survey design.

The existing literature shows that methodologies for index construction are based on the assumption that the data is collected through simple random sampling and no method of index construction has been reported that is capable of incorporating survey weights available in data

collected through surveys with planned sampling designs (complex survey designs), such as Household, Employment, Education, Income, Expenditure, Agricultural, Housing, Poverty alleviation and living conditions, Literacy, Enterprise, and Skill measurement surveys. This limitation results in biased estimates and rankings based on the existing methods of index construction for complex survey data which is collected through complex survey design. Notably, widely recognized large-scale surveys have complex survey designs involving stratification, unequal probabilities of selection, clustering, multi-stages and multi-phases like Demographic and Health Surveys (DHS), European Social Surveys, National Sample Surveys (NSS), National Family Health Surveys (NFHS), National Surveys of Children's Health (NSCH), and National Nutrition Surveys (NNS) exemplify real-life instances where this issue persists.

The existing PCA-based index methodology for survey data is based on the Ordinary Least Squares (OLS) estimator of variance covariance matrix which assumes that the data are collected using simple random sampling but most survey data are collected through complex survey design where the survey weight is attached with each sampled unit to make the sample representative of the population. The existing PCA (OLS-based approach) based index methodology for survey data is incapable of incorporating the survey weights in the calculation of indexes.

Ignoring survey weights in the development of estimates of a parameter in survey research can lead to biased and unreliable results. Survey weights are applied to adjust for the unequal probabilities of selection that occur in complex survey designs. These weights are essential for producing accurate and representative estimates of population parameters. Survey weights ensure that the sample accurately represents the target population and is designed to make the most efficient use of the data collected. Without survey weights, some segments of the population may be overrepresented or underrepresented in the sample. Survey weights help balance the sample to reflect the population's distribution, reducing this bias. Ignoring survey weights in survey research can compromise the quality and validity of the estimates, leading to biased, inefficient, and unrepresentative results. It's essential to use appropriate weighting techniques when analyzing survey data to ensure that the findings accurately reflect the target population.

Therefore, in this research study, we have proposed a new Survey Weighted PCA (SW-PCA based approach) based Index methodology for disaggregated survey-weighted data which produces the unbiased estimate of variance-covariance matrix, thereby producing unbiased eigenvalues, unbiased eigenvectors, resulting in the development of unbiased and efficient indices. By incorporating the survey weights and addressing the issue of multicollinearity among variables, we aim to overcome these challenges and provide a robust framework for constructing composite indices that accurately represent the underlying survey data.

3 Material and methods

3.1 Existing PCA (population variance covariance, PVC-PCA based approach) based index methodology for disaggregated population data

Consider a survey conducted where the population is distributed across various subpopulations, such as states within a country and unit level can be, e.g., individuals or households or farms etc. The observational recordings are taken from each individual or unit of the whole population. The researcher aims to compare the performance of these subpopulations. To facilitate this comparison, it is necessary to establish an index or measure that enables a meaningful assessment of subpopulation performance. For this, let us consider a finite population $U = (1, 2, \dots, k, \dots, N)$ of size N units having l subpopulations such that the h^{th} subpopulation N_h units and $\sum_{h=1}^l N_h = N$, ($h = 1, 2, \dots, l$) and $\mathbf{y} = (y_1, y_2, \dots, y_p)'$ be the set of p standardized variables which are used for index development. Let us suppose that the non-zero eigenvalues of the variance-covariance matrix Σ_{yy} of the variables under study for index development are $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_p$ and the corresponding eigenvectors are $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_p$. The Population variance covariance matrix for disaggregated population data is expressed as

$$\Sigma_{yy} = (N - 1)^{-1} \sum_h \sum_i (y_{hi} - \bar{Y}) (y_{hi} - \bar{Y})' \tag{1}$$

where, $\bar{Y} = \sum_h \sum_i y_{hi} / N$.

Let the population variance-covariance matrix Σ_{yy} is a real positive definite matrix. Let us suppose that the non-zero eigenvalues of are $\lambda_1 > \lambda_2 > \lambda_3 \dots > \lambda_p$ and the corresponding eigenvectors are $\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_p$. For distinct λ_j 's ($j = 1, 2, 3, \dots, p$), a $(p \times p)$ orthogonal matrix can be formed, where

$$\Gamma = [\gamma_1, \gamma_2, \gamma_3, \dots, \gamma_p]. \tag{2}$$

Γ matrix diagonalizes Σ_{yy} matrix such that

$$\Sigma_{yy} = \Gamma A \Gamma' \tag{3}$$

where $A = \text{diag}(\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p) = \Gamma \Sigma_{yy} \Gamma'$

Now let us consider an orthogonal transformation of \mathbf{y} such that

$$\mathbf{P} = \Gamma' \mathbf{y} \tag{4}$$

where PC_1, PC_2, PC_p are the p components of \mathbf{P} and are called principal components [28].

The composite index corresponding to i^{th} population unit of h^{th} subpopulation is given as

$$C_{hi} = \frac{\left[\sum_{j=1}^p \lambda_j PC_{hij} \right]}{\sum_{j=1}^p \lambda_j} \tag{5}$$

where the PC_{hij} 's are principal component scores corresponding to the population unit $\forall h = 1, 2, \dots, l; i = 1, 2, \dots, N_h; j = 1, 2, \dots, p$.

The average of C_{hi} 's within h^{th} subpopulation gives the composite index value for h^{th} subpopulation as

$$C_h = \sum_{i=1}^{N_h} C_{hi} / N_h \tag{6}$$

The composite index values of subpopulations are normalized as

$$CI_h = \frac{C_h - \min(C_h)}{\max(C_h) - \min(C_h)} \tag{7}$$

The ranking of l subpopulations is determined by the normalized composite index (CI_h) where a value of 1 represents the highest rank and a value of 0 represents the lowest rank.

3.2 PCA based index methodology for disaggregated survey (sample) data

Consider a survey conducted where the population is distributed across various subpopulations. When the population is large, the observational recordings from each individual or unit of the whole population (as in the case of PVC-PCA based index methodology for disaggregated population data) are expensive and time consuming, therefore a sample is drawn from the population. The sample from the population is generally collected using complex survey design as they are less costly and efficient as the selected sample is representative of the population. Since the selected sample is collected through a survey design, the survey weight is attached to each unit in the survey. When the researcher aims to compare the performance of these subpopulations, it is necessary to establish an index or measure that enables a meaningful assessment of subpopulation performance. Principal Component Analysis (PCA) is utilized to develop an index specifically tailored for disaggregated data to exclude the effect of multicollinearity present in the data. For PCA, the eigenvalues and eigenvectors of the variance-covariance matrix of the variables (Σ_{yy}) under study are the target parameters of interest.

For this, let us consider a finite population $U = (1, 2, \dots, k, \dots, N)$ of size N units having l subpopulations such that the h^{th} subpopulation has N_h units and $\sum_{h=1}^l N_h = N$, ($h = 1, 2, \dots, l$). Let s be a probabilistic sample of size n drawn from this population such that $\sum_{h=1}^l n_h = n$ where n_h is the number of units belonging to the h^{th} sub-population with the assumption that $n_h \neq 0$ and d_{hi} denotes the survey weight associated with i^{th} unit of the sample in h^{th} subpopulation such that $\sum_{h=1}^l \sum_{i=1}^{n_h} d_{hi} = 1$. Let $\mathbf{y} = (y_1, y_2, \dots, y_p)'$ be the p set of standardized indicator variables and $y_{hi} = (y_{hi1}, y_{hi2}, \dots, y_{hip})'$ be values of the variables \mathbf{y} corresponding to i^{th} sample unit of h^{th} subpopulation where, $h = 1, 2, \dots, l$ and $i = 1, 2, \dots, n_h$.

3.2.1 Existing PCA (OLS-PCA based approach) based index methodology for disaggregated survey data

Existing PCA-based index methodology for disaggregated survey data is based on the Ordinary Least Squares (OLS) estimator of $\sum_{yy} (\hat{\sum}_{yy})$ which assumes that the data is collected using a simple random sampling and does not include survey weight in its calculation. The ordinary least squares estimator of \sum_{yy} is given as

$$\hat{\sum}_{yy} = \mathbf{V}_{yys} = (n - 1)^{-1} \sum_h \sum_i (\mathbf{y}_{hi} - \bar{y}_s) (\mathbf{y}_{hi} - \bar{y}_s)' \quad (8)$$

where, $\bar{y}_s = \sum_h \sum_i \mathbf{y}_{hi}/n$. The OLS estimator of \sum_{yy} is used for the development of PCA-based index method on sampled data. For that let us assume that $\hat{\sum}_{yy}$ is a real positive definite matrix whose non-zero eigenvalues are $\lambda_{1yy} > \lambda_{2yy} > \lambda_{3yy} \dots > \lambda_{pyy}$ and the corresponding eigenvectors are $\gamma_{1yy}, \gamma_{2yy}, \gamma_{3yy} \dots \gamma_{pyy}$. For distinct $\lambda_{jyy}'s$ ($j = 1, 2, 3, \dots, p$), a $(p \times p)$ orthogonal matrix can be formed, where

$$\Gamma = [\gamma_{1yy}, \gamma_{2yy}, \gamma_{3yy} \dots \gamma_{pyy}] \quad (9)$$

Γ matrix diagonalizes $(\hat{\sum}_{yy})$ matrix such that

$$\hat{\sum}_{yy} = \Gamma A \Gamma' \quad (10)$$

where $A = \text{diag} = (\lambda_{1yy}, \lambda_{2yy}, \lambda_{3yy}, \dots, \lambda_{pyy}) = \Gamma' \hat{\sum}_{yy} \Gamma$.

Now let us consider an orthogonal transformation of \mathbf{y} such that

$$\mathbf{P} = \Gamma' \mathbf{y} \quad (11)$$

where $PC_{1yy}, PC_{2yy}, PC_{pyy}$ are the p components of \mathbf{P} and are called principal components [28].

The composite index corresponding to i^{th} sample unit of h^{th} subpopulation is given as

$$C_{hi} = \frac{\left[\sum_{j=1}^p \lambda_{jyy} PC_{hijyy} \right]}{\sum_{j=1}^p \lambda_{jyy}} \quad (12)$$

where the are principal component scores of $\hat{\sum}_{yy}$ corresponding to the sample unit $\forall h = 1, 2, \dots, l; i = 1, 2, \dots, n_h; j = 1, 2, \dots, p$.

The average of within h^{th} subpopulation gives the composite index value for h^{th} subpopulation as

$$C_h = \sum_{i=1}^{n_h} C_{hi}/n_h \quad (13)$$

The composite index values of subpopulations are normalized as

$$CI_h = \frac{C_h - \min(C_h)}{\max(C_h) - \min(C_h)} \quad (14)$$

The ranking of l subpopulations is determined by the normalized composite index (CI_h) where a value of 1 represents the highest rank and a value of 0 represents the lowest rank. The commonly used OLS-PCA based index methodology which relies on the Ordinary Least Squares (OLS) estimator, fails to incorporate survey weights in data collected through complex survey designs. This omission of survey weights results in a biased estimate of variance covariance matrix in commonly used OLS-PCA-based index methodology leading to biased eigenvalues, indices and thereby erroneous rankings of subpopulations. In the subsequent section, a Survey-Weighted (SW) estimator-based index method is proposed which includes the survey weight present in survey data.

3.2.2 Proposed survey weighted PCA (SW-PCA based approach) based index methodology for disaggregated survey data

The Survey Weighted (SW) estimator of Population Variance Covariance matrix given by Skinner et al. [29] and Smith et al. [30] is

$$\hat{\sum}_{yyw} = \mathbf{V}_{yys}^* = \sum_h \sum_i d_{hi} \mathbf{y}_{hi} \mathbf{y}_{hi}' - \bar{y}_s^* \bar{y}_s^{*'} \quad (15)$$

where, $\bar{y}_s^* = \sum_h \sum_i d_{hi} \mathbf{y}_{hi}$.

The SW estimator of $\sum_{yy} (\hat{\sum}_{yyw})$ is used for the development of SW-PCA based index methods. For that let us assume that the $\hat{\sum}_{yyw}$ is a real positive definite matrix whose non-zero eigenvalues are $\lambda_{1yyw} > \lambda_{2yyw} > \lambda_{3yyw} \dots \lambda_{pyyw}$ and the corresponding eigenvectors are $\gamma_{1yyw}, \gamma_{2yyw}, \gamma_{3yyw} \dots \gamma_{pyyw}$. For distinct $\lambda_{jyyw}'s$ ($j = 1, 2, 3, \dots, p$), a $(p \times p)$ orthogonal matrix can be formed, where

$$\Gamma = [\gamma_{1yyw}, \gamma_{2yyw}, \gamma_{3yyw} \dots \gamma_{pyyw}] \quad (16)$$

Γ matrix diagonalizes $\hat{\sum}_{yyw}$ matrix such that

$$\hat{\sum}_{yyw} = \Gamma A \Gamma' \quad (17)$$

where $A = \text{diag} (\lambda_{1yyw}, \lambda_{2yyw}, \lambda_{3yyw}, \dots, \lambda_{pyyw}) = \Gamma' \hat{\sum}_{yyw} \Gamma$

Now let us consider an orthogonal transformation of \mathbf{y} such that

$$\mathbf{P} = \Gamma' \mathbf{y} \quad (18)$$

where $PC_{1yyw}, PC_{2yyw}, PC_{pyyw}$ are the p components of \mathbf{P} and are called principal components [28].

The composite index corresponding to i^{th} sample unit of h^{th} subpopulation is given as

TABLE 1 Mean of variables considered for population data generation.

Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁	Y ₁₂
807.93	213.31	721.86	127.64	65.13	166.69	120.73	70.09	58.70	49.10	97.85	56.63

TABLE 2 Variance covariance matrix of variables considered for population generation.

Variables	Y ₁	Y ₂	Y ₃	Y ₄	Y ₅	Y ₆	Y ₇	Y ₈	Y ₉	Y ₁₀	Y ₁₁	Y ₁₂
Y ₁	263,931	34,252	76,232	19,551	11,007	26,131	19,112	6,828	6,776	6,280	3,673	7,081
Y ₂	34,252	22,964	41,505	9,415	4,393	4,090	4,918	3,003	2,995	2,189	1,594	2,829
Y ₃	76,232	41,505	620,633	54,028	14,295	8,103	18,088	16,563	8,402	11,975	10,450	15,495
Y ₄	19,551	9,415	54,028	14,862	3,476	2,290	3,434	1,884	1,883	1,560	896	1,835
Y ₅	11,007	4,393	14,295	3,476	3,891	3,192	2,433	973	1,508	982	555	1,601
Y ₆	26,131	4,090	8,103	2,290	3,192	28,266	4,843	3,285	2,466	2,187	3,389	2,588
Y ₇	19,112	4,918	18,088	3,434	2,433	4,843	6,531	1,726	1,520	1,361	924	1,772
Y ₈	6,828	3,003	16,563	1,884	973	3,285	1,726	4,607	887	1,316	1,763	1,731
Y ₉	6,776	2,995	8,402	1,883	1,508	2,466	1,520	887	1,852	717	733	890
Y ₁₀	6,280	2,189	11,975	1,560	982	2,187	1,361	1,316	717	2,733	1,582	1,184
Y ₁₁	3,673	1,594	10,450	896	555	3,389	924	1,763	733	1,582	101,389	2,261
Y ₁₂	7,081	2,829	15,495	1,835	1,601	2,588	1,772	1,731	890	1,184	2,261	6,203

TABLE 3 Stratum sample sizes.

Strata No.	1	2	3	4	5
Sample size	600	300	200	300	600

$$C_{hi} = \frac{\left[\sum_{j=1}^p \lambda_{jyyw} PC_{hijyyw} \right]}{\sum_{j=1}^p \lambda_{jyyw}} \tag{19}$$

where the PC_{hijyyw} 's are principal component scores of $\sum_{j=1}^p$ corresponding to the sample unit $\forall h = 1, 2, \dots, l; i = 1, 2, \dots, n_h; j = 1, 2, \dots, p$.

The average of C_{hi} 's within h^{th} subpopulation gives the composite index value for h^{th} subpopulation as

$$C_h = \sum_{i=1}^{n_h} C_{hi} / n_h \tag{20}$$

The composite index values of subpopulations are normalized as

$$CI_h = \frac{C_h - \min(C_h)}{\max(C_h) - \min(C_h)} \tag{21}$$

The ranking of l subpopulations is determined by the normalized composite index (CI_h) where a value of 1 represents the highest rank and a value of 0 represents the lowest rank. In the next section of the simulation study, we compare the existing OLS-PCA based index methodology for disaggregated survey data and proposed Survey Weighted PCA (SW-PCA) based Index methodology for disaggregated survey data with respect to PVC-PCA based index method on population data.

4 Simulation study

4.1 Population generation and sample extraction

To evaluate the empirical performance of the developed methodology, a simulation study has been conducted using artificially generated data. A population of $N = 100,000$ units was generated using multivariate normal distribution. The population consists of twelve indicator variables denoted by $\mathbf{y} = (y_1, y_2, \dots, y_{12})$ and one design variable z which is the sum of all the indicator variables. Thus, the population data becomes $\mathbf{x} = (\mathbf{y}', z)$ where the linearity and homoscedasticity assumptions of the variables are satisfied. The mean vector \mathbf{u}_y and variance-covariance matrix (Σ_{yy}) of twelve indicator variables were taken similar to that being estimated from the NSS 68th Round data for all of India on the twelve sets of variables. The details of the mean vector and variance covariance matrix of the variables are given in Tables 1, 2 respectively.

This population was then stratified into five strata of equal size based on the ordered z -values, the first stratum containing the 20,000 units with the smallest z -values and so on. A sample of size 2,000 units was drawn from this population following U-shaped allocation of sample size from the strata. The allocated sample sizes in the strata are presented in Table 3. The Monte Carlo simulation was run $S = 5,000$ times. A simulation study was carried out in R software.

Both OLS-PCA and SW-PCA based index were computed using the disaggregated sample data. The PVC-PCA based Index was also computed using the disaggregated population data. The performance of OLS-PCA and SW-PCA based indices on disaggregated sampled data is compared with the PVC-PCA based index developed from disaggregated population data.

4.2 Comparative analysis of proposed SW-PCA based index and the existing OLS-PCA based index methodology for disaggregated survey data with respect to PVC-PCA based index for disaggregated population data

4.2.1 Eigenvalues comparison

In the context of principal component analysis (PCA), eigenvalues are used to determine the amount of variance explained by each principal component. Therefore the eigenvalues are computed from ordinary least squares (OLS) estimator $\hat{\Sigma}_{yy}$, survey-weighted (SW) estimator $\hat{\Sigma}_{yyw}$ and the population parameter $\hat{\Sigma}_{yy}$ to compare the efficiency of existing OLS-PCA based index and SW-PCA based index on disaggregated survey data with respect to PVC-PCA based index on disaggregated population data. The eigenvalues of $\hat{\Sigma}_{yy}$ (used for OLS-PCA based approach) and $\hat{\Sigma}_{yyw}$ (used for SW-PCA based approach) are compared with the eigenvalues of population parameter Σ_{yy} (used for PVC-PCA based approach) as presented in Table 4. The results in Table 4 clearly illustrate that the first two principal components account for 83.92 percent of the total variation present in the population. This highlights the significant contribution of these components in the development of the index. Additionally, the standard deviation of the eigenvalues for the first two principal components is lower in the case of SW-PCA compared to OLS-PCA.

The bias of the eigenvalues of $\hat{\Sigma}_{yy}$ (OLS-PCA) and $\hat{\Sigma}_{yyw}$ (SW-PCA) estimator are compared with the eigenvalues of population parameter Σ_{yy} (PVC-PCA) in Figure 1.

Figure 1 shows that the bias of most of the eigenvalues of $\hat{\Sigma}_{yyw}$ (SW-PCA) estimator is lesser in comparison to the existing $\hat{\Sigma}_{yy}$ (OLS-PCA) estimator with respect to population parameter Σ_{yy} (PVC-PCA), especially for the eigenvalues which are having a major contribution to the total variation of the data. From the bias and standard deviation of eigenvalues of OLS-PCA and SW-PCA on sampled data with respect to PVC-PCA on population data, we can conclude that the SW-PCA method on sampled data addresses the issue of representativeness and its eigenvalues are nearer to eigenvalues of PVC-PCA based index method of population by incorporating the weights attached to each sampling unit. Since eigenvalues play a crucial role in index construction methods, Figure 1 and Table 4 lead to the conclusion that the SW-PCA-based index exhibits lower bias and higher efficiency compared to the existing OLS-PCA-based index for survey data referring PVC-PCA based index method on population data.

4.2.2 Eigenvectors comparison

In principal component analysis (PCA), eigenvectors are also generated by each principal component which are further used in the development of PCA scores. Therefore the eigenvectors are computed from the Ordinary Least Squares (OLS) estimator $\hat{\Sigma}_{yy}$, the Survey-Weighted (SW) estimator $\hat{\Sigma}_{yyw}$ and the population parameter $\hat{\Sigma}_{yy}$ to compare the efficiency of existing OLS-PCA

based index and SW-PCA based index on disaggregated survey data with respect to PVC-PCA based index on disaggregated population data. The eigenvectors of $\hat{\Sigma}_{yy}$ (used for OLS-PCA based approach) and $\hat{\Sigma}_{yyw}$ (used for SW-PCA based approach) are compared with the eigenvectors of population parameter Σ_{yy} (used for PVC-PCA based index on disaggregated population data) as presented in Supplementary Table 1. The bias of all the elements of eigenvectors of the first two principal components which contribute 83.92 percent of the total variation is lesser in the case of $\hat{\Sigma}_{yyw}$ (SW-PCA) in comparison to Σ_{yy} (OLS-PCA method) referring eigenvector of population parameter Σ_{yy} (PVC-PCA) indicating the efficiency of SW-PCA based index method. Table 5 presents the average bias and average standard deviation of the elements of eigenvectors and it can be seen from Table 5 that the bias and standard deviation of SW-PCA is lesser than OLS-PCA for all the eigenvectors. Therefore we can conclude that the SW-PCA method on sampled data addresses the issue of representativeness and its eigenvectors are nearer to the eigenvectors of the PVC-PCA based index method of population by incorporating the weights attached to each sampling unit.

4.2.3 Comparison of percentage of variance explained by principal components

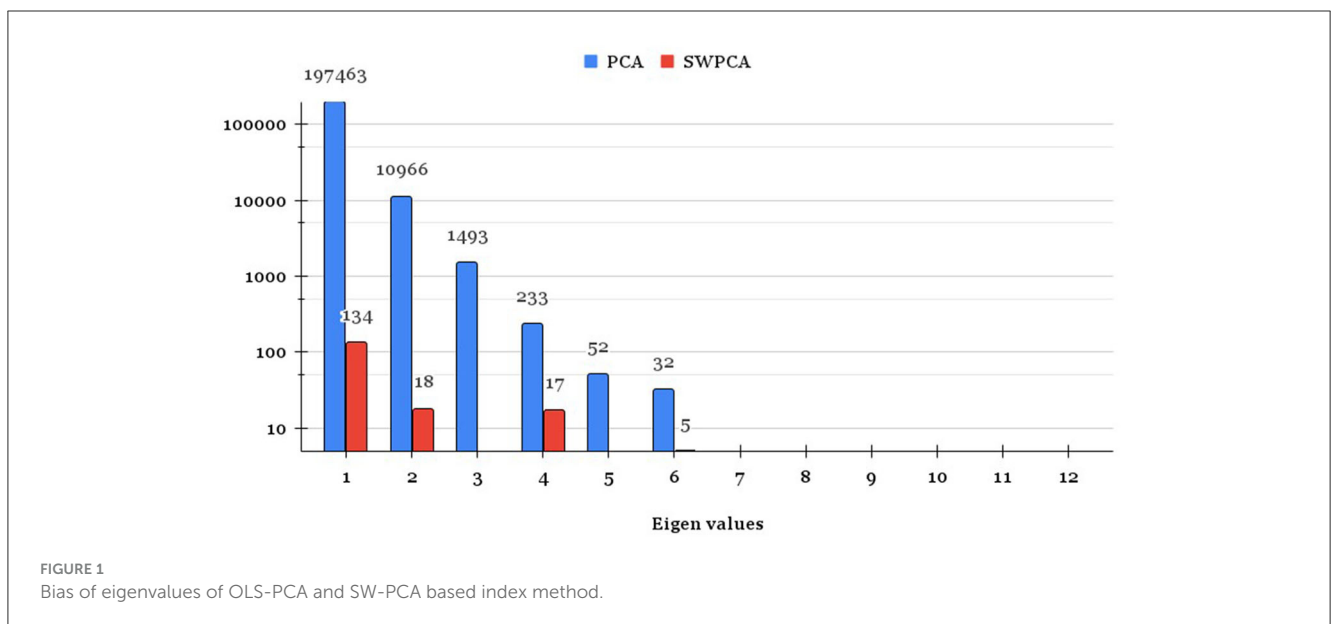
In this section, the percentage of variance explained by eigenvalues of existing OLS-PCA and proposed SW-PCA based method on disaggregated sampled data are compared with respect to PVC-PCA based index on population data. Table 6 and Figure 2 shows that the absolute bias of the percentage of variance explained by all the eigenvalues of Ordinary Least squares Estimator ($\hat{\Sigma}_{yy}$) (used for OLS-PCA based index on survey data) is higher in comparison to the bias of Survey Weighted estimator ($\hat{\Sigma}_{yyw}$) (used for SW-PCA based index on survey data) on survey data which indicates that percentage of variance explained by the eigenvalues of survey weighted estimator ($\hat{\Sigma}_{yyw}$) are more representative to the percentage of variance explained by the population variance covariance matrix Σ_{yy} (used for PVC-PCA index on population data). This observation suggests that SW-PCA is more efficient than the existing OLS-PCA based index for sampled data with respect to PVC-PCA based index method on population data, demonstrating its effectiveness in capturing the underlying structure of the data.

4.2.4 Comparison of efficiency of existing OLS-PCA and proposed SW-PCA based index methodology of survey data with respect to PVC-PCA based index methodology on population data

Developed indices based on existing OLS-PCA and proposed SW-PCA based index for disaggregated sampled survey data are compared with PVC-PCA index based on disaggregated population data using the criterion of percentage Relative Root Mean Square Error (RRMSE). RRMSE is the root mean squared error normalized by the root mean square value where each residual is scaled against the actual value. In this study, the residuals are the difference of

TABLE 4 Eigen value comparison of OLS-PCA and SW-PCA based index method on sampled data with respect to PVC-PCA based index on population data.

PVC-PCA Eigenvalues on population data		OLS-PCA Eigenvalues on sampled data			SW-PCA Eigenvalues on sampled data		
Mean	% of variance explained	Mean	Std Dev.	% of variance explained	Mean	Std Dev.	% of variance explained
651,412	60.27	848,875	20,361	65.75	651,545	15,409	60.29
255,664	23.65	266,630	8,416	20.65	255,682	8,351	23.66
101,041	9.35	102,534	3,160	7.94	100,895	3,329	9.34
26,865	2.49	27,098	840	2.10	26,882	904	2.49
19,328	1.79	19,380	612	1.50	19,273	661	1.78
7,651	0.71	7,684	233	0.60	7,656	256	0.71
6,246	0.58	6,238	193	0.48	6,226	212	0.58
4,096	0.38	4,082	125	0.32	4,092	138	0.38
3,422	0.32	3,392	107	0.26	3,398	118	0.31
2,258	0.21	2,254	70	0.17	2,249	77	0.21
1,905	0.18	1,882	59	0.15	1,885	64	0.17
973	0.09	964	31	0.07	964	34	0.09



the normalized composite index values calculated by the index methodologies for disaggregated sampled survey data (OLS-PCA and SW-PCA index for survey data) and PVC-PCA based index methodology on population data of *l* subpopulations. RRMSE of OLS-PCA and SW-PCA index for survey data examines the convergence of OLS-PCA and SW-PCA index with PVC-PCA based index respectively. RRMSE is defined by

$$RRMSE(\theta) = \sqrt{\frac{1}{S} \sum_{j=1}^S \left[\sum_{i=1}^l \left(\frac{\hat{\theta}_i - \theta_i}{\theta_i} \right)^2 \right]^*} \quad (22)$$

where $\hat{\theta}_i$ are the computed index values from disaggregated sampled data by existing OLS-PCA or proposed SW-PCA based index methodologies while θ_i are the index values of PVC-PCA based index methodology for disaggregated population data corresponding to *i*th strata/subpopulation respectively, *l* denotes total number of strata and *S* denotes the number of simulation run. The values of percentage RRMSE of different methods are reported in Table 7.

The results show that the value of percentage RRMSE is higher for the existing OLS-PCA based Index as compared to the proposed SW-PCA based Index on disaggregated sampled survey data. It signifies that the proposed SW-PCA based index on survey data is better representative and more converging to PVC-PCA based

TABLE 5 Average bias and standard deviation of the elements of OLS-PCA and SW-PCA eigenvectors with respect to PVC-PCA based index on population data.

Eigenvector	Average bias of the elements of OLS-PCA Eigenvectors	Average Std. Dev. of the elements of OLS-PCA Eigenvectors	Average bias of the elements of SW-PCA Eigenvectors	Average Std. Dev. of the elements of SW-PCA Eigenvectors
Eigenvector 1	-0.1644	0.0360	-0.0020	0.0242
Eigenvector 2	0.0902	0.0065	0.0001	0.0071
Eigenvector 3	0.0158	0.0726	0.0647	0.0956
Eigenvector 4	0.0482	0.0822	0.0329	0.1139
Eigenvector 5	0.0753	0.0357	0.0073	0.0696
Eigenvector 6	-0.1317	0.1574	0.0287	0.1588
Eigenvector 7	-0.3161	0.1975	-0.0727	0.1961
Eigenvector 8	-0.1515	0.1773	0.0295	0.1772
Eigenvector 9	-0.2006	0.1700	-0.0238	0.1711
Eigenvector 10	-0.2484	0.1963	-0.0479	0.1966
Eigenvector 11	-0.1363	0.1546	0.0166	0.1575
Eigenvector 12	-0.1479	0.1166	-0.0175	0.1177

TABLE 6 Bias of % of variance explained of PCA based method on disaggregated survey data.

Principal component	Bias of % of variance explained of PCA based method on survey data	Bias of % of variance explained of SW-PCA based method on survey data
PC1	5.48	0.02
PC2	3.00	0.01
PC3	1.41	0.01
PC4	0.39	0.00
PC5	0.29	0.01
PC6	0.11	0.00
PC7	0.10	0.00
PC8	0.06	0.00
PC9	0.06	0.01
PC10	0.04	0.00
PC11	0.03	0.01
PC12	0.02	0.00

index for the population in comparison to the existing OLS-PCA based index on survey data. In other words, the proposed SW-PCA based Index is more efficient than the existing OLS-PCA based Index for sampled data collected through complex survey design referring PVC-PCA based index on population data. Therefore, it is concluded from the simulation study that the proposed SW-PCA based Index should be used for the development of index for data collected through survey designs. The simulation study establishes that the survey-weighted PCA-based index for sample data is more

representative to the PCA based index on population data and thereby should be recommended for complex survey designs where the weight is attached to each sampling unit rather than the existing OLS estimator based PCA index for survey data.

4.3 Conclusion of simulation study

Basically, under an ideal situation, we should carry the complete enumeration of the population rather than a survey to get the exact estimate of population parameters which we can get from the PVC-PCA based index methodology on population data in our case study. But in practical situations, we carry the survey rather than the complete enumeration due to resource constraints. Under these situations, we can calculate different estimators like OLS-PCA and proposed SW-PCA based index of population parameter i.e., PVC-PCA based index in our case study. However it is recommended to use BLUE estimators (best linear unbiased estimators) having minimum divergence from the population parameter. In this Section 3.2.1, 3.2.2, 3.2.3 show that the SW-PCA based index has a lesser bias in comparison to OLS-PCA based index and Section 3.2.4 shows that the SW-PCA based index is converging more than OLS-PCA based index to PVC-PCA based index on population data i.e., population parameter.

5 Real data case study

5.1 Food consumption index for different states of India using NSS data (68th round)

In this section, we have applied the existing OLS-PCA based index methodology and our proposed SW-PCA based index methodology on Household Consumer Expenditure Survey (HCES), NSS 68th Round survey data collected through complex

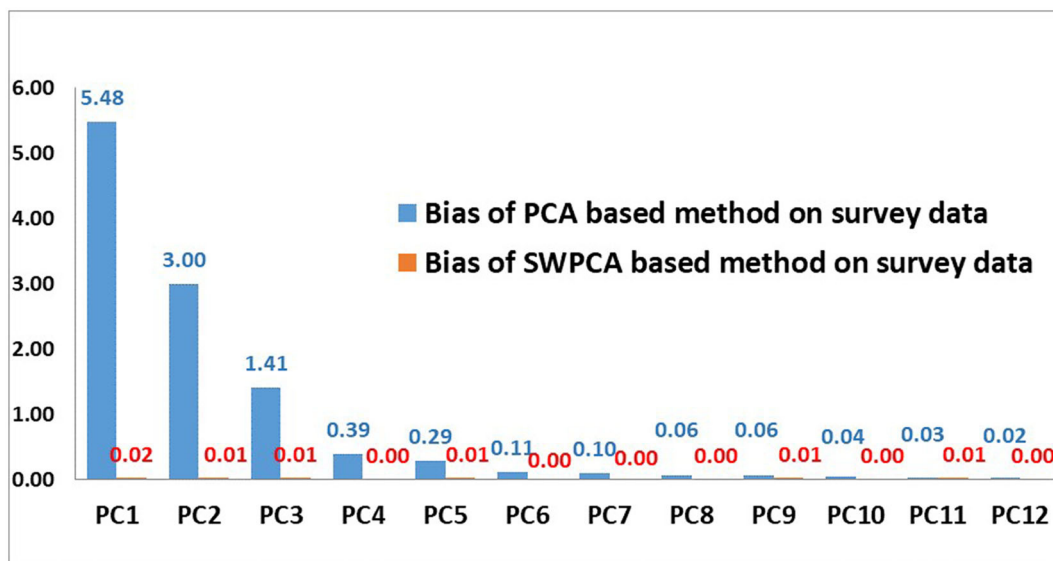


FIGURE 2 Bias of % of variance explained by principal components of PCA and SW-PCA based method.

TABLE 7 Percentage RRMSE of the existing OLS-PCA and proposed SW-PCA based index method for sampled/survey data with respect to PVC-PCA based index on population data.

Criterion	RRMSE of OLS-PCA based index	RRMSE of SW-PCA based index
RRMSE, %	22.84	19.46

design. NSS 68th Round data is collected from all the states of India using a Stratified Multistage Sampling Design. Households are the primary units of the survey and states are the subjects on which indices and ranking based on indices is developed.

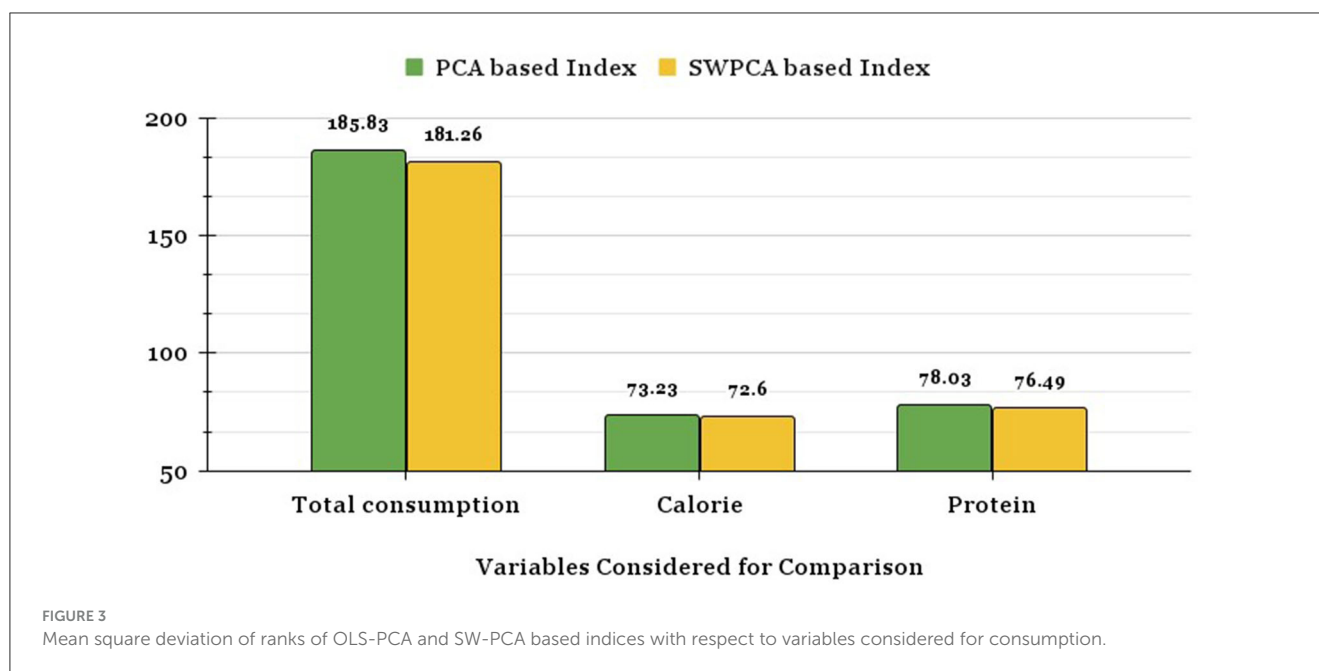
Household Consumer Expenditure Survey (HCES), NSS 68th Round data contains information on twelve variables of consumption, i.e., Cereals (Y1), Pulses & pulse products (Y2), Milk & milk products (Y3), Salt & sugar (Y4), Edible oil (Y5), Egg, fish & meat (Y6), Vegetables (Y7), Fruits (fresh) (Y8), Spices (Y9), Beverages (Y10), Served processed food (Y11), and Packaged processed food (Y12). These variables are assembled into a single quantity that has been termed as Food Consumption Index (FCI). The FCI may be used for ranking of the states within a country or ranking of districts within a state based on their consumption behavior. Therefore, FCIs are calculated for each of the sampled households within a state by combining the twelve variables of consumption using the PCA and SW-PCA based index. Then the FCIs of the households within a state are averaged to get the FCI score of the state. Finally, the FCI scores of the states are rescaled using the following formula

$$Rescaled\ FCI_i = \frac{FCI_i - \min(FCI_i)}{\max(FCI_i) - \min(FCI_i)}, \quad (23)$$

where, $Rescaled\ FCI_i$ is the rescaled FCI score of i^{th} state, $\min(FCI_i)$ and $\max(FCI_i)$ are the minimum and maximum FCI score of the states respectively. Thus, the ranking of the states of India is done based on rescaled FCI scores. The rescaled FCI scores and ranks of the states using OLS-PCA and SW-PCA based index have been provided in [Supplementary Table 2](#).

In this section, our aim is to show that there is a significant change in the ranking of states done through OLS-PCA and our proposed SW-PCA (Survey-Weighted estimator) based index methodology. Besides that, we have compared the ranking of states done through OLS-PCA and our proposed SW-PCA based index methodology with respect to Per Household Total Consumption (PHTC), Average Daily Intake Per Capita of Calories (ADIPCC), and Average Daily Intake Per Capita of Protein (ADIPCP) of the states of India. PHTC, ADIPCC, and ADIPCP are considered as references based on understanding that the ranking of states should converge with the PHTC, ADIPCC, and ADIPCP because the ranking of states is done based on consumption variables which are used to derive Per Household Total Consumption, Average Daily Intake Per Capita of Calories and Average Daily Intake Per Capita of Protein. More the convergence, the more efficient will be the index methodology.

To evaluate the performance of the developed methodology, we have used the mean square deviation of ranks of states between Rescaled FCI and variables considered for consumption like PHTC, ADIPCC, and ADIPCP both for OLS-PCA and SW-PCA index methodology for survey data. Per household total consumption has been calculated from NSSO 68th round data, and average daily intake per capita of calories and average daily intake per capita of protein are obtained from NSSO 68th round report released by the Government of India. The ranking of states with respect to these variables associated with consumption is given in [Supplementary Tables 3, 4](#).



The Mean square deviation of ranks is given as $\sum_{i=1}^n (R_i - V_i)^2/n$, where, R_i indicates the rank of i^{th} state through the Index and V_i indicates rank of i^{th} state with respect to variables considered for consumption like total consumption, per capita daily intake of calories and protein, and n is the total number of states used for comparison.

Figure 3 reveals that the mean square deviation of ranks in the case of SW-PCA based Index is smaller than the OLS-PCA based Index for each of the consumption variables considered for comparison. The ranking of states from SW-PCA based index method converges more to consumption variables i.e., Per Household Total Consumption, Average Daily Intake Per Capita of Calories and Average Daily Intake Per Capita of Protein. Therefore, the proposed SW-PCA based index is efficient in ranking the states in comparison to OLS-PCA based Index which does not utilize the survey information available in survey data.

Figure 4 shows the Food Consumption Index using OLS-PCA and SW-PCA based index methods in which the states are classified into four groups: (i) better performing states having food consumption index score range 0.75–1; (ii) moderate performing states having food consumption index score range 0.5–0.75; (iii) poorly performing states having food consumption index score range 0.25–0.50; and (iv) worst performing states having food consumption index score range 0.0–0.25. There are significant changes in the grouping of states when comparing both the index methods which is clear from Figure 4.

Figure 5 portrays the cluster dendrogram, which represents the hierarchal relationship between the states under consideration based on OLS-PCA and SW-PCA based indexes. The states can be grouped into three groups (low, medium, and high-performing states) based on the extent of food consumption. It was discovered that the cluster membership of 23 states/union territories remained constant, whereas the cluster membership of 12 states/union territories changed. Therefore, we can conclude

from the Figures 4, 5 that there are significant changes in the ranking of states due to the survey weights, and it is recommended to use the ranking of subpopulations based on SW-PCA based index.

Using the clustering of SW-PCA indices, the low-performing states and union territories in terms of the Food Consumption Index are Arunachal Pradesh, Goa, Himachal Pradesh, Jharkhand, Karnataka, Lakshadweep, Manipur, Meghalaya, Madhya Pradesh, Nagaland, Orissa, Sikkim, Tamil Nadu, and West Bengal. The mean value of the Food Consumption Index was 0.41 for the first group. Andhra Pradesh, Assam, Bihar, Gujarat, Jammu & Kashmir, Kerala, Puducherry, Punjab, and Uttarakhand are the medium-performing states with an average score of 0.63, while Andaman & Nicobar Islands, Chandigarh, Chhattisgarh, Daman & Diu, Delhi, Dadar & Nagar Haveli, Haryana, Maharashtra, Mizoram, Rajasthan, Tripura, and Uttar Pradesh are the best-performing states in terms of food consumption index, with an average food consumption index value of 0.41.

6 Conclusion

The research paper highlights a critical gap in the current methodology for constructing composite indices from survey data. The existing OLS-PCA based index methodology for survey data relies on the Ordinary Least Squares (OLS) estimator of the variance-covariance matrix, assuming that data is collected using simple random sampling, ignoring the complexities of real-world survey designs involving survey weights. The current OLS-PCA based index methodology for survey data does not incorporate these survey weights, leading to biased and unreliable results and compromising the quality and validity of survey findings.

In response to this challenge, the proposed Survey Weighted PCA (SW-PCA) based Index methodology represents a significant advancement. By explicitly accounting for survey weights and

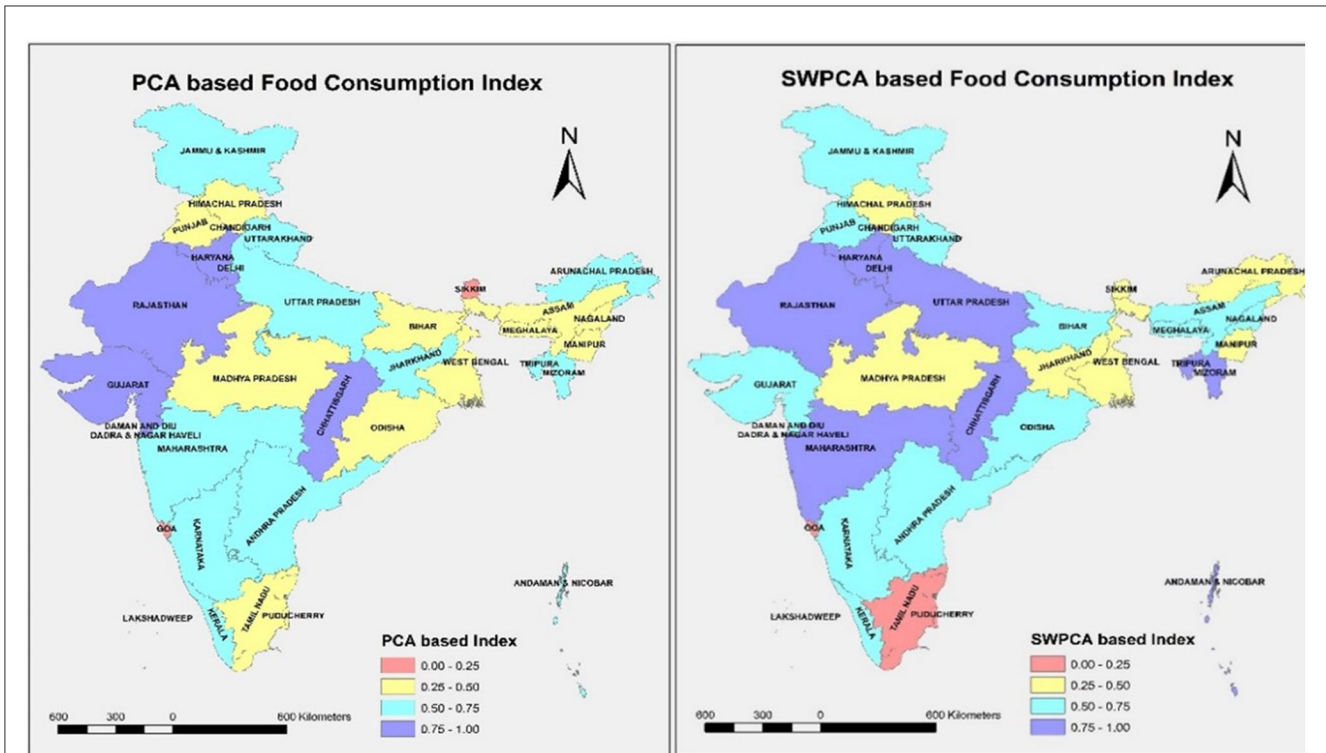


FIGURE 4 State-wise food consumption Index using OLS-PCA and SW-PCA based index methods.

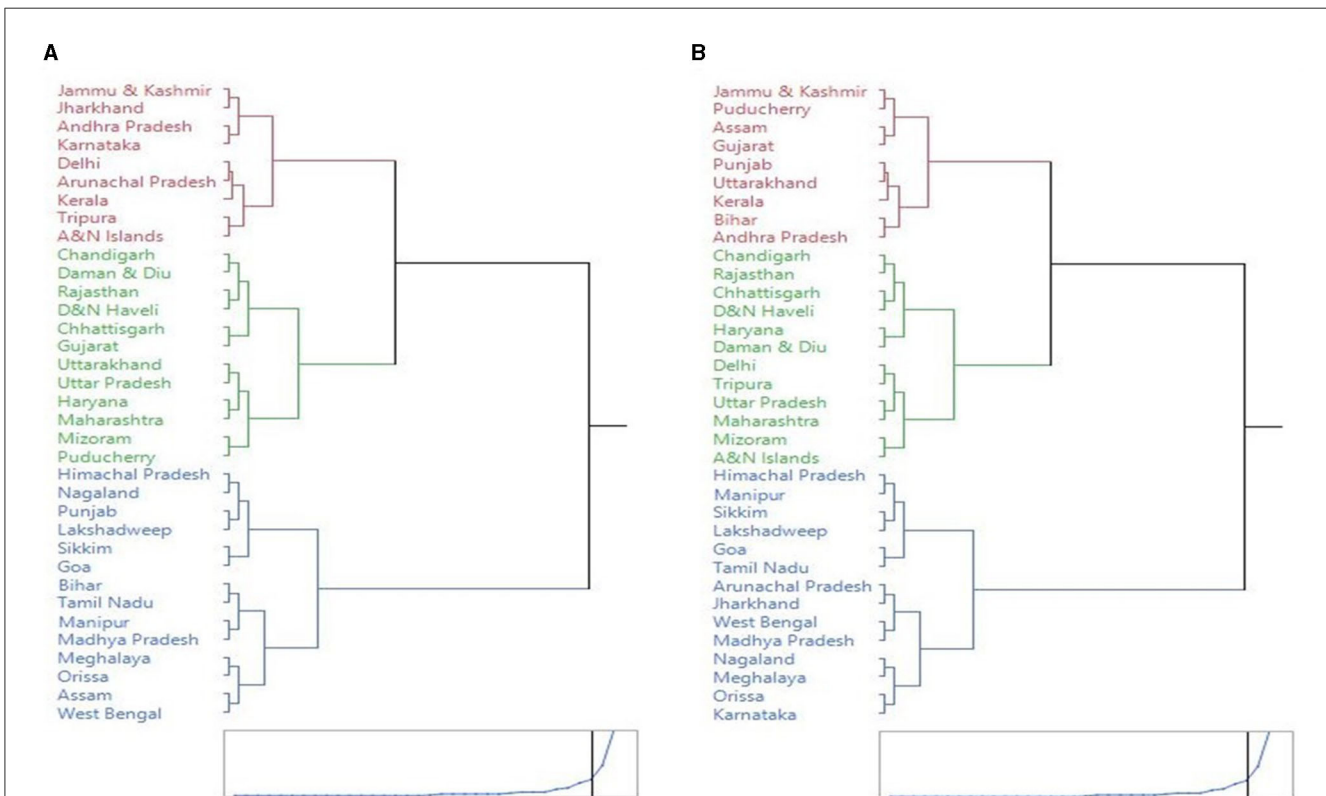


FIGURE 5 Cluster dendrogram of states based on PCA and SW-PCA based index ranking using minimum ward's distance. (A) PCA method. (B) SW-PCA method.

addressing multicollinearity issues, SW-PCA offers a more accurate and reliable way to construct composite indices from sample data collected through complex sampling designs. It ensures that the resulting indices provide an unbiased representation of the underlying population, improving the overall quality of survey-based research.

Our simulation study demonstrated that the SW-PCA based index outperforms indices that do not consider survey weights, indicating its higher efficiency. To validate the proposed methodology, we applied it to a real dataset to construct a food consumption index for different states in India. The results revealed a significant change and improvement in the ranking of states for the food consumption index when survey weights were incorporated in the index development process.

Based on the findings of this study, we can conclude that neglecting to incorporate survey weights in the construction of indices using survey data results in a loss of representativeness, leading to erroneous inferences and conclusions.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: [https://microdata.gov.in/nada43/index.php/catalog/127/study-description#:~:text=The%20NSS%2068th,\(ii\)%20Employment%20and%20Unemployment](https://microdata.gov.in/nada43/index.php/catalog/127/study-description#:~:text=The%20NSS%2068th,(ii)%20Employment%20and%20Unemployment).

Author contributions

DS: Conceptualization, Methodology, Project administration, Software, Visualization, Writing—original draft, Writing—review and editing. PB: Conceptualization, Data curation, Validation, Visualization, Writing—review and editing. RK:

Conceptualization, Investigation, Methodology, Supervision, Validation, Writing—review and editing. TA: Conceptualization, Validation, Visualization, Writing—review and editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fams.2023.1274530/full#supplementary-material>

References

- Wiesmann D. *A Global Hunger Index: Measurement concept, ranking of countries, and trends*. FCND discussion paper 212, IFPRI. (2006).
- Athreya VB, Bhavani RV, Anuradha G, Gopinath R, Velan AS. *Report on the state of food insecurity in rural India*. Published by M. S. Swaminathan Research Foundation. ISBN: 81-88355-06-2. (2008).
- Gentilini U, Webb P. How are we doing on poverty and hunger reduction? A new measure of country performance. *Food Policy*. (2008) 33:521–32. doi: 10.1016/j.foodpol.2008.04.005
- Hastings AD. *From human development to human security: A prototype, human security index*. UNESCAP working paper (WP/09/03). (2009).
- Athreya VB, Rukmani R, UNESCAP working paper. *Report on the state of food insecurity in urban India*. Published by M. S. Swaminathan Research Foundation. ISBN: 978-81-88355-21-1. (2010).
- Krishnamurthy PK, Lewis K, Choularton RJ. A methodological framework for rapidly assessing the impacts of climate risk on national-level food security through a vulnerability index. *Global Environ Change*. (2014) 25:121–32. doi: 10.1016/j.gloenvcha.2013.11.004
- Food insecurity and climate change technical report. (2015). Available online at: https://www.metoffice.gov.uk/food-insecurity-index/resources/HCVI_website_technical_report_v4.pdf (accessed September 22, 2023).
- Global food security index. (2016). *An annual measure of the state of global food security*. A report from the Economist Intelligence Unit, The Economist (www.economist.com).
- United Nations Development Programme (UNDP). *Human Development Report 2021/2022: Uncertain Times, Unsettled Lives*. Shaping Our Future in a Transforming World. (2022).
- Alkire S, Kanagaratnam U, Suppa N. *The global Multidimensional Poverty Index (MPI) 2023 country results and methodological note*. (2023). Available online at: <https://hdr.undp.org/system/files/documents/hdp-document/2023mpireportenpdf.pdf> (accessed September 22, 2023).
- Von Grebmer K, Bernstein J, Resnick D, Wiemers M, Reiner L, Bachmeier M, et al. *Global Hunger Index: Food Systems Transformation and Local Governance*. Bonn; Dublin: Welt Hunger Hilfe CONCERN Worldwide (2022).
- Narain P, Rai SC, Sarup S. Statistical evaluation on development on socio-economic front. *J Ind Soc Ag Stat*. (1991) 43:329–45.
- Narain P, Sharma SD, Rai SC, Bhatia VK. Regional dimensions of disparities in crop productivity in Uttar Pradesh. *J Ind Soc Ag Stat*. (2001) 54:62–79.
- Narain P, Sharma SD, Rai SC, Bhatia VK. Dimensions of regional disparities in socio-economic development of Madhya Pradesh. *J Ind Soc Ag Stat*. (2002) 55:88–107.
- Narain P, Sharma SD, Rai SC, Bhatia VK. Evaluation of economic development at micro level in Karnataka. *J Ind Soc Ag Stat*. (2003) 56:52–63.
- Narain P, Sharma SD, Rai SC, Bhatia VK. Estimation of socio economic development in hilly states. *J Ind Soc Ag Stat*. (2004) 58:126–35.
- Narain P, Sharma SD, Rai SC, Bhatia VK. Estimation of socio-economic development of different districts in Kerala. *J Ind Soc Ag Stat*. (2005) 59:48–55.

18. Narain P, Sharma SD, Rai SC, Bhatia VK. Statistical evaluation of social development at district level. *J Ind Soc Ag Stat.* (2007) 61:216–26.
19. Rai A, Sharma SD, Sahoo PM, Malhotra PK. Development of livelihood index for different agro-climatic zones of India. *Agric Econ Res Rev.* (2008) 21:173–82.
20. Dahal H. Factor analysis for soil test data: a methodological approach in environment friendly soil fertility management. *J Agric Environ.* (2007) 8:8–19. doi: 10.3126/aej.v8i0.722
21. Kumar M. *Study on development indices and their sensitivity analysis.* M.Sc. thesis, ICAR-Indian Agricultural Research Institute, New Delhi. (2008).
22. Raja TA, Tawheed Y. Statistical evaluation of socio-economic development with a new composite index of Kashmir valley. *Int Res J Agric Econ Stat.* (2014) 5:317–20. doi: 10.15740/HAS/IRJAES/5.2/317-320
23. Chao YS, Wu CJ. Principal component-based weighted indices and a framework to evaluate indices: results from the Medical Expenditure Panel Survey 1996 to 2011. *PLoS ONE.* (2017) 12:e0183997. doi: 10.1371/journal.pone.0183997
24. Senna LDD, Maia AG, Medeiros JDFD. The use of principal component analysis for the construction of the Water Poverty Index. *RBRH.* (2019) 24:84. doi: 10.1590/2318-0331.241920180084
25. Zhou X, Tang X, Zhang R. Impact of green finance on economic development and environmental quality: a study based on provincial panel data from China. *Environ Sci Pollut Res.* (2020) 27:19915–32. doi: 10.1007/s11356-020-08383-2
26. Lieberman-Cribbin W, Tuminello S, Flores RM, Taioli E. Disparities in COVID-19 testing and positivity in New York City. *Am J Prev Med.* (2020) 59:326–32. doi: 10.1016/j.amepre.2020.06.005
27. Mohammed A, Zayed T, Nasiri F, Bagchi A. Asset management-based resilience index formulation for pavements via principal components analysis. *Constr Innov.* (2022). doi: 10.1108/CI-04-2022-0083
28. Singh D. *Principal component analysis in modelling lactation milk yield in Haryana cows.* M.Sc. Thesis, Chaudhary Charan Singh Haryana Agricultural University, Hisar. (2009).
29. Skinner CJ, Holmes DJ, Smith TMF. The effect of sample design on principal component analysis. *J Am Stat Assoc.* (1986) 81:789–98. doi: 10.1080/01621459.1986.10478336
30. Smith TMF, Holmes DJ. Multivariate analysis. In: Skinner CJ, Holt D, Smith TMF, editors. *Analysis of Complex Surveys.* New York, NY, USA: John Wiley and Sons (1989). p. 165–190.