Social Indicators Research
https://doi.org/10.1007/s11205-018-2050-9

Estimation of Disaggregate-Level Poverty Incidence in Odisha Under Area-Level Hierarchical Bayes Small Area Model

Priyanka Anjoy[1] · **Hukum Chandra**[1] · **Pradip Basak**[1]

Accepted: 7 December 2018

Abstract
Sustainable development goal-1 of the United Nations is to end poverty in all its forms everywhere. The estimates of poverty related parameters obtained from large scale sample survey are often available at large domain level (e.g. state level). But, poverty rates are not uniformly distributed across the regions. The regional variations are masked in such large domain level estimates. However, for monitoring the progress of poverty alleviation programmes aimed at reduction of poverty often require micro or disaggregate level estimates. The traditional survey estimation approaches are not suitable for generating the reliable estimates at this level because of sample size problem. It is the main endeavor of Small Area Estimation (SAE) approach to produce micro level statistics with acceptable precision without incurring any extra cost and utilizing existing survey data. In this study, the Hierarchical Bayes approach of SAE has been applied to generate reliable and representative district level poverty incidence for the State of Odisha in India using the Household Consumer Expenditure Survey 2011–2012 data of National Sample Survey Office and linked with Population Census 2011. The results show the precise performance of model based estimates generated by SAE method to a greater extent than the direct survey estimates. A poverty map has also been produced to observe the spatial inequality in poverty distribution.

Keywords Small area estimate · Hierarchical Bayes · Poverty · Precision

✉ Priyanka Anjoy
anjoypriyanka90@gmail.com

Hukum Chandra
hchandra12@gmail.com

Pradip Basak
pradipbasak99@gmail.com

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

Published online: 14 December 2018

1 Introduction

A long stand complex societal issue like "Poverty" is the matter of huge global concern which demands relentless efforts at all levels to eradicate this social disease in all its forms and dimensions. Policy planning at national and cross-boundary level requires micro or regional level formulation and implementation to accelerate the progress of poverty alleviation programmes. In the line, there are many factors which converge to make poverty a complex and multidimensional phenomenon. Poverty is the state in which a person is not capable of maintaining socially acceptable living standards and the factors cumulatively determining this state are, lack of necessary money and materials for food, housing, clothes, land and other assets, absence of basic infrastructures like road, transport, water, health and sanitation facilities, lack of basic education or illiteracy. Moreover, in sociologic context poor people are those who are readily vulnerable to social exclusion, inhumane treatment, rudeness and also subjected to exploitation because they lack voice, power and independence (Sen 1981; Kabeer 1994; Narayana and Petesch 2002; Benjamin et al. 2013). According to United Nations (UN) facts for SDG-1, *767 million people live below the international poverty line of $1.90 a day and the overwhelming majority of people living below the poverty line belong to two regions: Southern Asia and sub-Saharan Africa. High poverty rates are often found in small, fragile, middle-income and conflict-affected countries, including China, India, Indonesia and Nigeria, are home to about half of the global poor.* Hence the goal set by UN is *by 2030, reduce at least by half the proportion of men, women and children of all ages living in poverty in all its dimensions* (United Nations 2017). To address the situation, substantial investment along with sound policy effort is required for taking poverty eradication actions at national, global as well as micro or local levels.

India has challenges of significant regional and social disparities in poverty. Moreover, poverty is concentrated in few states such as Bihar, Odisha, Uttar Pradesh, Jharkhand and Chhattisgarh. There are people who still live in streets and beg for food. Underprivileged children just leave school to earn food for livelihood; a large proportion of rural people live in unhygienic condition, even basic water, sanitation or medical facilities are not within their reach. Deprivation of women from equality in education is another cause of women backwardness and all these issues bind to make the poverty a complex phenomenon for India. About, half of the Indian population has agriculture-based economy but it is not substantial to feed their most of the basic needs which lead to the cases of farmers' suicides in a large number. Here, poverty has also intense coexistence with few social groups such as Scheduled Castes (SC) and Scheduled Tribes (ST). SDG approach in India to reduce the goal on national poverty require more attention to lessen poverty in poorer states as well as poverty gap also needed to be reduced amongst regions and also between different social groups; special consideration to women and children development in terms of their basic rights is also associated immediate concern. According to the Niti Aayog (then the Planning Commission) report in 2011–2012, India is home to about 269 million poor people with 216.5 million people residing in rural India which also manifests a significant gap in rural and urban poverty (Government of India 2017). The report has shown 25.7% poverty proportion in rural areas and 13.7% poverty proportion in urban areas. In view of dense rural poverty, necessary steps are demanded towards upliftment of rural people in terms of their basic livelihood, health, education, infrastructural facilities etc. Further, before going to the actual action calling poverty eradication programmes, measurement of poverty and its estimation is a crucial task for effective planning process.

In India, poverty estimates are produced separately for all the states for both rural and urban sectors based on Household Consumer Expenditure Survey (HCES) carried out by the National Sample Survey Office (NSSO), Government of India. The NSSO sampling designs used in rural and urban sectors are different. Indeed, estimation of rural poverty is more emphasized, as more than two-thirds of the population of the country lives in rural areas. The state specific poverty estimates provide an overall idea of poverty prone states and assists the administration in various policy formulation stages. However, it is to be noted that the estimates of poverty obtained from large scale sample survey at state and national level masks the regional level variations. As poverty rates are not uniformly distributed across the regions, state or national level estimates are not usable to represent the regional or micro level poverty incidences. Whereas, for monitoring the progress of poverty alleviation programmes aimed at reduction of poverty, often require local or disaggregate level estimates. But, in NSSO surveys, sample sizes are fixed in such a way that, reliable direct survey estimates (i.e. estimates obtained using domain-specific data and through traditional survey estimation method) can be obtained for planned domains (e.g., state) only, whereas sample sizes for unplanned domains (e.g., districts, taluk, block, municipalities, gram panchayats) either going to be very small or even zero for some domains. Such domains are also referred as small domains or “small areas” as the domain specific sample sizes are not enough to support reliable direct survey estimates (Rao 2003). Policy planners, administrators, government and other public agencies, private agency often require estimates at small domain levels for policy framing, fund disbursement, localized planning etc. It is the main endeavor of SAE approach to produce sound predictions of a target statistics for such small domains.

In India, several researchers have attempted to provide disaggregate level estimates of poverty using the HCES data across and within the states. Chaudhuri and Gupta (2009) have rendered the district level poverty estimates of India using HCES 2004–05 data of NSSO. However, the study acknowledged the limitation of large standard error observed for some districts which being the fundamental problem in determining local level estimates due to sample size problem. Coondoo et al. (2011) has described an approach for providing micro level poverty indices for two states of India namely, West Bengal and Madhya Pradesh. Their approach is based on subgroup decomposable property of poverty measure where sub-state level estimates are obtained by solving a system of linear equations. Major demerit of this approach is that it belongs to the class of synthetic indirect method. Synthetic estimators are known to be biased due to homogeneity assumption between the domains of interest. Chauhan et al. (2016) has studied intra and inter-regional disparities in poverty and inequality using three quinquennial rounds of HCES data of NSSO over two decades (1993–2012). Similar kind of investigation has been done by Mohanty et al. (2016) where the estimates are provided at district level within the region. However, in both of these studies poverty indicators are estimated fitting regression based fixed effect model. As a result, these estimates fail to represent dissimilarities across areas and the limitation is particularly being handled by SAE approach which takes into account the random area-specific effects and hence potentially explores the variability between areas. The rationale behind using SAE approach in the present study specifically motivated from the issue of glittering poverty in most of the parts of rural India and ineffectiveness of traditional direct estimation and synthetic estimation approach invoked in various studies stated above in measuring the poverty proportions at disaggregate or local levels. Recently Chandra et al. (2018a) employed SAE approach under an area-level small area model to generate the estimates of poverty incidence at district level in the State of Bihar in India using HCES 2011–2012 data of NSSO. This method of SAE is based on frequentist approach and uses an analytical expression of mean squared error (MSE) estimation. The MSE estimate is based on an approximation. This article, in particular, focuses on estimation of district-wise poverty incidence for the State of Odisha in India using the HCES 2011–2012 data of NSSO and linked with Population Census 2011 using Hierarchical Bayes (HB) approach of SAE and hence overcomes above described limitations.

Rest of the article is organized as follows. Next section describes about study area and poverty estimation approaches followed by description of survey and census data utilized in this study. Then some small area models are detailed which is implemented for poverty estimation along with HB approach for small area modeling. Empirical results are presented for various poverty estimation methods; finally the paper is concluded along with relevant remarks.

2 Study Area and Poverty Estimation Approach

The key consideration of the present paper is estimation of localized poverty incidence using appropriate small area model based methods. The poverty incidence is defined as the proportion of households with income below the poverty line, also referred as head count ratio (HCR). The HCR is a poverty indicator which measures the frequency of households under poverty line. The study area has been considered as Odisha state of India. Poverty incidence of rural Odisha is 35.69% against 25.70% in all India, according to the Niti Aayog report (then Planning Commission) during 2011–2012. It is the tenth largest state of union along with standing in fifth position in terms of poverty incidences. Poverty line (rural) of this state is Rs. 695, as set by Niti Aayog (then Planning Commission) during 2011–2012, which is lowest among all Indian states. There are various geographical or natural as well as social factors which has hindered the development of this state since many years. A large proportion of this state is under dense forest cover, hence prevalence of malaria is most common; people has developed forest based economy in many places as agriculture is being hampered by frequent flood or drought situations. Hilly regions of southern Odisha is a hindrance for settling up various infrastructural facilities. A large proportion of socially marginalized people (SC, ST population) have created social and economic disparity in most of the districts. Therefore, an attempt is made to obtain estimates of proportion of poor households at district level for rural areas of Odisha state using SAE approach by combining survey data from the HCES 2011–2012 of NSSO and the Population Census 2011. For this study, the HCES 2011–2012 of NSSO is the latest round of available survey. 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 Population Census 2011 data is used for the auxiliary variables and it can be accessed freely from the Census of India website.

In India, NSSO surveys are the main source of official statistics and these surveys are planned to produce statistics at state and national level. The sampling design used in the NSSO data is stratified multi-stage random sampling with districts as strata, villages as first stage units and households as the second stage units. The Monthly Per Capita Consumption Expenditure (MPCE) data from NSSO survey is used to define the living standard of a household. In particular, MPCE is used as an indicator for poverty incidence at state level. The state with lowest average MPCE is considered to be poorest. The MPCE is also used to compute state specific poverty line for both rural and urban sectors separately. The proportion of households lying below state specific poverty line is computed for each state for both rural and urban sectors. Through NSSO survey, thus efficient estimates of poverty proportion are obtained at state level for both rural and urban sectors but with the existing sampling design and sample size district level direct estimates are unstable, because within each district sample size is not large enough to provide district level estimates with adequate precision and reliability. Oversampling is not going to be a feasible approach, because that may leave other domains with small sample sizes as total sample sizes are fixed by the budget beforehand, hence incurring extra cost cannot be considered. SAE approach provides a unique way to deal with the sample sizes problem of micro levels and render stable estimates even at this level utilizing the data from already existing survey and census and quick to produce local level figures without extra budget and time constraints. Basically, it incorporates the idea of "borrowing strength" from related small areas or domains and thus increases the "effective sample sizes" of each domain. SAE techniques are generally model based methods, whereas statistical models are utilized to link the variable of interest with the auxiliary information e.g. Census or Administrative record. The concept of increasing effective sample size combining survey with census information through certain statistical models instead of oversampling makes the SAE quite distinctive and profitable.

Small area models are special case of linear mixed model and incorporate random area-specific effects which account for unstructured heterogeneity across areas beyond that is explained by auxiliary variables included in the fixed effect part of the model (Rao and Molina 2015). Based on the level of auxiliary information available, we differentiate between SAE methods based on unit-level models and those based on area-level models. In the former case these models are for the individual survey measurements and include area effects, while in the latter case these models are used to smooth out the variability in the unstable area-level direct survey estimates. Area-level modelling is typically used when unit-level data are unavailable, or, as is often the case, where model covariates (e.g. census variables) are only available at area-level. The Fay–Herriot (FH) model (Fay and Herriot 1979), is a widely used area-level model that assumes area-specific survey estimates are available, and that these follow an area-level linear mixed model with independent area random effects. This model can also accommodate survey weights in SAE by using the survey weighted direct estimates when fitting the linear mixed model. The application of FH model and its various extensions are widely available in various real life studies and literatures to solve the small domain estimation problems (Pratesi and Salvati 2009; Molina et al. 2009; Chandra et al. 2011, 2017; Chandra 2013; Portar et al. 2014; Pratesi and Salvati 2016). When survey data is not continuous, rather binary or count and in particular the target of inference is small area proportions rather than means or totals, then implementation of FH model which is based on linear mixed modeling framework may often lead to unrealistic estimates (Chandra et al. 2017, 2018b). Hence, the potential alternative is generalized linear mixed model (GLMM) for such data. The most commonly used GLMMs are the logistic-normal mixed model (i.e. GLMMs with logistic link function, also referred as the logistic linear mixed model) and the general Poisson-normal mixed model (i.e. GLMMs with log link function, also referred as the log linear mixed model). If the variable of interest is binary and the target of inference is a small area proportion, then the GLMM with logistic link function (i.e. the logistic linear mixed model) is commonly used. In this context, when only area-level data are available, an area-level version of a GLMM is considered for SAE, see for example, Chandra et al. (2011, 2017) and references therein. Unlike the FH model, this approach implicitly assumes simple random sampling with replacement within each area and ignores the survey weights. Unfortunately, this has the potential to seriously bias the estimates if the small area samples are seriously unbalanced with respect to key population characteristics, and consequently use of the survey weights appears to be inevitable for if one wishes to generate representative small area estimates. In frequentist approach of SAE, unlike estimation of survey weighted linear parameters like small area means and totals, there has been comparatively little research on estimation of survey weighted small area proportions under area-level small area models. In contrast, Bayesian framework of SAE, in particular HB approach of SAE, incorporates the survey weights in estimating small area proportions under area-level small area models, see in Liu et al. (2014). Further, both the FH as well as GLMM are based on some restrictive inbuilt assumptions on sampling variance; therefore two alternative models which are variant of area-level version of logistic-linear mixed model through relaxing its assumptions will be discussed further. In addition, incorporation of survey weight has also been considered to obtain robust estimates in terms of utilizing survey design information. Following the idea set out in Liu et al. (2014), the potential Bayesian analogues of all the four models are investigated in this paper. One of the strategic advantage in considering Bayesian approach is that, SAE methods are described by assuming particular probability distributions, which render the opportunities to analyze the uncertainties involved in the decision process. The range of Bayesian methods include Empirical Best Prediction and HB area-level and unit level models covered in varied studies related to small area literatures (Gelman 2006; Jiang and Lahiri 2006; You 2008; Souza et al. 2009; Ghosh et al. 2009; Liu et al. 2014; Lee et al. 2015). In particular, all the four models stated above will be explored in HB framework for modeling survey-weighted poverty proportions.

3 Data Description

In India each state are consist of districts and districts are important domains for planning process and policy formulation. NSSO surveys are usually designed to represent the whole nation or state and hence cannot guarantee adequate representation at the small domain level (e.g., districts or further disaggregation) within large areas. Particularly, in estimating quantities like poverty proportions, state level estimates are not able to represent regional scenario. Hence, SAE methodology sets an important step in deriving out micro or regional level estimates through borrowing strength from related sources. In the HCES 2011–2012 of NSSO used in this study, total of 2973 households were surveyed from 30 districts of Odisha. District specific sample size ranges from 64 to 160 with a median sample size of 95. Districts has been divided into three groups based on their sample sizes, 10 districts with sample size as 64; 8 districts with sample sizes 95 and 96; 12 districts with sample sizes 126, 128 and 160. District categories based on sample sizes are presented in Table 1.

The variable of interest at the unit (household) level in the published survey data file is binary, corresponding to whether a household is poor or not. In this context a household having MPCE below the state poverty line is defined as being poor. The poverty line used in this study is the same as that set by the Planning Commission, Government of India, for 2011–2012. The parameter of interest is then the proportion of poor households within each district. The state of Odisha poverty line for rural sector is Rs. 695. The parameter of interest is then the proportion of poor households or HCR within each district, which is also referred as poverty incidence or poverty rate. We now illustrate the theoretical framework used to produce small area estimates of the poverty incidence and their measure of precision. Let U denotes the finite population of interest of size N partitioned into m disjoint small areas, a sample s of size n is drawn from this population with a given survey

Table 1 Defining district categories based on sample sizes

District categories	Sample sizes	District name
Small districts (D1)	64	Jharsuguda, Sambalpur, Deogarh, Gajapati, Kandhamal, Boudh, Sonepur, Nuapada, Rayagada and Malkangiri
Medium districts (D2)	95, 96	Jagatsinghpur, Dhenkanal, Angul, Nayagarh, Khurda, Bolangir, Nawrangapur and Koraput
Large districts (D3)	126, 128, 160	Baragarh, Sundargarh, Keonjhar, Mayurbhanja, Balasore, Bhadrak, Kendrapara, Cuttack, Jajpur, Puri, Ganjam and Kalahandi

design. The set of population units in area i is denoted as U_i with known size N_i, such that $U = \cup_{i=1}^{m} U_i$ and $N = \sum_{i=1}^{m} N_i$. With N_i and n_i respectively being the population and sample size from small area i ($i=1,\ldots,m$), the units making up the sample in area i are denoted by s_i, so that $s = \cup_{i=1}^{m} s_i$ and $n = \sum_{i=1}^{m} n_i$. We assume, y_{ij} be the binary response for the target variable y for unit j ($j=1,\ldots,n_i$) in small area i. Our parameter of interest is small area (district) proportions $P_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$, which can be estimated employing direct survey estimator,

$$p_{iw} = \left(\sum_{j=1}^{n_i} w_{ij}\right)^{-1} \sum_{j=1}^{n_i} w_{ij} y_{ij}$$

where, w_{ij} denotes the survey weight provided that this weights are attached to each individual sampling unit (household) in the area (or district) i available from survey resources. The variance of p_{iw} can be expressed as,

$$\text{var}\left(p_{iw}\right) = \frac{P_i\left(1 - P_i\right)}{n_i} deff_i$$

where $deff_i$ represent the design effect reflecting the effect of complex sample design (Kish 1965). Considering negligible sampling fraction, design effect can be approximated as,

$$deff_i = n_i\left(\sum_{j=1}^{n_i} w_{ij}\right)^{-2} \sum_{j=1}^{n_i} w_{ij}^2.$$

The auxiliary variables (covariates) used in this study are drawn from the Indian Population Census of 2011. There are approximately 45 covariates available for this analysis. These covariates are only available as counts at district level. Therefore, a preliminary data analysis was carried out to select appropriate covariates for SAE modeling. For example, we first examine the correlation between different covariates and the target variable (direct survey estimates) and selected eight covariates namely proportion of SC population, proportion of ST population, female literacy rate, gender ratio, main working population ratio, marginal working population ratio, proportion of main female agricultural labourer and female cultivator and proportion of marginal female agricultural labourer and female cultivator. In addition, following Chandra et al. (2017, 2018b), we use Principal Component Analysis (PCA) to derive a composite score for selected group of variables namely main working population ratio, marginal working population ratio, proportion of main female agricultural labourer and female cultivator and proportion of marginal female agricultural labourer and female cultivator. The first principal component (denoted by G1) explained 81.48% of the variability in the selected group of variables, while adding the second component (denoted by G2) explained 92.47%. Using two indicators developed from PCA scores (i.e. G1 and G2) and four covariates from Population Census 2011 data (i.e. proportion of SC population, proportion of ST population, female literacy rate, gender ratio), we did some exploratory data analysis and modelling exercise. In particular, we fitted a generalised linear model using direct survey estimates of proportions of poor households as the response variable and the six variables (i.e. proportion of SC population, proportion of ST population, female literacy rate, gender ratio, G1 and G2) as potential covariates. The final selected model included intercept term and three covariates, proportion of SC population, proportion of ST population and G1, with residual deviance and Akaike Information Criterion (AIC) values of 329.48 and 473.43, respectively. This final model was used to produce district wise estimates of poverty incidence, i.e. estimates of the HCR. About 40% of the population of the Odisha belong to the SC (17.1%) and ST (22.8%) communities and a large scale social as well as economic disparity is resulted due to presence of this community. Their marginalization is prevalent in southern Odisha because of their forest dependence and physical isolation followed by northern and coastal region. Hence, inclusion of SC and ST population proportion as poverty indicator is significant both statistically as well as socio-economic point of view.

4 Small Area Models

4.1 FH and GLMM Model

Basic area-level FH model combines direct aggregate (district) level survey estimates with the available auxiliary variables obtained from various secondary sources, e.g., census or administrative records. Thus the model has two components,

(1) Sampling model for the direct survey estimates, and
(2) Linking model to incorporate area-specific auxiliary variables through linear regression framework.

For estimating small area proportions, survey-weighted proportions p_{iw} denotes the direct estimate for P_i, hence the sampling model for p_{iw} is expressed as follows,

$$p_{iw} = P_i + e_i;\ i = 1, \dots, m, \tag{1}$$

where, e_i's are independent sampling error assumed to have zero mean and known sampling variance σ^2_{ei}. Now, the linking model of P_i can be written as,

$$P_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i;\ i = 1, \dots, m, \tag{2}$$

where $\mathbf{x}_i$ represent area-specific covariates, $\boldsymbol{\beta}$ is the regression coefficient or fixed effect parameter and v_i being the area-specific random effect, independent and identically distributed as $E(v_i)=0$ and $Var(v_i) = \sigma^2_v$. Random area-specific effects are included in the linking model to account for between areas dissimilarities. Two random errors v_i and e_i are independent of each other within and across areas (districts). Now, to obtain the estimates of target parameter as well as MSE of the estimate it is customary to assume that two random error components follows normal distribution. However, a possible limitation of the model (1) is that when the target of inference is proportion, then assuming linear linking model with normal random effects may lead to unreliable and unrealistic estimates. Since, P_i takes value in the range 0 to 1, therefore to overcome the problem of unusual estimates (value <0 or >1), logistic or logit link function is preferred. The linking model is thus expressed as,

$$logit(P_i) = \ln\left\{P_i(1 - P_i)^{-1}\right\} = \mathbf{x}_i^T \boldsymbol{\beta} + v_i, \tag{3}$$

with $P_i = \exp\left(\mathbf{x}_i^T \boldsymbol{\beta} + v_i\right)\left\{1 + \exp\left(\mathbf{x}_i^T \boldsymbol{\beta} + v_i\right)\right\}^{-1} = \text{expit}\left(\mathbf{x}_i^T \boldsymbol{\beta} + v_i\right)$. When the small area proportion P_i is modeled using logit function, the estimates always falls within allowable range of (0, 1). But, both the FH as well as GLMM are based on strong inbuilt assumption sampling error has known design variances σ^2_{ei} which is obtained from direct variance estimates and customarily the sampling error e_i follows normal distribution. But, such postulation is quite restrictive in terms of normality assumption which may not hold for small sample sizes in domains. On the other side, unstable estimates may also be resulted because of including known sampling variances obtained using direct survey approach as direct variance estimates are not precise at all. Therefore, we will also discuss two alternative models which are the variant of logistic-linear mixed model through relaxing its assumptions. The new models postulates that sampling variances of the survey-weighted poverty proportions are unknown with the later one also drop the normality assumption of sampling error and replacing with beta distribution. So, all these four models are detailed using HB approach as follows.

4.2 Hierarchical Bayes Inference

In order to estimate small area proportion P_i, HB method is implemented employing Gibbs sampling approach. In the HB method, together with prior distribution of the parameters, prior of the hyper-parameters (model parameters) are also specified then inferences are made from the posterior distributions. Particularly, a parameter is estimated by posterior mean and posterior variance is taken as the measure of the error or uncertainty of the estimates. HB approach can effectively deal with complex small area models using Monte Carlo Markov Chain (MCMC), which overcomes the computational difficulties of high-dimensional integrations of posterior densities (You and Rao 2002).

Following Liu et al. (2014), we explore four HB models with known and unknown variance structures. The first model is FH model with known sampling variance of the survey-weighted small area proportion, provided that both the sampling and linking models has normal distributions (denoted by M1). The second (denoted by M2) and third (denoted by M3) one replaces logit-normal distribution in place of the normal distributions for linking model unlike first model. Thus these two models utilize unmatched sampling and linking models with the difference that in the second model sampling variance is assumed to be known, whereas in the third model it is unknown. Finally, fourth model (denoted by M4) is a variant of the third one which postulates non-normality of the sampling distributions and here, sampling model postulates beta type (beta-I) distribution having the desirable property of range (0, 1). These four models are described below.

M1: FH model with known sampling variance

Sampling model: $p_{iw} | P_i \sim N\left(P_i, \sigma_{ei}^2\right), \quad i = 1, \ldots, m$

Linking model: $P_i | \beta, \sigma_v^2 \sim N\left(x_i^T\beta, \sigma_v^2\right), \quad i = 1, \ldots, m$

M2: Logistic-normal mixed model with known sampling variance

Sampling model: $p_{iw} | P_i \sim N\left(P_i, \sigma_{ei}^2\right), \quad i = 1, \ldots, m,$

Linking model: $logit\left(P_i\right) | \boldsymbol{\beta}, \sigma_v^2 \sim N\left(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_v^2\right), \quad i = 1, \ldots, m.$

The assumption of known sampling variance in M1 and M2 may suffer from the drawback of unstable design variance term σ_{ei}^2.

M3: Logistic-normal mixed model with unknown sampling variance

Sampling model: $p_{iw} | P_i \sim N\left(P_i, \xi_i\right), \quad i = 1, \ldots, m,$

Linking model: $logit\left(P_i\right) | \boldsymbol{\beta}, \sigma_v^2 \sim N\left(\mathbf{x}_i^T\boldsymbol{\beta}, \sigma_v^2\right), \quad i = 1, \ldots, m.$

M4: Beta-logistic mixed model with unknown sampling variance

Sampling model: $p_{iw} | P_i \sim beta\left(a_i, b_i\right), \quad i = 1, \ldots, m,$

Linking model: $logit\left(P_i\right) | \boldsymbol{\beta}, \sigma_v^2 \sim N\left(\mathbf{x}_i^T \boldsymbol{\beta}, \sigma_v^2\right), \quad i = 1, \ldots, m.$

For M3 and M4, an approximate variance function ξ_i is used involving model parameter P_i, which is expressed as,

$$\xi_i = \frac{P_i\left(1 - P_i\right)}{n_i} deff_i$$

For model M4, in the spirit of Liu et al. (2014), the choice for parameters a_i and b_i are given as,

$$a_i = P_i\left(\frac{P_i\left(1 - P_i\right)}{\xi_i} - 1\right) = P_i\left(\frac{n_i}{deff_i} - 1\right)$$

$$b_i = \left(1 - P_i\right)\left(\frac{P_i\left(1 - P_i\right)}{\xi_i} - 1\right) = \left(1 - P_i\right)\left(\frac{n_i}{deff_i} - 1\right)$$

It is worth noting that choice of prior distributions plays a crucial role in Bayesian analysis, because inferences drawn from posterior densities depend on wide range of prior distributions. Improper prior densities such as usual choice of $\left(1/\sigma_v^2\right) \sim G(0.001, 0.001)$ can, but do not necessarily, lead to proper limiting posterior distributions. As a result, posterior inferences are sensitive to setting a small value like 0.001, indicated from the studies of Gelman (2006). Various non-informative prior distributions for σ_v^2 have been suggested in Bayesian literature including a uniform density on σ_v^2, see for example Gelman (2006),

Table 2 Summary of percent coefficient of variation generated by the different methods

Values	Direct	M1	M2	M3	M4
Minimum	10.25	10.03	10.09	10.31	10.61
Q1	16.90	14.11	15.82	13.60	14.86
Mean	24.51	21.74	21.63	19.36	19.58
Median	21.76	18.86	21.23	17.84	18.64
Q3	27.81	26.98	27.40	23.61	25.20
Maximum	60.19	57.03	34.31	36.96	32.01

Table 3 Relative gain in percent coefficient of variation over direct estimation approach

Districts categories	M1	M2	M3	M4
Overall	1.12	1.13	1.26	1.25
D1	1.13	1.13	1.26	1.25
D2	1.10	1.12	1.25	1.24
D3	1.12	1.14	1.26	1.25

Table 4 Average width of 95% credible interval along with MC error% (in bracket) for all the HB SAE methods

Districts	M1	M2	M3	M4
D1	0.221 (0.182)	0.236 (0.208)	0.219 (0.235)	0.223 (0.211)
D2	0.219 (0.181)	0.232 (0.207)	0.218 (0.234)	0.223 (0.210)
D3	0.215 (0.177)	0.229 (0.2)	0.215 (0.231)	0.218 (0.208)

Table 5 District-wise estimates of poverty proportions along with 95% confidence interval and percent coefficient of variation for the different methods

Districts	District-specific sample sizes	Direct estimation				Model based small area method (M3)			
		Estimates	Lower	Upper	%CV	Estimates	Lower	Upper	%CV
Baragarh	128	0.38	0.24	0.51	18.82	0.39	0.28	0.51	15.53
Jharsuguda	64	0.10	0.01	0.19	44.38	0.19	0.08	0.32	32.20
Sambalpur	64	0.44	0.24	0.64	23.29	0.39	0.25	0.55	19.58
Deogarh	64	0.54	0.35	0.72	17.90	0.52	0.36	0.66	14.82
Sundargarh	128	0.36	0.23	0.48	17.89	0.38	0.27	0.50	16.09
Keonjhar	128	0.41	0.27	0.55	16.88	0.41	0.31	0.54	14.74
Mayurbhanja	128	0.55	0.42	0.67	11.35	0.53	0.43	0.63	10.31
Balasore	128	0.27	0.12	0.42	27.88	0.27	0.18	0.38	19.96
Bhadrak	128	0.17	0.09	0.25	23.62	0.16	0.10	0.24	22.88
Kendrapara	126	0.06	0.00	0.12	50.10	0.10	0.05	0.18	34.12
Jagatsinghpur	96	0.18	0.07	0.29	31.97	0.17	0.10	0.26	24.26
Cuttack	128	0.10	0.02	0.19	42.69	0.12	0.06	0.20	27.90
Jajpur	128	0.14	0.06	0.22	28.41	0.14	0.08	0.21	24.31
Dhenkanal	96	0.01	0.00	0.03	60.19	0.12	0.04	0.22	36.96
Angul	95	0.09	0.02	0.16	38.15	0.16	0.08	0.27	31.31
Nayagarh	96	0.27	0.14	0.39	23.58	0.24	0.16	0.34	20.82
Khurda	96	0.22	0.10	0.34	27.60	0.19	0.11	0.29	23.86
Puri	128	0.26	0.14	0.38	23.11	0.23	0.15	0.35	21.11
Ganjam	160	0.19	0.10	0.29	25.25	0.23	0.14	0.33	20.83
Gajapati	64	0.48	0.32	0.64	16.80	0.50	0.37	0.62	13.07
Kandhamal	64	0.55	0.38	0.72	15.75	0.52	0.38	0.67	13.39
Boudh	64	0.67	0.53	0.80	10.25	0.59	0.45	0.72	11.80
Sonepur	64	0.42	0.20	0.65	27.06	0.38	0.23	0.55	20.92
Bolangir	96	0.39	0.23	0.55	20.41	0.38	0.27	0.50	15.53
Nuapada	64	0.54	0.36	0.72	16.95	0.53	0.38	0.66	14.13
Kalahandi	128	0.60	0.48	0.73	10.52	0.56	0.45	0.68	10.51
Rayagada	64	0.44	0.29	0.60	17.85	0.48	0.35	0.61	14.10
Nawrangapur	96	0.44	0.29	0.58	16.57	0.46	0.33	0.58	13.43
Koraput	96	0.63	0.50	0.77	10.98	0.59	0.47	0.71	10.88
Malkangiri	64	0.51	0.32	0.70	19.02	0.50	0.39	0.61	11.52

Table 6 District-wise estimates of poverty proportions along with 95% credible interval and CV% using M1, M2 and M4 model

Districts	M1	Lower	Upper	CV%	M2	Lower	Upper	CV%	M4	Lower	Upper	CV%
Baragarh	0.39	0.27	0.51	15.62	0.38	0.24	0.50	17.51	0.38	0.26	0.49	15.82
Jharsuguda	0.13	0.05	0.21	32.78	0.12	0.05	0.20	30.44	0.16	0.08	0.26	28.87
Sambalpur	0.39	0.24	0.54	20.36	0.39	0.22	0.57	23.09	0.39	0.24	0.55	20.76
Deogarh	0.52	0.37	0.66	14.68	0.52	0.35	0.70	16.78	0.52	0.35	0.67	15.86
Sundargarh	0.37	0.27	0.48	15.28	0.36	0.24	0.48	16.54	0.38	0.27	0.51	15.85
Keonjhar	0.41	0.29	0.52	15.06	0.40	0.27	0.53	16.42	0.41	0.30	0.54	15.05
Mayurbhanja	0.52	0.41	0.63	10.84	0.54	0.42	0.66	11.03	0.53	0.42	0.65	10.79
Balasore	0.28	0.15	0.40	23.15	0.25	0.13	0.40	26.13	0.27	0.16	0.39	21.85
Bhadrak	0.16	0.09	0.24	23.52	0.15	0.08	0.22	25.63	0.16	0.09	0.24	24.95
Kendrapara	0.07	0.01	0.12	42.88	0.07	0.03	0.12	33.17	0.09	0.04	0.15	30.85
Jagatsinghpur	0.17	0.07	0.27	30.39	0.16	0.06	0.27	32.82	0.17	0.09	0.26	25.99
Cuttack	0.11	0.03	0.19	38.41	0.10	0.04	0.18	34.31	0.12	0.06	0.18	26.45
Jajpur	0.14	0.06	0.21	28.06	0.13	0.05	0.21	30.95	0.14	0.07	0.22	26.00
Dhenkanal	0.02	0.00	0.03	57.03	0.03	0.01	0.04	27.65	0.07	0.03	0.11	32.01
Angul	0.11	0.05	0.17	29.58	0.11	0.05	0.17	27.54	0.14	0.08	0.21	25.28
Nayagarh	0.24	0.14	0.35	23.17	0.23	0.12	0.36	25.21	0.24	0.14	0.35	21.65
Khurda	0.19	0.08	0.29	29.48	0.18	0.07	0.28	31.99	0.19	0.09	0.30	27.78
Puri	0.24	0.12	0.35	23.74	0.23	0.11	0.35	26.99	0.23	0.13	0.34	23.05
Ganjam	0.22	0.13	0.32	21.63	0.20	0.12	0.28	22.09	0.22	0.14	0.31	20.25
Gajapati	0.49	0.36	0.63	14.07	0.49	0.34	0.63	15.04	0.50	0.35	0.65	14.99
Kandhamal	0.52	0.37	0.65	13.75	0.53	0.38	0.69	15.27	0.53	0.37	0.67	14.43
Boudh	0.60	0.47	0.72	10.83	0.63	0.50	0.77	10.37	0.61	0.47	0.74	11.44
Sonepur	0.40	0.23	0.56	21.90	0.38	0.21	0.59	25.60	0.38	0.22	0.56	22.82
Bolangir	0.39	0.26	0.52	17.37	0.37	0.23	0.53	20.38	0.38	0.25	0.51	17.03
Nuapada	0.52	0.38	0.66	14.05	0.53	0.37	0.70	15.93	0.53	0.38	0.68	14.82
Kalahandi	0.57	0.45	0.68	10.13	0.59	0.47	0.70	10.09	0.57	0.46	0.69	10.61
Rayagada	0.47	0.34	0.61	14.22	0.46	0.32	0.60	15.78	0.47	0.33	0.61	14.96
Nawrangapur	0.46	0.34	0.58	13.72	0.45	0.30	0.58	15.29	0.46	0.33	0.58	14.59
Koraput	0.59	0.47	0.71	10.03	0.61	0.49	0.74	10.48	0.60	0.47	0.73	11.15
Malkangiri	0.49	0.33	0.64	16.49	0.50	0.32	0.67	18.27	0.50	0.38	0.61	11.55

Souza et al. (2009) and reference therein. Non-informative prior distributions are intended to allow Bayesian inference for parameters about which not much is known beyond the data included in the analysis at hand. Following the idea given in Souza et al. (2009) and Gelman (2006), we have considered uniform prior for σ_v^2, that is Uniform (0, 1000) and distribution of $\boldsymbol{\beta}$ has been taken to be N(0, 10^6). HB small area proportion estimates are computed for all the four models using Metropolis–Hastings algorithm, drawing random samples from full conditional distributions of posterior quantities (You and Rao 2002). Gibbs sampling method is implemented to draw random samples from posterior densities. Finally, Posterior mean $E\left(P_i \mid p_{iw}\right)$ is taken as point estimate of P_i and posterior variance $V\left(P_i \mid p_{iw}\right)$ is taken as measure of variability. Full conditional distribution of all the four HB models under Gibbs sampler is provided in "Appendix".

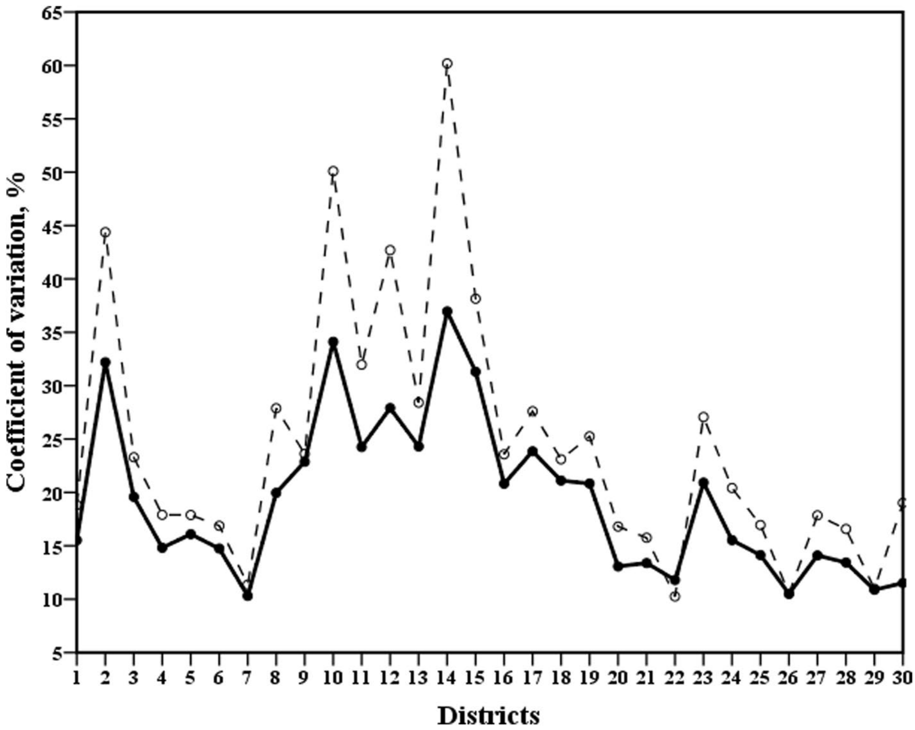

Fig. 1 District-wise percent coefficient of variation for direct (dash line, 0) and M3 method of SAE (solid line, •)

5 Empirical Analysis

For the HB estimation of small area proportion P_i, at first target binary variable y_{ij} is formed based on HCES data of NSSO from unit level household MPCE as described in Sect. 3. The appropriate auxiliary variates are obtained from Population Census (2011). To implement the Gibbs sampler, three independent chains are used each of length 10,000. The first 5000 iterations are deleted as "burn-in" periods. Potential scale reduction factor $\hat{R}$ determines the convergence success. Stationarity is attained when $\hat{R}=1$ (Rao 2003). Table 2 presents the summary of percentage coefficient of variation (%CV) for direct survey estimates as well as model based small area estimates generated by four different SAE methods defined under four small area models (M1–M4). Hereafter, for the sake of convenience, we also use the notation of M1–M4 to refer the SAE methods defined under small area models M1–M4. Note that the estimates with smaller values of %CV are considered to be reliable (i.e. smaller is better). Comparing all the HB models and corresponding small area estimates, it is to be noted that the precision level of M3 and M4 are quite improved over direct survey estimates. For direct estimates, the value of CV ranges from 10.25 to 60.19% with average of 24.51%, whereas, it varies from 10.31 to 36.96% with average of 19.36% for the M3 method and 10.61 to 32.01% with average of 19.58% for the

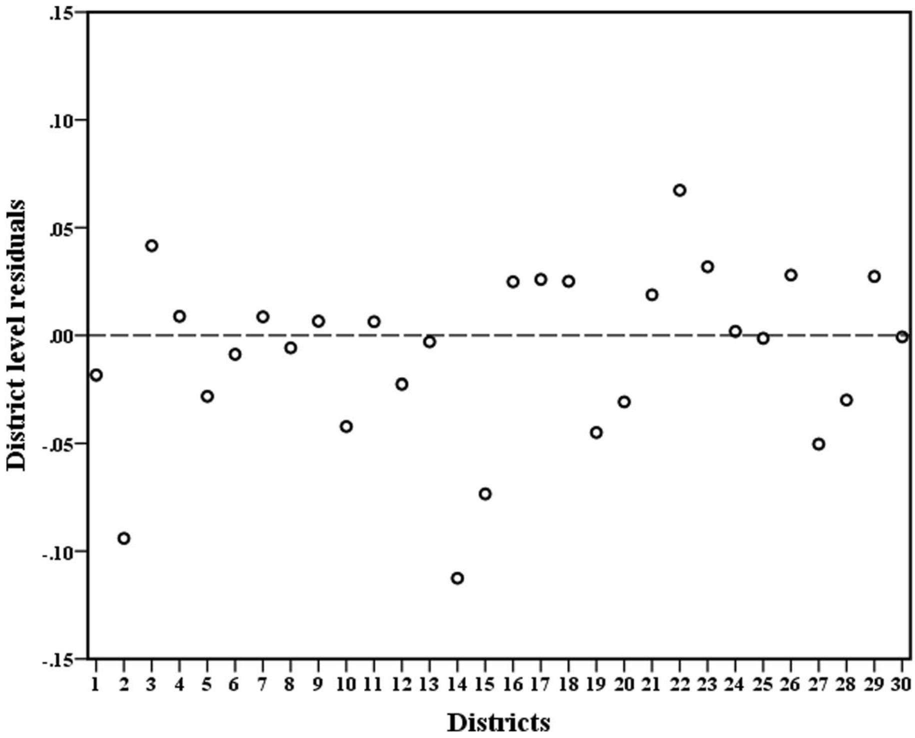

Fig. 2 Distribution of district-wise residuals generated by M3 method of SAE

M4 method. So, particularly the sharp reduction in maximum CV% compared to those of direct estimates is noticeable. Table 3 reports relative gain in percentage CV over different district categories. Relative gain is calculated as ratio of %CV of direct estimates over %CV of HB estimates. Improvement in average precision level in M3 method is slight over M2 approach of SAE which takes into account the assumption of known or fixed sampling variance in contrary to unknown sampling variance in M3. The results in Table 3 indicate that the M3 and M4 method has maximum improvement in relative gain over traditional direct estimation method. Critically examining all the model based estimates, it has been seen that for some districts M1 has produced inacceptable poverty estimates with lower bound as negative for e.g., Dhenkanal district with CV% as high as 57.03%. The results from M1 are based on linear linking model which is not suitable model for the proportion and hence such inacceptable estimates are often expected. Logistic linking model in M2, M3 and M4 is crucial for modeling proportions data. Further from Tables 2 and 3 we have noted M3 and M4 producing at par precision summary statistics with marginal improvement over M2. In onward discussions we choose to produce official poverty proportion estimates using M3 model, as it is comparatively easier to handle usual normality assumption in sampling model.

In Bayesian approach the true parameter is a random variable, thus our aim remains to capture the extent of uncertainty associated with the point estimate of the true parameter. Therefore, 95% credible interval limit is constructed such that the following holds,

$$\text{Prob}\left(\text{lb}\left(P_i\right) \leq P_i \leq \text{ub}\left(P_i\right)\right) = 0.95$$

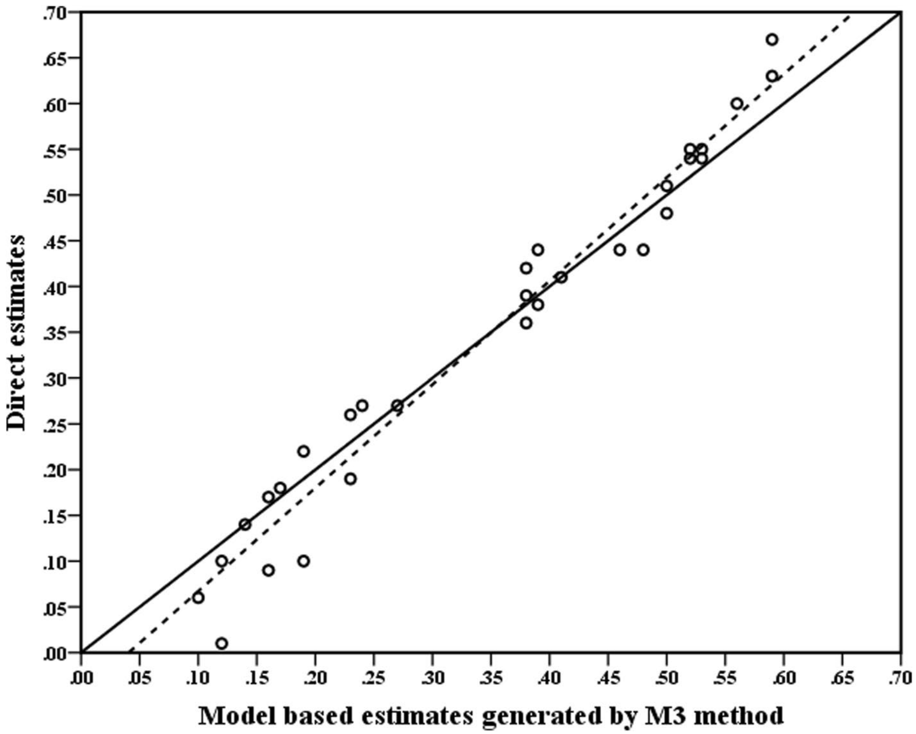

Fig. 3 Bias diagnostics plot between direct and model based estimates generated by M3 method with $y = x$ line (solid line) and regression line (dash line)

where, lb and ub are the lower bounds and upper bounds respectively. Table 4 reports the average values of 95% credible interval width along with Monte Carlo (MC) error (in %) over different district categories for all the four HB based SAE method (i.e. M1–M4) explored in this paper. From the results in Table 4, the mean width of the credible intervals for M2 is larger than those for M3 and M4. As it is expected that, for all the four models the average credible interval width declines with increasing sample sizes (D1 through D3), further variation in widths also declines with increasing sample size, also similar conclusions are indicated in Liu et al. (2014). The MC error% in M3 and M4 are higher than M1 and M2, which may be due to full conditional distributions in these models taking into account the uncertainty of the estimation of sampling variances.

Table 5 reports the district-wise estimates of poverty incidence (i.e. proportion of poor household) along with 95% confidence (credible) interval and %CV generated from direct and M3 methods. In Table 5, more than 20% CV for 16 districts has made the direct estimates in such districts highly unstable. For, Cuttack, Dhenkanal, Jharsuguda, and Kendrapara the CV was even more than 40%. A significant reduction in %CV has been achieved using the M3 method over traditional direct estimation method, thus has resulted stable and precise poverty estimates generated by M3 method. District-wise poverty proportion estimates obtained using M1, M2 and M4 model has been furnished in Table 6 for further reference.

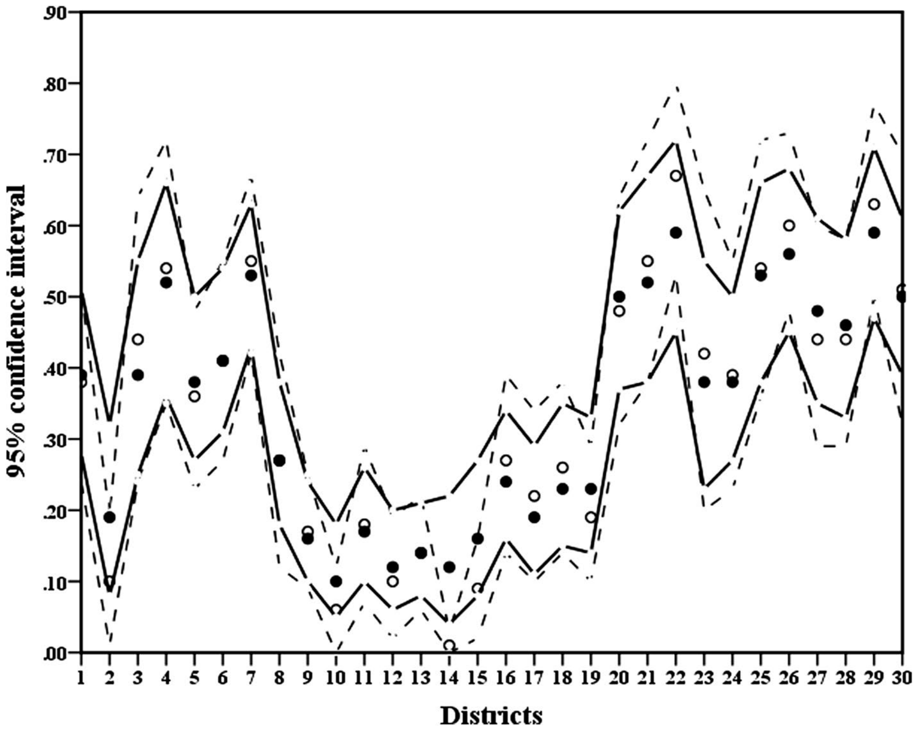

Fig. 4 District-wise 95% confidence interval (lower and upper) plot of poverty incidence for the direct estimates (dash line) and the model based estimates generated by M3 method (solid line). Direct estimate (dash, 0) and M3 estimate (solid, •)

6 Discussion

The paper focuses on estimation of small area poverty incidence and details the comparative assessment of direct estimation method and different small area model based approaches. Different HB models are evaluated competitively to extract out their relative advantages as well as disadvantages. Table 2 reflects the comparative precision levels of different estimates in terms of percentage CV. Table 3 reports the relative gain of HB SAE method over the direct estimation approach. In view of ineffectiveness of the traditional direct estimation method to predict small domain statistic with acceptable precision, model based small area methods are preferred. Estimation of such small domain figures is important from many aspects and it is the key potentiality of SAE approach to handle such situations proficiently. We have finally chosen M3 method to provide stable estimates of poverty proportions at district level; these M3 estimates has reliable %CV in most of the districts of Odisha. Figure 1 displays the district level values of % CV respectively implementing direct and M3 estimation methods.

The estimates obtained from M3 method are definitely useful for policy formulation, fund disbursement purpose in taking poverty eradication actions. Some bias diagnostics are also used to investigate if the model based M3 estimates are less extreme with the direct estimates. Figure 2 clearly revels that district-level residuals are randomly distributed. In

Table 7 District categories of Odisha based on quartile values of estimated poverty proportions using M3 model

Poverty proportion	Districts
< 0.19 (below Q1)	Bhadrak, Kendrapara, Jagatsinghpur, Cuttack, Jajpur, Dhenkanal and Angul
0.19–0.38 (Q1–Q2)	Jharsuguda, Sundargarh, Balasore, Nayagarh, Khurda, Puri, Ganjam, Sonepur and Bolangir
0.39–0.50 (Q2–Q3)	Baragarh, Sambalpur, Keonjhar, Gajapati, Rayagada, Nawrangapur and Malkangiri
> 0.50 (above Q3)	Mayurbhanja, Kandhamal, Boudh, Koraput, Nuapada, Deogarh and Kalahandi

Fig. 3, we plot M3 estimates on X-axis and direct estimates on Y-axis and look for the divergence of regression line from y = x and test for intercept = 0 and slope = 1. It shows, M3 estimates are less extreme when compared to direct survey estimates. Figure 4 shows the comparative illustration of 95% CIs of the model based M3 and the direct estimates. In general, 95% CIs for the direct estimates are wider than the 95% CIs for the M3 estimates. Further, 95% CIs for the M3 estimates are more precise and contain both direct and model based estimates of the poverty proportions. In Table 7, one attempt has been made to classify the districts based on estimated poverty incidences implementing M3 model. Districts have been categorized on the basis of quartile values of estimated poverty proportions; this type of classification may be useful for the administrators in taking effective financial and administrative decisions.

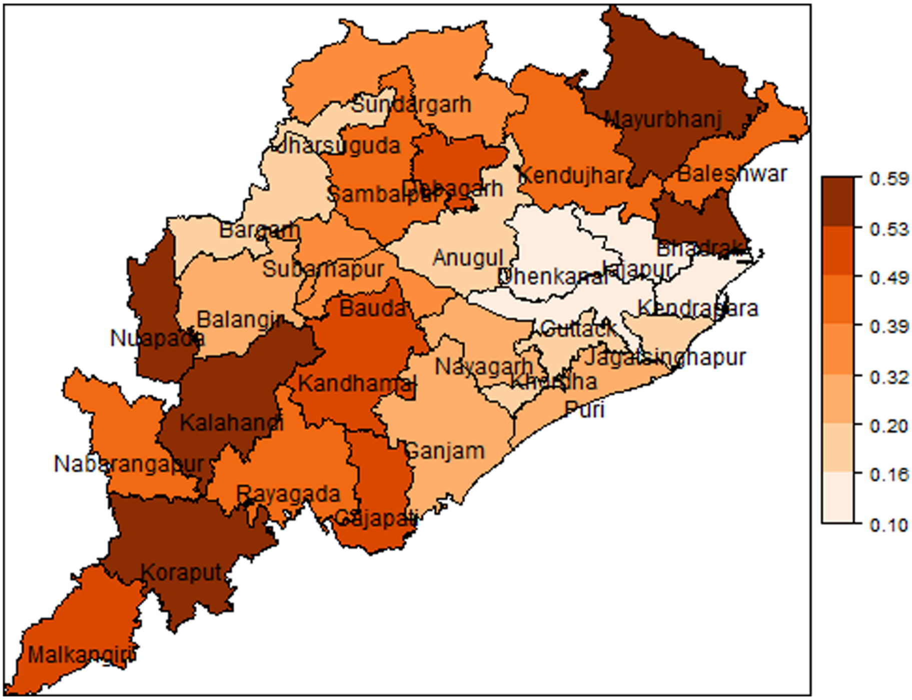

Fig. 5 Poverty map of Odisha showing spatial distribution of poverty incidence across districts generated by the M3 method of SAE

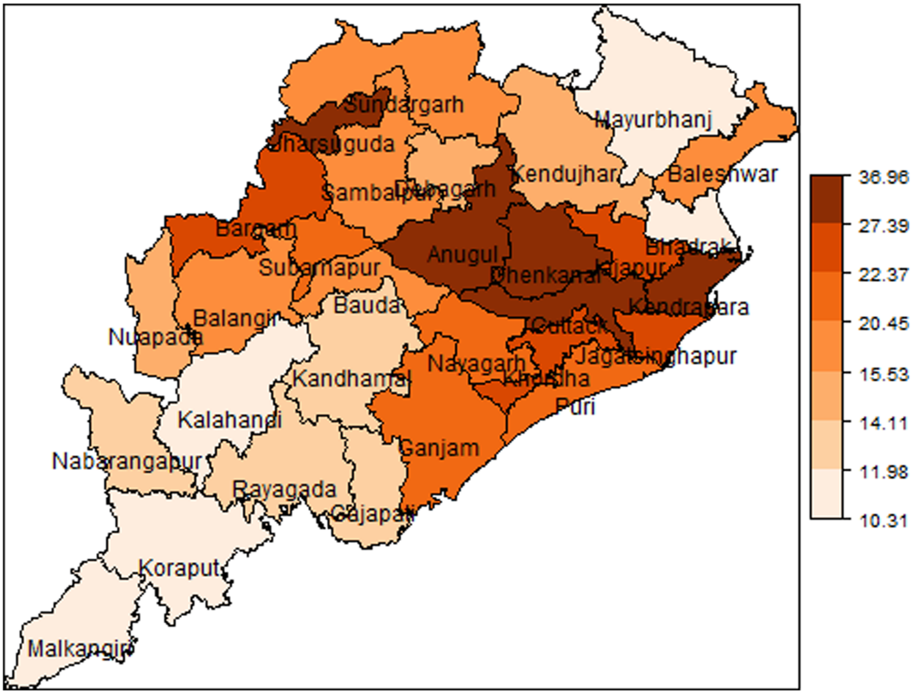

Fig. 6 Map of district-wise percentage coefficient of variation generated by the M3 method of SAE

According to Niti Aayog report (then Planning Commission) during 2011–2012, in rural Orissa the incidence of poverty is 35.69% as against 25.70% in all India. Hence, as per M3 estimates, there are 17 districts which are above average poverty incidence of Odisha, namely Sundargarh, Sonepur, Bolangir, Baragarh, Sambalpur, Keonjhar, Gajapati, Rayagada, Nawrangapur, Malkangiri, Mayurbhanja, Kandhamal, Boudh, Koraput, Nuapada, Deogarh and Kalahandi. There are total 18 districts which are above average poverty incidence of India, namely, Balasore, Sundargarh, Sonepur, Bolangir, Baragarh, Sambalpur, Keonjhar, Gajapati, Rayagada, Nawrangapur, Malkangiri, Mayurbhanja, Kandhamal, Boudh, Koraput, Nuapada, Deogarh and Kalahandi. Figure 5 shows spatial distribution of poverty incidence (measured as HCR or proportion of poor households) across districts of Odisha using M3 estimates. Darker regions are most poverty prone areas, which are basically northern and southern regions. The prime reasons behind higher poverty incidences are the prevalence of dense forest; frequent storm, flood and drought in southern areas hampering cultivation; geographical limitation due to presence of the high hilly regions of the Eastern Ghat is another region for some districts were completely isolated from the plains for several centuries owing to the non-existence of communication; high percentages of people belonging to socially marginalized classes like SC and ST. Both regional as well as social disparity has created significant spatial inequality in poverty distributions across districts of Odisha. Figure 6 is the map showing district-wise %CV for the M3 method. All these lead to the important remark that small area model based methods are potential a lot in capturing regional variations. They also may be expected to be useful in further disaggregate level studies, e.g., distribution of poverty crossing social groups with geographical districts.

7 Conclusion

The potentiality of SAE methodologies to generate reliable small domain inference is now quite established fact from varied theoretical researches, what needed is its real life implementation and applications. To strengthen the micro level planning, disaggregate level estimates are often required and small area models serve this purpose both adequately and efficiently. In this context, the current study also reflects a suitable example of why small area model based methods should be preferred. Poverty map produced from the model based small area estimates presents a quick view of the spatial inequality of poverty proportions across districts. Such spatial pattern of poverty incidence manifested in the map, is quite helpful to understand the concentration of poverty at regional level and assist the administrators in micro level planning. In India, a well-established statistical machinery like NSSO provide state and national level estimates on regular basis, however, it fail to render reliable local level estimates within the states using the same statistical design, as budget and timing constraints come to the front. SAE, basically tackle this constraints through its distinctive approach of "borrowing strength" from other small areas incorporating mixed modeling framework, in addition auxiliary variables from already available census or administrative records serve as potential source to yield small domain estimates with a good degree of precision. Present study utilizes this proficiency of SAE approach to measure poverty at localized level.

There are number of issues that warrant further investigation. In this paper we assume that district specific random effects are independent. Spatial dependence among neighbouring districts may also be taken into account within the same Bayes modeling framework to improve the model based estimates. The HB approach considered here assume a non-informative prior. Authors are grateful to one of the anonymous referee for drawing profound attention to the choice of particularly informative prior distributions followed by sensitivity analysis along with impact of such choices on corresponding model parameter estimates. The use of informative prior for the hyper-parameters needs to be examined in detail. For example, choice of improper or non-informative prior may be problematic due to small amount of data under various parameterization process and therefore selection of suitable distributions for the hyper-parameters with detailed check on posterior inferences can be a potential researchable issue.

UN has set the "No poverty" as the first SDG in view of its global significance in achieving the sustainable livelihood. Therefore, proper measurement of poverty phenomenon sets a crucial stage before implementation of various poverty eradication programmes. Added to that, poverty has other faces too. Food insecurity, proportion of malnourished and undernourished child, proportion of child lacking basic education, micro level female illiteracy proportion etc. are also needed to be adequately and suitable measured to assist the administration in taking the appropriate action eyeing no tolerance to this social threat, called "Poverty".

Acknowledgements The authors would like to acknowledge the valuable comments and suggestions of the Editor and two anonymous referees. These led to a considerable improvement in the paper.

Appendix

Full conditional distributions for the Gibbs sampler under four HB models are presented below. Let, $\tilde{p} = \left(p_{1w}, \dots, p_{mw}\right)^T$, $\mathbf{P} = \left(P_1, \dots, P_m\right)^T$, $\mathbf{X} = \left(\mathbf{x}_1^T, \dots, \mathbf{x}_m^T\right)^T$, $\mathbf{x}_i^T = \left(x_{i1}, \dots, x_{ik}\right)$, $\boldsymbol{\beta} = \left(\beta_1, \dots, \beta_k\right)^T$ and k represents number of auxiliary variates and usually x_{i1} is taken to be 1 $\forall$ i = 1, … , m.

The full conditional distributions for the M1 are given as,

(1)
$$P_i \mid \boldsymbol{\beta}, \sigma_v^2, \tilde{p} \sim N\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_{ei}^2} p_{iw} + \frac{\sigma_{ei}^2}{\sigma_v^2 + \sigma_{ei}^2} \mathbf{x}_i^T \boldsymbol{\beta}, \frac{\sigma_{ei}^2 \sigma_v^2}{\sigma_v^2 + \sigma_{ei}^2}\right);$$

(2)
$$\boldsymbol{\beta} \mid P_i, \sigma_v^2, \tilde{\mathrm{p}} \sim \mathrm{N}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{P}, \sigma_{\mathrm{v}}^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right);$$

(3)
$$\sigma_v^2 \mid \boldsymbol{\beta}, P_i, \tilde{p} \sim IG\left(a + \frac{m}{2}, b + \frac{\sum_{i=1}^{m}\left(P_i - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2}\right).$$

The full conditional distributions for the M2 are given as follows,

(1)
$$P_i \mid \boldsymbol{\beta}, \sigma_v^2, \tilde{p} \propto \frac{1}{P_i\left(1 - P_i\right)\sqrt{\sigma_{ei}^2\sigma_v^2}} \exp\left(-\frac{\left(p_{iw} - P_i\right)^2}{2\sigma_{ei}^2} - \frac{\left(\log it\left(P_i\right) - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2\sigma_v^2}\right);$$

(2)
$$\boldsymbol{\beta} \mid \mathrm{P_i}, \sigma_{\mathrm{v}}^2, \tilde{\mathrm{p}} \sim \mathrm{N}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{logit}(\mathbf{P}), \sigma_{\mathrm{v}}^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right);$$

(3)
$$\sigma_v^2 \mid \boldsymbol{\beta}, P_i, \tilde{p} \sim IG\left(a + \frac{m}{2}, b + \frac{\sum_{i=1}^{m}\left(\log it\left(P_i\right) - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2}\right).$$

The full conditional distributions for the M3 are given as below,

(1)
$$P_i \mid \boldsymbol{\beta}, \sigma_v^2, \tilde{p} \propto \frac{1}{P_i\left(1 - P_i\right)\sqrt{\xi_i\sigma_v^2}} \exp\left(-\frac{\left(p_{iw} - P_i\right)^2}{2\,\xi_i} - \frac{\left(\log it\left(P_i\right) - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2\sigma_v^2}\right);$$

(2)
$$\boldsymbol{\beta} \mid P_i, \sigma_{\mathrm{v}}^2, \tilde{\mathrm{p}} \sim \mathrm{N}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{logit}(\mathbf{P}), \sigma_{\mathrm{v}}^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right);$$

(3)
$$\sigma_v^2 \mid \boldsymbol{\beta}, P_i, \tilde{p} \sim IG\left(a + \frac{m}{2}, b + \frac{\sum_{i=1}^{m}\left(\log it\left(P_i\right) - \mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2}\right).$$

Let, $\delta_i = \frac{n_i}{deff_i} - 1$.

Then the full conditional distributions for the M4 are given as follows,

(1)
$$P_i \mid \boldsymbol{\beta}, \sigma_v^2, \tilde{p} \propto \frac{1}{P_i\left(1-P_i\right)\sqrt{\sigma_v^2}} \frac{p_{iw}^{P_i\delta_i-1}\left(1-p_{iw}\right)^{(1-P_i)\delta_i-1}}{\Gamma\left(P_i\delta_i\right)\Gamma\left(\left(1-P_i\right)\delta_i\right)} \exp\left(-\frac{\left(\log it\left(P_i\right)-\mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2\sigma_v^2}\right);$$

(2)
$$\boldsymbol{\beta} \mid P_i, \sigma_{\mathrm{v}}^2, \tilde{\mathrm{p}} \sim \mathrm{N}\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{logit}(\mathbf{P}), \sigma_{\mathrm{v}}^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right);$$

(3)
$$\sigma_v^2 \mid \boldsymbol{\beta}, P_i, \tilde{p} \sim IG\left(a+\frac{m}{2}, b+\frac{\sum_{i=1}^{m}\left(\log it\left(P_i\right)-\mathbf{x}_i^T\boldsymbol{\beta}\right)^2}{2}\right).$$

References

Benjamin, M., Stoker, T., & Suri, T. (2013). The economics of slums in the developing world. *Journal of Economic Perspectives, 27*(4), 187–210.

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. (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., Aditya, K., & Sud, U. C. (2018a). Localised estimates and spatial mapping of poverty incidence in the state of Bihar in India-an application of small area estimation techniques. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0198502.

Chandra, H., Salvati, N., & Chambers, R. (2017). Small area prediction of counts under a non-stationary spatial model. *Spatial Statistics, 20,* 30–56.

Chandra, H., Salvati, N., & Chambers, R. (2018b). Small area estimation under a spatially non-linear model. *Computational Statistics & Data Analysis, 126,* 19–38.

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 of Applied statistics, 38*(11), 2413–2432.

Chaudhuri, S., & Gupta, N. (2009). Levels of living and poverty patterns: A district-wise analysis for India. *Economic & Political Weekly, XLIV*(9), 94–110.

Chauhan, R. K., Mohanty, S. K., Subramanian, S. V., Parida, J. K., & Padhi, B. (2016). Regional estimates of poverty and inequality in India, 1993–2012. *Social Indicators Research, 127,* 1249–1296.

Coondoo, D., Majumder, A., & Chattopadhyay, S. (2011). District-level poverty estimation: A proposed method. *Journal of Applied Statistics, 38*(10), 2327–2343.

Fay, R. E., & Herriot, R. A. (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.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1,* 515–533.

Ghosh, M., Kim, D., Sinha, K., Maiti, T., Katzoff, M., & Parsons, V. L. (2009). Hierarchical and empirical Bayes small domain estimation of the proportion of persons without health insurance of minority subpopulations. *Survey Methodology, 35,* 53–66.

Government of India. (2017). *Millennium development goals—Final country report of India*. Central Statistics Office, Ministry of Statistics and Programme Implementation, Government of India. http://mospi.nic.in/sites/dfault/files/publication_reports/MDG_Final_Country_report_of_India_27nov17.pdf. Accessed 27 Nov 2017.

Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test, 15,* 111–999.

Kabeer, N. (1994). *"Beyond the poverty line: Measuring poverty and impoverishing measures" in reversed realities: Gender hierarchies in development thoughts* (pp. 136–162). London: Yerso.

Kish, L. (1965). *Survey sampling*. New York: Wiley.

Lee, D., Minton, J., & Pryce, G. (2015). Bayesian inference for the dissimilarity index in the presence of spatial autocorrelation. *Spatial Statistics, 11,* 81–95.

Liu, B., Lahiri, P., & Kalton, G. (2014). Hierarchical Bayes modeling of survey-weighted small area proportions. *Survey Methodology, 40,* 1–13.

Mohanty, S. K., Govil, D., Chauhan, R. K., Kim, R., & Subramanian, S. V. (2016). Estimates of poverty and inequality in the districts of India, 2011–2012. *Journal of Development Policy and Practice, 1*(2), 142–202.

Molina, I., Salvati, N., & Pratesi, M. (2009). Bootstrap for estimating the MSE of the spatial EBLUP. *Computational Statistics, 24,* 441–458.

Narayana, D., & Petesch, P. (2002). *Voices of the poor: From many lands World Bank*. New York: Oxford University Press.

Portar, A. T., Holan, S. H., Wikle, C. K., & Cressie, N. (2014). Spatial Fay–Herriot models for small area estimation with functional covariates. *Spatial Statistics, 10,* 27–42.

Pratesi, M., & Salvati, N. (2009). Small area estimation in presence of correlated random area effects. *Journal of Official Statistics, 25,* 37–53.

Pratesi, M., & Salvati, N. (2016). *Introduction on measuring poverty at local level using small area estimation methods. Pratesi/Analysis of poverty data by small area estimation* (pp. 1–18). New York: Wiley.

Rao, J. N. K. (2003). *Small area estimation*. New York: Wiley.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation* (2nd ed.). New York: Wiley.

Sen, A. K. (1981). *Poverty and famines: An essay on entitlements and deprivation*. Oxford: Clarendon Press.

Souza, D. F., Moura, F. A. S., & Migon, H. S. (2009). Small area population prediction via hierarchical models. *Survey Methodology, 35,* 203–214.

United Nations. (2017). *The sustainable development goals report*. http://www.un.org/sustainabledevelopment/poverty/. Accessed 15 Dec 2017.

You, Y. (2008). An integrated modeling approach to unemployment rate estimation for sub-provincial areas of Canada. *Survey Methodology, 34,* 19–27.

You, Y., & Rao, J. N. K. (2002). Small area estimation using unmatched sampling and linking models. *The Canadian Journal of Statistics, 30,* 3–15.