Statistics and Applications {ISSN 2452-7395 (online)}
Volume 18, No. 1, 2020 (New Series), pp 21–33

Estimation and Spatial Mapping of Incidence of Indebtedness in the State of Karnataka in India by Combining Survey and Census Data

Priyanka Anjoy, Hukum Chandra and Rajender Parsad
ICAR-Indian Agricultural Statistics Research Institute, New Delhi.

Received: 28 December 2019; Revised: 17 January 2020; Accepted: 20 March 2020

Abstract

Information about the household debt behaviour in different occupational categories is of key importance to the Governmental organization for taking effective policy measures targeting the vulnerable groups. This paper illustrates small area estimation (SAE) methodology to estimate proportion of indebted households in rural areas for the two major occupation categories- rural cultivator and rural non-cultivator as well as for both categories combined together across all the 30 districts of Karnataka state in India using the data of All India Debt and Investment Survey 2012-13 and population census 2011. The findings show that the district-level estimates of incidence of indebtedness obtained from SAE are more precise than the direct survey estimates. A spatial map has also been produced to observe the inequality in distribution of indebtedness within districts and in each occupational category across districts. Such maps are definitely useful for framing consistent policy actions and fund disbursement for the indebted household mass.

Key words: Small area estimation; Generalized linear mixed model; Indebtedness; Spatial Map.

---

1. Introduction

Agriculture plays an important role in the economy of Karnataka and it is the main occupation for more than 60% of population. Karnataka is a drought-prone region with a large proportion of wasteland and having the second largest arid zone in the country after Rajasthan. And due to these factors, the state has been facing agrarian distress with increasing incidence of farmers' suicides since 1997. In fact, the rate of farmer suicides in Karnataka has hit the highest level in a decade, topping the list after Maharashtra, highlighting agrarian distress in the state, according to the report Accidental Deaths and Suicides in India 2015 published by National Crime Records Bureau (NCRB). According to NCRB 2015 data, about 1,197 farmers committed suicides in Karnataka during 2014-15; the state was just behind Maharashtra and Telangana. The NCRB also found that about 79% suicides (946 out of 1,197) in Karnataka were due to bankruptcy or indebtedness. The pre-requisite for any effective policy approach taken in this regard is a proper statistical and economic framework that allows for an effective analysis and monitoring of farmers'

Corresponding Author: Priyanka Anjoy
E-mail: anjoypriyanka90@gmail.com

distress. Measure of disaggregated level indebtedness can be an important tool to the policy makers to mark certain region or group for upliftment and reduce the situation of agrarian distress or farmers' suicides. In this study we attempt to estimate such micro or disaggregated level incidence of indebtedness at micro or local level using the area level small area model.

Most of the large scale surveys are planned to produce reliable estimates at macro or higher geographical (e.g. national and state) level, and cannot be used directly to generate reliable micro or local (also referred to as small area) level estimates because of the small sample sizes (Rao and Molina 2015). This is because, large scale survey designed for a large population (*e.g.* national and state level) may select a small number of units or even no unit from the small area of interest (*e.g.* district or further disaggregation of district). Hence, sample sizes from small areas (or small domains) are too small to justify the use of traditional direct survey estimates. The underlying theory in the literature of survey sampling that helps in resolving the problem of smaller sample sizes is referred as small area estimation (SAE) technique. The technique is model-based methods that links the variable of interest from survey with the auxiliary information available from other data sources for small areas and hence increase the overall (effective) sample size and precision. In this paper we employ area level SAE technique to produce reliable estimates of the incidence of indebtedness among cultivators and non-cultivators categories as well as for both the categories combined in different districts of rural areas of Karnataka in India by linking data from the All-India Debt and Investment Survey (AIDIS) 2012-13 of National Sample Survey Office (NSSO), and the Population Census 2011. This work will enable us to obtain spatial distribution of incidence of indebtedness as well as regional inequality in such measure of indebtedness among the farm families and other families of rural areas in Karnataka. The rest of the paper has been organized into five sections. In Section 2, we discuss the data used in the paper. Section 3 provides an overview of SAE technique that has been used to generate incidence of indebtedness among occupational category by districts in Karnataka. In Section 4, we present diagnostic procedures to examine model assumptions and validate small area estimates including discussion about the results. Finally, Section 5 provides concluding remarks and some recommendations.

2. Data Description

This Section describes about data used in this analysis. In particular, the SAE analysis is based on the AIDIS 2012-13 data for rural areas of the State of Karnataka in India and the Population Census 2011. The sampling design used in the AIDIS 2012-13 data is stratified multi-stage random sampling with districts as strata, the census villages in the rural sector as first stage units and households as the ultimate stage units. For the state of Karnataka, there are a total of 2340 surveyed rural households (including both indebted and non-indebted) spread over 30 districts. The rural households are broadly classified into two types; namely; cultivator and non-cultivator households. As per the concepts and definitions of AIDIS, all rural households operating at least 0.002 hectare of land during the 365 days preceding the date of survey are treated as 'cultivator households'. On the other hand, all rural households operating no land or land less than 0.002 hectare are considered to be non-cultivator households. What follows, based on land holding size (LHS), we denote three categories of households: (i) LHS-A: All households (ii) LHS-C: Cultivator-households with LHS greater than 0.002 ha, and (iii) LHS-NC: Non cultivator-households with LHS less or equal to 0.002 ha. Here, the districts and district by household categories are small areas of interest. Table 1 presents the distribution of district-wise sample sizes for three categories of households. Across all the districts (*i.e.* LHS-A), the sample size ranges between a minimum of 55 households to a maximum of 112 with an average of 78 households. The sample sizes become too small if sub-grouped further by land holding size categories (i.e. district by cultivator and non-cultivator categories). That is, the sample size of rural cultivators (LHS-C) varies from a minimum of 23 to a maximum of 90 households across the 30 districts with an average of 49 households. And for non-cultivators (LHS-NC), the sample size varies from a minimum of 11 to a maximum of 51 households across the districts with an average of 29 households. Such small samples from the districts pose a challenge in deriving reliable direct estimates of indebtedness. Thus, SAE is an obvious choice to address this problem.

Table 1: Distribution of sample size by occupational categories across districts in rural Karnataka

District	All	Cultivator	Non-Cultivator	District	All	Cultivator	Non-Cultivator
Belgaum	112	67	45	Tumkur	112	90	22
Bagalkot	84	57	27	Kolar	56	45	11
Bijapur	112	85	27	Bangalore	56	23	33
Gulbarga	98	60	38	Bangalore Rural	56	34	22
Bidar	84	49	35	Mandya	112	85	27
Raichur	84	55	29	Hassan	84	63	21
Koppal	84	63	21	Dakshina Kannada	84	41	43
Gadag	56	31	25	Kodagu	56	35	21
Dharwad	56	28	28	Mysore	112	71	41
Uttara Kannada	56	32	24	Chamarajanagar	56	39	17
Haveri	84	52	32	Ramanagara	55	24	31
Bellary	112	72	40	Chikkaballapura	56	42	14
Chitradurga	84	33	51	Yadgir	56	44	12
Davanagere	84	58	26	Minimum	55	23	11
Shimoga	87	50	37	Maximum	112	90	51
Udupi	56	28	28	Average	78	49	29
Chikmagalur	56	27	29	Total	2340	1483	857

Two types of variables are utilized in SAE technique, the variable of interest and the auxiliary variable. As noticed in Section 1, the auxiliary (covariates) variables play an important role in SAE. The auxiliary variables for this analysis are available at district level from the Census 2011. The Population Census 2011 provides a number of covariates at district level that can be utilized for small area modeling. We therefore carried out a preliminary data analysis in order to define appropriate covariates for SAE modeling, using Principal Component Analysis (PCA) to derive composite scores for selected groups of variables. In particular, we carried out PCA separately on three groups of variables, all measured at district level and identified as P1, P2 and P3 below. The first group (P1) consisted of literacy rates by gender and proportions of worker population by gender. The first principal component (P11) for this group explained 61% of the variability, while adding the second principal component (P12) increased explained variability to 85%. The second group (P2) consisted of the proportions of main worker by gender, proportions of main cultivator by gender and proportions of main agricultural labourer by gender. The first principal component (P21) for this second group explained 48% of the variability in the P2 group, while adding the second component (P22) increased explained variability to 62%. Finally, the third group (P3) consisted of proportions of marginal cultivator by gender and proportions of marginal agriculture labourers by gender. The first principal component (P31) for this third group explained

37% of the variability in the P3 group, while adding the second component (P32) increased explained variability to 60%. Finally, three variables, P11, P21 and P31 that significantly explained the model with AIC value 51.59, are identified for the use in SAE analysis. In this paper, the *Y*-variable of interest is the indebted households, i.e. whether a household is in debt or not. A household is defined to be indebted if it has outstanding loan (from respective source) as on 30.06.2012. The target is to estimate the proportion of indebted household (*i.e.* the incidence of indebtedness) at the district (LHS-A) and district by household category (LHS-C and LHS-NC) level. Incidence of indebtedness (IOI) is defined as number of households with any one loan (from respective source) divided by all households in that population segment.

3. Methodological Framework

This Section describes the methodology used in the small area analysis considered in this paper. To begin with, we assume a finite population U of size N which is consisting of D non-overlapping and mutually exclusive small areas (or district in this paper). We assume that a sample s of size n is drawn from this population using a probability sampling method. Here, a subscript d has been used to denote quantities related to small area 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 quantities related to sample and non-sample parts of the population. Let y_{di} denotes the value of the variable of interest for unit $i\,(i=1,...,N_d)$ in area d. The variable of interest, with values y_{di}, is binary (*e.g.* $y_{di}=1$ if i^{th} household is in debt and 0 otherwise) in area d, the aim is to estimate the small area population count, $y_d=\sum_{i\in U_d}y_{di}$, or equivalently the small area proportion, $P_d=N_d^{-1}y_d$, in area d. The standard direct survey estimator (hereafter denoted by DIR) for P_d is, $p_{dw}=\sum_{i\in s_d}\tilde{w}_{di}y_{di}$ where $\tilde{w}_{di}=w_{di}\big/\sum_{i\in s_d}w_{di}$ is the normalized survey weight with $\sum_{i\in s_d}\tilde{w}_{di}=1$ and w_{di} is the survey weight for unit i in area d. The estimated design-based variance of DIR is approximated by $v(p_{dw})\approx\sum_{i\in s_d}\tilde{w}_{di}(\tilde{w}_{di}-1)(y_{di}-p_{dw})^2$, with the simplifications $w_{di}=a_{di}^{-1}$, $a_{di,di}=a_{di}$ and $a_{di,dj}=a_{di}a_{dj}, i\neq j$, where a_{di} is the first order inclusion probability of unit i in area d and $a_{di,dj}$ is the second order inclusion probability of units i and j in area d. Under simple random sampling (SRS), $w_{di}=N_d n_d^{-1}$ and DIR is then $p_d=n_d^{-1}y_{sd}$, with estimated variance $v(p_d)\approx n_d^{-1}p_d(1-p_d)$, where $y_{sd}=\sum_{i\in s_d}y_{di}$ denotes the sample count in area d. Similarly, $y_{rd}=\sum_{i\in s_r}y_{di}$ denotes the non-sample count in area d. If the sampling design is informative, this SRS-based version of DIR may be biased. Furthermore, DIR is based on area-specific sample data and can therefore be very imprecise when the area specific sample size is small or may even be impossible to compute if this sample size is zero. However, model-based SAE procedures that 'borrow strength' via a common statistical model for all the small areas can be used to address this problem. If we ignore the sampling design, the sample count y_{sd} in area (*i.e.* district) d can be assumed to follow a Binomial distribution with parameters n_d and π_d, i.e. $y_{sd}\sim \text{Bin}(n_d,\pi_d)$, where π_d is the probability of occurrence of an event for a population unit in area d or the probability of prevalence in area d.

Similarly, for the non-sample count, $y_{rd} \sim Bin(N_d - n_d, \pi_d)$. Further, y_{sd} and y_{rd} are assumed to be independent binomial variables with π_d being a common success probability.

Let $\mathbf{x}_d$ be the *k*-vector of covariates for area *d* from available data sources. Following Chandra *et al.* (2011) the model linking the probability π_d with the covariates $\mathbf{x}_d$ is the logistic linear mixed model of the form

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

with $\pi_d = \exp(\mathbf{x}_d^T\boldsymbol{\beta} + u_d)\left\{1+\exp(\mathbf{x}_d^T\boldsymbol{\beta} + u_d)\right\}^{-1}$. 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 captures the area dissimilarities. We assume that u_d's are independently and normally distributed with mean zero and variance σ_u^2. Here, we observe that model (1) relates the area level proportions (direct estimates) from the survey data to the area level covariates. 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. Model (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, 2013; Chandra *et al.*, 2017). Under model (1), an empirical predictor (EP) of the population count y_d in area *d* is

$$\hat{y}_d^{EP} = y_{sd} + \hat{y}_{rd} = y_{sd} + \left(N_d - n_d\right)\left[\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\left(1+\exp(\mathbf{x}_d^T\hat{\boldsymbol{\beta}} + \hat{u}_d)\right)^{-1}\right]. \tag{2}$$

An estimate of the corresponding proportion in area *d* is obtained as $\hat{p}_d^{EP} = N_d^{-1}\hat{y}_d^{EP}$**.** 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 $\mathbf{u} = (u_1, ..., u_D)^T$. We can observe that the parameters $\boldsymbol{\beta}$ and σ_u^2 are the same for every area; i.e., they can be estimated using the data from all small areas. We use an iterative procedure that combines the Penalized Quasi-Likelihood (PQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u}$ with REML estimation of σ_u^2 to estimate unknown parameters (Chandra *et al.*, 2011).

The mean squared error (MSE) estimates are computed to assess the reliability of estimates and also to construct the confidence interval (CI). The MSE estimate of (2) is:

$$mse(\hat{p}_d^{EP}) = M_1(\hat{\sigma}_u^2) + M_2(\hat{\sigma}_u^2) + 2M_3(\hat{\sigma}_u^2)\,. \tag{3}$$

Following Chandra *et al.* (2011) we define few notations to express different components of (3). We denote by $\hat{\mathbf{V}}_s = diag\left\{n_d\hat{p}_d^{EP}(1-\hat{p}_d^{EP})\right\}$ and $\hat{\mathbf{V}}_r = diag\left\{(N_d - n_d)\hat{p}_d^{EP}(1-\hat{p}_d^{EP})\right\}$ 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 $\mathbf{X}$=$(\mathbf{x}_1^T, ..., \mathbf{x}_D^T)^T$ 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 $\hat{\mathbf{T}}_{22} = \hat{\mathbf{T}} + \hat{\mathbf{T}}\hat{\mathbf{V}}_s\mathbf{X}\hat{\mathbf{T}}_{11}\mathbf{X}^T\hat{\mathbf{V}}_s^T\hat{\mathbf{T}}$. Under model (1), the 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 $M_3(\hat{\sigma}_u^2) = trace\left(\hat{\nabla}_i\hat{\mathbf{\Sigma}}\hat{\nabla}'_j v(\hat{\sigma}_u^2)\right)$ with $\hat{\mathbf{\Sigma}} = \hat{\mathbf{V}}_{sd} + \hat{\phi}\mathbf{I}_D\hat{\mathbf{V}}_{sd}\hat{\mathbf{V}}_{sd}^T$. Let us write $\Delta = \mathbf{A}\hat{\mathbf{T}}$ and $\hat{\nabla}_i = \partial(\Delta_i)/\partial\phi\big|_{\phi=\hat{\phi}} = \partial(\boldsymbol{A}_i\hat{\mathbf{T}})/\partial\sigma_u^2\Big|_{\sigma_u^2=\hat{\sigma}_u^2}$, where $\boldsymbol{A}_i$ is the i^{th} row of the matrix $\boldsymbol{A}$. 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 $\hat{\sigma}_u^2$. This term also depends upon whether we use ML or REML estimate of $\hat{\sigma}_u^2$. We use REML estimates for $\hat{\sigma}_u^2$ and where $v(\hat{\sigma}_u^2) = 2\left((\hat{\sigma}_u^2)^{-2}(D-2t_1) + (\hat{\sigma}_u^2)^{-4}t_{11}\right)^{-1}$ with $t_1 = (\hat{\sigma}_u^2)^{-1}trace(\hat{\mathbf{T}}_{22})$ and $t_{11} = trace(\hat{\mathbf{T}}_{22}\hat{\mathbf{T}}_{22})$.

4. Results

The estimation of district level estimates of indebted household for cultivators, non-cultivators and their combined category has been carried out by using direct and model-based methods. In the present study, two types of diagnostics measures are employed: (i) the model diagnostics, and (ii) the diagnostics for the small area estimates. The model diagnostics have been applied to verify model assumptions. The second diagnostics have been applied to validate reliability of the model-based small area estimates.

In model (2), the random area specific effects $u_d\,(d=1,..,D)$ have been 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 not significantly different from the regression line $y=0$, where under model (2), the area level residuals are defined as $r_d = \hat{\eta}_d - \mathbf{x}_d^T\hat{\boldsymbol{\beta}}$. The histogram and q-q plot are used to examine the normality assumption. Figure 1 presents the histogram of the district-level residuals, distribution of the district-level residuals and normal q-q plot of the district-level residuals. Besides these graphical methods for checking normality, Shapiro-Wilk (SW) test (*i.e.* test based on uncertainty measurement in terms of p-value) has been performed. The p-value from SW test indicates the chance that the sample comes from a normal distribution. Typically, if p-value is less than 0.05 we can conclude that the sample deviates from normality. Table 2 reports the results of SW test.

Table 2: Shapiro-Wilk (SW) test result for the occupational categories

Occupational category	SW statistic	p-value
All	0.991	0.996
Cultivator	0.986	0.946
Non-Cultivator	0.961	0.332

For assessing validity and reliability of the model-based small area estimates, we applied a set of diagnostics. The values for the model-based small area estimates derived from the fitted model should be consistent with the unbiased direct survey estimates. 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. Again the model-based small area estimates should have mean squared errors significantly lower than the variances of the corresponding direct survey estimates. For this purpose, we consider three commonly used diagnostics, *viz*. the bias diagnostics, percentage coefficient of variation (CV %) 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:

$$Bias = D^{-1}\left(\sum\nolimits_{d=1}^{D} Direct\ estimate_d\right) - D^{-1}\left(\sum\nolimits_{d=1}^{D} Model\ based\ estimate_d\right) \text{ and}$$

$$RE = D^{-1}\left(\sum\nolimits_{d=1}^{D} \frac{Direct\ estimate_d - Model\ based\ estimate_d}{Direct\ estimate_d}\right).$$

The values of Bias and RE are given in Table 3. The diagnostic results in Table 3 reveal that model-based small area estimates are consistent with the direct survey estimates.

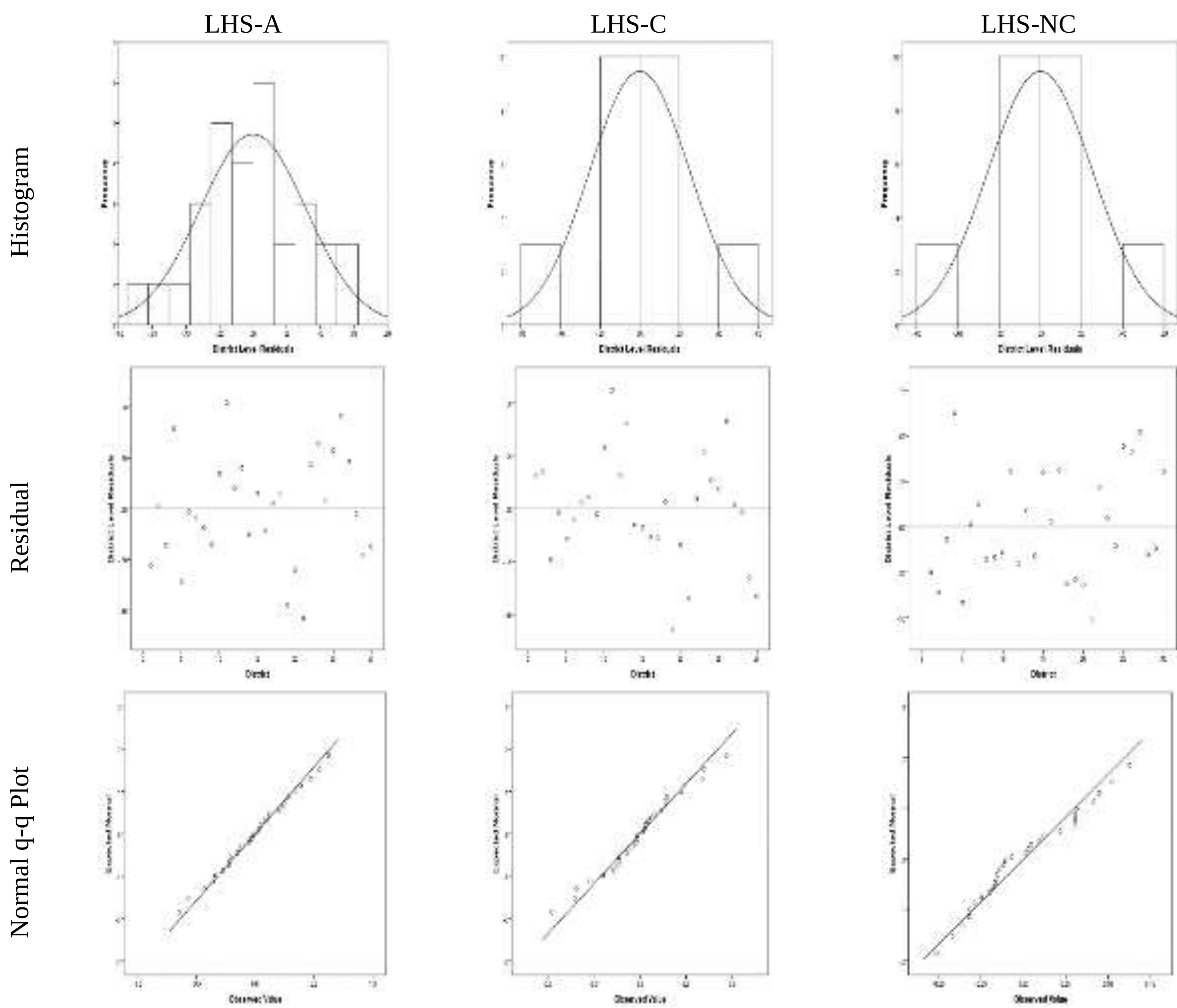
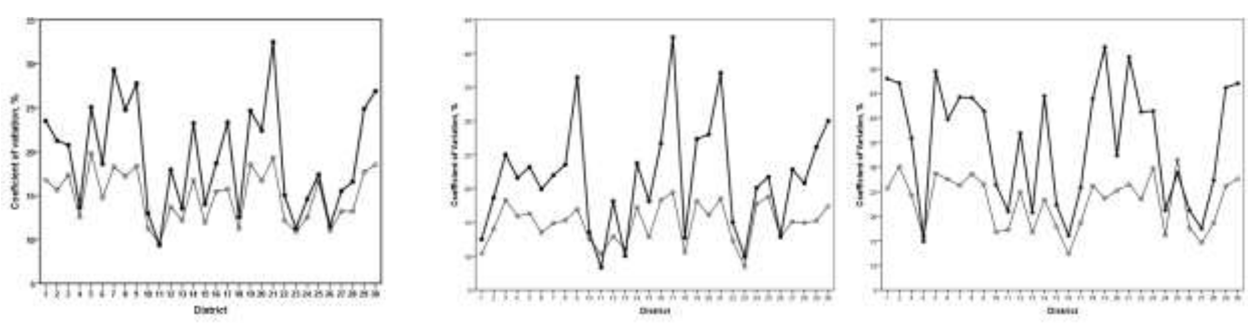


Figure 1: Histogram, distribution and normal q-q plots of the district-level residuals for model based SAE estimates of incidence of indebtedness

Table 3: Bias diagnostics for sample districts

Occupation Category	Bias	RE
All	-0.004	-0.042
Cultivator	-0.006	-0.057
Non-Cultivator	-0.003	-0.131

We compute %CV to assess the improved precision of the model-based estimates compared to the direct survey estimates. Estimates with large CVs are considered unreliable. The average (minimum, maximum) values of CV of direct and model-based (*i.e.* EP) estimates of indebtedness are 19.44% (9.35%, 32.45%) and 14.96% (9.56%, 19.82%), respectively. Similarly, the average (minimum, maximum) values of CV of direct and model-based estimates for cultivators and non-cultivators are 21.61% (8.33%, 42.38%) and 14.88% (8.42%, 19.46%); 35.11% (14.89%, 54.41%) and 22.94% (12.3%, 31.53%), respectively. The district-wise distribution of percentage CV of the model-based estimates and the direct estimates for cultivator and non-cultivator as well as their combined category is shown in Figure 2. These plots show that model-based estimates have a higher degree of reliability as compared to the direct estimates. In general, 95% CIs for the direct estimates are wider than the 95% CIs for the model-based estimates. 95% CIs for the model-based estimates are more precise and contain both direct and model-based estimates of the incidence of indebtedness. The districts-wise estimates of proportion of indebted households along with 95% CIs for the 30 districts of Karnataka are presented in Table 4. The district-wise estimates of proportion of indebted households generated by EP method range between 31.5 to 60.7 % with an average of 46.8%. Similarly, the estimates of proportion of indebted households by occupational categories within districts ranges between 39.7 to 70.2% with an average of 53.3% for cultivators and 24 to 72.9% with average of 39.6% for non-cultivators (Table 4). The maximum proportion of indebted cultivator households (0.70) is reported in Hassan while Udupi (0.73) in case of non-cultivator households. Overall, the maximum incidence of indebtedness (0.61) is found to be in district Haveri. In about 25 out of 30 districts, the incidence of indebtedness is higher among cultivator households as compared to non-cultivator households. The spatial mapping of the incidence of indebtedness among occupational categories (cultivators and non-cultivators) and also for their combined category is shown in Figure 3. Such mapping is useful in microscopic identification of location as well as extent of indebtedness of occupationally differentiated indebted households.



LHC-A LHS-C LHS-NC

Figure 2: District-wise coefficient of variation (%) for the direct (solid line,●) and the model-based SAE estimates (thin line, o) of the incidence of indebtedness

5. Conclusions

The Census in India, like in other countries, usually has limited 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 socio-economic 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. Due to emphasis on disaggregate level Sustainable Development Goal indicators, Government of India as well as different State Governments are now struggling with generation of disaggregate level statistics. The SAE is only indispensable alternative to meet the growing demand for such disaggregated level statistics needed for decentralized policy planning. Using SAE method to link data from the AIDIS 2012-13 and the Population Census 2011, we have derived district level estimates of incidence of indebtedness among cultivators and non-cultivators categories as well as for both the categories combined in different districts of rural areas of Karnataka in India and mapped them to show the spatial variability in incidence of indebtedness at district level. The results might be useful for the program managers and policy planners to implement their policy and interventions effectively.

The use of the diagnostic measure *e.g.* coefficient of variation and the comparison with direct estimates confirm that the model-based district level estimates are robust and provide reliable district level estimates of incidence of indebtedness. The results further confirm that the state level estimates of incidence of indebtedness reported in the AIDIS 2012-13 report mask the district level heterogeneity in rural areas of Karnataka. In particular, this study uncovers the district level incidence of indebtedness in rural areas of Karnataka with their accuracy measures. The region wise picture of indebtedness depicts that southern Karnataka is having higher cases of farm indebtedness, which may be due to more dependence on informal source of credit in this region. Cultivator households need credit on a continuous basis for meeting their working capital needs, hence limited formal source of credit may lead to rising chances of farm indebtedness in this category. It is noteworthy that the AIDIS data used in this study is based on reference year 2012-13 which is almost seven years old. Obvious question arises that the present scenario would be different from what emerges from this study. But, AIDIS is the only regular source to obtain unit level data pertaining to farm indebtedness and the AIDIS-2012-13 is the latest available data for this purpose. Since there is no other recent and updated data available, the estimates generated based on this data is expected to be used as recent information by policy and research analyst and Government departments.

Acknowledgements

The authors express their gratefulness to the reviewer and the editor for their invaluable comments and suggestions, which have helped us to improve the paper substantially.

References

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

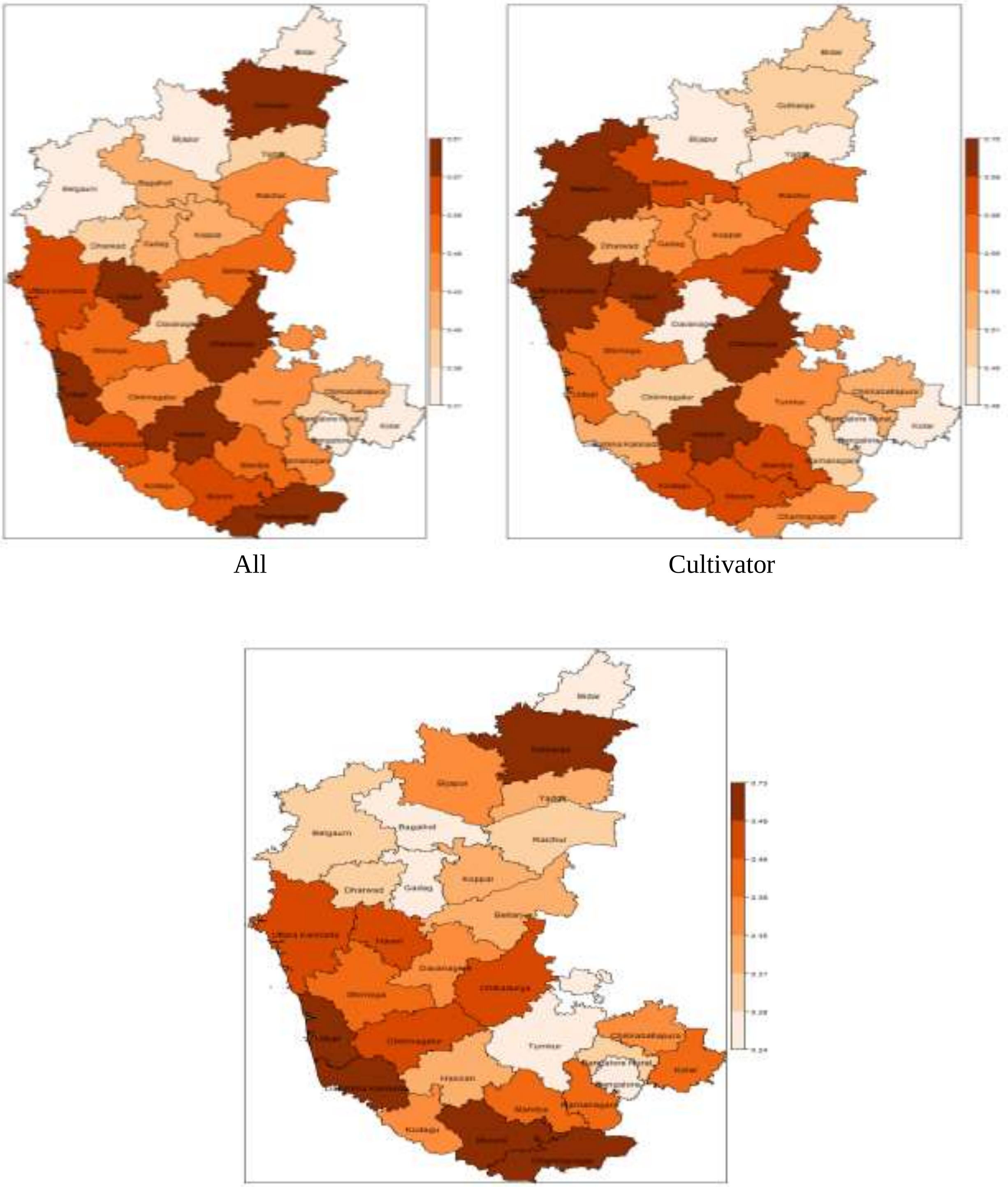
Chandra, H., Salvati, N. and Sud, U. C. (2011). Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India – An application of small area estimation Technique. *Journal of Applied Statistics*, **38**, 2413–2432.

Chandra, H. (2013). Exploring spatial dependence in area-level random effect model for disaggregate-level crop yield estimation. *Journal of Applied Statistics*, **40**, 823-842.

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

Fay R. E. and 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.

Rao J. N. K. and Molina I. (2015). *Small Area Estimation*. 2nd Edition. John Wiley and Sons.



All Cultivator

Non-Cultivator

Figure 3: Maps of the incidence of indebtedness in rural Karnataka

Table 4: District and occupational category wise estimates of incidence of indebtedness in rural Karnataka along with 95% confidence interval (Lower and Upper) for the direct and model-based small area (EP) method

Category	Districts	Direct			EP		
		Estimate	Lower	Upper	Estimate	Lower	Upper
All	Belgaum	0.34	0.18	0.49	0.38	0.25	0.50
	Bagalkot	0.43	0.25	0.61	0.43	0.30	0.56
	Bijapur	0.34	0.20	0.48	0.38	0.25	0.51
	Gulbarga	0.65	0.48	0.83	0.57	0.43	0.71
	Bidar	0.27	0.14	0.40	0.33	0.20	0.46
	Raichur	0.43	0.27	0.58	0.43	0.31	0.56
	Koppal	0.42	0.18	0.66	0.42	0.27	0.58
	Gadag	0.38	0.20	0.57	0.41	0.27	0.55
	Dharwad	0.36	0.16	0.55	0.40	0.25	0.54
	Uttara Kannada	0.58	0.44	0.73	0.56	0.44	0.68
	Haveri	0.67	0.55	0.80	0.61	0.49	0.72
	Bellary	0.49	0.32	0.66	0.48	0.35	0.61
	Chitradurga	0.61	0.45	0.77	0.57	0.43	0.70
	Davanagere	0.37	0.20	0.53	0.40	0.27	0.54
	Shimoga	0.52	0.38	0.67	0.51	0.39	0.63
	Udupi	0.55	0.35	0.76	0.58	0.40	0.75
	Chikmagalur	0.46	0.25	0.66	0.46	0.32	0.60
	Tumkur	0.48	0.37	0.60	0.48	0.37	0.58
	Kolar	0.26	0.14	0.39	0.35	0.22	0.47
	Bangalore	0.32	0.18	0.46	0.39	0.26	0.52
	Bangalore Rural	0.22	0.08	0.37	0.32	0.20	0.43
	Mandya	0.56	0.40	0.73	0.53	0.41	0.66
	Hassan	0.62	0.48	0.76	0.58	0.46	0.71
	Dakshina Kannada	0.55	0.39	0.70	0.55	0.41	0.68
	Kodagu	0.61	0.41	0.82	0.54	0.37	0.72
	Mysore	0.63	0.49	0.77	0.56	0.44	0.68
	Chamarajanagar	0.64	0.44	0.83	0.57	0.42	0.71
	Ramanagara	0.46	0.31	0.60	0.46	0.34	0.58
	Chikkaballapura	0.35	0.18	0.53	0.42	0.27	0.56
	Yadgir	0.34	0.16	0.52	0.38	0.24	0.52
Cultivator	Belgaum	0.62	0.46	0.77	0.60	0.48	0.72
	Bagalkot	0.61	0.39	0.84	0.56	0.41	0.71
	Bijapur	0.35	0.18	0.52	0.42	0.27	0.58
	Gulbarga	0.50	0.29	0.71	0.49	0.33	0.64
	Bidar	0.41	0.23	0.60	0.46	0.32	0.61
	Raichur	0.53	0.33	0.74	0.55	0.40	0.69
	Koppal	0.55	0.31	0.79	0.54	0.39	0.70
	Gadag	0.54	0.29	0.80	0.53	0.37	0.69
	Dharwad	0.45	0.13	0.78	0.52	0.35	0.69
	Uttara Kannada	0.69	0.51	0.88	0.60	0.45	0.74
	Haveri	0.79	0.66	0.92	0.65	0.52	0.79
	Bellary	0.64	0.41	0.86	0.58	0.43	0.73
	Chitradurga	0.83	0.67	0.99	0.69	0.54	0.84
	Davanagere	0.42	0.22	0.61	0.44	0.29	0.59

	Shimoga	0.54	0.35	0.73	0.55	0.41	0.69
	Udupi	0.49	0.23	0.75	0.55	0.36	0.75
	Chikmagalur	0.38	0.06	0.69	0.46	0.28	0.63
	Tumkur	0.55	0.41	0.69	0.55	0.44	0.66
	Kolar	0.26	0.12	0.40	0.40	0.26	0.54
	Bangalore	0.44	0.20	0.69	0.51	0.35	0.68
	Bangalore Rural	0.26	0.07	0.45	0.41	0.26	0.56
	Mandya	0.59	0.42	0.76	0.57	0.43	0.70
	Hassan	0.73	0.59	0.88	0.70	0.59	0.82
	Dakshina Kannada	0.57	0.34	0.79	0.52	0.34	0.71
	Kodagu	0.62	0.36	0.89	0.56	0.36	0.77
	Mysore	0.66	0.49	0.82	0.56	0.42	0.70
	Chamarajanagar	0.55	0.30	0.80	0.55	0.39	0.71
	Ramanagara	0.51	0.30	0.71	0.51	0.36	0.65
	Chikkaballapura	0.41	0.20	0.62	0.51	0.36	0.67
	Yadgir	0.32	0.13	0.51	0.45	0.30	0.61
Non-Cultivator	Belgaum	0.22	0.01	0.42	0.29	0.14	0.43
	Bagalkot	0.14	0.01	0.27	0.24	0.10	0.38
	Bijapur	0.31	0.09	0.54	0.35	0.19	0.52
	Gulbarga	0.77	0.55	1.00	0.58	0.40	0.76
	Bidar	0.13	0.00	0.25	0.28	0.12	0.43
	Raichur	0.27	0.06	0.48	0.29	0.13	0.45
	Koppal	0.36	0.05	0.67	0.34	0.16	0.51
	Gadag	0.22	0.03	0.40	0.28	0.12	0.44
	Dharwad	0.27	0.05	0.48	0.30	0.15	0.46
	Uttara Kannada	0.45	0.22	0.68	0.49	0.33	0.65
	Haveri	0.52	0.30	0.73	0.46	0.31	0.62
	Bellary	0.28	0.08	0.47	0.32	0.16	0.47
	Chitradurga	0.50	0.29	0.70	0.49	0.33	0.64
	Davanagere	0.31	0.04	0.58	0.38	0.20	0.55
	Shimoga	0.51	0.28	0.73	0.46	0.30	0.62
	Udupi	0.71	0.49	0.94	0.73	0.55	0.91
	Chikmagalur	0.53	0.26	0.79	0.47	0.30	0.64
	Tumkur	0.21	0.03	0.39	0.28	0.14	0.43
	Kolar	0.30	-0.02	0.62	0.39	0.21	0.57
	Bangalore	0.24	0.09	0.39	0.30	0.15	0.45
	Bangalore Rural	0.13	0.00	0.26	0.28	0.14	0.43
	Mandya	0.48	0.09	0.86	0.41	0.22	0.59
	Hassan	0.33	0.06	0.60	0.32	0.13	0.50
	Dakshina Kannada	0.52	0.31	0.74	0.56	0.38	0.73
	Kodagu	0.60	0.26	0.93	0.37	0.14	0.60
	Mysore	0.59	0.35	0.84	0.50	0.33	0.67
	Chamarajanagar	0.78	0.51	1.05	0.62	0.44	0.80
	Ramanagara	0.40	0.19	0.61	0.42	0.27	0.57
	Chikkaballapura	0.29	0.03	0.55	0.35	0.17	0.54
	Yadgir	0.46	0.04	0.88	0.34	0.16	0.53