Agricultural Economics Research Review 2018, 31 (1), 29-44
DOI: 10.5958/0974-0279.2018.00003.4

Localized estimates of the incidence of indebtedness among rural households in Uttar Pradesh: an application of small area estimation technique

Hukum Chandra

ICAR-Indian Agricultural Statistics Research Institute, New Delhi-110012, India

Abstract Socioeconomic conditions exhibit significant variation at higher levels of spatial disaggregation, yet these have not received much attention in the economic and statistical investigations in terms of their identification and prioritization for targeting of efforts and investment. Applying small area estimation (SAE) approach, in this paper we have generated estimates of the incidence of indebtedness among social groups across districts of Indian state of Uttar Pradesh. The findings show that the estimates from SAE are more precise and representative of the spatial heterogeneity in the socioeconomic conditions than do the direct survey estimates. Such estimates of spatial inequality in the incidence of indebtedness provide useful information to financial institutions to develop financial products and services suited to the marginalized regions and social groups; and to policymakers to take appropriate measures to address the problem of agrarian distress which is often attributed to the indebtedness.

Keywords Socio-economics, Indebtedness, Small area estimates, Diagnostics, Spatial mapping

JEL classification C810, C890

1 Introduction

As economic planning becomes more decentralized, the need for micro-level statistics on socioeconomic conditions, infrastructure and institutions cannot be undermined. Micro-level statistics is essential to target social and spatial heterogeneity in the programmes and strategies aimed at alleviating the inter-personal and inter-regional inequalities. However, our understanding of the spatial and social inequalities at this level of disaggregation is severely hampered by the lack of availability of statistics. The available statistics at national or state level mask variations which is available at local levels, restricting designing and targeting of heterogeneity at higher levels of spatial disaggregation, and the scope for monitoring and evaluation of parameters locally within and across administrative units. The macro developmental planning often leads to sub-optimal interventions and outcomes because it often focuses on groups or areas that require minimal efforts to achieve the developmental goals, neglecting those areas or groups for whom these are needed the most. Nonetheless, local developmental goals cannot be set and monitored in the absence of baseline information. The existing data, based on national level socioeconomics surveys, produce estimates that are representative of the macro-geographical units and cannot be used directly to produce reliable micro or local (also referred as small area) level estimates because of the small sample sizes (Pfeffermann 2002; Molina & Rao 2015).

Conducting surveys at higher levels of spatial disaggregation to generate local statistics is time-consuming and also costly. We, therefore, need special techniques that can generate estimates at micro-level utilizing the already available survey data. Small area estimation (SAE) is such a technique that can produce reliable estimates at micro-level. The technique is a model-based method that links the variable of interest from survey with the auxiliary information available

*Corresponding author: hchandra12@gmail.com

from other data sources for small areas. Depending on the availability of auxiliary information, small area models are of two broad types. One, the area level random effect models that are applied when the auxiliary information is available only at area level. These relate small area direct survey estimates to area-specific auxiliary information (Fay & Herriot 1979). Two, the nested error unit level regression models, proposed originally by Battese et al. (1988) that relate unit values of a variable to the unit-specific auxiliary information.

In this paper, we apply SAE technique to produce model-based estimates of the incidence of indebtedness among social groups in different districts of Uttar Pradesh by linking information from the All-India Debt and Investment Survey 2012-13 (AIDIS) and the Census 2011. *Here,* we consider small areas as districts, and social groups within districts. Such estimates provide useful information to financial institutions to develop financial products and services for the marginalized regions and social groups at a higher level of spatial disaggregation; and to policymakers to take appropriate measures to address the problem of agrarian distress which is often attributed to indebtedness.

Rest of the paper is organized as follows. In section 2, we discuss data used in this paper. Section 3 provides an overview of SAE technique that we use to generate incidence of indebtedness among social groups across districts. In section 4, we introduce diagnostic procedures to examine model assumptions and validate small area estimates, and discuss the results. Concluding remarks are made in the last section.

2 Data

Uttar Pradesh is the most populous state of Indian Union having 828 persons per sq.km of geographical area. The state accounts for 16.2% of the country's total population with considerable spatial and social heterogeneity. Agriculture is the main occupation of majority of the population in the state. Also, it has a large number of poor belonging to various social groups, viz. scheduled caste (SC), scheduled tribe (ST), other backward castes (OBCs), and other castes (Others). The scheduled castes/tribes are at the bottom of the hierarchy of Indian social structure, followed by the OBCs and other or upper castes. Thus, it is reasonable for us to apply SAE technique to generate estimates of incidence of indebtedness at district level among social groups in Uttar Pradesh. For the purpose, we require two types of variables: (i) the variable of interest for which small area estimates are required. Our variable of interest in the incidence of indebtedness, that is, the proportion of indebted households in a social group or in a district. We extract information on indebtedness from AIDIS 2012-13. A household is defined as indebted if it has an outstanding loan *(from respective source)* as on June 30, 2012. (ii) auxiliary (covariates) variables that are available for the population. The covariates that we in this analysis are available at district level in the Census 2011. However, there could be an issue of concordance of information between Census 2011 and AIDIS 2012-13. But, the covariates that we use from Census are unlikely to change much in such a shorter period.

The estimates of indebtedness in AIDIS are based on stratified multi-stage random sampling; districts as strata, census villages as first stage units and households as ultimate stage units. In Uttar Pradesh, there are a total 8587 surveyed households (including both indebted and non-indebted) spread over 71 districts. Table 1 presents sample sizes of different social groups, and the sample sizes of social groups across districts

Table 1. Social groups and their sample sizes in AIDIS 2012-13 in Uttar Pradesh

Social group	Description of households type	Sample size			
		Min	Max	Average	Total
All	All	56	195	121	8587
ST	Scheduled tribe	1	19	3	81
SC	Scheduled caste	4	88	33	2327
OBC	Other Backward class	20	118	64	4528
Others	Upper castes	4	55	24	1651

Source: AIDIS, 2012-13.

are shown in table 2. Across districts, the sample size ranges between 56 to 195 with an average of 121. The sample sizes become too small if sub-grouped further by social groups. For example, the sample size of ST varies from 1 to 19 across district with an average of 3. Further, the sample comes only from 27 districts from 81 households Theoretically, SAE technique enables us to produce estimates for the districts that do not have any sample of this social group. Practically, this social group does not require estimates of indebtedness in these districts. Furthermore, of the 27 districts, 19 districts have sample size of not more than

Table 2. Distribution of sample size by social group across districts in Uttar Pradesh

District	Total	ST	SC	OBC	Others	District	Total	ST	SC	OBC	Others
Saharanpur	140	0	35	90	15	Lalitpur	56	1	4	45	6
Muzaffarnagar	168	0	36	99	33	Hamirpur	56	0	7	34	15
Bijnor	154	0	42	76	36	Mahoba	56	0	15	34	7
Moradabad	168	0	50	85	33	Banda	84	0	22	53	9
Rampur	112	1	17	72	22	Chitrakoot	56	0	12	40	4
Jyotiba Phule Nagar	84	0	29	34	21	Fatehpur	140	0	34	65	41
Meerut	105	0	26	58	21	Pratapgarh	168	0	39	82	47
Baghpat	61	0	14	32	15	Kaushambi	84	0	38	31	15
Ghaziabad	83	1	25	44	13	Allahabad	193	4	50	118	21
Gautam B. Nagar	56	0	18	22	16	Barabanki	168	1	72	76	19
Bulandshahar	140	0	35	50	55	Faizabad	112	0	27	46	39
Aligarh	140	0	37	60	43	Ambedkar Nagar	112	0	50	32	30
Hathras	84	0	35	36	13	Sultanpur	168	1	40	86	41
Mathura	98	2	28	49	19	Bahraich	168	0	30	94	44
Agra	138	0	27	60	51	Shrawasti	56	0	6	22	28
Firozabad	112	2	21	81	8	Balrampur	112	0	13	62	37
Etah	84	0	14	47	23	Gonda	168	0	38	76	54
Mainpuri	83	1	31	47	4	Siddharth Nagar	153	2	26	90	35
Budaun	168	0	25	104	39	Basti	140	0	40	84	16
Bareilly	168	0	26	118	24	Sant Kabir Nagar	82	0	29	47	6
Pilibhit	112	2	24	71	15	Mahrajganj	140	1	26	96	17
Shahjahanpur	140	3	29	94	14	Gorakhpur	168	4	52	99	13
Kheri	168	17	45	72	34	Kushinagar	168	1	41	103	23
Sitapur	168	0	88	63	17	Deoria	168	0	35	101	32
Hardoi	168	1	49	87	31	Azamgarh	192	0	83	88	21
Unnao	140	3	45	41	51	Mau	112	1	41	51	19
Lucknow	84	0	49	20	15	Ballia	168	3	18	115	32
Rae Bareli	168	0	61	60	47	Jaunpur	195	0	63	107	25
Farrukhabad	84	0	8	51	25	Ghazipur	167	0	36	115	16
Kannauj	84	1	13	59	11	Chandauli	112	0	47	49	16
Etawah	84	0	18	39	27	Varanasi	108	2	36	56	14
Auraiya	83	2	35	35	11	Sant R. Nagar	84	0	26	38	20
Kanpur Dehat	84	0	10	61	13	Mirzapur	140	1	44	81	14
Kanpur Nagar	84	0	25	40	19	Sonbhadra	84	19	28	22	15
Jalaun	84	0	29	51	4	Kashiram Nagar	84	1	24	49	10
Jhansi	84	3	36	33	12	Total	8587	81	2327	4528	1651

Source: AIDIS 2012-13.

2. Thus, extremely small samples from these districts pose a challenge in deriving reliable estimates of indebtedness. Likewise, for other social groups too, the sample size in many districts is less than 10. The direct survey estimates based on such small samples cannot be reliable. SAE is an obvious choice to address this problem.

The Census 2011 provides a number of covariates at district level that can be utilized for small area modelling. For a select group of similar variables viz., gender-wise main workers, gender-wise main cultivators and gender-wise main agricultural labourers, we derive a composite score applying the principal component analysis (PCA), and refer it as working group index. The first principal component (WG1) explains 49% of the variability in the selected group of variables, and the second component (WG2) explains 67% of the variation. Using these working group indicators and the remaining covariates from Census, we undertake some exploratory analysis like correlation between different covariates and modelling exercise. Finally, three variables, viz. proportion of female workers in the total female population (FWP), proportion of literate females (FLR) and working group index (WG1) that significantly explain the model, are identified for the use in SAE analysis.

3 Analytical framework

Let us assume that population U of size N consists of D non-overlapping and mutually exclusive small areas (or areas), and a sample s of size n is drawn from this population. We use a subscript d to index values belonging to small area d ($d = 1,\ldots,D$). Let U_d and s_d be the population and sample of sizes N_d and n_d in small area d, respectively such that $U = \bigcup_{d=1}^{D} U_d$, $N = \sum_{d=1}^{D} N_d$, $s = \bigcup_{d=1}^{D} s_d$ and $n = \sum_{d=1}^{D} n_d$. We use subscript s and r respectively to denote values of sample and non-sample parts of the population. The value of the variable of interest y (which is the number of indebted household) in a small area (district in our case) d is defined by y_d, and we denote by y_{sd} and y_{rd} the sample and the non-sample counts of indebted households in small area d, respectively. The variable of interest (i.e. sample count of binary variable) y_{sd} has a binomial distribution with parameters n_d and p_d, denoted by $y_{sd} \sim Bin(n_d, p_d)$, where p_d is the probability of a household being indebted in small area d, often termed as the probability of 'success'. Similarly, for the non-sample count, $y_{rd} \sim Bin(N_d - n_d, p_d)$. Further, y_{sd} and y_{rd} are assumed to be independent binomial variables with p_d being a common success probability. Let $\mathbf{x}_d$ be the k-vector of covariates for area d from available data sources. The model linking probabilities of success p_d with covariates $\mathbf{x}_d$ is the logistic linear mixed model (see Breslow & Clayton 1993; Saei & Chambers 2003) as follows:

$$logit(p_d) = \ln\left\{p_d(1-p_d)^{-1}\right\} = \eta_d = \mathbf{x}_d^T\boldsymbol{\beta} + u_d, \quad (1)$$

with $p_d = \exp(\mathbf{x}_d^T\boldsymbol{\beta} + u_d)\left\{1 + \exp(\mathbf{x}_d^T\boldsymbol{\beta} + u_d)\right\}^{-1}$.

Here $\boldsymbol{\beta}$ is the k-vector of regression coefficients, often known as fixed effect parameters, and u_d is the area-specific random effect that capture the between area heterogeneity. We assume that u_d is independent and normally distributed with mean zero and variance σ_u^2. Here, we observe that equation (1) relates the area (or district) level proportions (direct estimates) from the survey data to the area (or district) level covariates. This type of model is often referred to as 'area-level' model in SAE terminology (see, Molina & Rao 2015). This model was originally used by Fay & Herriot (1979) for predicting mean per capita income in small geographical areas (<500 persons) within counties in the United States. The Fay and Herriot (FH) method for SAE is based on area level linear mixed model and their approach is applicable to a continuous variable. Equation (1), a special case of a generalized linear mixed model (GLMM) with logit link function, is suitable for modelling discrete data, particularly the binary variables (*Chandra et al. 2011;* Johnson *et al. 2010*). Let T_d denotes the total number of indebted households in area (or district) d. We can express the total count as $T_d = y_{sd} + y_{rd}$, where y_{sd} (sample count) is known and y_{rd} (non-sample count) is unknown. Therefore, an empirical best predictor (EBP) estimate $\hat{T}_d$ of the total count in area (or district) d is obtained by replacing y_{rd} by its predicted value in equation (1):

$$\hat{T}_d^{EBP} = y_{sd} + \hat{y}_{rd} = y_{sd} + \left(N_d - n_d\right)$$
$$\left[\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\left(1 + \exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\right)^{-1}\right]$$

An estimate of the proportion of indebted households (i.e. *incidence of indebtedness*) in a small area d is then defined as:

$$\hat{p}_d^{EBP} = N_d^{-1}\hat{T}_d^{EBP} = N_d^{-1}\left\{y_{sd} + (N_d - n_d)\left[\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\left(1+\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\right)^{-1}\right]\right\} \quad (2)$$

In many small area applications, we come across situations where the areas do not have sample data, i.e. $n_d = 0$. Eventually, the traditional survey estimation approaches do not provide solution to this problem. The SAE can be used to derive estimates for such areas. In particular, for $n_d = 0$, the synthetic-type estimator of T_d is defined as $\hat{T}_d^{Syn} = N_d\left\{\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}})\left(1+\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}})\right)^{-1}\right\}$. The *incidence of indebtedness* is estimated by $\hat{p}_d^{Syn} = \exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}})\left(1+\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}})\right)^{-1}$. It is obvious that in order to compute the small area estimates by equation (2), we require estimates of the unknown parameters $\boldsymbol{\beta}$ and $\mathbf{u}$. We use an iterative procedure that combines the Penalized Quasi-Likelihood (PQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u} = (u_1,\ldots,u_D)^T$ with restricted maximum likelihood (REML) estimation of σ_u^2 to estimate unknown parameters (see, Chandra et al. 2011; Breslow & Clayton 1993; Saei & Chambers 2003).

We now turn to estimation of mean squared error (MSE) for small predictors given by equation (2). The MSE estimates are computed to assess the reliability of estimates and also to construct the confidence interval (CI) for the estimates. Following Chandra et al. (2011), the mean squared error estimate of (2) is:

$$mse(\hat{p}_d^{EBP}) = M_1(\hat{\sigma}_u^2) + M_2(\hat{\sigma}_u^2) + 2M_3(\hat{\sigma}_u^2) \quad (3)$$

For simplicity and ease of implementation, we define few notations to express different components of the MSE (Eq.3). We denote by $\hat{\mathbf{V}}_s = diag\left\{n_d\hat{p}_d^{EBP}(1-\hat{p}_d^{EBP})\right\}$ and $\hat{\mathbf{V}}_r = diag\left\{(N_d - n_d)\hat{p}_d^{EBP}(1-\hat{p}_d^{EBP})\right\}$ the diagonal matrices defined by the corresponding variances of the sample and non-sample parts, respectively. We then define $\mathbf{A} = \left\{diag(N_d^{-1})\right\}\hat{\mathbf{V}}_r$, $\mathbf{B} = \left\{diag(N_d^{-1})\right\}\hat{\mathbf{V}}_{rd}\mathbf{X} - \mathbf{A}\hat{\mathbf{T}}\hat{\mathbf{V}}_s\mathbf{X}$ and $\hat{\mathbf{T}} = \left(\hat{\sigma}_u^2\mathbf{I}_D + \hat{\mathbf{V}}_s\right)^{-1}$, where $\mathbf{X}=(\mathbf{x}_1^T,\ldots,\mathbf{x}_D^T)^T$ is a $D \times k$ matrix, and $\mathbf{I}_D$ is an identity matrix of order D. We further write $\hat{\mathbf{T}}_{11} = \left\{\mathbf{X}^T\hat{\mathbf{V}}_s\mathbf{X} - \mathbf{X}^T\hat{\mathbf{V}}_s\hat{\mathbf{T}}\hat{\mathbf{V}}_s\mathbf{X}\right\}^{-1}$ and $\hat{\mathbf{T}}_{22} = \hat{\mathbf{T}} + \hat{\mathbf{T}}\hat{\mathbf{V}}_s\mathbf{X}\hat{\mathbf{T}}_{11}\mathbf{X}^T\hat{\mathbf{V}}_s^T\hat{\mathbf{T}}$. With these notations, assuming model (1) holds, the various components of MSE estimate are:

$M_1(\hat{\sigma}_u^2) = \mathbf{A}\hat{\mathbf{T}}\mathbf{A}^T$, $M_2(\hat{\sigma}_u^2) = \mathbf{B}\hat{\mathbf{T}}_{11}\mathbf{B}^T$, and $M_3(\hat{\sigma}_u^2) = trace\left(\hat{\nabla}_i\hat{\Sigma}\hat{\nabla}'_j v(\hat{\sigma}_u^2)\right)$

with $\hat{\Sigma} = \hat{\mathbf{V}}_{sd} + \hat{\phi}\mathbf{I}_D\hat{\mathbf{V}}_{sd}\hat{\mathbf{V}}_{sd}^T$. Let us write $\Delta = \mathbf{A}\hat{\mathbf{T}}$ and $\hat{\nabla}_i = \partial(\Delta_i)/\partial\phi\big|_{\phi=\hat{\phi}} = \partial(A_i\hat{\mathbf{T}})/\partial\sigma_u^2\big|_{\sigma_u^2=\hat{\sigma}_u^2}$, where A_i is the i^{th} row of the matrix A. Here $v(\hat{\sigma}_u^2)$ is the asymptotic covariance matrix of the estimate of variance component $\hat{\sigma}_u^2$, which can be evaluated as the inverse of the appropriate Fisher information matrix for $\hat{\sigma}_u^2$. Further, this also depends upon whether we use ML or REML estimate for $\hat{\sigma}_u^2$. We use REML estimates for $\hat{\sigma}_u^2$ and then $v(\hat{\sigma}_u^2) = 2\left((\hat{\sigma}_u^2)^{-2}(D-2t_1) + (\hat{\sigma}_u^2)^{-4}t_{11}\right)^{-1}$ with $t_1 = (\hat{\sigma}_u^2)^{-1}trace(\hat{\mathbf{T}}_{22})$ and $t_{11} = trace(\hat{\mathbf{T}}_{22}\hat{\mathbf{T}}_{22})$. The MSE estimates of (7) is a special case of (3) when $n_d = 0$ and it is given by $mse(\hat{p}_d^{Syn}) = \left[diag\left\{\hat{p}_d^{Syn}(1-\hat{p}_d^{Syn})\right\}\right]\hat{\sigma}_u^2\mathbf{I}_D\left[diag\left\{\hat{p}_d^{Syn}(1-\hat{p}_d^{Syn})\right\}\right]^T$.

4 Empirical results

4.1 Diagnostic measures

Generally, two types of diagnostics measures are suggested and employed in SAE application; (i) the model diagnostics, and (ii) the diagnostics for the small area estimates. The model diagnostics are applied to verify model assumptions. The other diagnostics are used to validate reliability of the model-based small area estimates. In equation (1), the random area specific effects u_d are assumed to have a normal distribution with mean zero and fixed variance σ_u^2. If the model assumptions are satisfied, then the area (or district) level residuals are expected to be randomly distributed and

Table 3. Shapiro-Wilk test result for different social groups

Social group	Test statistics value	p-value
All	0.9839	0.4985
ST	0.9820	0.9042
SC	0.9891	0.7982
OBC	0.9679	0.1661
Others	0.9754	0.1762

not significantly different from the regression line *y=0;* whereas, from equation (1) the area (or district) level residuals are defined as $r_d = \hat{\eta}_d - \mathbf{x}_d^T\hat{\boldsymbol{\beta}}\,(d=1,\ldots,D)$. The histogram and q-q plots are used to examine the normality assumption. Figure 1 presents the histogram of the district-level residuals (left-hand side plots), distribution of the district-level residuals and normal q-q plots of the district-level residuals for different social groups. Besides these graphical methods for checking data normality, we also perform Shapiro-Wilk (SW) test of normality (i.e., test based on uncertainty measurement in terms of p-value). The p-value from SW test indicates the chance that the sample comes from a normal distribution. Typically, a value of 0.05 is used as cutoff, i.e. if p-value is less than 0.05 we can conclude that the sample deviates from normality. Table 3 reports the results of SW test for different social groups.

The plots in figure 1 suggest that the model diagnostics are fully satisfied with the data that we have used in this analysis. For example, figure 1 shows that the randomly distributed district level residuals and the line of fit does not significantly differ from the line *y=0*, as expected in all the plots. The q-q plots as well as histograms also confirm the normality assumption. The SW test results in table 3 show large p-values for all the social groups. Hence, we conclude that the data we utilized in this paper are likely to be normally distributed. This conclusion is supplemented by the results of the model diagnostics in figure 1.

For assessing validity and reliability of the model-based small area estimates, we must use a set of diagnostics. Model-based small area estimates should be (a) consistent with unbiased direct survey estimates, and (b) more precise than direct survey estimates. The values for the model-based small area estimates derived from the fitted model should be consistent with the unbiased direct survey estimates, wherever these are available. In other words, these 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. The model-based small area estimates should have mean squared errors significantly lower than the variances of the corresponding direct survey estimates (see, Chandra et al. 2011; Brown et al. 2001). For this purpose, we consider three commonly used diagnostics, viz. the bias diagnostics, coefficient of variation (CV in %) and 95% confidence intervals for the small area estimates. We compute bias between average value of direct and model estimates (Bias) and average relative difference between direct and model estimates (RE) as:

$$Bias = D^{-1}\left(\sum_d \text{Direct estimate}_d\right) \text{-} D^{-1}\left(\sum_d \text{Model based estimate}_d\right),$$

and $$RE = D^{-1}\sum_d \left\{\frac{\text{Direct estimate}_d \text{ - Model based estimate}_d}{\text{Direct estimate}_d}\right\}.$$

The values of Bias and RE are given in table 4. We also apply Goodness of fit (GoF) diagnostic. This tests whether the direct and model-based estimates are statistically different. The null hypothesis is that the direct and model-based estimates are statistically equivalent. The alternative is that the direct and model-based estimates are statistically different. The GoF diagnostic is computed using the Wald statistic for every model-based estimate:

$$W = \sum_d \left\{\frac{\left(\text{Direct estimate}_d \text{ - Model based estimate}_d\right)^2}{\hat{Var}(\text{Direct estimate}_d) + \hat{MSE}(\text{Model based estimate}_d)}\right\}.$$

The value from the test statistic is compared against the value from a chi square distribution with *D=71* degrees of freedom which is 91.670 at 5% level of significance. The GoF diagnostic results are presented in table 5. The diagnostic results in table 4 and 5 reveal that model-based small area estimates are consistent with the direct survey estimates.

Table 4. Bias diagnostics for sample districts

Social group	Bias	RE
All	0.000	0.013
ST#	-0.002	-0.226
SC	0.008	0.098
OBC	-0.002	0.027
Others	-0.016	0.031

#Based on estimates of 9 districts only.

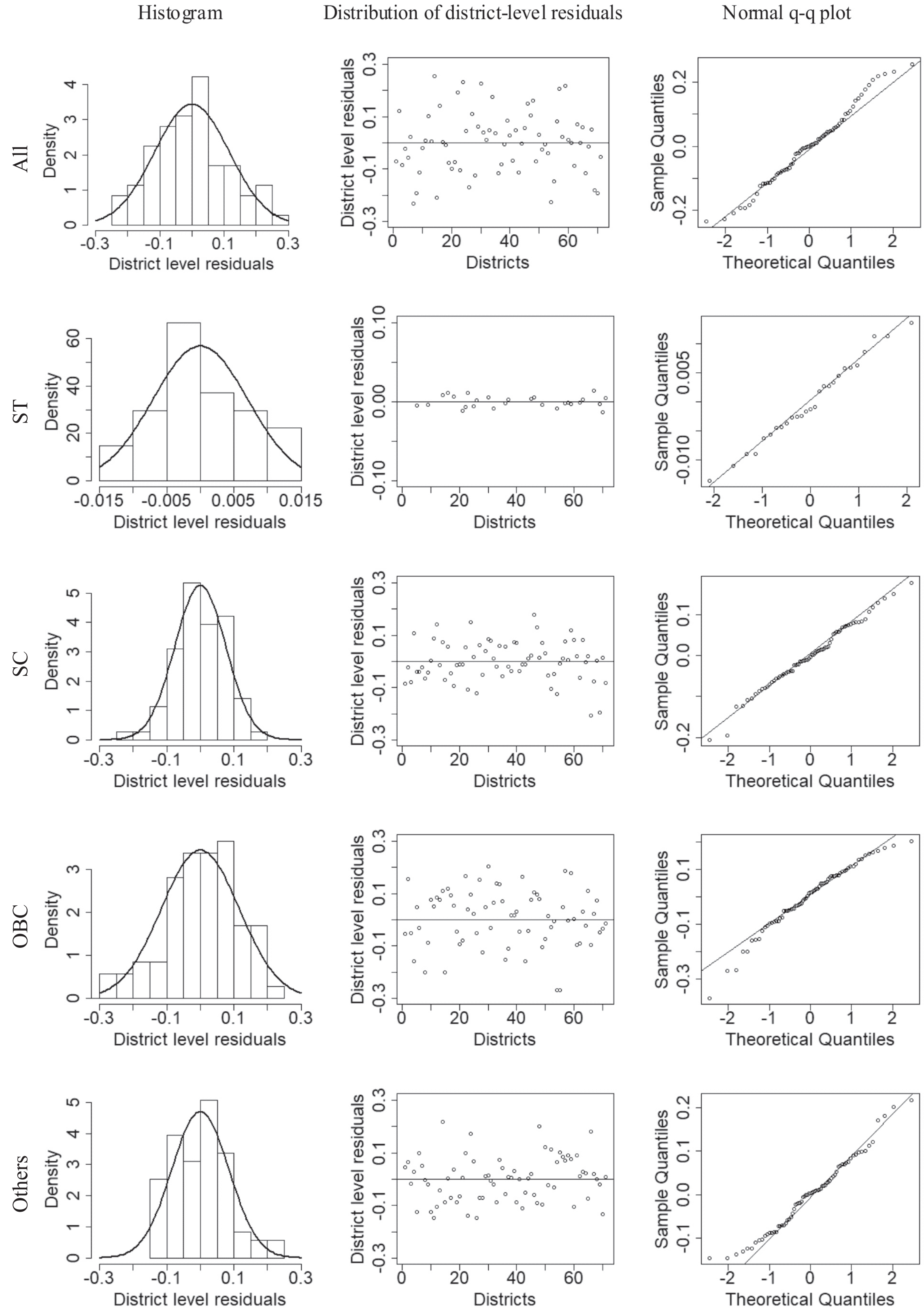


Figure 1. Histogram, distribution and normal q-q plots of the district-level residuals for model based SAE estimates of incidence of indebtedness

Table 5. Goodness of fit diagnostic results

Social group	Goodness of Fit[#]
All	27.304
ST	1.226*
SC	54.436
OBC	37.968
Others	60.681

#A smaller value (less than 91.670) indicates no statistically significant difference between model-based and direct estimates.

*Based data from 8 districts with non-zero variance of direct estimate.

We compute CV to assess the improved precision of the model-based estimates compared to the direct survey estimates. CVs show sampling variability as a percentage of the estimate. Estimates with large CVs are considered unreliable (*Chandra et al. 2011;* Johnson *et al. 2010*). The district-wise distribution of CV of the model-based estimates and the direct estimates for SC, OBC and Others as well as their combined category are shown in figure 2. These plots show that model-based estimates have a higher degree of reliability as compared to the direct estimates. As expected, the relative performance of model-based estimates improves when sample size decreases (see, figure 2). The CV of ST group is not plotted in figure

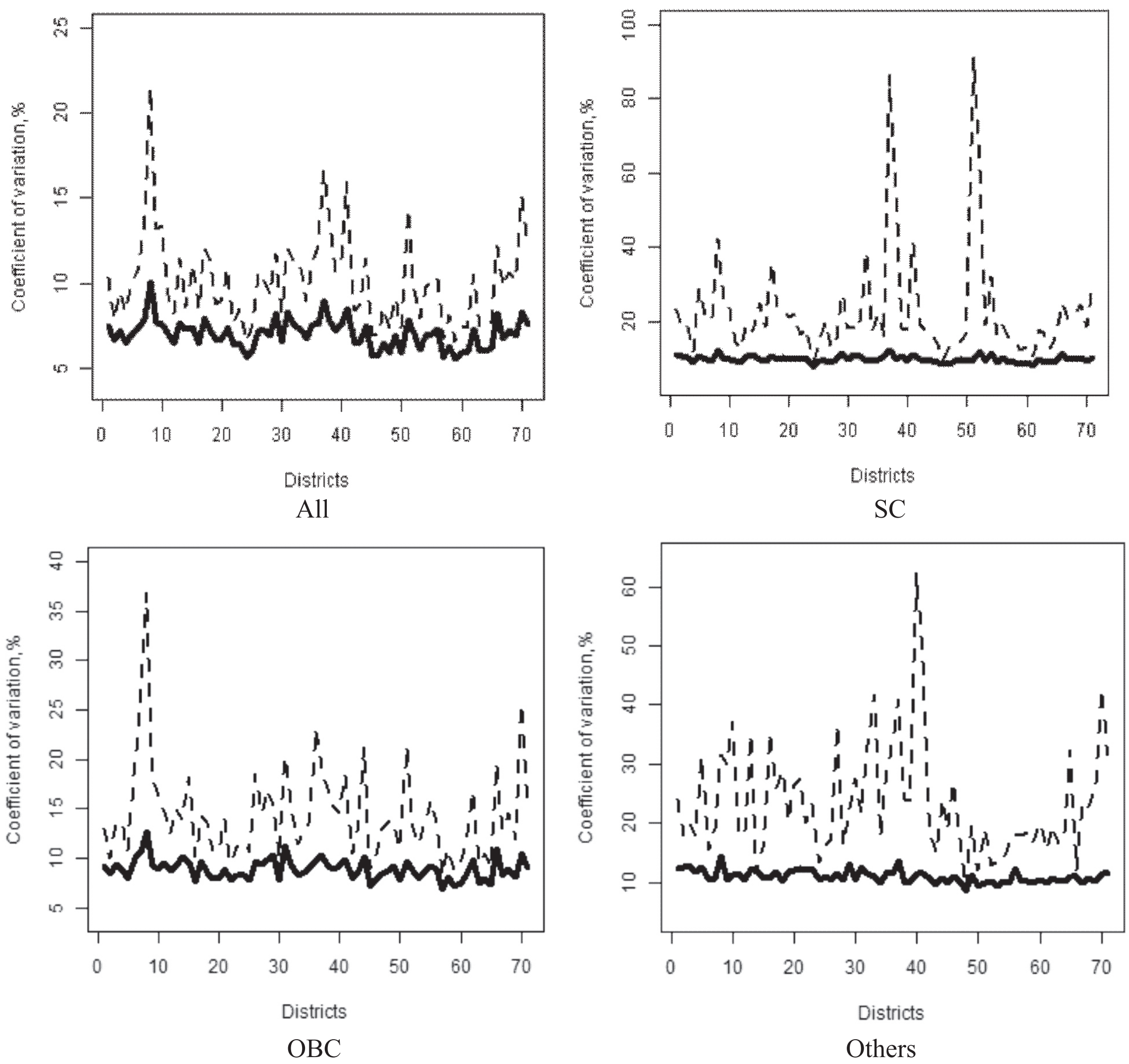


Figure 2. District-wise coefficient of variation (%) for the direct (dash line) and the model-based SAE estimates (solid line) of the incidence of indebtedness

2 because of few estimates only. But, conclusion for ST group is similar as for other social groups. In case of ST, district wise sample sizes are very small. For example, out of 27 sample districts, sample of sizes are just 1, 2, 3 and 4 in 13, 6, 4 and 2 districts respectively. Hence, it is difficult to generate valid direct survey estimates of incidence of indebtedness in such districts. Further, it is not possible to estimate variance (or standard error) of direct estimate with sample size 1. However, SAE method provides reasonable estimates along with estimates of variance (or standard error) for such districts. The model-based estimates for these districts have reasonable values of CV and within the acceptable limit (appendix table 1).

The districts-wise 95% confidence intervals (CIs) of the model-based and the direct estimates are reported in appendix table 1. Figure 3 shows the comparative illustration of 95% CIs of the model-based and the direct estimates. The 95% CIs for the direct estimates are calculated assuming a simple random sample generated the weighted proportions. Obviously, this ignores the effects of differential weighting and clustering within districts that would further inflate the

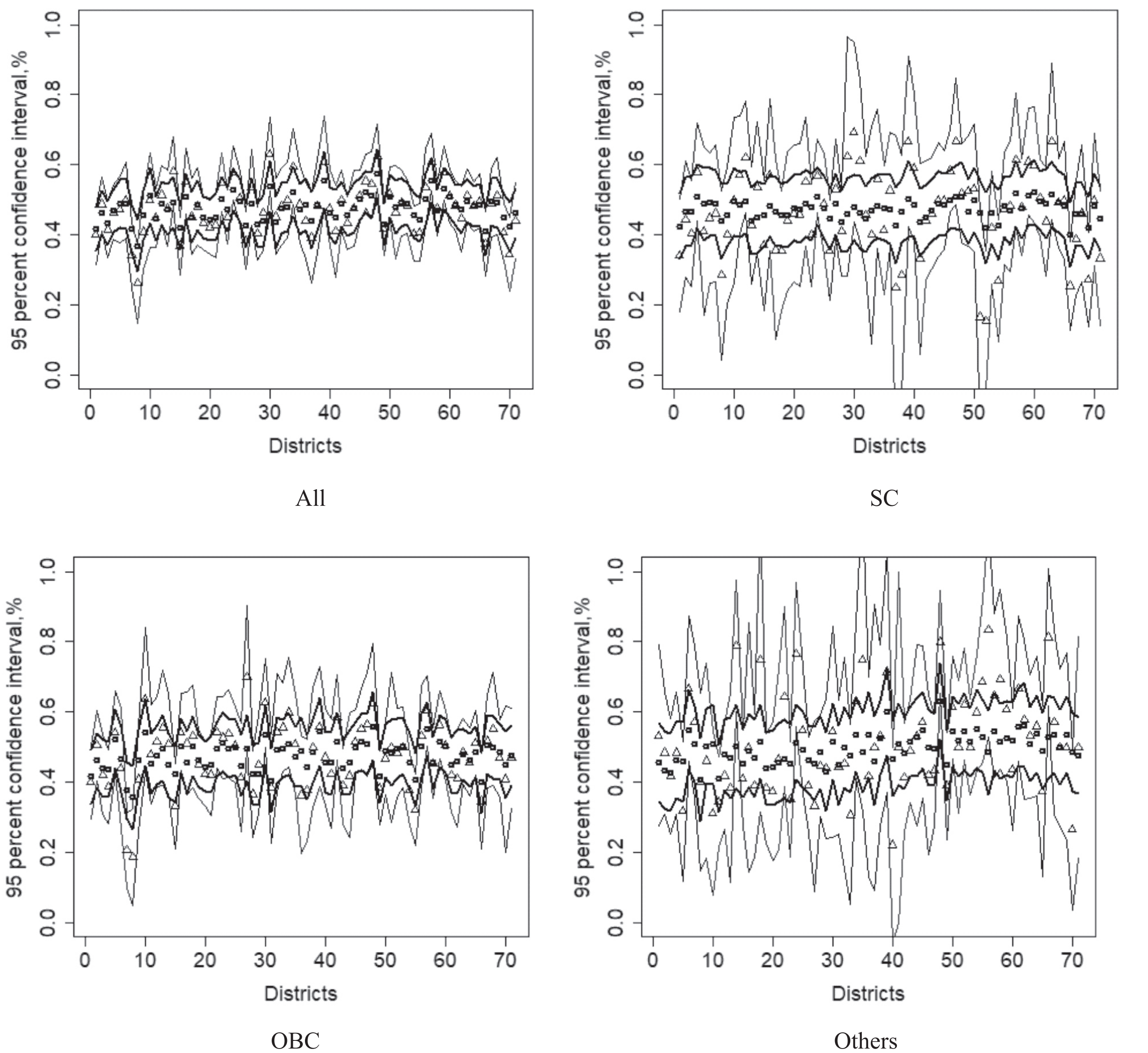
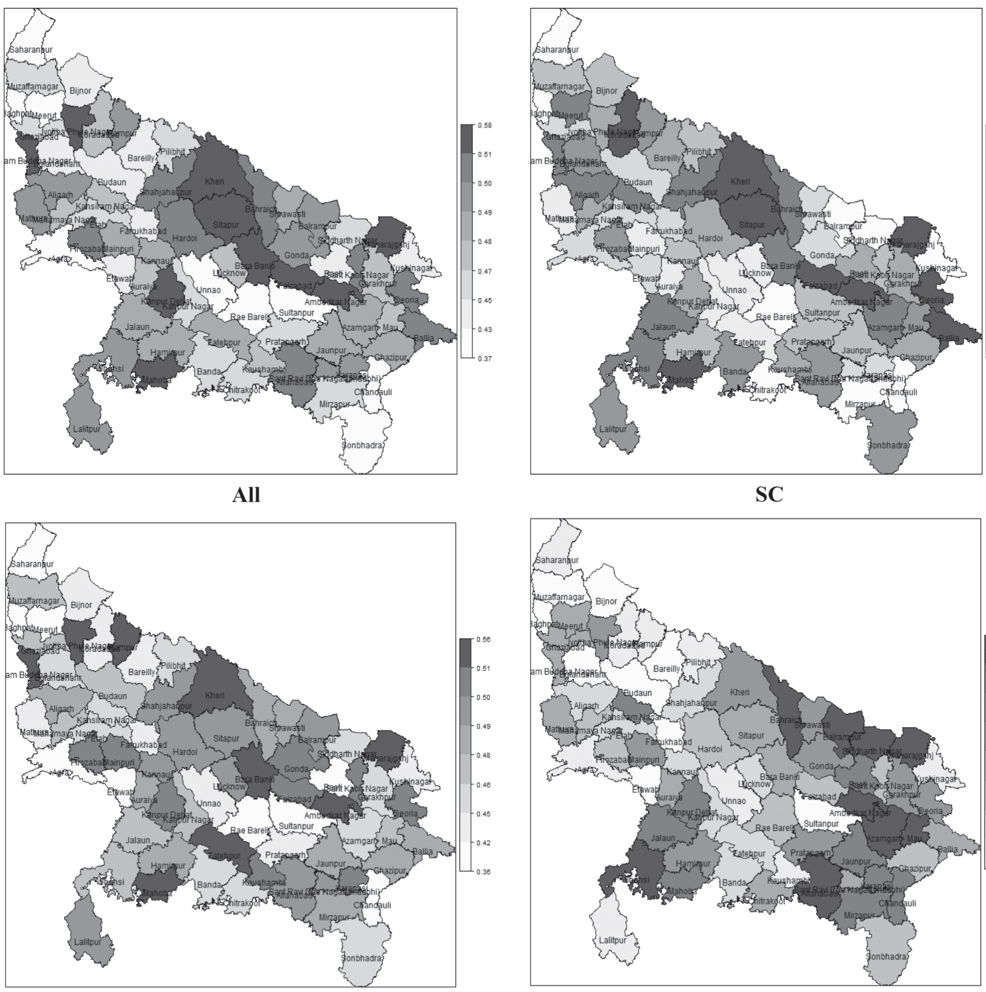


Figure 3. District-wise 95 % confidence interval plots for the direct estimates (dash line) and the model-based SAE estimates (solid line) of the incidence of indebtedness. Direct estimate (dash, ◊) and model based SAE estimate (Solid,°)

true standard errors of the direct estimates. The standard errors of the direct estimates are too large and therefore the estimates are unreliable. In general, 95% CIs for the direct estimates are wider than the 95% CIs for the model-based estimates. Further, 95% CIs for the model-based estimates are more precise and contain both direct and model-based estimates of the incidence of indebtedness. The 95% CIs of ST group is not shown in figure 3. For ST group, in few districts direct estimate has unacceptable and invalid confidence limit, like negative lower bound. This is a well-known problem in direct estimation when area specific sample sizes are very small. This problem also arises when all observations are either 0 or 1, i.e. sample proportion is either 0 or 1. In contrast, the model-based estimates perform well. Further, relative performance of 95% CIs for the model-based estimates improve as sample size decreases (see appendix table 1).

4.2 Discussion

The small area estimates diagnostic measures based on CV and 95% CIs clearly prove reliability and stability of the model-based estimates generated by SAE method over those obtained from direct survey estimates. The results reported in appendix table 1 clearly show the degree of inequality with respect to distribution of indebted rural households in different districts as well as across social groups. The spatial mapping of the incidence of indebtedness among social groups (SC, OBC and Others) and also for their combined category is shown in figure 4. This mapping is useful in identifying location of socially differentiated indebted households. The district wise values of the model-based estimate of indebted households range between 37 to 58 % with an average of 48 %. The model-based estimate of indebted



All SC OBC Others

Figure 4. Maps of the incidence of indebtedness in Uttar Pradesh

households by social groups within districts ranges between 40 to 52 % with average of 47% for SC; 36 to 56% with average of 47 % for OBC; 41 to 63% with average of 50 % for Other castes; and 8 to 72 % with average of 38% for ST (appendix table 1).

Appendix table 2 also provides estimates as well as their lower (Lower) and upper (Upper) bounds at 95% CI for districts and regions, i.e. Eastern, Central, Western and South. A critical review of the results in appendix table 1 shows that in many districts the lower bound of 95% confidence interval is negative and upper bounds greater than one that result in practically impossible and inadmissible values of CI for direct estimates. For example, the CI of direct estimates for Shrawasti and Balrampur districts of Eastern region and Lalitpur and Hamirpur districts of Southern region in SC are negative. Similar is the case of Banda district in Southern region. The CI of direct estimates for Maharajganj, Sant Kabir Nagar and Chandauli districts in eastern region, Mainpuri district in western region, and Jalaun district in southern region exceeds one. In contrast, the model estimate of these districts with precise CI and reasonable CV are still reliable. A similar problem, but in other situation is observed when there is no variability in the sample data for a district. For example, districts where all y values in the sample are one and direct estimate is one, the estimated SE is 0. That is, CI with extreme sample values provides very little information. These abnormalities with direct estimation are seen in many districts for ST group. The average (minimum and maximum) values of percent CV of direct and model-based estimates of indebted are 10.16 (6.6 and 21.47) and 7.03 (5.60 and 10.11), respectively. Similarly, the average (minimum and maximum) values of percent CV of direct and model-based estimates for SC, OBC and Others categories are 22.91 (9.29 and 91.29) and 9.88 (7.85 and 12.54), 14.44 (8.26 and 36.80) and 8.89 (6.83and 12.71) and 23.52 (9.13 and 62.36) and 10.97 (8.59 and 14.54), respectively. It is, thus, evident that model-based SAE method brings gain in efficiency in district level estimates. The SAE can be used as cost effective and efficient approach for generating reliable micro level statistics using the existing survey data combining with auxiliary information from other sources. The results clearly indicate the advantage of using SAE technique to cope up the small sample size problem in producing estimates or reliable confidence intervals.

5 Concluding remarks

In recent years, SAE has received considerable attention of researchers and practitioners, resulting in tremendous theoretical literature on the topic. The theoretical research in statistics can be advantageous for the society in true sense if it is applied and implemented in solving real life problems. For example, theory of SAE method for estimation of proportions for small areas is well developed, yet its applications in agricultural or social sciences are not much. In India, the Census is usually limited in its scope in collection of data; it focuses mainly on basic social and demographic information and that too at decennial interval. On the other hand, NSSO conducts regular surveys on a number of socioeconomic indicators, but their utility is restricted to generate national and state level estimates, but not administrative units below state because of small sample sizes for such units. In this paper, we illustrate application of SAE techniques to generate reliable and representative estimates of indebtedness among social groups at district level by combining data from AIDIS 2012-13 with data from Census 2011.

The diagnostic measures reveal that SAE estimates precise and representative of the social and spatial heterogeneity at a higher level of spatial disaggregation. The method and application discussed in this paper can be used as guideline for producing reliable, timely and cost-effective estimates using survey data. The Government of India is pacing considerable emphasis on micro planning for achieving a balanced economic development, and through SAE techniques is possible to generate information various social and economic aspects at a higher level of spatial disaggregation that can be used in prioritization and targeting of efforts and investments.

Acknowledgement

The work presented in this paper was carried out under ICAR-National Fellow project at ICAR- Indian Agricultural Statistics Research Institute, New Delhi, India.

References

Battese, G.E., Harter, R.M. & Fuller, W.A. (1988). An error component model for prediction of county crop areas using survey and satellite data. Journal of the American Statistical Association, 83, 28-36.

Breslow, N.E. & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 88, 9-25.

Brown, G., Chambers, R., Heady, P. & Heasman, D. (2001). Evaluation of small area estimation methods - an application to unemployment estimates from the UK LFS. In: Proceedings of the symposium 'achieving data Quality in a statistical agency: a methodological perspective'. Statistics Canada, Quebec, Canada, 16-19, October.

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.

Fay, R.E. & Herriot, R.A. (1979). Estimation of income from small places: an application of James-Stein procedures to census data. Journal of the American Statistical Association, 74, 269-277.

Johnson, F.A., Chandra, H., Brown, J. & Padmadas, S. (2010). Estimating district-level births attended by skilled attendants in ghana using demographic health survey and census data: an application of small area estimation technique. Journal Official Statistics, 26(2), 341-359.

Pfeffermann, D. (2002). Small area estimation: new developments and directions. International Statistical Review, 70, 125-143.

Rao, J.N.K. & Molina I. (2015). Small Area Estimation, John Wiley and Sons, New York.

Saei, A. & Chambers R. (2003). Small area estimation under linear and generalized linear mixed models with time and area effects. Working Paper M03/15, Southampton Statistical Sciences Research Institute, University of Southampton, United Kingdom.

Received: 15 February 2017; Accepted: 17 April 2018

Appendix table 1. District and social group-wise direct and model-based estimates of incidence of indebtedness

Region	District	All						SC					
		Direct			Model-based			Direct			Model-based		
		Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper
Eastern	Maharajganj	0.61	0.48	0.74	0.56	0.48	0.64	0.67	0.42	0.91	0.50	0.39	0.61
	Pratapgarh	0.43	0.36	0.51	0.46	0.40	0.52	0.46	0.30	0.62	0.47	0.38	0.56
	Kaushambi	0.48	0.37	0.59	0.48	0.41	0.55	0.50	0.34	0.66	0.49	0.40	0.58
	Allahabad	0.51	0.44	0.58	0.50	0.45	0.56	0.50	0.36	0.64	0.49	0.40	0.57
	Faizabad	0.54	0.45	0.64	0.51	0.45	0.58	0.67	0.49	0.85	0.51	0.42	0.60
	Ambedkar Nagar	0.63	0.53	0.72	0.58	0.51	0.64	0.52	0.38	0.66	0.51	0.41	0.61
	Sultanpur	0.42	0.34	0.49	0.43	0.37	0.49	0.53	0.37	0.68	0.47	0.38	0.56
	Bahraich	0.52	0.44	0.59	0.51	0.45	0.57	0.53	0.35	0.72	0.50	0.41	0.59
	Shrawasti	0.46	0.33	0.60	0.48	0.40	0.55	0.17	-0.14	0.47	0.47	0.37	0.56
	Balrampur	0.49	0.40	0.59	0.49	0.42	0.56	0.15	-0.05	0.35	0.42	0.32	0.52
	Gonda	0.48	0.41	0.56	0.49	0.43	0.55	0.42	0.26	0.58	0.46	0.37	0.55
	Siddharth Nagar	0.41	0.33	0.48	0.46	0.39	0.52	0.27	0.10	0.44	0.43	0.33	0.53
	Basti	0.41	0.32	0.49	0.44	0.38	0.50	0.48	0.32	0.63	0.48	0.39	0.57
	Sant Kabir Nagar	0.54	0.43	0.65	0.50	0.43	0.58	0.48	0.30	0.67	0.47	0.37	0.57
	Gorakhpur	0.48	0.40	0.55	0.47	0.41	0.53	0.48	0.34	0.62	0.48	0.39	0.56
	Kushinagar	0.58	0.50	0.65	0.53	0.47	0.59	0.61	0.46	0.76	0.51	0.43	0.60
	Deoria	0.51	0.43	0.59	0.51	0.45	0.57	0.60	0.43	0.77	0.52	0.43	0.62
	Azamgarh	0.48	0.41	0.56	0.49	0.43	0.54	0.51	0.40	0.62	0.50	0.41	0.58
	Mau	0.45	0.35	0.54	0.47	0.40	0.54	0.44	0.28	0.59	0.49	0.39	0.59
	Ballia	0.51	0.43	0.59	0.50	0.44	0.56	0.67	0.44	0.89	0.52	0.42	0.61
	Jaunpur	0.48	0.41	0.55	0.48	0.42	0.54	0.49	0.37	0.62	0.49	0.41	0.58
	Ghazipur	0.50	0.42	0.57	0.49	0.43	0.55	0.50	0.33	0.67	0.48	0.39	0.57
	Chandauli	0.38	0.28	0.47	0.41	0.34	0.48	0.26	0.13	0.38	0.40	0.31	0.49
	Varanasi	0.49	0.39	0.59	0.50	0.43	0.56	0.39	0.23	0.55	0.46	0.37	0.55
	Sant Ravi Das Nagar	0.51	0.40	0.62	0.49	0.42	0.56	0.46	0.27	0.66	0.46	0.37	0.55
	Mirzapur	0.41	0.32	0.49	0.45	0.39	0.51	0.27	0.14	0.41	0.42	0.33	0.51
	Sonbhadra	0.35	0.24	0.45	0.42	0.35	0.49	0.50	0.31	0.69	0.48	0.39	0.58
Western	Saharanpur	0.40	0.32	0.48	0.42	0.36	0.48	0.34	0.18	0.50	0.43	0.33	0.52
	Muzaffarnagar	0.49	0.41	0.57	0.46	0.40	0.52	0.44	0.28	0.61	0.47	0.37	0.57
	Bijnor	0.42	0.34	0.50	0.43	0.37	0.50	0.40	0.25	0.56	0.47	0.37	0.56
	Moradabad	0.46	0.39	0.54	0.47	0.41	0.53	0.58	0.44	0.72	0.51	0.42	0.60
	Rampur	0.47	0.38	0.57	0.49	0.42	0.56	0.41	0.17	0.65	0.49	0.39	0.59
	Jyotiba Phule Nagar	0.50	0.39	0.61	0.49	0.42	0.56	0.45	0.26	0.63	0.49	0.40	0.59
	Meerut	0.34	0.25	0.44	0.42	0.35	0.49	0.46	0.27	0.66	0.49	0.40	0.58
	Baghpat	0.26	0.15	0.37	0.37	0.30	0.44	0.29	0.04	0.53	0.44	0.33	0.55
	Ghaziabad	0.41	0.30	0.52	0.46	0.39	0.53	0.40	0.20	0.60	0.46	0.37	0.55
	G.B.Nagar	0.50	0.37	0.63	0.51	0.44	0.59	0.50	0.26	0.74	0.50	0.40	0.59
	Bulandshahr	0.45	0.37	0.53	0.45	0.39	0.51	0.57	0.40	0.74	0.49	0.40	0.58
	Aligarh	0.51	0.43	0.60	0.49	0.43	0.55	0.62	0.46	0.78	0.50	0.41	0.59
	Mahamaya Nagar	0.48	0.37	0.59	0.47	0.40	0.55	0.43	0.26	0.60	0.44	0.35	0.54
	Mathura	0.58	0.48	0.68	0.49	0.42	0.56	0.54	0.35	0.72	0.45	0.35	0.55
	Agra	0.37	0.29	0.45	0.42	0.36	0.48	0.37	0.18	0.56	0.46	0.37	0.55
	Firozabad	0.55	0.46	0.65	0.51	0.44	0.58	0.57	0.36	0.79	0.48	0.39	0.57
	Etah	0.45	0.34	0.56	0.45	0.38	0.52	0.36	0.10	0.61	0.47	0.37	0.56
	Mainpuri	0.48	0.37	0.59	0.49	0.42	0.56	0.35	0.18	0.53	0.46	0.37	0.54
	Budaun	0.43	0.36	0.51	0.45	0.39	0.51	0.44	0.24	0.64	0.46	0.36	0.55

contd...

	Bareilly	0.42	0.35	0.50	0.44	0.38	0.50	0.46	0.27	0.66	0.48	0.38	0.57
	Pilibhit	0.43	0.34	0.52	0.45	0.38	0.52	0.46	0.25	0.66	0.48	0.38	0.57
	Shahjahanpur	0.55	0.47	0.63	0.50	0.44	0.57	0.55	0.37	0.74	0.49	0.40	0.59
	Farrukhabad	0.46	0.36	0.57	0.44	0.37	0.51	0.63	0.28	0.97	0.46	0.35	0.56
	Kannauj	0.63	0.53	0.74	0.54	0.47	0.61	0.69	0.44	0.95	0.48	0.39	0.57
	Etawah	0.45	0.34	0.56	0.44	0.36	0.51	0.61	0.38	0.84	0.47	0.36	0.57
	Auraiya	0.48	0.37	0.59	0.48	0.40	0.55	0.46	0.29	0.63	0.45	0.35	0.54
	Kansiram Nagar	0.44	0.33	0.55	0.46	0.39	0.53	0.33	0.14	0.53	0.45	0.36	0.54
Central	Kheri	0.45	0.38	0.53	0.47	0.41	0.54	0.40	0.25	0.55	0.48	0.38	0.58
	Sitapur	0.58	0.50	0.65	0.53	0.47	0.59	0.57	0.46	0.67	0.51	0.43	0.59
	Hardoi	0.51	0.43	0.58	0.50	0.44	0.56	0.49	0.35	0.63	0.48	0.39	0.57
	Unnao	0.39	0.30	0.47	0.43	0.37	0.49	0.36	0.21	0.50	0.45	0.36	0.53
	Lucknow	0.54	0.43	0.64	0.49	0.42	0.56	0.53	0.39	0.67	0.49	0.40	0.58
	Rae Bareli	0.40	0.33	0.48	0.43	0.37	0.49	0.41	0.28	0.54	0.44	0.35	0.52
	Kanpur Dehat	0.50	0.39	0.61	0.48	0.41	0.55	0.40	0.09	0.71	0.47	0.38	0.56
	Kanpur Nagar	0.60	0.49	0.70	0.52	0.45	0.59	0.56	0.36	0.76	0.48	0.39	0.57
	Fatehpur	0.50	0.42	0.58	0.49	0.43	0.55	0.44	0.27	0.61	0.45	0.36	0.54
	Bara Banki	0.55	0.48	0.63	0.52	0.46	0.58	0.58	0.47	0.70	0.50	0.42	0.59
Southern	Jalaun	0.49	0.38	0.60	0.47	0.40	0.54	0.41	0.23	0.60	0.48	0.38	0.57
	Jhansi	0.44	0.33	0.55	0.49	0.41	0.56	0.53	0.36	0.69	0.48	0.37	0.58
	Lalitpur	0.39	0.26	0.52	0.44	0.36	0.52	0.25	-0.18	0.68	0.43	0.32	0.54
	Hamirpur	0.48	0.35	0.62	0.49	0.41	0.56	0.29	-0.06	0.63	0.47	0.38	0.56
	Banda	0.48	0.37	0.59	0.47	0.39	0.54	0.59	0.38	0.80	0.49	0.39	0.58
	Chitrakoot	0.41	0.28	0.54	0.45	0.38	0.53	0.33	0.06	0.61	0.44	0.34	0.53
	Mahoba	0.61	0.52	0.69	0.56	0.49	0.62	0.62	0.42	0.81	0.52	0.42	0.61

Appendix table 1 contd.

Region	District	OBC						Others					
		Direct			Model-based			Direct			Model-based		
		Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper
Eastern	Maharajganj	0.56	0.39	0.73	0.55	0.45	0.64	0.71	0.37	1.06	0.60	0.48	0.72
	Pratapgarh	0.39	0.28	0.50	0.44	0.36	0.52	0.49	0.34	0.64	0.52	0.42	0.62
	Kaushambi	0.42	0.24	0.60	0.46	0.36	0.55	0.53	0.28	0.79	0.53	0.42	0.64
	Allahabad	0.52	0.42	0.61	0.50	0.43	0.57	0.57	0.36	0.79	0.55	0.44	0.66
	Faizabad	0.57	0.42	0.71	0.51	0.42	0.59	0.44	0.28	0.59	0.50	0.40	0.59
	Ambedkar Nagar	0.63	0.45	0.80	0.56	0.46	0.66	0.80	0.65	0.95	0.63	0.52	0.74
	Sultanpur	0.38	0.28	0.49	0.42	0.34	0.49	0.39	0.24	0.54	0.45	0.35	0.55
	Bahraich	0.47	0.37	0.57	0.49	0.41	0.57	0.61	0.47	0.76	0.55	0.44	0.65
	Shrawasti	0.50	0.29	0.71	0.48	0.39	0.58	0.50	0.31	0.69	0.52	0.42	0.62
	Balrampur	0.48	0.36	0.61	0.50	0.41	0.58	0.62	0.46	0.78	0.55	0.43	0.66
	Gonda	0.50	0.39	0.61	0.50	0.42	0.58	0.50	0.36	0.64	0.52	0.42	0.61
	Siddharth Nagar	0.38	0.28	0.48	0.46	0.38	0.54	0.60	0.43	0.77	0.55	0.44	0.67
	Basti	0.32	0.22	0.42	0.41	0.33	0.48	0.69	0.46	0.92	0.53	0.43	0.63
	Sant Kabir Nagar	0.53	0.39	0.68	0.50	0.41	0.59	0.83	0.53	1.14	0.49	0.37	0.60
	Gorakhpur	0.45	0.35	0.55	0.46	0.38	0.53	0.69	0.44	0.95	0.52	0.41	0.62
	Kushinagar	0.56	0.47	0.66	0.52	0.44	0.59	0.61	0.41	0.81	0.53	0.42	0.63
	Deoria	0.50	0.41	0.60	0.50	0.43	0.58	0.44	0.26	0.61	0.52	0.41	0.63
	Azamgarh	0.42	0.32	0.53	0.45	0.37	0.53	0.67	0.46	0.87	0.56	0.45	0.67
	Mau	0.41	0.27	0.55	0.46	0.37	0.55	0.58	0.35	0.81	0.56	0.44	0.68
	Ballia	0.49	0.39	0.58	0.48	0.41	0.55	0.53	0.35	0.71	0.51	0.41	0.61
	Jaunpur	0.46	0.36	0.55	0.46	0.39	0.54	0.56	0.36	0.76	0.54	0.43	0.65
	Ghazipur	0.51	0.42	0.61	0.49	0.42	0.56	0.38	0.13	0.62	0.49	0.38	0.60

contd...

	Chandauli	0.35	0.21	0.48	0.40	0.31	0.49	0.81	0.62	1.01	0.53	0.41	0.65
	Varanasi	0.52	0.38	0.65	0.51	0.42	0.59	0.57	0.31	0.84	0.54	0.43	0.64
	Sant Ravi Das Nagar	0.55	0.39	0.71	0.50	0.41	0.59	0.50	0.28	0.72	0.50	0.39	0.61
	Mirzapur	0.47	0.36	0.58	0.49	0.41	0.56	0.50	0.23	0.77	0.54	0.43	0.64
	Sonbhadra	0.41	0.20	0.62	0.45	0.36	0.54	0.27	0.04	0.50	0.49	0.38	0.60
Western	Saharanpur	0.40	0.30	0.50	0.42	0.34	0.49	0.53	0.28	0.79	0.46	0.34	0.57
	Muzaffarnagar	0.51	0.40	0.61	0.46	0.39	0.54	0.48	0.31	0.66	0.44	0.32	0.55
	Bijnor	0.42	0.31	0.53	0.44	0.36	0.52	0.42	0.25	0.58	0.43	0.32	0.54
	Moradabad	0.39	0.28	0.49	0.44	0.36	0.52	0.48	0.31	0.66	0.46	0.36	0.57
	Rampur	0.54	0.42	0.66	0.52	0.44	0.61	0.32	0.12	0.52	0.46	0.35	0.58
	Jyotiba Phule Nagar	0.44	0.27	0.61	0.47	0.37	0.56	0.67	0.46	0.87	0.55	0.44	0.66
	Meerut	0.21	0.10	0.31	0.38	0.30	0.46	0.57	0.36	0.79	0.51	0.40	0.62
	Baghpat	0.19	0.05	0.33	0.36	0.27	0.45	0.40	0.15	0.65	0.41	0.29	0.53
	Ghaziabad	0.41	0.26	0.56	0.46	0.38	0.55	0.46	0.19	0.74	0.50	0.40	0.61
	G.B. Nagar	0.64	0.43	0.84	0.54	0.45	0.64	0.31	0.08	0.54	0.51	0.40	0.62
	Bulandshahr	0.48	0.34	0.62	0.45	0.37	0.54	0.35	0.22	0.47	0.41	0.32	0.51
	Aligarh	0.52	0.39	0.65	0.48	0.39	0.56	0.42	0.27	0.57	0.48	0.38	0.58
	Mahamaya Nagar	0.56	0.39	0.72	0.50	0.41	0.59	0.38	0.11	0.65	0.47	0.35	0.58
	Mathura	0.51	0.37	0.65	0.45	0.36	0.54	0.79	0.60	0.98	0.50	0.38	0.62
	Agra	0.33	0.21	0.46	0.42	0.34	0.50	0.41	0.27	0.55	0.46	0.36	0.55
	Firozabad	0.54	0.43	0.65	0.50	0.43	0.58	0.50	0.15	0.85	0.49	0.38	0.59
	Etah	0.51	0.36	0.66	0.46	0.37	0.55	0.39	0.19	0.59	0.47	0.36	0.58
	Mainpuri	0.53	0.39	0.68	0.50	0.42	0.59	0.75	0.32	1.18	0.52	0.41	0.62
	Budaun	0.45	0.35	0.55	0.46	0.39	0.54	0.38	0.23	0.54	0.44	0.34	0.54
	Bareilly	0.42	0.33	0.51	0.45	0.37	0.52	0.38	0.18	0.57	0.44	0.34	0.55
	Pilibhit	0.42	0.31	0.54	0.45	0.37	0.53	0.47	0.21	0.72	0.46	0.35	0.57
	Shahjahanpur	0.54	0.44	0.65	0.50	0.42	0.57	0.64	0.39	0.90	0.47	0.35	0.58
	Farrukhabad	0.45	0.31	0.59	0.43	0.34	0.51	0.44	0.24	0.64	0.43	0.32	0.54
	Kannauj	0.63	0.50	0.75	0.54	0.45	0.62	0.55	0.25	0.85	0.52	0.41	0.62
	Etawah	0.38	0.23	0.54	0.41	0.32	0.50	0.44	0.25	0.64	0.45	0.34	0.56
	Auraiya	0.54	0.37	0.71	0.49	0.40	0.58	0.45	0.15	0.75	0.50	0.38	0.62
	Kansiram Nagar	0.47	0.33	0.61	0.48	0.39	0.56	0.50	0.18	0.82	0.48	0.37	0.59
Central	Kheri	0.53	0.41	0.65	0.51	0.43	0.60	0.35	0.19	0.52	0.45	0.34	0.57
	Sitapur	0.54	0.41	0.67	0.50	0.42	0.58	0.76	0.56	0.97	0.51	0.41	0.62
	Hardoi	0.51	0.40	0.61	0.50	0.42	0.58	0.55	0.37	0.73	0.49	0.39	0.60
	Unnao	0.41	0.26	0.57	0.45	0.36	0.53	0.39	0.26	0.53	0.46	0.37	0.56
	Lucknow	0.70	0.50	0.90	0.50	0.40	0.59	0.33	0.09	0.58	0.45	0.35	0.56
	Rae Bareli	0.37	0.24	0.49	0.42	0.34	0.51	0.45	0.30	0.59	0.49	0.38	0.59
	Kanpur Dehat	0.56	0.43	0.68	0.50	0.42	0.58	0.31	0.05	0.56	0.48	0.37	0.58
	Kanpur Nagar	0.60	0.45	0.75	0.51	0.42	0.60	0.63	0.41	0.85	0.54	0.43	0.64
	Fatehpur	0.58	0.46	0.71	0.52	0.43	0.60	0.41	0.26	0.57	0.48	0.38	0.58
	Bara Banki	0.55	0.44	0.67	0.52	0.44	0.59	0.42	0.19	0.65	0.50	0.39	0.61
Southern	Jalaun	0.51	0.37	0.65	0.47	0.39	0.56	0.75	0.32	1.18	0.49	0.37	0.60
	Jhansi	0.36	0.20	0.53	0.49	0.39	0.59	0.42	0.13	0.70	0.54	0.41	0.66
	Lalitpur	0.38	0.23	0.52	0.44	0.35	0.53	0.50	0.09	0.91	0.46	0.33	0.59
	Hamirpur	0.50	0.33	0.67	0.49	0.40	0.58	0.53	0.28	0.79	0.53	0.42	0.63
	Banda	0.47	0.33	0.61	0.46	0.37	0.54	0.22	-0.05	0.50	0.47	0.36	0.58
	Chitrakoot	0.43	0.27	0.58	0.46	0.37	0.55	0.50	0.00	1.00	0.51	0.39	0.62
	Mahoba	0.60	0.50	0.70	0.55	0.48	0.63	0.65	0.42	0.88	0.54	0.44	0.65

Appendix table 2: District and social group-wise direct and model-based estimates of incidence of indebtedness in Uttar Pradesh

Region	District	ST					
		Direct			Model-based		
		Estimate	Lower	Upper	Estimate	Lower	Upper
Eastern	Allahabad	0.25	-0.18	0.68	0.16	0.00	0.32
	Sultanpur	0.00	0.00	0.00	0.44	0.16	0.72
	Siddharth Nagar	0.00	0.00	0.00	0.46	0.08	0.84
	Gorakhpur	0.25	-0.18	0.68	0.28	0.08	0.48
	Kushinagar	0.00	0.00	0.00	0.26	0.12	0.40
	Mau	0.00	0.00	0.00	0.08	0.04	0.20
	Ballia	0.33	-0.21	0.88	0.24	0.06	0.42
	Varanasi	1.00	1.00	1.00	0.30	0.10	0.50
	Mirzapur	0.00	0.00	0.00	0.27	0.07	0.47
	Sonbhadra	0.11	-0.04	0.25	0.18	0.04	0.40
Western	Rampur	0.00	0.00	0.00	0.46	0.22	0.70
	Ghaziabad	0.00	0.00	0.00	0.37	0.19	0.55
	Mathura	1.00	1.00	1.00	0.56	0.12	1.00
	Firozabad	1.00	1.00	1.00	0.42	0.22	0.62
	Mainpuri	1.00	1.00	1.00	0.36	0.20	0.52
	Pilibhit	0.00	0.00	0.00	0.56	0.28	0.84
	Shahjahanpur	0.33	-0.21	0.88	0.55	0.21	0.89
	Kannauj	1.00	1.00	1.00	0.40	0.18	0.62
	Auraiya	0.00	0.00	0.00	0.46	0.16	0.76
	Kansiram Nagar	1.00	1.00	1.00	0.54	0.30	0.78
Central	Kheri	0.47	0.23	0.71	0.40	0.00	0.80
	Hardoi	0.00	0.00	0.00	0.57	0.33	0.81
	Unnao	0.33	-0.21	0.88	0.27	0.09	0.45
	Bara Banki	1.00	1.00	1.00	0.41	0.17	0.65
Southern	Jhansi	0.33	-0.21	0.88	0.39	0.01	0.77
	Lalitpur	1.00	1.00	1.00	0.72	0.38	1.06
	Mahoba	0.00	0.00	0.00	0.25	0.07	0.43