Centre for Statistical and Survey Methodology

The University of Wollongong

Working Paper

19-10

Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India-an application of small area estimation technique

Hukum Chandra, Nicola Salvati and U.C. Sud

Copyright © 2008 by the Centre for Statistical & Survey Methodology, UOW. Work in progress, no part of this paper may be reproduced without permission from the Centre.

Centre for Statistical & Survey Methodology, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Disaggregate-level estimates of indebtedness in the state of Uttar Pradesh in India-an application of small area estimation technique

Hukum Chandra

Centre for Statistical and Survey Methodology, University of Wollongong, Wollongong, Australia.

E-mail: hchandra@uow.edu.au

Nicola Salvati

Dipartimento di Statistica e Matematica Applicata all'Economia, University of Pisa, Italy,

E-mail: salvati@ec.unipi.it

U. C. Sud

Indian Agricultural Statistics Research Institute, New Delhi, India

Abstract

The National Sample Survey Organisation (NSSO) surveys are the main source of official statistics in India and generate a range of invaluable data at the macro level (e.g. state and national level). However, the NSSO data cannot be used directly to produce reliable estimates at the micro level (e.g. district or further disaggregate level) due to small sample sizes. There is a rapidly growing demand of such micro level statistics in India as the country is moving from centralized to more decentralized planning system. In this article we employ small area estimation (SAE) techniques to derive model-based estimates of proportion of indebted households at district or at other small area levels in the State of Uttar Pradesh in India by linking data from the Debt-Investment Survey 2002-03 of NSSO and the Population Census 2001 and the Agriculture Census 2003. Our results show that the model-based estimates are precise and representative. For many small areas it is even not possible to produce estimates using sample data alone. The model based estimates generated using SAE are still reliable for such areas. The estimates are expected to provide invaluable information to policy-analysts and decision-makers.

Key words: indebted households, NSSO survey, census, small area estimation, proportion.

1. Introduction

In recent years, the thrust of planning process has shifted from macro to micro level. There is demand by the administrators and policy planners for reliable estimates of various parameters at the micro level. In view of the demands of modern time the thrust of research efforts has also shifted to development of precise estimators for small areas. An offshoot of this development is that various small area estimation techniques are being proposed by the researchers for implementation. In India there is great emphasis on district level planning. For example, the efforts to develop databases required for planning and decision-making at lower than the State level, were initiated quite some time back with the Planning Commission of Government of India setting up a “Working Group on Districts planning” in September, 1982. The Working Group in its report clearly highlighted the data requirement for planning and decision-making at the district level. However, it was found that though a lot of data are collected, processed and published for the country as a whole or for individual states, not much disaggregation of the data for sub-state level is done. India has been in an advantageous position due to availability of regular data through National Sample Survey Organisation (NSSO) surveys. The NSSO surveys are planned to generate statistics at state and national level. There is no regular flow of estimates at further below level, e.g., at the districts level. Indeed, the NSSO surveys provide reliable state and national level estimates; they can not be used to derive reliable direct estimates at the district level owing to small sample sizes which lead to high levels of sampling variability (see [7] and [8]). Due to the lack of statistics at this level, proper planning, fund allocation and also monitoring of various plans is likely to suffer.

Although in the Indian context, 'district' is a very important domain for the planning process, we do not have surveys to produce estimates at these levels. At the same time, it is also true that conducting any such surveys aimed at this level is going to be very costly and time consuming job. Using the state level survey (e.g., NSSO surveys) data to derive the direct estimates at district or smaller domain level, we may end up with very small sample sizes in these domains which may result in very unstable estimates for these domains. A solution to this problem is to consider small area estimation (SAE) techniques. The SAE techniques aim at producing reliable estimates for such domains with small sample sizes by borrowing strength from data of other domains. The SAE techniques are generally based on model-based methods. The idea is to use statistical models to link the variable of interest with auxiliary information, e.g. Census and Administrative data, for the small areas to define model-based estimators for these areas. Such small area models can be classified into two broad types:

(i) Area level random effect models, which are used when auxiliary information is available only at area level. They relate small area direct estimates to area-specific covariates (Fay and Herriot [4]) and

(ii) Nested error unit level regression models, proposed originally by Battese, Harter and Fuller [2]. These models relate the unit values of a study variable to unit-specific covariates.

We adopt the area level model since covariates are available only at the area level. In this article we employ small area estimation techniques to derive model-based estimates of proportion of indebted households at small area levels in the State of Uttar Pradesh in India by linking data from the Debt-Investment Survey 2002-03 of NSSO and the Population

Census 2001 and the Agriculture Census 2003. Small areas are defined as the different districts and district by land holding classes of State of Uttar Pradesh in India. The article illustrates how the NSSO and Census data can be combined to derive reliable estimates for the proportion of indebted household at the district level. The rest of the paper is organised as follows. In Section 2 we describe the data used for the analysis and in Section 3 we present an overview of the methodology used for the analysis. Section 4 discusses the diagnostic procedures for examining the model assumptions and validating the small area estimates and describes the results. Section 5 finally sets out the main conclusions.

2. Data

In this article we adopt an area level small area model to derive the small area level estimates (see [4]). Two types of variables are required for this analysis.

(i) The variable of interest for which small area estimates are required is drawn from the Debt-Investment Survey 2002-03 of NSSO. We used 59th round data of NSSO for rural areas on Debt and Investment survey conducted for the calendar year 2002-03 in the State of Uttar Pradesh in India. The target variable used for the study was indebted households. A household is defined to be indebted if it has outstanding loan as on 30.6.02. The parameter of interest is the proportion of indebted household at the district and district by holding size level.

(ii) The auxiliary (covariates) variables known for the population are drawn from the Population Census 2001 and the Agriculture Census 2003. It is noteworthy that use of covariates from the 2001 Population Census and the 2003 Agriculture Census to model indebted household from the NSSO survey may raise issues of comparability.

However, the covariates used in this study are not expected to change significantly over a short period of time. There were 158 covariates available from the Population Census 2001 and the Agriculture Census 2003 to consider for the modelling.

Out of these, suitable covariates were selected for the analysis as follows. We first examined the correlation of all these covariates with the target variable and then selected the covariates with reasonably good correlation with the target variable. This was followed by step-wise regression analysis. Finally, two variables the Crop loan distributed (Indian Rupees in lakhs) in Rabi season (Rabi) and Female Agricultural Labor (AL_F) were identified for the further analysis which significantly explained the model.

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. There are total of 11,814 households (i.e. number of surveyed households which includes both indebted and non-indebted households) from the 69 districts of the Uttar Pradesh. The average land holding size is 1.41 hectare. The district specific sample size varies from 55 to 340 with average sample size of 171. The district specific sample size becomes very small if we consider further sub-grouping of the districts (e.g., district by land holding classes). Based on land holding size in hectare (hereafter ha) the households are classified into five different holding classes as set out in Table 1. These are the standard classification of land holding classes in India.

Our aim is to estimate proportion of indebted households at district level for different land holding classes as well as for all classes combined together. Therefore, we define different districts (Cat0) and districts by land holding classes (Cat1-Cat5) of the State of Uttar Pradesh as the small areas of interest, see Table 1 for the definition. The district and district by land holding class-wise sample sizes for the NSSO data used in this analysis are presented in Table 2. The most striking point in Table 2 is that the sample size 0 and 1 can be seen in many districts or small areas. For example, category 5 (Cat 5) has 9 districts with sample of size 0. For these districts it is not possible to generate the direct estimates using tradition sample survey estimation approaches. Among the six categories defined above, the last three categories Cat3 to Cat 5 have very small average sample sizes (averaged over the different districts) of 18, 9 and 3 respectively. Further, there are many other districts in Cat 4 and Cat 5 with sample of size 1. It is again difficult to derive reliable estimates and their standard errors for such districts. Indeed, SAE is an obvious answer to these problems. The SAE techniques provide reliable estimates for the districts having small or even no sample data ([8]). The underlining theory of SAE has been illustrated in next Section.

3. An overview of the methodology

We now set out the small area estimation techniques used to produce the model-based estimates and their measure of precision. To start, we first fix our notation. Throughout, we use a subscript *d* to index the quantities belonging to small area d $(d = 1,...,D)$, where D is the number of small areas (or areas) in the population. The subscript s and *r* are used for denoting the quantities related to the sample and non-sample parts of the population. So that n_d and N_d represent the sample and population sizes in small area d, respectively. The value of variable of interest *y* (which is the number of indebted household) in the area *d* is defined by y_d and we denote by y_{sd} and y_{rd} the sample and non-sample counts of indebted households in area *d*. Indeed, the variable of interest y_{sd} has a Binomial distribution with parameters n_d and π_d, denoted by $y_{sd} \sim Bin(n_d, \pi_d)$, where π_d is the probability of an indebt household in area *d*, often termed as the probability of a 'success'. Similarly, $y_{rd} \sim Bin(N_d - n_d, \pi_d)$. Further, y_{sd} and y_{rd} are assumed to be independent Binomial variables with π_d being a common success probabilities. Recall that in model-based small area estimation the survey data is supplemented by the availability of auxiliary information from various sources, e.g., Census and Administrative records. Let $\mathbf{x}_d$ be the *k*-vector of the covariates for area *d* from the previous sources. The model linking the probabilities of success π_d with the covariates $\mathbf{x}_d$ is the logistic linear mixed model (see [3], [6] and [9]) given by

$$logit(\pi_d) = \ln\left\{\frac{\pi_d}{1-\pi_d}\right\} = \eta_d = \mathbf{x}'_d\boldsymbol{\beta} + u_d\,,\ (d = 1,...,D)\,, \qquad (1)$$

where $\boldsymbol{\beta}$ is the *k*-vector of regression coefficient often known as fixed effect parameters and u_d is the area-specific random effect that accounts for between area dissimilarity beyond that explained by the auxiliary variables included in the fixed part of the model. We assume that u_d's are independent and normally distributed with mean zero and variance φ. Under model (1), we get

$$\pi_d = \exp(\eta_d)\{1+\exp(\eta_d)\}^{-1} = \exp(\mathbf{x}'_d\boldsymbol{\beta} + u_d)\{1+\exp(\mathbf{x}'_d\boldsymbol{\beta} + u_d)\}^{-1}.$$

It is evident that model (1) relates the area level proportions to area level covariates. This type of model is often referred to as 'area-level' model in SAE terminology, see for example [8]. Such type of model was originally used by Fay and Herriot [4] for the prediction of mean per-capita income (PCI) in small geographical areas (less than 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. In contrast, model (1) is a special case of a generalized linear mixed model (GLMM) with logit link function (see [3]) and suitable for discrete, particularly binary variable. It is noteworthy that the FH model is not applicable in such cases. Saei and Chambers[9] and Manteiga et al.[6] described this model in the context of SAE. By definition, the means of y_{sd} and y_{rd} given u_d under model (1) are:

$$E\left(y_{sd} \mid u_d\right) = n_d \pi_d = n_d \left[\exp(\mathbf{x}_d' \boldsymbol{\beta} + u_d)\left(1 + \exp(\mathbf{x}_d' \boldsymbol{\beta} + u_d)\right)^{-1}\right] \tag{2}$$

$$E\left(y_{rd} \mid u_d\right) = \left(N_d - n_d\right)\pi_d = \left(N_d - n_d\right)\left[\exp(\mathbf{x}_d' \boldsymbol{\beta} + u_d)\left(1 + \exp(\mathbf{x}_d' \boldsymbol{\beta} + u_d)\right)^{-1}\right]. \tag{3}$$

Let T_d denotes the total number of indebt households in small area *d*. We can write $T_d = y_{sd} + y_{rd}$, where the first term y_{sd}, the sample count is known whereas the second term y_{rd}, the non-sample count, is unknown. Therefore, an estimate $\hat{T}_d$ of the total number of indebted households in area *d* is obtained by replacing y_{rd} by its predicted value under the model (1). That is,

$$\hat{T}_d = y_{sd} + \hat{y}_{rd} = y_{sd} + \left(N_d - n_d\right)\left[\exp(\mathbf{x}_d' \hat{\boldsymbol{\beta}} + \hat{u}_d)\left(1 + \exp(\mathbf{x}_d' \hat{\boldsymbol{\beta}} + \hat{u}_d)\right)^{-1}\right]. \tag{4}$$

Often we come across the situations when small areas do not have sample data at all (Table 1 and 2). That is $n_d = 0$ and $y_{sd} = 0$. For example, in the NSSO data for Cat5 there are 9 districts with $n_d = 0$. Eventually traditional survey estimation approaches do not provide solution to this problem. In contrast, SAE can be used to derive estimates for such areas. In particular, for the small areas with $n_d = 0$, we use synthetic-type estimator for computing T_d defined as

$$\hat{T}_d^{Syn} = N_d \left\{\exp(\mathbf{x}_d' \hat{\boldsymbol{\beta}})\left(1 + \exp(\mathbf{x}_d' \hat{\boldsymbol{\beta}})\right)^{-1}\right\}. \tag{5}$$

An estimate of proportion of indebted households p_d in a small area *d* is obtained as

$$\hat{p}_d = \frac{\hat{T}_d}{N_d} = \frac{1}{N_d}\left\{ y_{sd} + \left(N_d - n_d\right)\left[\exp(\mathbf{x}'_d\hat{\boldsymbol{\beta}} + \hat{u}_d)\left(1+\exp(\mathbf{x}'_d\hat{\boldsymbol{\beta}} + \hat{u}_d)\right)^{-1}\right]\right\}. \tag{6}$$

Similarly, for areas with $n_d = 0$, proportion is estimated by

$$\hat{p}_d^{Syn} = \exp(\mathbf{x}'_d\hat{\boldsymbol{\beta}})\left(1+\exp(\mathbf{x}'_d\hat{\boldsymbol{\beta}})\right)^{-1}. \tag{7}$$

It is obvious that in order to compute the estimates given by equation (4) to (7), we require estimates of the unknown parameters $\boldsymbol{\beta}$ and $\mathbf{u}$. A major difficulty in use of logistic linear mixed model (LLMM) for SAE is the estimation of unknown model parameters $\boldsymbol{\beta}$ and $\mathbf{u}$ since the likelihood function for LLMM often involves high dimensional integrals (computed by integrating a product of discrete and normal densities, which has no analytical solution) which are difficult to evaluate numerically. We used an iterative procedure that combines the Penalized Quasi-Likelihood (PQL) estimation of $\boldsymbol{\beta}$ and $\mathbf{u} = (u_1, ..., u_D)'$ with restricted maximum likelihood (REML) estimation of ϕ to estimate these unknown parameters. Detailed description of the approach can be followed from [6, 9].

We now turn to estimation of mean squared error (MSE) for predictors given by equation (6) and (7). The MSE estimates are computed to assess the reliability of estimates and also to construct the confidence interval (CI) for the estimates. The mean squared error estimate of (6) under model (1) is (see [6, 9]) given by

$$mse(\hat{p}_d) = m_1(\hat{\phi}) + m_2(\hat{\phi}) + 2m_3(\hat{\phi}). \tag{8}$$

The first two components m_1 and m_2 constitute the largest part of the overall MSE estimates in (8). These are the MSE of the best linear unbiased predictor (BLUP)-type estimator when ϕ is known ([8]). The third component m_3 is the variability due to the estimate of ϕ. For simplicity, we used few notations to write the analytical expression of various components of the mean squared error (8). We denote by $\hat{\mathbf{V}}_{sd} = diag\left\{n_d \hat{p}_d (1-\hat{p}_d)\right\}$ and $\hat{\mathbf{V}}_{rd} = diag\left\{(N_d - n_d)\hat{p}_d(1-\hat{p}_d)\right\}$, the diagonal matrices defined by the corresponding variances of the sample and non-sample part respectively. Similarly, we define $\mathbf{A} = \left\{diag(N_d^{-1})\right\}\hat{\mathbf{V}}_{rd}$, $\mathbf{B} = \left\{diag(N_d^{-1})\right\}\hat{\mathbf{V}}_{rd}\boldsymbol{X}_r - \mathbf{A}\hat{\boldsymbol{T}}_s\hat{\mathbf{V}}_{sd}\boldsymbol{X}_s$ and $\hat{\boldsymbol{T}}_s = \left(\phi^{-1}\mathbf{I}_D + \hat{\mathbf{V}}_{sd}\right)^{-1}$, where $\boldsymbol{X}_s$ and $\boldsymbol{X}_r$ are the sample and non-sample part of auxiliary information and $\mathbf{I}_D$ is an identity matrix of order *D*. We further write $\hat{\mathbf{T}}_{(1)} = \left\{\boldsymbol{X}_s'\hat{\mathbf{V}}_{sd}\boldsymbol{X}_s - \boldsymbol{X}_s'\hat{\mathbf{V}}_{sd}\hat{\boldsymbol{T}}_s\hat{\mathbf{V}}_{sd}\boldsymbol{X}_s\right\}^{-1}$ and $\hat{\mathbf{T}}_{(2)} = \hat{\boldsymbol{T}}_s + \hat{\boldsymbol{T}}_s\hat{\mathbf{V}}_{sd}\boldsymbol{X}_s\hat{\mathbf{T}}_{(1)}\boldsymbol{X}_s'\hat{\mathbf{V}}_{sd}'\hat{\boldsymbol{T}}_s$. With these notations, assuming model (1) holds, the various components of equation (8) are

$$m_1(\hat{\phi}) = \mathbf{A}\hat{\boldsymbol{T}}_s\mathbf{A}',$$

$$m_2(\hat{\phi}) = \mathbf{B}\hat{\mathbf{T}}_{(1)}\mathbf{B}', \text{ and}$$

$$m_3(\hat{\phi}) = trace\left(\hat{\nabla}_i\hat{\boldsymbol{\Sigma}}\hat{\nabla}_j' v(\hat{\phi})\right) \text{ with } \hat{\boldsymbol{\Sigma}} = \hat{\mathbf{V}}_{sd} + \hat{\phi}\mathbf{I}_D\hat{\mathbf{V}}_{sd}\hat{\mathbf{V}}_{sd}'.$$

Here $v(\hat{\phi})$ is the asymptotic covariance matrix of the estimates of variance components $\hat{\phi}$, which can be evaluated as the inverse of the appropriate Fisher information matrix for $\hat{\phi}$. Note that this also depends upon whether we are using maximum likelihood (ML) or restricted maximum likelihood (REML) estimates for $\hat{\phi}$. We used REML estimates for $\hat{\phi}$, then $v(\hat{\phi}) = 2\left(\hat{\phi}^{-2}(D-2t_1) + \hat{\phi}^{-4}t_{11}\right)^{-1}$ with $t_1 = \hat{\phi}^{-1}trace(\hat{\mathbf{T}}_{(2)})$ and $t_{11} = trace(\hat{\mathbf{T}}_{(2)}\hat{\mathbf{T}}_{(2)})$. Let us write $\Delta = \mathbf{A}\hat{\boldsymbol{T}}_s$ and $\hat{\nabla}_i = \partial(\Delta_i)/\partial\phi\big|_{\phi=\hat{\phi}} = \partial(\boldsymbol{A}_i\hat{\boldsymbol{T}}_s)/\partial\phi\big|_{\phi=\hat{\phi}}$, where $\boldsymbol{A}_i$ is the i^{th} row of the matrix $\boldsymbol{A}$. The MSE estimates of (7) is a special case of (8) when $n_d = 0$, given as

$$mse(\hat{p}_d^{Syn}) = \left[diag\left\{\hat{p}_d(1-\hat{p}_d)\right\}\right]\hat{\phi}\mathbf{I}_D\left[diag\left\{\hat{p}_d(1-\hat{p}_d)\right\}\right]' \tag{9}$$

The numerical results reported in Sections 4 are obtained using R version 2.9.2.

4. Empirical results

4.1 *Diagnostic procedures*

Generally two types of diagnostics procedures are tested in small area estimation, the model diagnostics and the diagnostics for the small area estimates, see for example[1]. The first diagnostics are used to verify the assumptions of underlying model and the second diagnostics are applied to validate the reliability of the model-based small area estimates. The random area effects $u_d (d = 1,...,D)$ in model (1) are assumed to have a normal distribution with mean zero and variance φ. If the model assumptions are satisfied then the district level residuals are expected to be randomly distributed and not significantly different from the regression line *y=0,* where under model (1), the area level residuals are defined as $r_d = \hat{\eta}_d - \mathbf{x}'_d\hat{\boldsymbol{\beta}}$. The distribution of the district level residuals (left side plots) and q-q plots (right side plots) for Cat0 to Cat5 data are shown in Figure 1. The Figure 1 clearly reveals 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 also confirm the normality assumption. Therefore the model diagnostics are fully satisfied for the data.

To validate the reliability of the model-based small area estimates we used the bias diagnostics, coefficient of variation (CV) and computed the 95 percent confidence intervals. The bias diagnostics are used to investigate if the model-based estimates are less extreme when compared to the direct survey estimates, when it is available [5]. In addition, if direct estimates are unbiased, their regression on the true values should be linear and correspond to the identity line. If model-based estimates are close to the true values the regression of the direct estimates on the model-based estimates should be similar [1]. We plot direct estimates on Y-axis and model-based estimates on X-axis and we look for divergence of regression line from $Y = X$ and test for intercept = 0 and slope = 1 (see for example [1]). The bias scatter plots of the direct estimates against the model-based estimates for Cat0 to Cat5 data are set out in Figure 2. The results for bias test are given in Table 3. It is noteworthy that the model based estimates used in bias tests are based on synthetic model. It is meaningful because it overcomes the shrinkage effect and shows that the deterministic part of the model gives unbiased predictions as do the direct estimates. The bias diagnostic results in Table 3 clearly show that only the slope for cat4 fails this diagnostic. The plots show that the model-based estimates are less extreme when compared to the direct estimates, demonstrating the typical SAE outcome of shrinking more extreme values towards the average. It has to be noted that districts with extreme direct estimates are mainly those with small sample sizes. Such cases were observed more in the plots belonging to Cat3 to Cat 5.

We computed the coefficient of variation (CV) to assess the improved precision of the model-based estimates compared to the direct estimates. The CVs show the sampling variability as a percentage of the estimate. Estimates with large CVs are considered unreliable (i.e. smaller is better). There are no internationally accepted tables available that allow us to judge what is "too large" ([1] and [5]). Figure 3 presents the district-wise distribution of the percentage CV of model based estimates and direct estimates for all six categories (Cat0-Cat5) considered in the analysis. The estimated CVs show that model-based estimates have a higher degree of reliability when compared to the (non-zero) direct estimates. We note that the average sample size for the districts become smaller as we move from Cat 0 to Cat 5. As expected, relative performance of model based estimates are better as sample size decreases (see Figure 3). Particularly, for Cat 5 direct estimates have very high CV. The model based estimates still perform well. It is interesting to note that for Cat 5 out of 69 districts there are 9 districts with no sample data. For these 9 districts we cannot produce the direct estimates, however, model based estimates generated for these districts have reasonably good CV values and that to within the acceptable limit (Table 4).

In Table 5 we present the districts-wise 95% confidence intervals of the model-based and the direct estimates. The 95% confidence intervals (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 true standard errors of the direct estimates. The standard errors of the direct estimates are too large and therefore the estimates are unreliable. Note that for many districts we can even not produce the confidence intervals due to unavailability of standard errors.

4.2 *Discussions*

The small area estimates diagnostic measures clearly depict that the model-based estimates (i.e. the estimates generated by the SAE approach) are reliable and more stable than the corresponding direct estimates (Figure 3). Table 5 presents the direct estimates and model-based estimates along with 95% confidence intervals for the State of Uttar Pradesh for five different land-holding classes as well as combined. These results show the degree of inequality with respect to distribution of indebted households in different districts as well as between various land-holding classes. The most interesting point is the model based estimates for districts where there is very small (e.g. $n_d = 1$ or 2) or no sample information. So it is not feasible to have direct estimates and their CI for such cases. In Table 5 there are many districts where there is no direct estimate and their 95% confidence interval. This leaves us with no way except SAE. Table 4 presents the model-based estimates for 9 districts of Cat 5 with $n_d =0$. These estimates are reliable with CVs below 5%. Note that these estimates can be biased if synthetic assumption is violated.

A critical review of Table 5 shows that in many districts the lower bound (Lower) of 95% confidence interval is negative and upper bound (Upper) is greater than 1.0 which results in practically impossible and inadmissible values of CI for direct estimates. For example, the CI of direct estimates for Chitrakut in Cat 2 and Mathura in Cat 3 exceeds 1.0. In contrast, the model estimate of Chitrakut and Mathura with precise CI and reasonable CV percent are still reliable. A similar problem, but in other situation was observed when there was no variability in the sample data of district. For example see Lalitpur in Cat 2 where all y values in sample were 1 and estimated direct proportion was 1.0 and estimated SE was 0. That is CI with extreme sample value provides very little information. These abnormalities with direct estimation were seen in many districts when we observed Cat 4 and 5. In Cat 4 there are 31 out of 69 districts where direct estimation can not even define proper CI. We note that in Cat 5 more than half districts have sample of size 0, 1 or 2 and therefore the problem with direct estimates is worst. Out of 69 districts there are only 3 districts where CI for direct estimates is even defined. In such circumstance, SAE plays an important role in generating micro level statistics. The results clearly show the advantage of using SAE technique to cope up the small sample size problem in producing the estimates or reliable confidence intervals.

These estimates can definitely be useful for resource allocation and policy decision-making relating the indebtedness. The land holding class specific estimates have added advantages for policy planning and resource allocation based on farm category wise. These estimates are also helpful in identifying the districts/regions or farm categories with higher level of indebtedness. For example, in the budget year 2008-09, Govt of India announced the creation of a farmers' debt relief fund. This scheme waives the debts of farmers in general and small and marginal farmers in particular. The implementation of this scheme will definitely need the estimates generated in this study. The concerned Govt department can use these estimates to allocate the fund to various districts according to the proportion of indebted farmers.

5. Conclusions

A great deal of theoretical research has been done for the SAE. This is the time for their real life applications and implementation. The method for estimation of proportions for small areas is well developed ([6 and 9]), however, there is limited application in the area of agricultural or social sciences. Further, there is rarely any application to the Indian data. In this article we demonstrate the application of SAE techniques to estimate the district level statistics of indebtedness for different land holding classes as well as all classes combined together using survey and census data. The diagnostic procedures clearly confirm that the model-based district level estimates for different land holding classes as well as all classes combined together have reasonably good precision. The SAE method has also generated reliable estimates for the districts with no or very small sample sizes such as 1 or 2. This application of small area analysis is the first of its kind with most popular NSSO data in India to estimate the proportions at disaggregate levels. In India, Censuses are usually limited as they tend to focus mainly on the basic socio-demographic and economic data and not available for every time period. The NSSO survey, on the other hand, contributes to providing estimates at the State and National level. They do not provide sub-state level statistics. However, it is known that regional and national estimates usually mask variations (heterogeneity) at the sub-state or district level and render little information for micro level planning and allocation of resources.

These days a lot of emphasis is being given to micro level planning in India. District is an important domain for planning process in the country and therefore availability of district level statistics is vital for monitoring of policy and planning. For example, Govt of India UNDP project on "Capacity Development for District Planning". It expects decentralised planning to improve effectiveness of development programmes. This study produces reliable statistics at micro level using existing surveys and other already available secondary data and can be seen as an indicative example for further applications. Such micro level statistics can be generated without conducting separate survey for this purpose and unlike Census regular estimates can be produced from regular existing surveys. Govt of India currently has number of schemes (for example, Indira Awaas Yojana, Pradhan Mantri Gram Sadak Yojana, Mahatma Gandhi National Rural Employment Guarantee Scheme etc.) for rural areas since the rural development in India is one of the most important factors for the growth of the Indian economy. These disaggregate level estimates are useful for implementation of these schemes.

References

[1] R. Ambler, D. Caplan, R. Chambers, M. Kovacevic, and S. Wang, *Combining unemployment benefits data and LFS data to estimate ILO unemployment for small areas: an application of a modified Fay-Herriot method*. Proceedings of the Int. Assoc. of Survey Stat., Meeting of the ISI, Seoul, August 2001.

[2] G. E. Battese, R. M. Harter, and W. A. Fuller, *An error component model for prediction of county crop areas using survey and satellite data,* J. of the Amer. Stat. Assoc. 83 (1988), pp. 28-36.

[3] N. E. Breslow and D. G. Clayton, *Approximate inference in generalized linear mixed models,* J. of the Amer. Stat. Assoc. 88(1993), pp.9-25.

[4] R. E. Fay and R. A. Herriot, *Estimation of income from small places: an application of james-stein procedures to census data,* J. of the Amer. Stat. Assoc. 74(1979), pp. 269-277.

[5] F.A. Johnson, H. Chandra, J. J. Brown, and S. Padmadas, *District-level estimates of institutional births in ghana: application of small area estimation technique using census and DHS data*, J. of Off. Stat. 26 (2010), pp. 341–359.

[6] G.W. Manteiga, M.J. Lombardìa, I. Molina, D. Morales, and L. Santamarìa, *Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed mode*l, Comput. Stat. & Data Anal. 51(2007), pp. 2720-2733.

[7] D. Pfeffermann, *Small area estimation: new developments and directions*. Int. Stat. Rev. 70(2002), pp.125-143.

[8] J.N.K. Rao, *Small Area Estimation*. Wiley Series in Survey Methodology, John Wiley and Sons Inc, 2003.

[9] A. Saei and R. Chambers, *Small area estimation under linear and generalized linear mixed models with time and area effects, Meth. W.P. No. M03/15*(2003), Southampton Statistical Sciences Research Institute, University of Southampton, UK.

Table 1. Definition of land holding classes

Category	Descriptions	Land holding size (ha)	Sample size			
			Min	Max	Average	Total
Cato	All	All land holding size	55	340	171	11814
Cat1	Marginal	less than 1 ha	18	231	115	7952
Cat2	Small	≥ l and < 2 ha	5	54	26	1800
Cat3	Semi-medium	≥ 2 ha and < 4 ha	1	36	18	1229
Cat4	Medium	≥4 ha and < 10 ha	1	20	9	628
Cat5	Large	≥ l0 ha	0	10	3	205

Table 2. Distribution of districts-wise sample size.

District	Cat0	Cat 1	Cat2	Cat3	Cat4	Cat5
Saharanpur	168	111	16	26	12	3
Muzaffar Nr	223	146	21	35	17	4
Bijnor	224	148	44	18	10	4
Moradabad	224	153	36	23	9	3
Rampur	112	83	16	10	2	1
J.B.P. Nagar	112	74	14	18	4	2
Meerut	168	116	22	16	13	1
Baghpat	55	37	11	1	5	1
Ghaziabad	153	100	28	17	4	4
Bulad Shahar	277	187	34	29	19	8
Aligarh	223	140	33	31	15	4
Hathras	56	36	9	6	5	0
Mathura	167	109	29	17	11	1
Agra	168	121	17	18	12	0
Firozabad	111	72	19	17	3	0
Etah	223	149	40	25	9	0
Mainpuri	111	77	18	7	8	1
Badaun	280	196	47	23	13	1
Bareilly	224	147	42	24	7	4
Pilibhit	112	72	12	20	7	1
Shahjahanpur	167	119	26	16	5	1
Kheri	279	175	49	29	20	6
Sitapur	280	175	54	30	16	5
Hardoi	224	145	38	27	11	3
Unnao	224	153	38	20	11	2
Lucknow	112	76	27	6	2	1
Raibarely	224	131	48	26	14	5
Farukhabad	98	63	20	9	6	0
Kannauj	112	76	22	8	5	1
Etawah	111	71	21	14	4	1
Auraya	112	75	21	11	3	2
Kanpur Dehat	222	149	31	31	6	5
Kanpur Nr	56	40	12	3	1	0
Jalaun	112	62	16	16	14	4
Jhanshi	112	56	22	21	9	4
Lalitpur	56	19	5	21	7	4
Hamirpur	56	19	8	14	11	4
Mahoba	56	18	11	9	13	5
Banda	112	64	17	17	13	1
Chitrakut	56	36	10	5	3	2
Fatehpur	168	109	29	17	9	4
Pratapgarh	223	151	40	19	12	1
Kaushambi	137	99	24	9	3	2
Allahabad	307	221	38	24	18	6
Barabanki	224	156	37	19	10	2
Faizabad	112	72	18	14	7	1
Ambedker Nr	168	116	27	15	5	5
Sultanpur	280	196	31	35	13	5
Bahraich	196	136	20	19	11	10
Srawasti	84	61	9	10	4	0
Balrampur	168	112	26	21	5	4
Gonda	224	153	31	19	19	2
Sidharth Nr	168	106	32	17	8	5
Basti	196	136	21	21	9	9
S.Kabir Nr	84	65	10	5	4	0
Maharajganj	168	119	26	13	6	4
Gorakhpur	280	210	36	22	10	2
Kushi Nr	294	212	50	19	13	0
Deoria	224	170	24	15	13	2
Azamgarh	340	231	51	36	16	6
Mau	112	75	19	9	6	3
Ballia	224	176	14	21	12	1
Jaunpur	336	227	54	31	15	9
Ghazipur	227	162	35	21	6	3
Chandauli	111	76	14	10	7	4
Varanasi	168	127	22	13	4	2
St. Ravidas Nagar	111	85	10	12	3	1
Mizapur	168	117	16	17	8	10
Shanbhadra	140	80	32	12	13	3
Total	11814	7952	1800	1229	628	205

Nr= Nagar

Table 3. Bias diagnostics test for Cat0- Cat5.

Category	Parameters	Estimate	Std error	t	Prob>|t|
Cat0	Intercept	-0.001	0.128	-0.010	0.992
	Model based estimate	1.009	0.239	4.222	0.000
Cat1	Intercept	-0.020	0.179	-0.112	0.911
	Model based estimate	1.035	0.352	2.938	0.005
Cat2	Intercept	-0.063	0.175	-0.359	0.721
	Model based estimate	1.121	0.301	3.727	0.000
Cat3	Intercept	-0.061	0.256	-0.237	0.814
	Model based estimate	1.112	0.449	2.475	0.016
Cat4	Intercept	0.108	0.888	0.122	0.904
	Model based estimate	0.839	1.391	0.604	0.548
Cat5	Intercept	.129	0.310	0.416	0.679
	Model based estimate	0.820	0.471	1.741	0.086

Table 4. District-wise model-based estimates for the districts of Cat 5 with no sample data.

District	Estimate	CV,%	95% Confidence Interval	
			Lower	Upper
Hathras	0.76	2.07	0.73	0.79
Agra	0.74	2.22	0.71	0.77
Firozabad	0.77	2.00	0.74	0.80
Etah	0.67	2.85	0.63	0.71
Farukhabad	0.77	2.01	0.74	0.80
Kanpur Nr	0.69	2.67	0.65	0.73
Srawasti	0.74	2.20	0.71	0.78
S.Kabir Nr	0.71	2.51	0.67	0.74
Kushi Nr	0.50	4.31	0.46	0.54

Table 5. District-wise model-based and direct estimates of proportion of indebted households.

		Cat0						Cat1					
		Direct			Model-based			Direct			Model-based		
Region	District	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper
Western	Saharanpur	0.60	0.52	0.67	0.60	0.54	0.65	0.59	0.49	0.68	0.58	0.51	0.65
	Muzaffarnagar	0.61	0.55	0.68	0.60	0.55	0.66	0.58	0.49	0.66	0.57	0.51	0.64
	Bijnor	0.55	0.48	0.61	0.56	0.51	0.61	0.51	0.43	0.59	0.52	0.46	0.59
	Moradabad	0.56	0.49	0.62	0.57	0.52	0.62	0.51	0.43	0.59	0.53	0.46	0.59
	Rampur	0.55	0.45	0.64	0.57	0.51	0.63	0.52	0.41	0.63	0.53	0.46	0.61
	J.B.P.Nr	0.52	0.43	0.61	0.52	0.46	0.58	0.47	0.36	0.59	0.48	0.40	0.56
	Meerut	0.59	0.51	0.66	0.58	0.53	0.64	0.56	0.47	0.65	0.55	0.48	0.62
	Baghpat	0.58	0.45	0.71	0.55	0.49	0.62	0.62	0.46	0.78	0.54	0.45	0.62
	Ghaziabad	0.55	0.47	0.63	0.55	0.50	0.61	0.50	0.40	0.60	0.51	0.44	0.58
	Bulad Shahar	0.56	0.50	0.62	0.56	0.51	0.60	0.56	0.49	0.63	0.55	0.49	0.60
	Aligarh	0.59	0.52	0.65	0.60	0.55	0.65	0.53	0.45	0.61	0.55	0.48	0.62
	Mathura	0.55	0.42	0.67	0.56	0.51	0.62	0.48	0.38	0.57	0.51	0.44	0.58
	Hathras	0.68	0.60	0.75	0.57	0.51	0.64	0.67	0.51	0.82	0.54	0.46	0.63
	Agra	0.63	0.55	0.70	0.58	0.53	0.64	0.61	0.52	0.70	0.56	0.50	0.63
	Firozabad	0.52	0.43	0.62	0.54	0.48	0.60	0.47	0.36	0.59	0.49	0.42	0.57
	Etah	0.52	0.45	0.59	0.55	0.50	0.60	0.50	0.42	0.58	0.52	0.46	0.59
	Farukhabad	0.52	0.43	0.61	0.54	0.48	0.60	0.43	0.31	0.55	0.48	0.40	0.56
	Mainpuri	0.57	0.51	0.63	0.55	0.49	0.61	0.49	0.38	0.61	0.49	0.42	0.57
	Badaun	0.55	0.49	0.61	0.56	0.51	0.61	0.53	0.46	0.60	0.53	0.47	0.59
	Bareilly	0.63	0.54	0.73	0.61	0.56	0.66	0.61	0.53	0.69	0.58	0.52	0.65
	Pilibhit	0.57	0.50	0.64	0.56	0.50	0.62	0.56	0.44	0.67	0.53	0.45	0.61
	Shahjahanpur	0.68	0.62	0.73	0.62	0.56	0.67	0.65	0.56	0.73	0.59	0.52	0.65
	Kannauj	0.50	0.44	0.56	0.53	0.47	0.59	0.46	0.35	0.57	0.49	0.41	0.56
	Etawah	0.51	0.45	0.58	0.52	0.46	0.58	0.49	0.38	0.61	0.49	0.41	0.57
	Auraya	0.50	0.44	0.56	0.51	0.45	0.57	0.48	0.37	0.59	0.48	0.41	0.56
Central	Kheri	0.58	0.49	0.68	0.58	0.54	0.63	0.56	0.49	0.63	0.56	0.50	0.62
	Sitapur	0.56	0.50	0.63	0.56	0.51	0.61	0.55	0.48	0.63	0.54	0.48	0.60
	Hardoi	0.46	0.37	0.56	0.50	0.45	0.55	0.43	0.35	0.52	0.46	0.40	0.53
	Unnao	0.59	0.50	0.69	0.57	0.52	0.62	0.59	0.51	0.67	0.55	0.49	0.61
	Lucknow	0.51	0.42	0.60	0.52	0.46	0.58	0.46	0.35	0.57	0.48	0.40	0.55
	Raibarely	0.51	0.42	0.61	0.52	0.47	0.57	0.45	0.36	0.54	0.48	0.41	0.54
	Kanpur Dehat	0.47	0.41	0.54	0.50	0.45	0.55	0.43	0.35	0.51	0.46	0.40	0.52
	Kanpur Nr	0.52	0.39	0.65	0.53	0.46	0.59	0.50	0.34	0.66	0.50	0.42	0.58
	Fatehpur	0.59	0.50	0.68	0.54	0.49	0.59	0.58	0.48	0.67	0.53	0.46	0.60
Southern	Jalaun	0.63	0.53	0.72	0.58	0.53	0.64	0.53	0.41	0.66	0.53	0.45	0.61
	Jhanshi	0.55	0.42	0.67	0.54	0.48	0.60	0.46	0.33	0.60	0.49	0.41	0.57
	Lalitpur	0.71	0.58	0.84	0.57	0.50	0.63	0.63	0.41	0.85	0.51	0.42	0.60
	Hamirpur	0.52	0.39	0.65	0.52	0.46	0.59	0.53	0.30	0.76	0.50	0.41	0.59
	Mahoba	0.48	0.39	0.57	0.52	0.45	0.58	0.28	0.06	0.49	0.46	0.37	0.55
	Banda	0.55	0.43	0.68	0.53	0.47	0.59	0.52	0.39	0.64	0.50	0.42	0.58
	Chitrakut	0.61	0.53	0.68	0.54	0.48	0.61	0.56	0.39	0.72	0.51	0.42	0.59
Eastern	Pratapgarh	0.44	0.38	0.51	0.47	0.42	0.52	0.36	0.29	0.44	0.42	0.35	0.48
	Kaushambi	0.57	0.49	0.65	0.53	0.47	0.58	0.59	0.49	0.68	0.53	0.46	0.60
	Allahabad	0.57	0.52	0.63	0.54	0.50	0.59	0.60	0.54	0.67	0.56	0.51	0.62
	Barabanki	0.57	0.51	0.64	0.56	0.51	0.61	0.52	0.44	0.60	0.52	0.46	0.58
	Faizabad	0.48	0.39	0.57	0.50	0.44	0.56	0.43	0.32	0.55	0.47	0.39	0.54
	Ambedker Nr	0.48	0.40	0.55	0.50	0.44	0.55	0.47	0.37	0.56	0.48	0.41	0.54
	Sultanpur	0.51	0.45	0.57	0.50	0.46	0.55	0.47	0.40	0.54	0.47	0.41	0.53
	Bahraich	0.51	0.44	0.58	0.52	0.47	0.57	0.49	0.40	0.57	0.49	0.43	0.56
	Srawasti	0.51	0.41	0.62	0.52	0.46	0.58	0.46	0.33	0.59	0.48	0.40	0.55
	S.Kabir Nr	0.61	0.53	0.68	0.54	0.48	0.60	0.60	0.48	0.72	0.52	0.45	0.60
	Kushi Nagar	0.58	0.52	0.65	0.55	0.50	0.59	0.56	0.49	0.63	0.53	0.48	0.59
	Balrampur	0.46	0.39	0.54	0.48	0.43	0.54	0.43	0.34	0.52	0.45	0.38	0.52
	Gonda	0.54	0.47	0.61	0.53	0.48	0.58	0.56	0.48	0.64	0.54	0.47	0.60
	Sidharth Nr	0.43	0.32	0.53	0.47	0.41	0.52	0.38	0.28	0.47	0.43	0.36	0.50
	Basti	0.58	0.50	0.65	0.55	0.50	0.60	0.52	0.44	0.61	0.51	0.44	0.57
	Maharajganj	0.47	0.41	0.53	0.49	0.44	0.54	0.45	0.36	0.54	0.47	0.40	0.53
	Gorakhpur	0.52	0.46	0.58	0.52	0.47	0.57	0.52	0.46	0.59	0.51	0.46	0.57
	Deoria	0.54	0.47	0.61	0.54	0.49	0.59	0.55	0.47	0.62	0.53	0.47	0.59
	Azamgarh	0.47	0.42	0.52	0.49	0.44	0.53	0.45	0.38	0.51	0.46	0.41	0.52
	Mau	0.58	0.49	0.67	0.54	0.48	0.60	0.52	0.41	0.63	0.50	0.42	0.57
	Ballia	0.56	0.49	0.62	0.53	0.48	0.59	0.57	0.49	0.64	0.54	0.48	0.60
	Jaunpur	0.47	0.42	0.53	0.48	0.44	0.53	0.44	0.37	0.50	0.45	0.39	0.50
	Ghazipur	0.43	0.37	0.50	0.47	0.42	0.52	0.43	0.35	0.50	0.45	0.39	0.51
	Chandauli	0.40	0.30	0.49	0.47	0.41	0.53	0.30	0.20	0.41	0.41	0.34	0.48
	Varanasi	0.51	0.43	0.58	0.51	0.45	0.56	0.50	0.42	0.59	0.49	0.42	0.56
	St. Ravidas Nr	0.48	0.38	0.57	0.51	0.44	0.57	0.51	0.40	0.61	0.49	0.42	0.57
	Mizapur	0.44	0.36	0.51	0.47	0.41	0.52	0.46	0.37	0.55	0.47	0.40	0.54
	Shanbhadra	0.36	0.28	0.44	0.44	0.38	0.49	0.25	0.15	0.35	0.38	0.31	0.45

Table 5. District-wise model-based and direct estimates of proportion of indebted households(contd.).

		Cat2						Cat3					
		Direct			Model-based			Direct			Model-based		
Region	District	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper
Western	Saharanpur	0.50	0.25	0.75	0.65	0.58	0.73	0.62	0.42	0.81	0.66	0.57	0.75
	Muzaffarnagar	0.67	0.46	0.87	0.63	0.55	0.71	0.69	0.53	0.84	0.66	0.55	0.76
	Bijnor	0.64	0.49	0.78	0.63	0.55	0.70	0.50	0.26	0.74	0.61	0.53	0.68
	Moradabad	0.72	0.57	0.87	0.64	0.56	0.71	0.52	0.31	0.73	0.64	0.55	0.72
	Rampur	0.69	0.45	0.92	0.64	0.59	0.70	0.50	0.17	0.83	0.63	0.56	0.69
	J.B.P.Nr	0.57	0.30	0.84	0.59	0.52	0.67	0.56	0.32	0.79	0.54	0.45	0.64
	Meerut	0.73	0.54	0.92	0.64	0.58	0.70	0.56	0.31	0.81	0.63	0.55	0.70
	Baghpat	0.55	0.24	0.85	0.61	0.56	0.67	1.00			0.58	0.52	0.64
	Ghaziabad	0.68	0.50	0.85	0.62	0.56	0.69	0.59	0.35	0.83	0.60	0.52	0.67
	Bulad Shahar	0.53	0.36	0.70	0.60	0.53	0.66	0.55	0.37	0.74	0.59	0.51	0.67
	Aligarh	0.73	0.57	0.88	0.66	0.58	0.74	0.68	0.51	0.84	0.67	0.57	0.77
	Mathura	0.66	0.48	0.83	0.63	0.56	0.69	0.82	0.64	1.01	0.63	0.55	0.70
	Hathras	0.78	0.49	1.07	0.61	0.55	0.67	0.50	0.06	0.94	0.57	0.50	0.64
	Agra	0.71	0.48	0.93	0.62	0.56	0.68	0.72	0.51	0.94	0.59	0.51	0.67
	Firozabad	0.63	0.41	0.85	0.62	0.55	0.68	0.53	0.28	0.77	0.57	0.49	0.66
	Etah	0.63	0.47	0.78	0.65	0.57	0.72	0.56	0.36	0.76	0.64	0.55	0.73
	Farukhabad	0.75	0.56	0.94	0.62	0.56	0.69	0.44	0.10	0.79	0.58	0.51	0.65
	Mainpuri	0.72	0.51	0.94	0.61	0.54	0.68	0.86	0.58	1.14	0.56	0.48	0.64
	Badaun	0.60	0.45	0.74	0.63	0.56	0.71	0.65	0.45	0.85	0.61	0.53	0.70
	Bareilly	0.71	0.58	0.85	0.65	0.58	0.72	0.67	0.47	0.86	0.63	0.55	0.72
	Pilibhit	0.58	0.29	0.87	0.62	0.57	0.68	0.65	0.44	0.86	0.59	0.50	0.67
	Shahjahanpur	0.69	0.51	0.87	0.63	0.57	0.69	0.75	0.53	0.97	0.61	0.53	0.68
	Kannauj	0.50	0.29	0.71	0.62	0.55	0.68	0.50	0.13	0.87	0.58	0.51	0.65
	Etawah	0.48	0.26	0.70	0.60	0.52	0.67	0.64	0.38	0.90	0.55	0.46	0.64
	Auraya	0.52	0.30	0.74	0.59	0.52	0.66	0.55	0.24	0.85	0.54	0.46	0.63
Central	Kheri	0.59	0.45	0.73	0.63	0.55	0.71	0.69	0.52	0.86	0.64	0.55	0.73
	Sitapur	0.50	0.37	0.63	0.61	0.53	0.68	0.57	0.39	0.75	0.60	0.51	0.68
	Hardoi	0.47	0.31	0.63	0.59	0.52	0.66	0.56	0.36	0.75	0.56	0.48	0.65
	Unnao	0.58	0.42	0.74	0.59	0.52	0.65	0.60	0.38	0.82	0.57	0.49	0.64
	Lucknow	0.63	0.44	0.82	0.58	0.51	0.65	0.67	0.25	1.08	0.55	0.48	0.62
	Raibarely	0.50	0.36	0.64	0.55	0.47	0.63	0.69	0.51	0.87	0.57	0.48	0.66
	Kanpur Dehat	0.55	0.37	0.73	0.60	0.53	0.67	0.52	0.34	0.69	0.57	0.48	0.66
	Kanpur Nr	0.58	0.29	0.87	0.58	0.53	0.63	0.33	-0.32	0.99	0.56	0.51	0.61
	Fatehpur	0.69	0.52	0.86	0.52	0.44	0.59	0.41	0.17	0.65	0.52	0.44	0.60
Southern	Jalaun	0.50	0.25	0.75	0.61	0.56	0.67	0.75	0.53	0.97	0.61	0.53	0.68
	Jhanshi	0.64	0.43	0.84	0.59	0.53	0.65	0.52	0.30	0.74	0.57	0.49	0.64
	Lalitpur	1.00	1.00	1.00	0.59	0.54	0.65	0.71	0.52	0.91	0.56	0.47	0.64
	Hamirpur	0.63	0.27	0.98	0.58	0.53	0.63	0.29	0.04	0.53	0.55	0.48	0.62
	Mahoba	0.45	0.15	0.76	0.58	0.53	0.64	0.78	0.49	1.07	0.56	0.49	0.62
	Banda	0.59	0.35	0.83	0.55	0.49	0.61	0.47	0.23	0.72	0.54	0.47	0.62
	Chitrakut	0.80	0.54	1.06	0.58	0.52	0.63	0.80	0.41	1.19	0.55	0.49	0.61
Eastern	Pratapgarh	0.60	0.45	0.75	0.53	0.46	0.61	0.63	0.41	0.85	0.54	0.46	0.62
	Kaushambi	0.54	0.34	0.75	0.52	0.45	0.59	0.44	0.10	0.79	0.52	0.45	0.59
	Allahabad	0.42	0.26	0.58	0.48	0.39	0.58	0.54	0.34	0.75	0.53	0.42	0.64
	Barabanki	0.65	0.49	0.80	0.58	0.50	0.65	0.68	0.47	0.90	0.59	0.51	0.66
	Faizabad	0.50	0.26	0.74	0.55	0.50	0.61	0.57	0.30	0.84	0.55	0.47	0.62
	Ambedker Nr	0.41	0.22	0.60	0.55	0.48	0.61	0.53	0.27	0.79	0.54	0.47	0.62
	Sultanpur	0.55	0.37	0.73	0.51	0.44	0.58	0.63	0.47	0.79	0.53	0.43	0.63
	Bahraich	0.45	0.23	0.67	0.57	0.51	0.62	0.53	0.30	0.76	0.56	0.49	0.64
	Srawasti	0.56	0.21	0.90	0.58	0.52	0.64	0.70	0.40	1.00	0.55	0.47	0.62
	S.Kabir Nr	0.70	0.40	1.00	0.56	0.50	0.62	0.40	-0.08	0.88	0.53	0.46	0.59
	Kushi Nagar	0.58	0.44	0.72	0.50	0.41	0.59	0.84	0.67	1.01	0.53	0.44	0.62
	Balrampur	0.50	0.30	0.70	0.52	0.45	0.59	0.48	0.26	0.70	0.53	0.44	0.61
	Gonda	0.55	0.37	0.73	0.57	0.51	0.64	0.26	0.06	0.47	0.55	0.48	0.63
	Sidharth Nr	0.44	0.26	0.61	0.53	0.45	0.60	0.53	0.28	0.77	0.53	0.45	0.61
	Basti	0.62	0.41	0.83	0.56	0.50	0.62	0.67	0.46	0.87	0.55	0.47	0.63
	Maharajganj	0.62	0.42	0.81	0.53	0.46	0.59	0.38	0.11	0.66	0.54	0.46	0.61
	Gorakhpur	0.53	0.36	0.69	0.54	0.47	0.61	0.41	0.20	0.62	0.55	0.47	0.64
	Deoria	0.58	0.38	0.78	0.57	0.52	0.63	0.60	0.34	0.86	0.57	0.50	0.64
	Azamgarh	0.47	0.33	0.61	0.52	0.44	0.61	0.56	0.39	0.72	0.55	0.45	0.65
	Mau	0.74	0.53	0.94	0.57	0.50	0.63	0.67	0.34	0.99	0.54	0.47	0.61
	Ballia	0.57	0.30	0.84	0.53	0.47	0.59	0.48	0.26	0.70	0.54	0.45	0.62
	Jaunpur	0.65	0.52	0.78	0.53	0.45	0.62	0.52	0.34	0.69	0.53	0.43	0.62
	Ghazipur	0.43	0.26	0.59	0.53	0.45	0.60	0.48	0.26	0.70	0.54	0.46	0.62
	Chandauli	0.64	0.38	0.90	0.56	0.50	0.62	0.30	0.00	0.60	0.53	0.46	0.60
	Varanasi	0.45	0.24	0.67	0.56	0.49	0.63	0.46	0.18	0.74	0.53	0.45	0.61
	St. Ravidas Nr	0.50	0.17	0.83	0.59	0.52	0.66	0.33	0.05	0.61	0.54	0.45	0.63
	Mizapur	0.25	0.03	0.47	0.51	0.44	0.57	0.29	0.07	0.52	0.52	0.43	0.60
	Shanbhadra	0.47	0.29	0.64	0.52	0.44	0.59	0.58	0.29	0.87	0.52	0.44	0.59

Table 5. District-wise model-based and direct estimates of proportion of indebted households(contd.).

		Cat4						Cat5					
		Direct			Model-based			Direct			Model-based		
Region	District	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper	Estimate	Lower	Upper
Western	Saharanpur	0.83	0.61	1.05	0.69	0.56	0.83	0.33	-0.32	0.99	0.62	0.39	0.85
	Muzaffarnagar	0.71	0.48	0.93	0.68	0.54	0.82	0.75	0.26	1.24	0.55	0.28	0.81
	Bijnor	0.60	0.28	0.92	0.65	0.53	0.77	0.75	0.26	1.24	0.69	0.52	0.86
	Moradabad	0.67	0.34	0.99	0.67	0.55	0.79	1.00			0.63	0.43	0.83
	Rampur	0.50	-0.48	1.48	0.66	0.55	0.77	1.00			0.70	0.56	0.84
	J.B.P.Nr	0.75	0.26	1.24	0.63	0.49	0.76	1.00			0.79	0.64	0.93
	Meerut	0.62	0.34	0.89	0.66	0.53	0.78	1.00			0.68	0.54	0.82
	Baghpat	0.20	-0.19	0.59	0.62	0.50	0.74	1.00			0.75	0.63	0.87
	Ghaziabad	0.75	0.26	1.24	0.65	0.54	0.76	0.50	-0.07	1.07	0.73	0.57	0.88
	Bulad Shahar	0.63	0.41	0.85	0.65	0.52	0.77	0.63	0.27	0.98	0.66	0.44	0.87
	Aligarh	0.73	0.50	0.96	0.69	0.55	0.83	0.25	-0.24	0.74	0.60	0.35	0.86
	Mathura	0.55	0.24	0.85	0.65	0.53	0.78	0.00	0.00	0.00	0.65	0.51	0.79
	Hathras	0.80	0.41	1.19	0.64	0.52	0.76				0.76	0.73	0.79
	Agra	0.50	0.20	0.80	0.63	0.50	0.76				0.74	0.71	0.77
	Firozabad	1.00	1.00	1.00	0.65	0.53	0.77				0.77	0.74	0.80
	Etah	0.33	0.01	0.66	0.65	0.52	0.78				0.67	0.63	0.71
	Farukhabad	0.83	0.51	1.16	0.65	0.53	0.77				0.77	0.74	0.80
	Mainpuri	0.63	0.27	0.98	0.63	0.49	0.77	1.00			0.79	0.65	0.92
	Badaun	0.46	0.18	0.74	0.64	0.51	0.77	1.00			0.72	0.59	0.85
	Bareilly	0.43	0.03	0.82	0.65	0.53	0.78	0.75	0.26	1.24	0.70	0.51	0.88
	Pilibhit	0.57	0.18	0.97	0.64	0.52	0.76	0.00			0.75	0.63	0.88
	Shahjahanpur	1.00	1.00	1.00	0.67	0.55	0.78	1.00			0.73	0.61	0.86
	Kannauj	1.00	1.00	1.00	0.66	0.54	0.78	1.00			0.76	0.64	0.89
	Etawah	0.50	-0.07	1.07	0.62	0.49	0.76	1.00			0.79	0.65	0.93
	Auraya	0.67	0.01	1.32	0.63	0.50	0.76	0.50	-0.48	1.48	0.77	0.63	0.92
Central	Kheri	0.60	0.38	0.82	0.66	0.53	0.80	0.67	0.25	1.08	0.61	0.37	0.85
	Sitapur	0.75	0.53	0.97	0.66	0.54	0.78	0.80	0.41	1.19	0.68	0.50	0.86
	Hardoi	0.55	0.24	0.85	0.63	0.50	0.75	0.67	0.01	1.32	0.73	0.59	0.87
	Unnao	0.64	0.34	0.93	0.64	0.52	0.76	1.00			0.70	0.57	0.82
	Lucknow	0.00	0.00	0.00	0.62	0.50	0.74	1.00			0.74	0.62	0.87
	Raibarely	0.86	0.67	1.05	0.66	0.54	0.78	0.40	-0.08	0.88	0.55	0.34	0.77
	Kanpur Dehat	0.83	0.51	1.16	0.65	0.53	0.76	0.60	0.12	1.08	0.73	0.56	0.89
	Kanpur Nr	1.00	1.00	1.00	0.64	0.54	0.74				0.69	0.65	0.73
	Fatehpur	0.67	0.34	0.99	0.62	0.49	0.75	0.75	0.26	1.24	0.57	0.37	0.77
Southern	Jalaun	0.93	0.79	1.07	0.68	0.57	0.80	1.00	1.00	1.00	0.68	0.51	0.84
	Jhanshi	0.78	0.49	1.07	0.65	0.53	0.76	0.75	0.26	1.24	0.70	0.54	0.85
	Lalitpur	0.57	0.18	0.97	0.63	0.50	0.75	1.00	1.00	1.00	0.76	0.60	0.91
	Hamirpur	0.82	0.58	1.06	0.65	0.53	0.77	0.25	-0.24	0.74	0.70	0.54	0.86
	Mahoba	0.46	0.18	0.74	0.61	0.48	0.74	0.80	0.41	1.19	0.72	0.56	0.89
	Banda	0.77	0.53	1.01	0.64	0.52	0.76	1.00	1.00	1.00	0.64	0.53	0.75
	Chitrakut	0.33	-0.32	0.99	0.62	0.51	0.73	0.50	-0.48	1.48	0.71	0.58	0.84
Eastern	Pratapgarh	0.58	0.29	0.87	0.62	0.49	0.75	1.00	1.00	1.00	0.59	0.46	0.72
	Kaushambi	0.67	0.01	1.32	0.62	0.50	0.74	0.50	-0.48	1.48	0.60	0.44	0.75
	Allahabad	0.67	0.44	0.89	0.63	0.48	0.79	0.33	-0.08	0.75	0.42	0.14	0.70
	Barabanki	0.90	0.70	1.10	0.67	0.55	0.78	0.50	-0.48	1.48	0.59	0.44	0.74
	Faizabad	0.71	0.35	1.08	0.63	0.52	0.75	1.00	1.00	1.00	0.65	0.54	0.76
	Ambedker Nr	0.60	0.12	1.08	0.63	0.51	0.74	0.80	0.41	1.19	0.64	0.45	0.82
	Sultanpur	0.69	0.43	0.95	0.63	0.50	0.76	0.40	-0.08	0.88	0.55	0.33	0.77
	Bahraich	0.82	0.58	1.06	0.65	0.54	0.77	0.50	0.17	0.83	0.63	0.39	0.88
	Srawasti	0.75	0.26	1.24	0.63	0.51	0.75				0.74	0.71	0.78
	S.Kabir Nr	0.75	0.26	1.24	0.62	0.51	0.74				0.71	0.67	0.74
	Kushi Nagar	0.54	0.26	0.82	0.61	0.47	0.75				0.50	0.46	0.54
	Balrampur	0.60	0.12	1.08	0.62	0.50	0.74	1.00	1.00	1.00	0.59	0.40	0.78
	Gonda	0.63	0.41	0.85	0.63	0.51	0.76	0.00	0.00	0.00	0.66	0.54	0.79
	Sidharth Nr	0.63	0.27	0.98	0.62	0.50	0.75	0.80	0.41	1.19	0.60	0.40	0.80
	Basti	0.89	0.67	1.11	0.65	0.53	0.76	0.78	0.49	1.07	0.68	0.46	0.90
	Maharajganj	0.50	0.06	0.94	0.62	0.50	0.74	0.50	-0.07	1.07	0.56	0.37	0.76
	Gorakhpur	0.60	0.28	0.92	0.63	0.51	0.75	1.00	1.00	1.00	0.56	0.40	0.72
	Deoria	0.31	0.05	0.57	0.61	0.48	0.73	0.50	-0.48	1.48	0.64	0.51	0.77
	Azamgarh	0.56	0.31	0.81	0.62	0.49	0.76	0.67	0.25	1.08	0.53	0.29	0.77
	Mau	0.67	0.25	1.08	0.63	0.51	0.75	0.67	0.01	1.32	0.70	0.55	0.85
	Ballia	0.50	0.20	0.80	0.61	0.49	0.74	1.00	1.00	1.00	0.58	0.44	0.71
	Jaunpur	0.33	0.09	0.58	0.59	0.45	0.72	0.44	0.10	0.79	0.62	0.38	0.87
	Ghazipur	0.50	0.06	0.94	0.62	0.50	0.74	0.33	-0.32	0.99	0.57	0.39	0.74
	Chandauli	0.71	0.35	1.08	0.63	0.51	0.75	1.00	1.00	1.00	0.70	0.53	0.86
	Varanasi	1.00	1.00	1.00	0.63	0.51	0.75	0.50	-0.48	1.48	0.73	0.58	0.87
	St. Ravidas Nr	0.33	-0.32	0.99	0.62	0.48	0.75				0.78	0.64	0.91
	Mizapur	0.63	0.27	0.98	0.62	0.49	0.75	0.50	0.17	0.83	0.56	0.28	0.84
	Shanbhadra	0.54	0.26	0.82	0.61	0.48	0.74	0.33	-0.32	0.99	0.61	0.44	0.79

Figure 1. Distribution of the district level residuals (left hand side plots) and normal q-q plot of the district level residuals (right hand side plots) for Cat0 (up) to Cat 5(down).

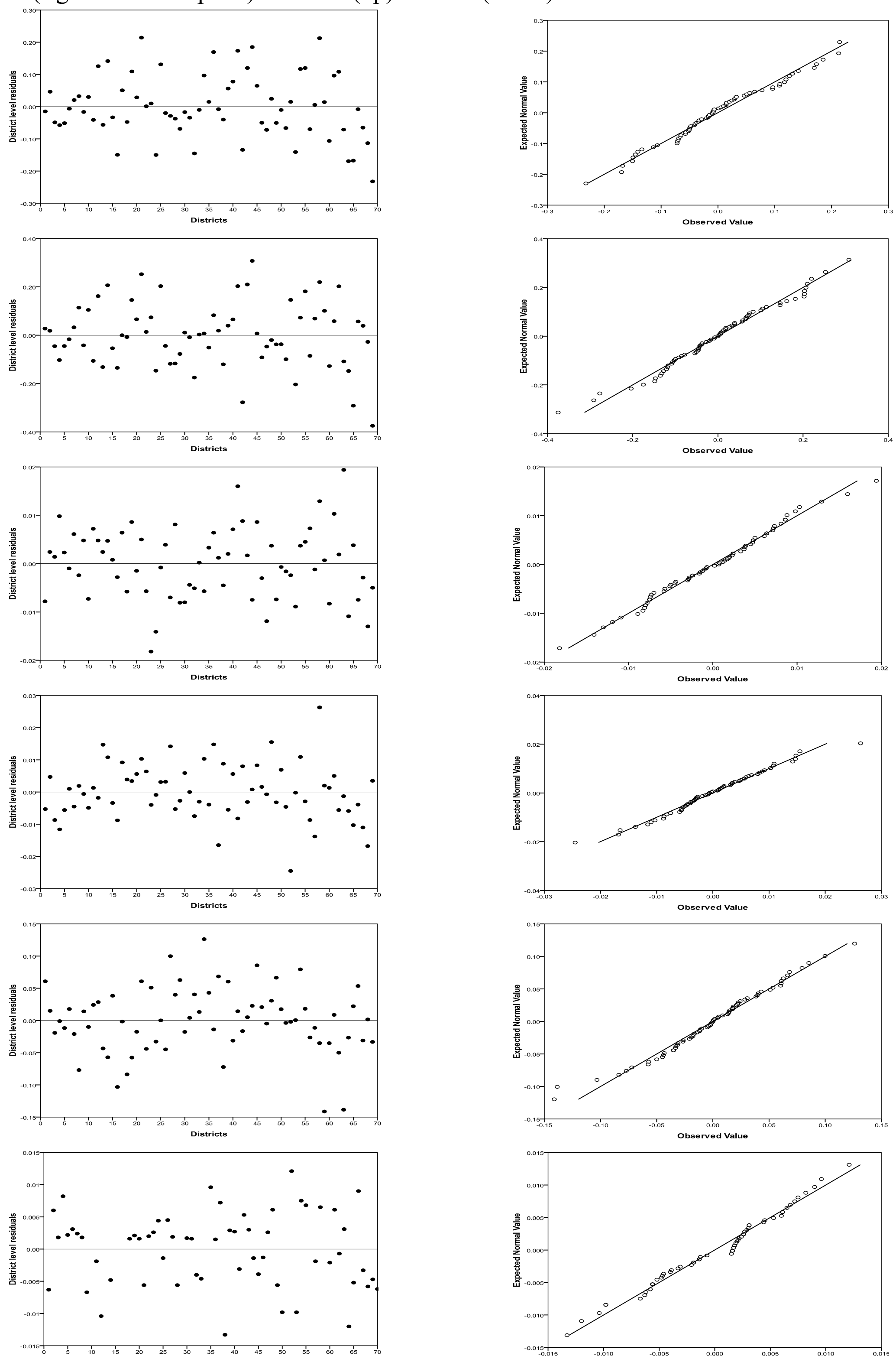

Figure 2. Bias diagnostics plots with $Y = X$ line (solid) and regression line (dotted) for Cat0- Cat5.

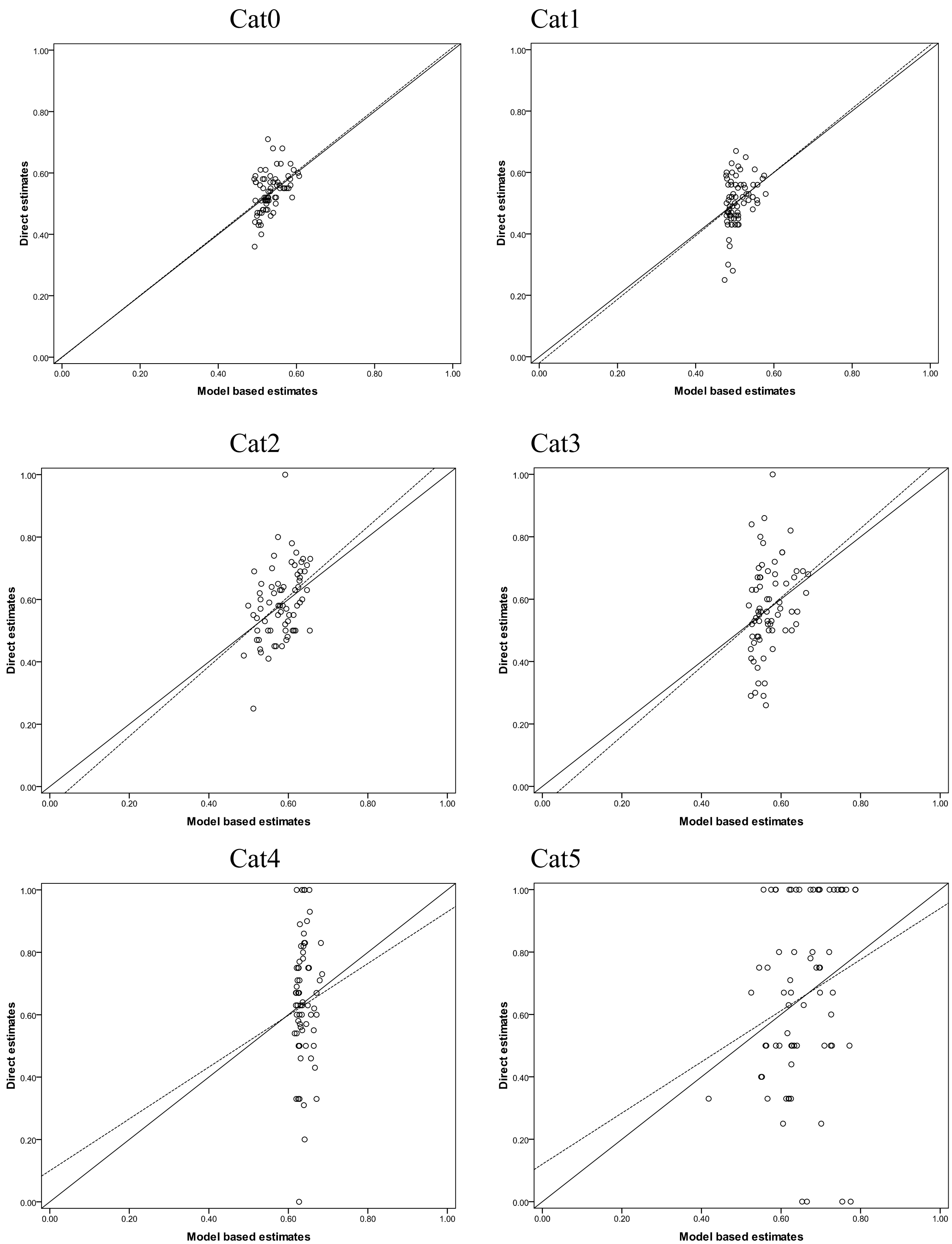

Figure 3. District-wise coefficient of variation for direct (solid line) and model-based estimate (dash line) for Cat0- Cat5.