Social Indicators Research (2021) 154:623–646
https://doi.org/10.1007/s11205-020-02573-8

ORIGINAL RESEARCH

Measuring and Mapping Disaggregate Level Disparities in Food Consumption and Nutritional Status via Multivariate Small Area Modelling

Saurav Guha[1] · **Hukum Chandra**[1]

Accepted: 16 November 2020 / Published online: 3 January 2021

Abstract

Although India has progressed significantly on several health outcomes but the state of food and nutrition security in the country still requires sustained efforts to accelerate achievement. Existing data based on socio-economic surveys conducted by National Sample Survey Office (NSSO) produce precise measures of food and nutrition security status at state and national level. However, these surveys cannot be used directly to produce reliable district or further smaller domain level estimates because of small sample sizes which lead to high level of sampling variability. Decentralized administrative planning system in India demands the availability of disaggregate (e.g. district) level statistics for target oriented effective policy planning and monitoring, as food and nutrition security is often unevenly distributed among the subsets of relatively small areas. But, due to lack of district level estimates, the mapping and analyse related to food and nutrition security measures are restricted to state and national level. As a result, disaggregate level dissimilarity and variability existing in food and nutrition security are often masked. This article delineates multivariate small area estimation (SAE) technique to obtain reliable and representative estimates of food consumption and nutrition status at district level for the rural areas of state of Uttar Pradesh in India by combining latest round of available Household Consumer Expenditure Survey 2011–2012 data of NSSO and the Indian Population Census 2011. The empirical evidence indicate that the estimates generated by SAE approach are reliable and representative. Spatial maps showing district level inequality in distribution of food and nutrition security in Uttar Pradesh is also produced. The disaggregate level estimates and spatial maps of food and nutrition security are directly relevant to sustainable development goal indicator 2.1.2—severity of food insecurity. The estimates and maps of food insecurity indictors are anticipated to offer irreplaceable information to administrative decision-makers and policy experts for identifying the regions requiring more attention. Government of India has recently launched number of schemes for the benefit of rural population in the country and these estimates will be useful for fund allocation as well as in the monitoring of these schemes.

Keywords Household consumer expenditure survey · NSSO · Census · Multivariate small-area estimation · Joint modelling

Extended author information available on the last page of the article

1 Introduction

After the Green Revolution in India, agricultural productivity and overall food production has increased significantly in the country. Consequently, India had a surplus stock of cereals for the first time ever with a national focus on calorie support to all people, especially for those from lower income groups (MoSPI and WFP 2019). In the following decades, as the economy continued to grow, the country experienced a significant decline in poverty levels. Despite this remarkable feat, the rate of malnutrition in India remains stubbornly high. The food security together with enriched nutrition is among the greatest priority of the Government of India to achieve the Sustainable Development Goal 2 (UNDP 2015). In India, National Sample Survey Office (NSSO) of Ministry of Statistics and Program Implementation, Government of India is the nodal authority to collect Household Consumer Expenditure Survey (HCES) data for producing estimates related to different food insecurity indicators for both rural and urban sectors at the state and the national level. Despite being highly crucial, the estimates of food insecurity indicators along with the disparities in food consumption and nutrition intake are unavailable further down the state level viz. district or further level of disaggregation in India. In the present circumstances, the growing interests of the scientists, government organization, policy makers and public agencies are concentrated in obtaining the local level statistical synopses. These local level areas or domains, better known as small areas or small domains are formed by cross-classification of several demographic and topographic variables that includes small topographic areas (e.g. districts) or small demographic groups (e.g. land category, social groups, religion, age-sex groups) or cross classifying both. On the other hand, in the existing survey data viz. HCES of NSSO, the small areas may have very small or even zero sample sizes and direct estimation in such cases may lead to large sampling error. The SAE methodology provides a viable and cost effective solution this problem of small sample sizes (Rao and Molina 2015). The SAE methods borrow strength from other external data sources viz. areas, time periods etc. to generate precise and representative estimates.

The SAE methods are mostly based on model-based approaches and the main idea behind this is to make a link between the target variable and the auxiliary information through statistical models. This may leads to describe the model-based estimators for these small areas. The area level or the unit level models are generally used in SAE based on the availability of the level of auxiliary information. Area level models are utilized when the model covariates (e.g. census variables) are available only in aggregate. The Fay–Herriot model (Fay and Herriot 1979) is a widely used area level model in SAE that assumes area-specific survey estimates are available, and that these follow an area level linear mixed model with area random effects, Chandra (2013) and Chandra et al (2015). Extensive work may be found in the literature where researchers have dealt with the problem of small area estimation by applying the Fay–Herriot model. The uncertainty of model-based small area estimators was studied by Prasad and Rao (1990), Datta and Lahiri (2000), González-Manteiga et al. (2010), Datta et al. (2011). Rao and Yu (1994) extended the Fay–Herriot model with time series and cross-sectional data while Marhuenda et al. (2013) and Morales et al. (2015) studied the spatiotemporal version of the Fay–Herriot model. Fay (1987) and Datta et al. (1991) introduced the multivariate Fay–Herriot model and Benavent and Morales (2016) extended it by considering different covariance structure of the random effects.

Time and again, there is a need of estimating correlated measures like food insecurity, unemployment or poverty indicators. Multivariate models often take into account for the correlation of several variables and typically fit to this kind of situations (Benavent and

Morales 2016). Unlike the Fay–Herriot model (Fay and Herriot 1979), the multivariate Fay–Herriot model (MFH) considers joint modelling of more than one target variable taking into consideration different covariance arrangements between the vectors of the target variables and the vector of random effects. Several small area applications for estimation poverty, food insecurity and other socio-economics parameters have also been described in Indian data, see for example, Anjoy et al. (2019), Chandra et al. (2011), Islam et al. (2019) and references therein. However, these applications are based on use of univariate small area modelling ignoring the correlation between related variables of interest. Moreover, surveys are generally multivariate in nature and collect more than one target variables (e.g., HCES of NSSO). In small area estimation problem where there is scarcity of sample size within the areas, exploitation of correlation between the variables can offer an advantage in producing the reliable estimates for small area parameters (Rao and Molina 2015).

An attempt has been made in this paper to produce district level estimates of disparities in food consumption and nutrient intake for rural areas of Uttar Pradesh through joint modelling of the three target variables related to disparities in food consumption and nutrient intake under a MFH model approach. Uttar Pradesh is the most populous state in the country as well as the most populous country subdivision in the world. The state accounts for about 16.16% of India's and about 2.9% of world's population with an area of 243,290 square km that equals to 6.88% of India's total geographical area. It holds the third largest economy of the country but it has a large number of people living below the poverty line. The mainstream occupation of majority of population in the state is based on agriculture and according to the Population Census 2011 about 78% of the people reside in rural areas. Therefore, it seems reasonable to consider rural areas of Uttar Pradesh to generate the district level estimates of disparities in food consumption and nutrition status using SAE techniques. The rest of the paper is organized as follows. The HCES data and the population census data is described in Sect. 2. In Sect. 3 we introduced the MFH model while the result and discussion part is discussed in Sect. 4. Finally some important conclusions are drawn in Sect. 5.

2 Data and Model Specification

In this section, the primary sources of the data used in multivariate SAE application is introduced. We utilized the 2011–2012 HCES data of the NSSO for rural districts of Uttar Pradesh and the Population Census data of 2011. These data are used for estimating the disparities in food and nutrition intake at district level in Uttar Pradesh. The 2011–2012 HCES is the latest round of available survey being used of policy analysis in India. The NSSO survey data is not freely downloadable but it can be obtained from the NSSO, Ministry of Statistics and Programme Implementation, Government of India (http://mospi.nic.in/). The HCE surveys of NSSO is carried out at regular intervals as part of its "rounds" and normally a year is taken to be the duration of each of these "rounds". A representative sample of households is randomly selected through a suitable sampling design and the surveys are carried out by interviewing the selected household. The entire geographical area of India is eventually covered through these surveys of NSSO. Stratified multi-stage random sampling is used as the sampling design in the 2011–2012 HCES. In this survey, districts are selected as strata with villages as first stage units and households as second stage units. The 2011–2012 HCES of NSSO is designed to generate reliable estimates at state and national level for both the rural and urban sectors of the country. But this survey data cannot directly be used to generate reliable estimates at district level, because within each district sample size is not large enough to provide district level estimates with adequate precision and reliability. Although, district is always being a very crucial part of the planning process in the country, there are no surveys conducted to produce district level estimates in India and this leads to limit the policy interventions at the district or even further lower level.

The 2011–2012 HCES data of the NSSO comprised 5915 households from the rural areas in 71 districts of Uttar Pradesh. The sample sizes of all the surveyed districts ranged from 32 to 128 with average of 83. This survey provides information on quantity and value of more than 142 food items with a reference period of last 30 days for a few food items and last 7 days for the rest food items for rural areas in Uttar Pradesh. Table 1 apparently reveals that these districts comprised relatively small sample sizes with an average sampling fraction of 0.00023. On account of the constraint of small sample size, it is not possible to produce precise and reliable direct estimates at district level and subsequently leads to producing large standard errors from this survey (Chandra et al. 2011 and Rao and Molina 2015). An attempt has been made in this paper to address this issue of small sample size in obtaining district level estimates from the 2011–2012 HCES data. The multivariate SAE approach has been adopted to handle this issue by incorporating relevant auxiliary information from 2011 Population Census data.

To estimate the disparities in nutrient intakes, the suggested intake of food items has been transformed into calorie, fat and proteins. The quantities of food recorded as consumed by the household are converted into the equivalent amounts of energy, protein and fat on the basis of a Nutrition Chart largely based on an ICMR publication (Gopalan et.al. 1991) which gives the energy, protein and fat content per unit of different foods in the Indian diet. It needs to be said, however, that the actual intake of nutrients depends on how these foods are actually processed and/or cooked in the surveyed households, Government of India. (2014). Usually the amount of total calorie, protein and fat intake for any food item is calculated form the quantities consumed as reported by the sample households. One of the constraints of the NSSO statistics is that the records on meals intake is at the household level, therefore we can't encompass the element of intra-household disparities of food consumption. In our analysis all the estimates are averaged as per capita on the family degree. We have taken three target variables for jointly model the disparities in food consumption and nutrition level using multivariate SAE approach. The target variables at the household level in the 2011–2012 HCES data are Y_1: Average calorie intake (Kcal), Y_2: Average protein intake (Protein) and Y_3: Average fat intake (Fat) per person per day. Average dietary energy intake per person per day in rural India is 2400 kilocalorie (Kcal), as defined by the Ministry of Health and Family Welfare, Government of India. This paper aims to estimate the disparities in food consumption and nutrition level of rural households in Uttar Pradesh by jointly model the target variables viz. Kcal, Protein and Fat at small area level.

Table 1 Details of sample size, sampled food insecure households and sampling fraction in 2011–2012 HCES dataset

Particulars	Minimum	Maximum	Average	Total
Sample size	32	128	83	5915
Food insecure households	10	111	53	3778
Sampling fraction	0.00015	0.00032	0.00023	0.01647

3 Theoretical Framework

This section briefly describes SAE method applied in the estimation of district level inequality in distribution of food and nutrition security. Let the population is divided into D small areas or areas (districts in our application) and let there are M number of target variables of the study. Here D is the total number of small areas in the population while M is the number of target variables of the study. Throughout, a subscript $d(d = 1, \ldots, D)$ is used to index the quantities belonging to small area d and a subscript $m(m = 1, \ldots, M)$ is used to denote the target variable m under the study. Let $y_{dm}(m = 1, \ldots, M)$ be an unbiased direct survey estimator of an unobservable population parameter (for example, the population mean) Y_{dm} of the variable m for small area d. Let $\mathbf{x}_{dm}$ be a p_m-vector of known auxiliary variables for area d that are related to the population mean Y_{dm} for target variable m. These area-specific auxiliary variables are typically obtained from secondary data sources viz. the population census or administrative registers. Let us denote $\boldsymbol{Y}_d$ be the d-vector population mean of target variables of the study and $\mathbf{y}_d = \left(y_{1m}, \ldots, y_{dm}\right)^T$ be a vector of direct survey estimators of $\boldsymbol{Y}_d$. Following Benavent and Morales (2016), an area level Fay–Herriot model (Fay and Herriot 1979) for more than one target variables is

$$\boldsymbol{y}_d = \boldsymbol{Y}_d + \boldsymbol{\varepsilon}_d \text{ and } \boldsymbol{Y}_d = \mathbf{X}_d\boldsymbol{\beta}+\mathbf{u}_d \tag{1}$$

In SAE literature, this model in (1) is often referred to the multivariate version of the Fay–Herriot model. The first stage accounts for the sampling variability of the survey estimates $\mathbf{y}_d$ of true area means $\boldsymbol{Y}_d$ and the second stage links the true area means $\boldsymbol{Y}_d$ to a matrix of known auxiliary variables $\mathbf{X}_d = diag\left(\mathbf{x}_{d1}, \ldots, \mathbf{x}_{dM}\right)_{M\times p}$ with $p = \sum_{m=1}^{M} p_m$. The model (1) can be written as an area-level random effect model given by

$$\mathbf{y}_d = \mathbf{X}_d\boldsymbol{\beta} + \mathbf{u}_d + \boldsymbol{\varepsilon}_d,\ d = 1, \ldots, D \tag{2}$$

Here $\boldsymbol{\beta} = \left(\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_m\right)'_{p\times 1}$ and $\boldsymbol{\beta}_m$ is a p_m- vector of unobservable fixed effect parameters. The vector of random area effects $\mathbf{u}_d$ are independent and identically distributed with $\mathbf{u}_d \overset{ind}{\sim} N\left(0, \mathbf{V}_{u_d}\right)$ and vectors of independent sampling errors $\boldsymbol{\varepsilon}_d$ follows $\boldsymbol{\varepsilon}_d \sim N\left(0, \mathbf{V}_{\varepsilon_d}\right)$. The two errors $\mathbf{u}_d$ and $\boldsymbol{\varepsilon}_d$ are independent of each other within and across areas with covariance matrices $\mathbf{V}_{\varepsilon_d}$ are known and $\mathbf{V}_{u_d}$ depend on unknown parameters $\left(\theta_1, \ldots, \theta_M\right)$. Aggregating D-area-level models, the model (2) can be written in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \tag{3}$$

where $\mathbf{y} = \mathrm{col}\left(\mathbf{y}_d; 1 \le d \le D\right)$ is the $DM \times 1$ vector of direct survey estimates, $\mathbf{X} = \mathrm{col}\left(\mathbf{X}_d\,; 1 \le d \le D\right)$ is the $DM \times p$ matrix of covariates, $Z = \mathrm{col}'(Z_d; 1 \le d \le D)$ is the known covariates of dimension $DM \times DM$ characterizing differences among the small areas, $\mathbf{u} = \mathrm{col}\left(\mathbf{u}_d\,; 1 \le d \le D\right)$ is the $DM \times 1$ vector of random area effects and $\boldsymbol{\varepsilon} = \mathrm{col}\left(\boldsymbol{\varepsilon}_d\,; 1 \le d \le D\right)$ is the $DM \times 1$ vector of sampling errors with $\mathbf{u} \sim N\left(0, \mathbf{V}_u\right)$ and $\boldsymbol{\varepsilon} \sim N\left(0, \mathbf{V}_\varepsilon\right)$. In general, $\mathbf{Z}$ is given by a matrix whose dth column $\mathbf{Z}_d, d = 1, \ldots, D$, is an indicator variable which takes the value 1 if a unit is in area d and is zero otherwise. In particular, in model (3) $\mathbf{Z}$ is a diagonal matrix of order $DM \times DM$. Moreover, it is assumed that the random area effects $\mathbf{u}$ are distributed independently of the sampling errors $\boldsymbol{\varepsilon}$ with $\mathbf{u} \sim N\left(0, \mathbf{V}_u\right)$ and $\boldsymbol{\varepsilon} \sim N\left(0, \mathbf{V}_\varepsilon\right)$ where $\mathbf{V}_u = \mathrm{diag}\left(\mathbf{V}_{ud}\,; 1 \le l \le D\right)$ is the covariance matrix of random area effects and $\mathbf{V}_\varepsilon = \mathrm{diag}\left(\mathbf{V}_{\varepsilon d}\,; 1 \le l \le D\right)$ is the matrix of design variances.

We now consider two particularizations of the model (3) to obtain model-based small area estimates. The First predictor based on univariate Fay–Herriot model (UFH) considers $\mathbf{V}_{u_d} = \mathrm{diag}\left(\sigma^2_{um}; 1 \le m \le M\right), \mathbf{V}_{\epsilon_d} = \mathrm{diag}\left(\sigma^2_{\epsilon dm}; 1 \le m \le M\right), \quad d = 1, \ldots, D$ and we assume $\sigma^2_{\epsilon dm}$'s are known. For the second predictor based on multivariate Fay–Herriot model (MFH), $\mathbf{V}_{u_d} = \mathrm{diag}(\sigma^2_{um}; 1 \le m \le M)$, $d = 1, \ldots, D$ and we assume a known but not necessarily diagonal matrix $\mathbf{V}_\epsilon$, i.e. sampling errors are not independent with each other in this case. For both the predictors, the number of unknown parameters to be estimated is equal to M with $\theta_m = \sigma^2_{um}$, $m = 1, \ldots, M$. Under the model (3), $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $Var(\mathbf{y}) = \mathbf{V}_y = \mathbf{V}_u + \mathbf{V}_\epsilon = diag\left(\mathbf{V}_{yd}; 1 \le d \le D\right)$, with $\mathbf{V}_u = \mathbf{Z}' \mathbf{V}_u \mathbf{Z}$ and $\mathbf{V}_{yd} = \mathbf{V}_{ud} + \mathbf{V}_{\epsilon d}$, $d = 1, \ldots, D$. Here, $\mathbf{V}_y$ depends on M unknown variance component parameters given by $\boldsymbol{\theta} = \left(\theta_1, \ldots, \theta_M\right)$ and the restricted maximum likelihood (REML) method is often used to estimate $\boldsymbol{\theta}$(Benavent and Morales 2016). Replacing the estimated values $\hat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$ in $\mathbf{V}_u$ to obtain $\hat{\mathbf{V}}_u = \mathbf{V}_u(\hat{\boldsymbol{\theta}})$ and $\hat{\mathbf{V}}_y = \hat{\mathbf{V}}_u + \mathbf{V}_\epsilon$, the multivariate version of empirical best linear unbiased predictors (EBLUP) of $\boldsymbol{Y}$ is defined as

$$\hat{Y}_{\mathrm{MFH}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} \tag{4}$$

Here, empirical the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and the EBLUP of $\mathbf{u}$ are obtained as $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}' \hat{\mathbf{V}}_y^{-1} \mathbf{X}\right)^{-1} \mathbf{X}' \hat{\mathbf{V}}_y^{-1} \mathbf{y}$ and $\hat{\mathbf{u}} = \hat{\mathbf{V}}_u \mathbf{Z}' \hat{\mathbf{V}}_y^{-1} \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$ respectively. In small area applications, the mean squared error (MSE) estimates are desirable to measure the precision of estimates and also to construct the confidence interval for the estimates. The MSE estimate of multivariate version of EBLUP (4) is obtained using the MSE estimation given by Benavent and Morales (2016).

4 Results and Discussions

4.1 Model Fitting

The auxiliary variables used in this analysis are taken from the 2011 Population Census of India. The multivariate SAE approach based on area level models were applied to obtain the small area estimates as these auxiliary variables are available only as counts at district level. In the 2011 Population Census data, a total of almost 30 such auxiliary variables are available for usage in this analysis. Prior to the determination of suitable covariates for multivariate SAE modelling, an exploratory data analysis has been done for selection of few auxiliary variables. Principal Component Analysis (PCA) was also employed to obtain composite scores for some selected sets of auxiliary variables. In particular, we performed PCA separately on two sets of auxiliary variables and all measured at district level. These two sets of auxiliary variables are noted as P1 and P2 below. The first set (P1) comprised the proportions of main cultivators by gender, proportions of main workers by gender and proportions of main agricultural laborers by gender. 44% of the variability in the P1 set was explained by the first principal component (P11) for P1 and explained variability was increased to 69% by adding the second principal component (P12). The second set (P2) comprised the proportions of marginal agriculture laborers by gender and the proportions of marginal cultivator by gender. 52% of the variability in the set P2 was explained by the first principal component (P21) for P2 and explained variability was increased to 90% by adding the second principal component (P22).

First, we fitted the MFH Model mentioned in Sect. 3 using direct survey estimates of Kcal, Protein and Fat as the three response variables and the four principal component scores P11, P12, P21, P22 with some other selected auxiliary variables from the 2011 Population Census data as suitable covariates. The final selected model included four covariates namely proportional scheduled caste population (SC), literacy rate (Lit), index for main worker population (P11) and index for marginal worker population (P21). Table 2 present the estimated regression parameters for the three response variables Kcal, Protein and Fat. Noting the signs of the estimates of regression parameters we conclude that districts having larger population proportion in covariates SC, Lit, and P11 and smaller population proportion in P21 covariate have greater Kcal intake. On the other hand, districts having larger population proportion in Lit, and P11 and smaller population proportion in covariates SC and P21 have greater Protein and Fat intake. The variance component parameter estimates for MFH model are given by $\hat{\sigma}^2_{u1} = 6333.90$, $\hat{\sigma}^2_{u2} = 7.74$ and $\hat{\sigma}^2_{u3} = 20.04$. The MFH model is employed using four significant covariates to generate the estimates of disparities in food consumption and nutrition level (i.e. disparities in Kcal, Protein and Fat intake) at district level for rural areas in Uttar Pradesh. The 2011–2012 HCES data of NSSO 68th round and the 2011 Population Census data of India is utilized for this analysis. In what follows, some crucial diagnostic measures are described to examine the model assumptions and validate the empirical performances of the MFH model. Following Brown et al. (2001), two types of diagnostics measures are employed to verify model assumptions viz. (1) the model diagnostics, and (2) the small area estimates diagnostics. The other diagnostic measures are used to validate the reliability of the model-based multivariate small area estimates of Kcal, Protein and fat obtained by MFH model.

4.2 Diagnostic Measures

Now for each of the target variable, the corresponding random area specific effects under MFH model given in (3) are assumed to be normally distributed with zero mean and fixed variance σ^2_{um}, $m = 1, 2, 3$. The district level residuals are expected to be randomly distributed around zero if the assumptions of the underlying model are satisfied. The normality assumption can be examined by using normal probability (q-q) plot. Figure 1 demonstrates the normal probability (Q–Q) plots of district-level residuals for the three target variables viz. Kcal, Protein and Fat. Further, the Shapiro–Wilk test was also performed to assess the normality assumption of the district specific random effects. If a dataset does not possess

Table 2 Regression parameters, standard error and *p*-values for Kcal, Protein and Fat

Variables		Intercept	SC	Lit	P11	P21
Kcal	Estimate	1601.10	314.13	201.66	1473.70	−352.42
	Standard error	203.35	247.52	211.94	410.34	132.96
	p-value	0.0000	0.0204	0.0341	0.0003	0.0080
Protein	Estimate	35.92	−0.29	16.98	66.22	−20.54
	Standard error	6.46	8.19	6.89	13.07	4.32
	p-value	0.0000	0.0972	0.0137	0.0000	0.0000
Fat	Estimate	20.49	−31.79	37.05	64.65	−49.73
	Standard error	9.22	11.99	9.73	18.96	6.39
	p-value	0.0262	0.0080	0.0001	0.0007	0.0000

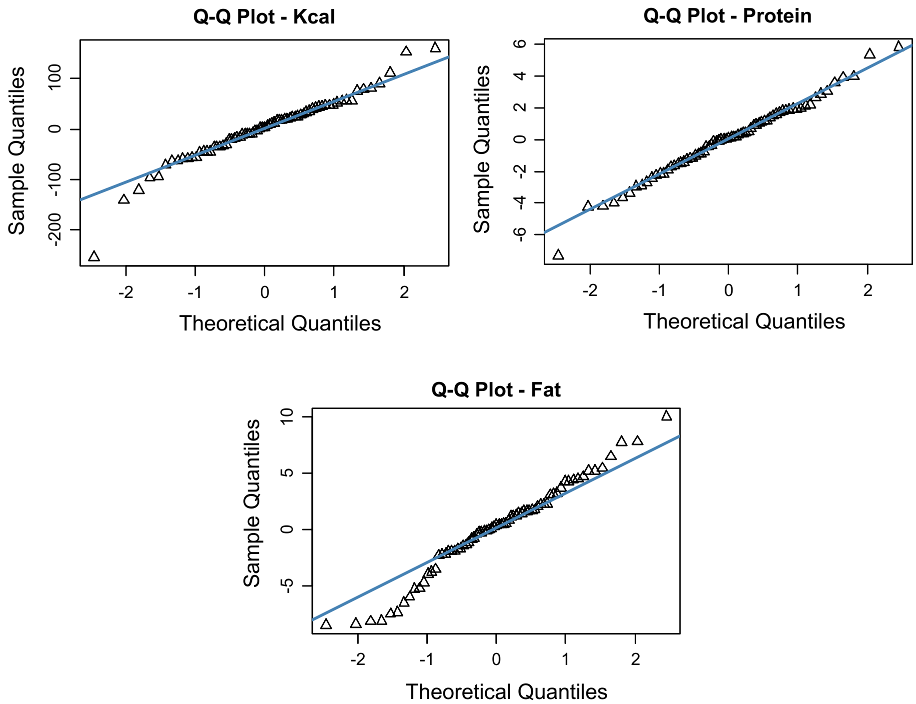

Fig. 1 Normal q-q plots of the district-level residuals for Kcal, Protein and Fat

normality then the p-value of the Shapiro–Wilk test will be $< <0.05$. The Shapiro–Wilk test is executed using the R function *shapiro.test()* and the summary of the test is given in Table 3. The p-values of the Shapiro–Wilk test were 0.475, 0.722 and 0.112 for Kcal, Protein and Fat respectively. Moreover, the Q–Q plot also provide evidence in favor of the model assumption of normality and the Shapiro–Wilk p-values are greater than 0.05 which leads to the conclusion that the district specific random effects are expected to be normally distributed.

The reliability as well as the validity of the model-based multivariate small area estimates are evaluated by considering a set of commonly used diagnostics measures. Following Brown et al. (2001) and Chandra et al. (2011), these diagnostics are based on the argument that model-based small area estimates should be (a) consistent with unbiased direct survey estimates, i.e., they should provide an approximation to the direct survey estimates that is consistent with these values being “close” to the expected values of the direct estimates; and (b) more precise than direct survey estimates, as evidenced by lower mean squared error estimates, i.e., the model-based small area estimates should have mean

Table 3 Summary of Shapiro–Wilk test along with p-values of district-level residuals obtained under MFH model for Kcal, Protein and Fat

Variables	Test statistic value	p-value
Kcal	0.985	0.634
Protein	0.988	0.797
Fat	0.984	0.526

squared errors significantly lower than the variances of corresponding direct survey estimates. We have selected the following measures viz. the bias diagnostic, the percent coefficient of variation (CV) diagnostic and the 95 percent confidence interval (CI) diagnostic. In addition, we implemented a calibration diagnostic where the model-based estimates are aggregated to higher level and compared with direct survey estimates at this level, Chandra et al. (2011). Here direct estimates are defined as the survey weighted direct estimates.

The bias diagnostic examine the validity whereas the CV and CI examine the precision of the model-based multivariate small area estimates. The bias diagnostic measure is established following the idea of Chandra et al. (2011). Being unbiased of the true values of the target population, the direct survey estimates' regression on the true population values should appear to be linear and relate to the identity line. The regression of the direct survey estimates on the model-based small area estimates should be analogous if the model-based estimates are adjacent to these true population values. Therefore, the direct survey estimates in the y-axis vs. model-based estimates in the x-axis are plotted and we observed the departure of the model-based estimates from the fitted values of the regression line. The bias diagnostic plots are given in Fig. 2 in which direct survey estimates (Y-axis) are plotted against corresponding model based MFH estimates (X-axis) and tested for divergence of the fitted least squares regression line (thick line) from the line of equality $Y=X$ line (thin line). Figure 2 reveals that the model-based small area estimates are not as much of extreme to the direct survey estimates, indicating the typical SAE outcome of shrinking more extreme values towards the average and the R^2 value were given by 0.72, 0.74 and

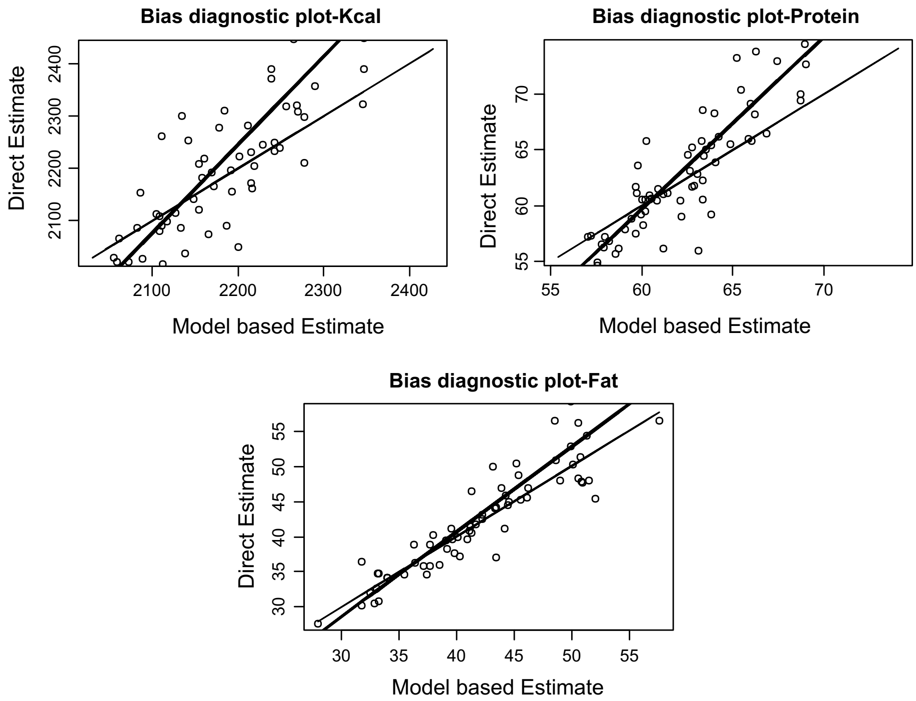

Fig. 2 Bias diagnostic plot with $y=x$ line (thin line) and regression line (solid line) for Kcal, Protein and Fat consumption for rural areas in Uttar Pradesh: Model based MFH estimates versus direct estimates

0.77 for the target variable Kcal, Protein and Fat respectively. Overall, this bias diagnostic measures indicate that the model based multivariate small area estimates likely to be consistent with direct survey estimates.

Next, we examined the magnitude to which the model-based multivariate small area estimates of Kcal, protein and Fat improved in precision than the UFH and direct estimates. Model-based multivariate small area estimates with small CVs are considered reliable. Table 4 presents a summary of percentage CVs of the direct estimates and the model-based multivariate small area estimates of the target variables Kcal, Protein and Fat. Figure 3 reports the District-wise root MSE while Fig. 4 presents the District-wise CV for the direct and MFH estimates for the all the three target variables. The CVs of the direct estimates are larger than the model based estimates for Kcal, Protein and Fat. Table 4 and Fig. 4, clearly indicate that direct estimates of all the three target variables are truly unstable with CVs ranging from 1.88 to 8.80% with a mean value of 3.87% for Kcal, 1.99–8.14% with a mean value of 3.94% for Protein and 3.42–18.02% with a mean value of 7.07% for Fat. On the other hand, the CVs of the model based estimates vary from 1.59 to 3.22% with a mean value of 2.40% for Kcal, 1.61–3.78% with a mean value of 2.44% for Protein and 2.98–8.56% with a mean value of 5.26% for Fat. The relative performance of the model based multivariate small area estimates for all the target variables has improved with decreasing sample sizes of the districts when compared to the direct estimates. Thus, these model based estimates are more precise and reliable and indicate the disparity in food and nutrition intake level much better than the direst estimates. Figure 5 displays the 95 percent confidence intervals (CIs) produced by direct and model based estimates while the corresponding widths of CIs are demonstrated in Fig. 6. It can be established from Figs. 5 and 6 that 95% CIs of the model based estimates are much narrower than that of the direct estimates. In addition, we also investigated the aggregation property of the district level estimates generated by model-based SAE method at higher level of aggregation (e.g., Regional and State level). The state-level estimates of Kcal, protein and fat is derived by

$$\hat{Y}_i = \sum_{j=1}^{D} N_j \hat{Y}_{ij} \Big/ \sum_{j=1}^{D} N_j \quad i = 1,\ 2,\ 3 \text{ and } j = 1, \ldots, D$$

where $\hat{Y}_{ij}$ denote the estimate of Kcal, protein and fat intake for $i = 1,\ 2,\ 3$ and district j with N_j being the population size of the jth district. We grouped the districts in four regions viz. Eastern, Western, Central, and Southern regions and examined the aggregation property. Regional and state level estimates of Kcal, protein and fat intake are reported in

Table 4 Summary of the distribution of percentage coefficients of variation (% CV) for the direct and model-based estimates of Kcal, Protein and Fat

Values	Kcal			Protein			Fat		
	Direct	UFH	MFH	Direct	UFH	MFH	Direct	UFH	MFH
Minimum	1.88	1.82	1.59	1.99	1.94	1.61	3.42	3.31	2.98
Q1	3.01	2.79	2.18	3.09	2.92	2.16	5.38	5.04	4.35
Median	3.53	3.22	2.39	3.67	3.39	2.40	6.55	5.99	5.10
Mean	3.87	3.38	2.40	3.94	3.52	2.44	7.07	6.20	5.26
Q3	4.36	3.74	2.61	4.58	4.08	2.69	8.22	7.05	6.05
Maximum	8.80	5.44	3.22	8.14	6.11	3.78	18.02	10.29	8.56

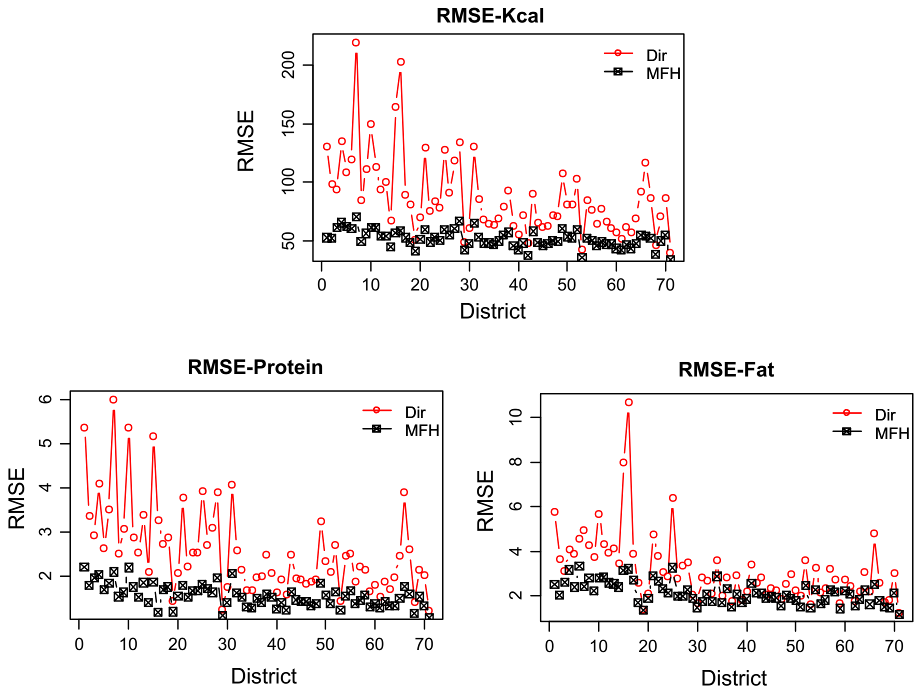

Fig. 3 District specific (increased sample size) Root MSE of direct and MFH estimators of Kcal, Protein and Fat in Uttar Pradesh

Table 5. When we compare the model-based SAE estimates with the direct estimates, we found that the SAE estimates are very close to the direct estimates in both the state and regional level.

Figures 7, 8, 9 demonstrate three maps indicating the MFH estimates of Kcal, Protein and Fat in different districts in the rural areas of Uttar Pradesh. These map provides the district-wise degree of inequality and reveals the distribution of the consumption of the food and nutrition intake. The results given in Tables 6, 7, 8 supplement these maps and report the district-wise direct and MFH estimates along with their 95% confidence interval and CV. The results of Kcal estimates indicates that the eastern region of Uttar Pradesh are having a low level of calorie intake while central part of Uttar Pradesh indicate highest level of calorie intake followed by western region. In case of Protein and Fat consumption, the results indicate an east–west divide in the distribution. For instance, western part of Uttar Pradesh seems to have high level of Protein and Fat intake while Eastern part indicate low level of Protein and Fat intake. These results may provide useful information to the policy maker for effective policy formulation and financial resolutions.

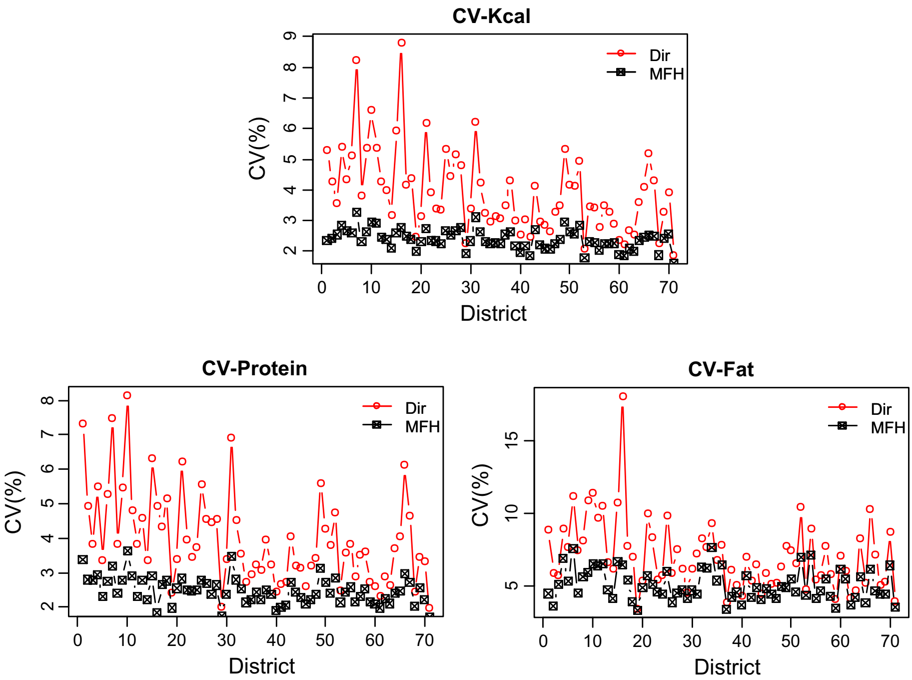

Fig. 4 District specific percentage coefficient of variation (CV) of direct and MFH estimators of Kcal, Protein and Fat in Uttar Pradesh. Districts are arranged in increasing order of sample size

5 Conclusions

In this paper, we initially summarize the empirical best linear unbiased predictor under multivariate Fay-Harriot model (MFH) for small area means. We then applied the MFH method in the 2011–2012 HCES data of NSSO, India to estimate the disparities in food consumption and nutrition level and to produce spatial maps related to Kcal, Protein and Fat intake in the districts of rural areas in Uttar Pradesh, India. We used the 2011 Population Census of India to collect the auxiliary variables for this analysis. Efficient estimation of correlated measures like food insecurity, nutritional consumption disparities are often required multivariate modelling approach which takes into account for the correlation between the target variables. In this study, all the target variables are jointly modelled using multivariate SAE to capture the inherent correlation between them. The improvement over univariate method based estimation is achieved in terms of MSE and CV of the district level estimates of Kcal, Protein and Fat intake in rural sector of Uttar Pradesh, India. The nutrition intake across the districts of Uttar Pradesh can help to stimulate the discussion about the drivers of hunger in this state. The empirical results so obtained, were assessed by various diagnostic measures and revealed that the model-based multivariate SAE method defined by MFH provide significant gains in efficiency in obtaining district level estimates of Kcal. Protein and Fat which in turn measures the disparities in food consumption and nutrition level. The MFH estimates based spatial maps indicate the evidence of unequal distribution of food consumption and nutrition level across the districts of rural areas of Uttar Pradesh, India.

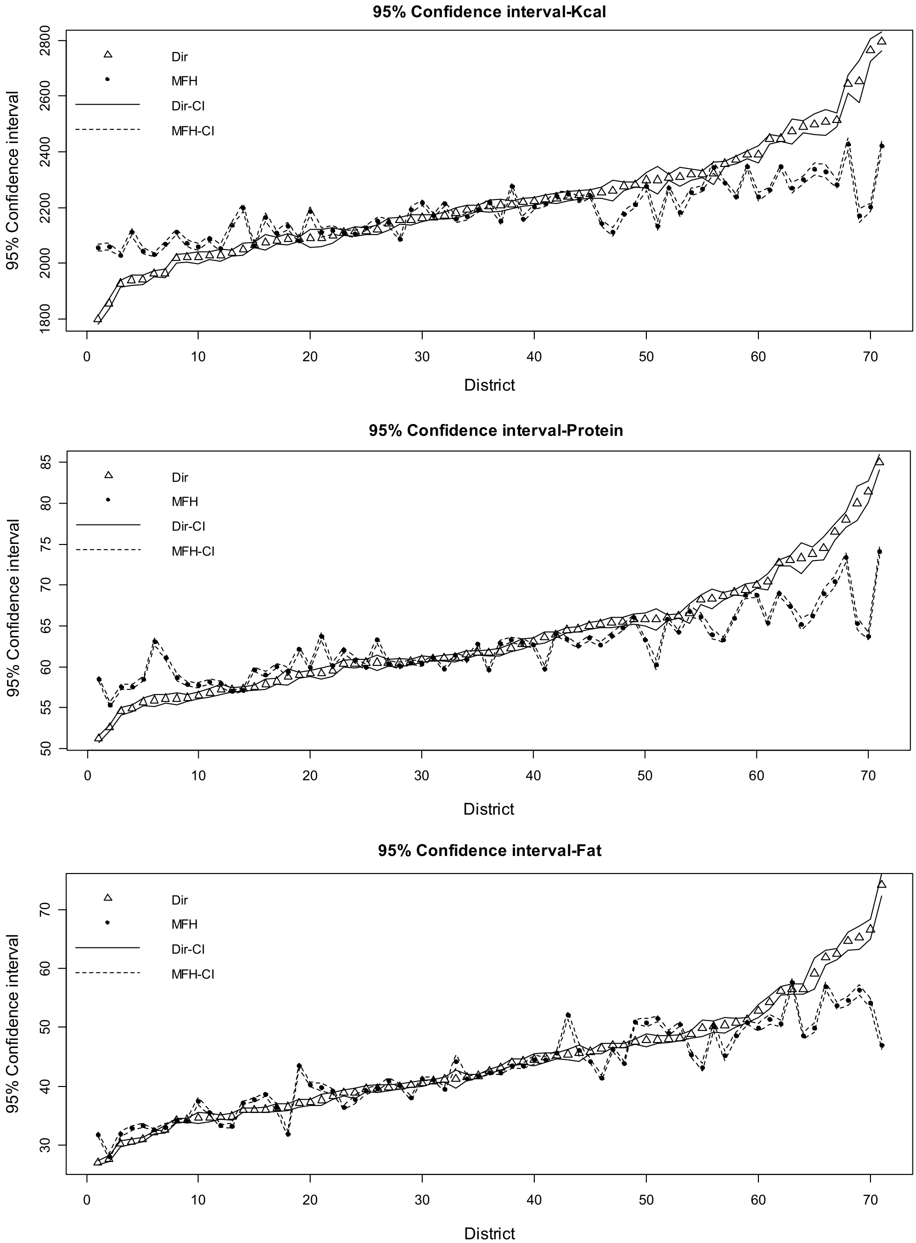

Fig. 5 District-wise 95% nominal confidence interval for the direct and model based MFH estimates for the Kcal, Protein and Fat intake in Uttar Pradesh. Districts are arranged in increasing order of direct estimates

This analysis undoubtedly established the advantages of SAE approach to deal with the problem of small sample sizes in obtaining precise and cost effective disaggregate or local level estimates along with the confidence intervals from existing survey data. Moreover, this analysis also illustrates the benefit of using multivariate small area

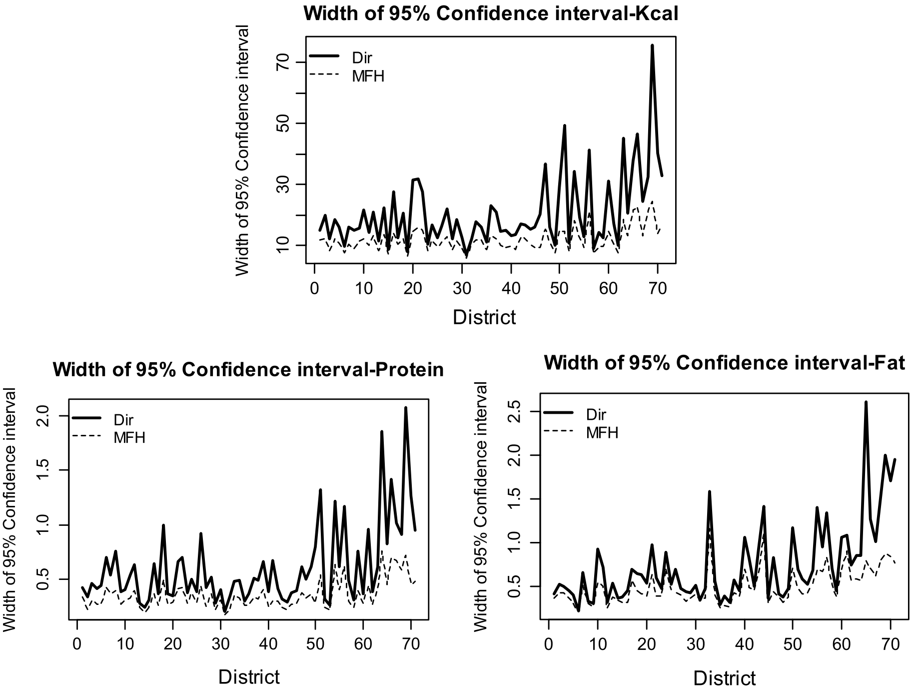

Fig. 6 District-wise width of 95% nominal confidence interval for the direct and model based MFH estimates for the Kcal, Protein and Fat intake in Uttar Pradesh. Districts are arranged in increasing order of direct estimates

Table 5 Aggregated level estimates of Kcal, Protein and Fat generated by different SAE methods

Region	Kcal			Protein			Fat		
	Direct	FH	MFH	Direct	FH	MFH	Direct	FH	MFH
State	3264.75	3244.64	3214.01	93.60	92.94	92.01	64.73	63.45	62.87
Eastern	2732.56	2738.69	2750.24	76.28	76.40	76.67	46.79	46.65	46.93
Western	3579.29	3532.90	3476.95	104.70	103.23	101.63	81.32	78.25	77.38
Central	2777.57	2772.79	2743.07	78.78	78.71	77.73	54.55	53.26	52.56
Southern	7874.05	7735.85	7470.71	233.71	229.29	221.61	154.45	155.17	148.73

Estimates are aggregated over 71 districts at state level as well as four regional levels

estimation over the univariate case by modelling the target variables jointly through multivariate Fay Harriot models. This study also reveals that large proportion of the rural sector of Uttar Pradesh's population is undernourished and below the recommended calorie intake of Government of India. Therefore, there is a massive need to build up some accord on the standards for least calorie and nutrition intake necessity as these factors makes disarray with respect to the seriousness of craving and undernourishment. In India, the surveys conducted by NSSO, Government of India are aimed to produce national and state level estimates which does not reflect the actual scenario

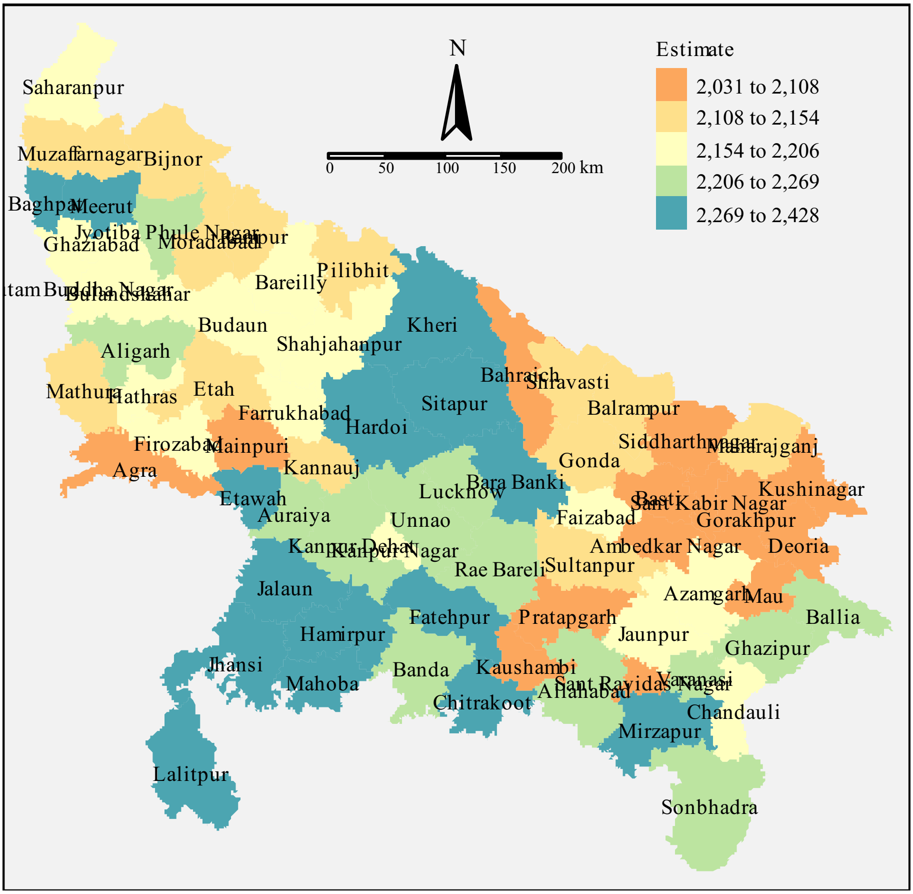

Fig. 7 Model based MFH estimates showing the spatial distribution of Kcal consumption by District in Uttar Pradesh

at the micro level (e.g. district level). The Government of India is pacing considerable emphasis on micro level planning for achieving a balanced economic development including food security. The district is an important domain for planning process in the country and therefore availability of district level statistics is vital to monitoring of policy and planning. This study produces reliable statistics at district level through SAE techniques that can be used in prioritization and targeting of efforts and investments. By implementing SAE technique, we are able to address the small sample size

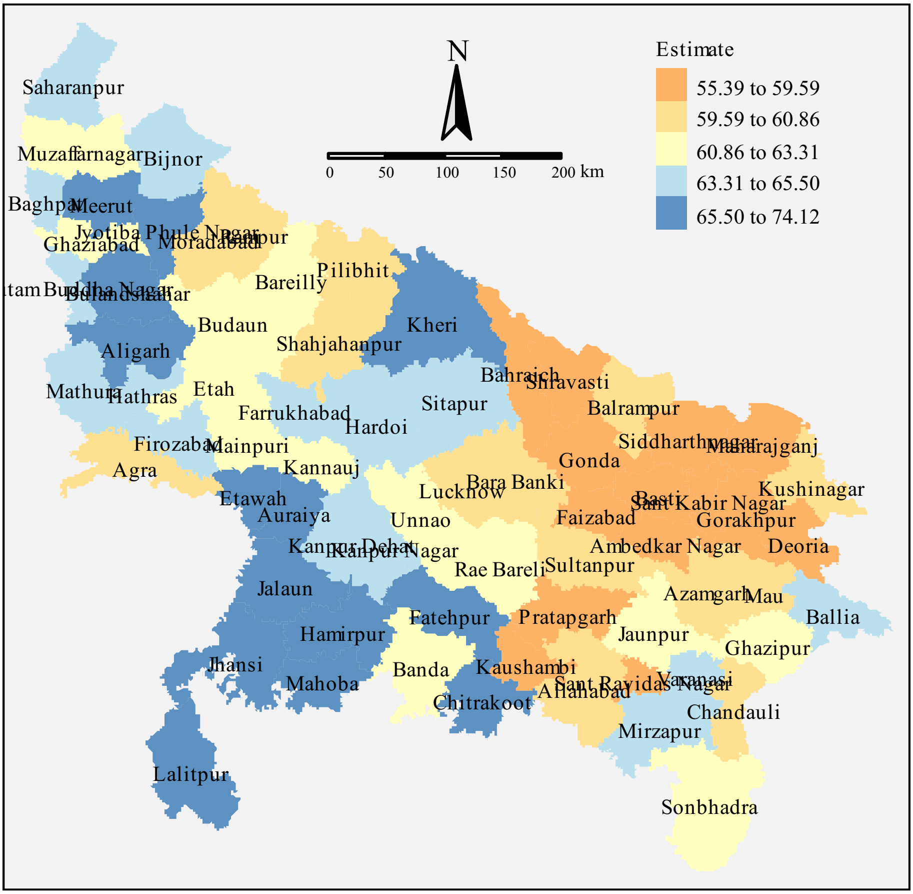

Fig. 8 Model based MFH estimates showing the spatial distribution of Protein consumption by District in Uttar Pradesh

problem in producing the cost effective and reliable disaggregate level estimates and confidence intervals from existing survey data by combining auxiliary information from different published sources with direct survey estimates. The estimates and spatial maps generated by this study can be used by different Departments and Ministries in Government of India as well as International organizations in their policy planning to formulate effective action plans relevant to sustainable development goal indicator 2.1.2—severity of food insecurity.

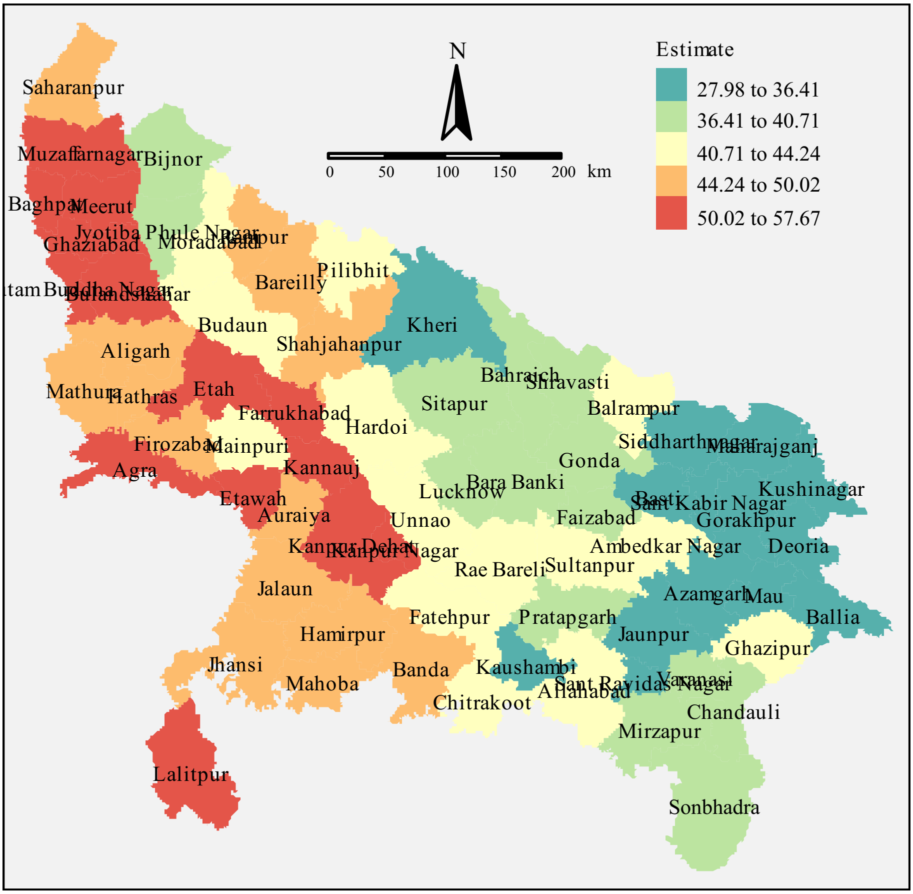

Fig. 9 Model based MFH estimates showing the spatial distribution of Fat consumption by District in Uttar Pradesh

Table 6 Direct and MFH estimates along with 95% confidence interval (95% CI) and percentage coefficient of variation (CV) of Kcal by district in rural areas of Uttar Pradesh

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Saharanpur	96	2155	2136	2174	4.32	2193	2182	2205	2.63
Muzaffarnagar	128	2037	2026	2049	3.29	2138	2130	2146	2.23
Bijnor	96	2115	2103	2128	3.00	2127	2118	2137	2.18
Moradabad	128	2110	2099	2120	2.91	2109	2101	2117	2.26
Rampur	64	2099	2071	2127	5.38	2118	2102	2133	2.92
Jyotiba Phule Nr	64	2205	2182	2228	4.26	2219	2206	2232	2.46
Meerut	64	2514	2489	2538	4.01	2284	2271	2297	2.39
Baghpat	32	2473	2428	2519	5.29	2269	2251	2288	2.35
Ghaziabad	64	2120	2104	2137	3.19	2154	2143	2165	2.11
Gautam B. Nr	32	2310	2276	2344	4.29	2184	2166	2202	2.41
Bulandshahr	96	2197	2186	2208	2.54	2191	2183	2200	1.95
Aligarh	96	2371	2356	2385	3.05	2238	2229	2248	2.17
Hathras	64	2764	2724	2804	5.93	2202	2188	2216	2.61
Mathura	64	2299	2249	2348	8.80	2135	2120	2149	2.78
Agra	96	1963	1953	1972	2.47	2031	2023	2038	1.86
Firozabad	63	2209	2188	2230	3.83	2154	2142	2166	2.32
Etah	64	2141	2120	2163	4.17	2148	2135	2161	2.48
Mainpuri	64	1856	1836	1875	4.37	2060	2048	2072	2.39
Budaun	96	2181	2163	2199	4.13	2158	2146	2169	2.71
Bareilly	95	2277	2261	2293	3.50	2178	2167	2189	2.54
Pilibhit	64	2081	2068	2094	2.46	2109	2098	2119	1.99
Shahjahanpur	96	2223	2210	2236	2.95	2201	2191	2211	2.21
Kheri	128	2447	2437	2457	2.38	2347	2339	2354	1.88
Sitapur	128	2357	2348	2366	2.21	2289	2282	2296	1.86
Hardoi	128	2308	2297	2318	2.69	2270	2262	2278	2.08
Unnao	96	2162	2150	2175	2.87	2216	2207	2226	2.08
Lucknow	64	2233	2216	2250	3.14	2243	2230	2256	2.32
Rae Bareli	128	2282	2272	2292	2.53	2211	2203	2219	1.99
Farrukhabad	64	2090	2058	2122	6.19	2187	2172	2202	2.73
Kannauj	64	1939	1920	1958	3.93	2112	2100	2124	2.34
Etawah	64	2490	2469	2511	3.38	2300	2287	2313	2.33
Auraiya	64	2319	2299	2338	3.37	2256	2244	2269	2.25
Kanpur Dehat	64	2390	2358	2421	5.33	2239	2224	2254	2.67
Kanpur Nagar	64	2050	2028	2072	4.44	2201	2187	2214	2.52
Jalaun	64	2298	2269	2327	5.16	2277	2262	2291	2.67
Jhansi	64	2796	2763	2829	4.81	2422	2406	2439	2.77
Lalitpur	32	2643	2610	2675	3.57	2428	2406	2449	2.54
Hamirpur	32	2506	2460	2553	5.39	2330	2307	2353	2.83
Mahoba	32	2499	2461	2536	4.35	2338	2316	2359	2.68
Banda	64	2172	2160	2184	2.27	2215	2204	2225	1.92
Chitrakoot	32	2323	2282	2364	5.12	2345	2324	2366	2.61

Table 6 (continued)

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Fatehpur	96	2389	2377	2402	2.66	2347	2337	2356	2.06
Pratapgarh	128	1928	1916	1940	3.62	2031	2022	2039	2.36
Kaushambi	64	1798	1783	1813	3.39	2057	2045	2069	2.33
Allahabad	128	2249	2233	2265	4.11	2243	2233	2252	2.46
BaraBanki	96	2210	2195	2224	3.28	2277	2266	2287	2.25
Faizabad	63	2075	2047	2102	5.37	2165	2151	2179	2.62
Ambedkar Nr	96	2027	2013	2041	3.50	2088	2078	2098	2.39
Sultanpur	128	2254	2233	2274	5.18	2143	2133	2152	2.53
Bahraich	96	2021	2000	2043	5.34	2059	2047	2071	2.96
Shrawasti	64	2091	2059	2123	6.23	2112	2095	2128	3.11
Balrampur	63	2261	2224	2297	6.61	2112	2096	2127	2.94
Gonda	128	2018	2003	2033	4.32	2113	2104	2122	2.50
Siddharthnagar	96	1941	1924	1957	4.18	2045	2034	2056	2.64
Basti	96	1964	1948	1980	4.14	2069	2058	2080	2.57
Sant Kabir Nr	64	2029	2008	2050	4.23	2055	2042	2068	2.62
Mahrajganj	96	2086	2065	2106	4.94	2134	2122	2146	2.83
Gorakhpur	128	2087	2079	2095	2.26	2083	2076	2090	1.87
Kushinagar	128	2154	2141	2166	3.29	2087	2078	2095	2.42
Deoria	96	2067	2058	2075	2.07	2062	2055	2069	1.77
Azamgarh	128	2220	2204	2235	3.92	2160	2151	2170	2.56
Mau	64	2114	2097	2131	3.25	2105	2093	2117	2.30
Ballia	96	2447	2429	2464	3.48	2264	2254	2275	2.32
Jaunpur	128	2166	2159	2173	1.88	2172	2166	2178	1.60
Ghazipur	127	2231	2217	2244	3.50	2215	2206	2223	2.24
Chandauli	64	2192	2175	2208	2.98	2169	2157	2181	2.24
Varanasi	96	2246	2230	2261	3.44	2229	2219	2239	2.27
Bhadohi	64	2021	2006	2037	3.14	2073	2061	2084	2.28
Mirzapur	96	2321	2308	2334	2.80	2269	2259	2278	2.04
Sonbhadra	64	2240	2223	2257	3.08	2248	2236	2261	2.24
Kanshiram Nr	32	2653	2577	2729	8.25	2171	2147	2196	3.26

Nr Nagar

Table 7 Direct and MFH estimates along with 95% confidence interval (95% CI) and percentage coefficient of variation (CV) of Protein by district in rural areas of Uttar Pradesh

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Saharanpur	96	62.30	62.74	63.33	3.97	63.33	2.47	2.24	2.51
Muzaffarnagar	128	59.00	60.11	62.15	3.62	62.15	2.13	1.98	2.53
Bijnor	96	63.90	63.67	64.02	3.24	64.02	2.07	1.93	2.37
Moradabad	128	60.60	60.47	60.54	2.73	60.54	1.66	1.58	2.15
Rampur	64	59.50	59.71	60.16	4.82	60.16	2.87	2.53	2.91
Jyotiba Phule Nr	64	65.80	65.50	66.02	3.84	66.02	2.53	2.27	2.30
Meerut	64	73.80	71.11	66.26	4.58	66.26	3.38	2.83	2.79
Baghpat	32	73.30	69.76	65.19	7.31	65.19	5.36	3.79	3.39
Ghaziabad	64	61.80	62.54	62.85	3.36	62.85	2.08	1.93	2.23
Gautam B. Nr	32	68.30	67.21	64.00	4.94	64.00	3.37	2.83	2.80
Bulandshahr	96	66.00	65.98	65.83	2.45	65.83	1.62	1.54	1.90
Aligarh	96	72.70	71.47	69.02	2.66	69.02	1.93	1.81	2.01
Hathras	64	81.40	71.28	63.75	6.33	63.75	5.16	3.62	2.92
Mathura	64	65.80	65.21	63.31	4.94	63.31	3.25	2.74	1.85
Agra	96	57.50	58.08	59.68	2.74	59.68	1.58	1.51	2.06
Firozabad	63	65.40	64.96	63.81	3.84	63.81	2.51	2.27	2.40
Etah	64	62.80	63.04	63.03	4.34	63.03	2.73	2.42	2.67
Mainpuri	64	55.90	58.26	63.12	5.16	63.12	2.88	2.53	2.77
Budaun	96	61.50	61.22	60.87	4.05	60.87	2.49	2.27	2.71
Bareilly	95	65.20	64.50	62.73	3.07	62.73	2.00	1.87	2.23
Pilibhit	64	59.20	59.39	59.96	2.43	59.96	1.44	1.39	1.98
Shahjahanpur	96	60.50	60.57	59.99	3.21	59.99	1.94	1.83	2.43
Kheri	128	69.10	68.22	65.96	2.60	65.96	1.80	1.70	2.09
Sitapur	128	66.20	65.71	64.22	2.32	64.22	1.53	1.48	1.98
Hardoi	128	64.50	64.29	63.40	2.89	63.40	1.86	1.76	2.26
Unnao	96	60.40	60.97	62.09	3.15	62.09	1.91	1.79	2.29
Lucknow	64	60.40	60.71	60.85	3.41	60.85	2.06	1.93	2.55
Rae Bareli	128	64.60	64.22	62.55	2.65	62.55	1.71	1.63	2.11
Farrukhabad	64	60.50	62.29	63.34	6.23	63.34	3.77	3.05	2.83
Kannauj	64	56.10	57.87	61.15	3.97	61.15	2.22	2.05	2.49
Etawah	64	73.00	71.41	67.45	3.47	67.45	2.53	2.28	2.46
Auraiya	64	68.20	67.71	66.18	3.73	66.18	2.54	2.29	2.51
Kanpur Dehat	64	70.40	69.02	65.42	5.58	65.42	3.93	3.13	2.78
Kanpur Nagar	64	59.20	61.27	63.79	4.56	63.79	2.70	2.40	2.68
Jalaun	64	69.40	69.04	68.72	4.46	68.72	3.10	2.67	2.37
Jhansi	64	85.00	79.94	74.12	4.57	74.12	3.89	3.19	2.64
Lalitpur	32	76.50	74.71	70.40	3.83	70.40	2.93	2.59	2.78
Hamirpur	32	74.50	72.24	68.93	5.50	68.93	4.10	3.25	2.95
Mahoba	32	78.00	76.02	73.36	3.37	73.36	2.63	2.37	2.30
Banda	64	61.70	62.05	62.77	2.00	62.77	1.23	1.20	1.75
Chitrakoot	32	66.50	67.34	66.83	5.28	66.83	3.51	2.92	2.75

Table 7 (continued)

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Fatehpur	96	70.00	69.79	68.74	2.62	68.74	1.83	1.74	2.08
Pratapgarh	128	52.60	53.60	55.39	3.72	55.39	1.96	1.83	2.40
Kaushambi	64	51.20	52.94	58.48	3.41	58.48	1.75	1.67	2.38
Allahabad	128	60.50	61.09	60.43	4.07	60.43	2.46	2.22	2.47
BaraBanki	96	58.20	59.04	60.08	3.20	60.08	1.86	1.75	2.21
Faizabad	63	56.10	57.56	58.72	5.47	58.72	3.07	2.63	2.78
Ambedkar Nr	96	56.20	56.82	57.89	3.44	57.89	1.93	1.81	2.38
Sultanpur	128	63.60	62.26	59.77	6.12	59.77	3.89	3.10	2.96
Bahraich	96	57.90	57.94	59.05	5.60	59.05	3.24	2.79	3.12
Shrawasti	64	58.80	58.96	59.45	6.92	59.45	4.07	3.28	3.47
Balrampur	63	65.80	62.40	60.24	8.14	60.24	5.35	3.81	3.64
Gonda	128	55.70	56.77	58.52	4.66	58.52	2.60	2.33	2.72
Siddharthnagar	96	54.60	55.40	57.58	4.29	57.58	2.34	2.15	2.73
Basti	96	54.90	55.60	57.57	3.79	57.57	2.08	1.93	2.40
Sant Kabir Nr	64	57.20	57.28	57.94	4.52	57.94	2.59	2.32	2.80
Mahrajganj	96	56.80	57.62	58.18	4.74	58.18	2.69	2.41	2.84
Gorakhpur	128	57.30	57.28	57.18	2.45	57.18	1.40	1.36	2.03
Kushinagar	128	61.70	60.94	59.63	3.46	59.63	2.14	1.99	2.57
Deoria	96	57.20	57.17	57.03	2.50	57.03	1.43	1.38	2.14
Azamgarh	128	61.10	60.72	59.73	3.33	59.73	2.03	1.90	2.21
Mau	64	60.50	60.15	60.18	3.56	60.18	2.15	2.00	2.53
Ballia	96	68.60	66.90	63.31	3.59	63.31	2.46	2.23	2.43
Jaunpur	128	61.00	61.03	61.16	1.99	61.16	1.22	1.18	1.72
Ghazipur	127	63.10	63.15	62.65	3.51	62.65	2.22	2.04	2.30
Chandauli	64	60.90	60.88	60.40	2.75	60.40	1.68	1.60	2.14
Varanasi	96	65.50	65.36	64.84	3.84	64.84	2.51	2.27	2.58
Bhadohi	64	56.50	56.73	57.78	2.97	57.78	1.68	1.60	2.22
Mirzapur	96	65.00	64.76	63.54	2.89	63.54	1.88	1.77	2.16
Sonbhadra	64	61.10	61.44	61.41	3.25	61.41	1.98	1.86	2.43
Kanshiram Nr	32	80.00	70.48	65.27	7.49	65.27	5.99	3.92	3.20

Nr Nagar

Table 8 Direct and MFH estimates along with 95% confidence interval (95% CI) and percentage coefficient of variation (CV) of Fat by district in rural areas of Uttar Pradesh

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Saharanpur	96	48.00	48.29	49.06	6.11	49.06	2.93	2.63	4.22
Muzaffarnagar	128	47.60	48.69	50.97	5.76	50.97	2.74	2.49	4.24
Bijnor	96	35.90	36.78	37.13	5.06	37.13	1.82	1.74	4.52
Moradabad	128	41.70	41.65	41.71	4.06	41.71	1.69	1.63	3.43
Rampur	64	44.50	43.51	44.47	9.68	44.47	4.31	3.49	6.36
Jyotiba Phule Nr	64	37.60	40.88	39.84	10.50	39.84	3.95	3.25	6.50
Meerut	64	62.50	58.20	53.78	6.60	53.78	4.12	3.36	4.72
Baghpat	32	65.30	59.75	56.36	8.83	56.36	5.77	4.17	4.45
Ghaziabad	64	56.50	55.31	57.67	6.15	57.67	3.47	2.99	4.14
Gautam B. Nr	32	61.90	58.03	57.03	5.89	57.03	3.65	3.10	3.56
Bulandshahr	96	51.30	50.82	50.81	4.32	50.81	2.22	2.07	3.66
Aligarh	96	48.80	47.84	45.42	6.96	45.42	3.40	2.92	5.68
Hathras	64	74.30	52.95	47.06	10.73	47.06	7.97	4.61	6.67
Mathura	64	59.20	47.86	49.96	18.02	49.96	10.67	4.93	6.42
Agra	96	48.20	48.05	50.57	5.37	50.57	2.59	2.37	4.17
Firozabad	63	52.80	50.86	49.98	8.11	49.98	4.28	3.44	5.58
Etah	64	50.20	49.85	50.16	7.70	50.16	3.87	3.23	5.40
Mainpuri	64	37.10	39.50	43.42	6.95	43.42	2.58	2.37	3.87
Budaun	96	44.00	43.66	43.40	6.46	43.40	2.84	2.59	4.84
Bareilly	95	45.80	45.52	44.27	3.87	44.27	1.77	1.70	3.37
Pilibhit	64	40.80	41.02	41.14	3.42	41.14	1.40	1.36	3.28
Shahjahanpur	96	46.90	46.53	46.28	4.43	46.28	2.08	1.96	4.04
Kheri	128	38.80	39.32	36.35	7.05	36.35	2.73	2.49	6.14
Sitapur	128	40.20	39.76	37.99	6.01	37.99	2.42	2.26	5.47
Hardoi	128	43.10	42.92	42.26	4.13	42.26	1.78	1.71	3.64
Unnao	96	39.70	39.92	40.99	5.88	40.99	2.33	2.17	4.78
Lucknow	64	39.50	39.24	39.14	5.32	39.14	2.10	1.99	4.83
Rae Bareli	128	46.90	45.16	43.93	4.67	43.93	2.19	2.06	4.17
Farrukhabad	64	47.80	49.35	50.83	9.96	50.83	4.76	3.69	5.67
Kannauj	64	45.40	47.71	52.12	8.29	52.12	3.77	3.17	5.10
Etawah	64	56.20	54.14	50.62	5.42	50.62	3.05	2.71	4.56
Auraiya	64	50.80	50.04	48.67	5.70	48.67	2.89	2.61	4.40
Kanpur Dehat	64	64.70	55.49	54.58	9.85	54.58	6.37	4.26	5.97
Kanpur Nagar	64	47.90	48.35	51.49	5.85	51.49	2.80	2.53	3.85
Jalaun	64	44.90	45.52	44.52	7.47	44.52	3.36	2.92	4.47
Jhansi	64	56.50	54.49	48.54	6.16	48.54	3.48	3.05	4.72
Lalitpur	32	54.30	53.31	51.37	5.74	51.37	3.12	2.79	5.08
Hamirpur	32	45.60	46.73	46.16	8.92	46.16	4.07	3.37	6.85
Mahoba	32	50.40	49.37	45.18	7.68	45.18	3.87	3.26	5.29
Banda	64	45.20	45.47	45.60	4.53	45.60	2.05	1.94	4.14
Chitrakoot	32	41.20	44.19	44.22	11.13	44.22	4.58	3.62	7.55

Table 8 (continued)

District	Sample size	Direct				MFH			
		Estimate	95% CI		CV	Estimate	95% CI		CV
			Lower	Upper			Lower	Upper	
Fatehpur	96	44.20	44.49	43.47	5.09	43.47	2.25	2.11	4.39
Pratapgarh	128	37.20	36.93	40.29	8.22	40.29	3.06	2.71	5.59
Kaushambi	64	26.90	28.08	31.72	6.13	31.72	1.65	1.59	4.67
Allahabad	128	42.50	42.28	42.26	5.19	42.26	2.21	2.06	3.83
BaraBanki	96	35.90	36.67	37.67	5.20	37.67	1.87	1.78	4.14
Faizabad	63	34.60	35.85	37.43	10.85	37.43	3.75	3.14	5.90
Ambedkar Nr	96	40.50	39.89	41.32	6.30	41.32	2.55	2.34	4.96
Sultanpur	128	46.40	42.37	41.36	10.30	41.36	4.78	3.67	6.12
Bahraich	96	38.30	37.88	39.15	7.71	39.15	2.95	2.67	4.85
Shrawasti	64	39.60	38.54	39.69	7.17	39.69	2.84	2.59	4.38
Balrampur	63	49.90	43.28	43.22	11.36	43.22	5.67	4.14	6.52
Gonda	128	36.00	36.70	38.56	7.15	38.56	2.57	2.36	4.62
Siddharthnagar	96	30.50	31.14	32.87	7.41	32.87	2.26	2.12	5.48
Basti	96	30.90	31.66	33.28	6.50	33.28	2.01	1.90	4.57
Sant Kabir Nr	64	32.60	32.62	33.04	8.22	33.04	2.68	2.44	6.30
Mahrajganj	96	34.60	34.42	35.53	10.40	35.53	3.60	3.10	6.95
Gorakhpur	128	34.20	34.08	34.02	5.01	34.02	1.71	1.65	4.41
Kushinagar	128	34.70	34.54	33.29	5.27	33.29	1.83	1.75	4.39
Deoria	96	34.10	34.18	34.04	4.76	34.04	1.62	1.57	4.32
Azamgarh	128	34.80	34.93	33.21	8.73	33.21	3.04	2.70	6.38
Mau	64	27.60	28.48	27.98	7.68	27.98	2.12	2.00	6.18
Ballia	96	36.40	37.33	31.81	8.91	31.81	3.24	2.84	7.10
Jaunpur	128	32.10	32.55	32.55	3.90	32.55	1.25	1.23	3.53
Ghazipur	127	41.50	42.07	41.23	7.76	41.23	3.22	2.81	5.48
Chandauli	64	38.80	38.24	37.71	9.31	37.71	3.61	3.07	7.61
Varanasi	96	39.90	41.12	40.10	5.43	40.10	2.17	2.04	4.09
Bhadohi	64	30.20	30.98	31.81	6.74	31.81	2.03	1.92	5.34
Mirzapur	96	41.10	40.86	39.52	5.69	39.52	2.34	2.17	4.61
Sonbhadra	64	36.30	36.44	36.43	7.84	36.43	2.85	2.58	6.42
Kanshiram Nr	32	66.70	58.17	54.13	7.40	54.13	4.93	3.82	4.47

Nr Nagar

References

Anjoy, P., Chandra, H., & Basak, P. (2019). Estimation of disaggregate-level poverty incidence in Odisha under area-level hierarchical bayes small area model. *Social Indicator Research, 144,* 251–273.

Benavent, R., & Morales, D. (2016). Multivariate Fay-Herriot models for small area estimation. *Computational Statistics and Data Analysis, 94,* 372–390.

Brown, G., Chambers, R., Heady, P., & Heasman, D. (2001). Evaluation of small area estimation methods: An application to unemployment estimates from the UK LFS. *Proceedings of statistics Canada symposium 2001*. Achieving data quality in a statistical agency: A methodological perspective.

Census. (2011). Primary census abstracts, registrar general of India, Ministry of home affairs, Government of India. http://www.censusindia.gov.in/2011census/population_enumeration.html

Chandra, H., Salvati, N., & Sud, U. C. (2011). Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India-an application of small area estimation technique. *Journal Applied Statistics, 38*(11), 2413–2432.
Chandra, H. (2013). Exploring spatial dependence in area level random effect model for disaggregate level crop yield estimation. *Journal of Applied Statistics, 40*(4), 823–842.
Chandra, H., Salvati, N., & Chambers, R. (2015). A spatially nonstationary Fay-Herriot model for small area estimation. *Journal of Survey Statistics and Methodology, 3*(2), 109–135.
Datta, G.S., Fay, R.E., & Ghosh, M. (1991). Hierarchical and empirical Bayes multivariate analysis in small area estimation. *In: Proceedings of bureau of the census 1991 annual research conference*, US bureau of the census, Washington, DC, 63–79.
Datta, G. S., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica, 10,* 613–627.
Datta, G., Kubokawa, T., Molina, I., & Rao, J. N. K. (2011). Estimation of mean squared error of model-based small area estimators. *TEST, 20*(2), 367–388.
Fay, R. E., & Herriot, R. (1979). Estimates of income for small places: An application of James stein procedures to census data. *Journal of the American Statistical Association, 74,* 269–277.
Fay, R. E. (1987). Application of multivariate regression of small domain estimation. In R. Platek, J. N. K. Rao, C. E. Särndal, & M. P. Singh (Eds.), *Small area statistics* (pp. 91–102). New York: Wiley.
González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2010). Small area estimation under Fay-Herriot models with nonparametric estimation of heteroscedasticity. *Statistical Modelling, 10*(2), 215–239.
Gopalan, C., Rama Sastri, B. V., & Balasubramanian, S. C. (1991). *Nutritive values of Indian foods*. Hyderabad: National Institute of Nutrition.
Government of India. (2014). *Nutritional intake in India, 2011–12*. 68th Round, Report No. 560, National Sample Survey Organisation (NSSO). Ministry of Statistics and Programme Implementation. New Delhi.
Islam, S., & Chandra, H. (2019). Small area estimation combining data from two surveys. *Communications in Statistics—Simulation and Computation*. https://doi.org/10.1080/03610918.2019.1588308.
Marhuenda, Y., Molina, I., & Morales, D. (2013). Small area estimation with spatio-temporal Fay-Herriot models. *Computational Statistics and Data Analysis, 58,* 308–325.
Morales, D., Pagliarella, M. C., & Salvatore, R. (2015). Small area estimation of poverty indicators under partitioned area-level time models. *SORT, 39*(1), 19–34.
MosPI, Government of India & WFP (2019). *Food and nutrition security analysis*, *India, 2019*. Ministry of Statistics and Programme Implementation & the World Food Programme. http://mospi.nic.in/sites/default/files/publication_reports/document%281%29.pdf
Prasad, N. G. N., & Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association, 85,* 163–171.
Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). New York: Wiley.
Rao, J. N. K., & Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *The Canadian Journal of Statistics*, *22*(4), 511–528.
UNDP. (2015). Sustainable Development Goals. https://www.undp.org/content/undp/en/home/sustainable-development-goals.html

Affiliations

Saurav Guha[1] · **Hukum Chandra**[1]

✉ Saurav Guha
saurav.iasri@gmail.com

Hukum Chandra
hchandra12@gmail.com

[1] ICAR-Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi, India