IARIW 2025

IARIW 2025

Thursday, October 2 & Friday, October 3

Comparing Income Estimates Using Different Small Area Estimation Techniques: Evidence from Indian Household Survey

Preeti Preeti
(Indian Institute of Technology Roorkee)

Bharat Diwakar
(Indian Institute of Technology Roorkee)

preeti@hs.iitr.ac.in

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on "Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies" October 2-3, 2025

Session 4A: Income Measurement in Developing Countries
Time Slot: Friday, October 3, 8:00-9:30 AM

Comparing Income Estimates Using Different Small Area Estimation Techniques: Evidence from Indian Household Survey[1]

Preeti[2] **Bharat Diwakar**[3]

Abstract

The disaggregated data plays a vital role in policy formulation and assessing their efficacy, which leads policymakers to use micro-level information. However, surveys are primarily provide statistics at the national and state levels, and disaggregated estimates often suffer from high sampling variability due to limited sample sizes. Small Area Estimation (SAE) addresses this issue by producing more reliable estimates for small domains through the integration of survey data with external sources. In the absence of unit-level information, area-level models are commonly applied, with the Fay-Herriot (FH) model is a widely used frequentist method that treats model parameters as fixed. In contrast, the Hierarchical Bayesian (HB) model incorporates prior knowledge, making inference more intuitive. In this study, we attempt to conduct an empirical analysis of both models to evaluate their applicability in India. We generate district-level estimates for rural areas of Chhattisgarh (located in central India) for the years 2004 and 2011, a region characterized by high poverty and low per capita income. The results show that average household income increased over the period, but southern and some northern districts remain in the lower-income range. The study also identifies key factors - inaccessibility of health services, lack of education, and poor infrastructure behind this uneven distribution and offers policy recommendations aimed at fostering more equitable growth in the state.

Keywords: Direct estimators, Fay-Herriot model, Hierarchical Bayesian model, Income distribution

---

[1] This is the preliminary work, please do not cite without permission.

[2] Research Scholar, Department of Humanities and Social Sciences, IIT Roorkee. Email: preeti@hs.iitr.ac.in .

[3] Corresponding author. Assistant Professor, Department of Humanities and Social Sciences, IIT Roorkee. Email: d.bharat@hs.iitr.ac.in.

1. Introduction

India has recently gained global attention by becoming a 4 trillion-dollar economy,[4] reflecting its growing economic potential. These estimates appear convincing to some extent, as the population and consumption are continuously rising in the country; however, improvements are not uniformly visible at ground level. The recent World Bank report "Poverty and Equity Brief: April 2025" claims a decline in poverty from 16.2% in 2011-12 to 2.3% in 2022-23, and a fall in the Gini coefficient of consumption inequality from 0.28 to 0.25 over the same period. But the above numbers raise a critical question regarding this progress. There is no doubt that the consumption expenditure pattern of people has improved over the years, but many are struggling to catch up with the basic necessities. It is important to study the income inequality and consumption inequality separately. Because consumption may follow income trends, but as the income keeps increasing, it often leads to higher savings rather than a proportionate increase in consumption. According to the World Inequality Database,[5] income inequality has worsened between 2004 and 2023, as the value of Gini coefficient rose from 0.52 to 0.62 during the same period. Similarly, Bharti et al. (2024) found that between 2015 and 2023, the top 1% held 22.6% of income and 40.6% of wealth, while the bottom 50% had almost nothing. The picture becomes even more unequal when we examine the pattern at the state or district level, as states are facing significant inter and intra-state disparity (Roy, 2017; Mohanty et al., 2016). Moreover, rural areas are more vulnerable than their urban counterparts (Subramanian, 2019). Therefore, these findings force us to rethink the inequality and its implications on development.

The distinction between inequality (uneven resource and opportunity distribution) and inequity (disparities from poor service provision) is essential, as both lead to deprivation and poverty. The concept of inequality is covered under the Sustainable Development Goal (SDG)-10, which aims

[4] Press Information Bureau, Government of India. https://www.pib.gov.in/PressNoteDetails.aspx?NoteId=154660#:~:text=India%20became%20the%204th%20largest,the%20lowest%20since%202018%E2%80%9319.

[5] Poverty and Equity Berief: April 2025. http://documents.worldbank.org/curated/en/099722104222534584

to boost the income of the bottom 40% at a rate faster than the national average, while SDG 8 focuses on inclusive and sustainable growth through productivity and reduced income gaps. To achieve these, the Government of India came up with several initiatives like Make in India, Start-up India, Skill India and Digital India, which specifically target poverty by providing employment opportunities (Taunk & Nimbalkar, 2021). In alignment with these goals, the Aspirational Districts and Blocks Programme, under the motto Sabka Saath Sabka Vikas, targets socio-economic development through cooperative federalism and performance-based rankings to reduce inequality.

Districts play a vital role in shaping a state's growth. But income data in India face challenges in coverage, availability, and quality (Bharti et al., 2024). Consequently, researchers frequently use consumption expenditure as a proxy for income (Subramanian & Jayaraj, 2013; Chancel & Piketty, 2019; UNU-WIDER, 2021; Tyagi, 2023). On the other hand, even having data, we may face issues related to low sample sizes at the district level. The various national surveys often lack the granularity to produce reliable district-level statistics due to limited sample sizes (Chandra et al., 2011). Therefore, direct estimators under a design-based framework may fail to produce reliable results for small areas (Anjoy & Chandra, 2019). Several studies have attempted to estimate income distribution in India by integrating different data sources. Banerjee and Piketty (2005) used tax and national accounts data for 1922-2000. Later on, Chancel and Piketty 2019 conducted the most comprehensive study; they used national accounts, tax tabulation and household survey datasets and estimated income inequality for a long time period, 1922 to 2014. Their work was extended by Somanchi (2023) for the period 2014 to 2022. They combined several large survey datasets like the Periodic Labour Force Survey (PLFS), the Consumer Pyramid Household Survey (CPHS) and tax estimates. Similarly, Bharti et al. (2024) estimated income distribution for the years 2014 to 2023. All the studies highlighted similar findings that the top 1 % held the highest share during the British era. Between 1950 and 1980, it came down because of the adoption of the socialist policy of the government. But around the year 2000, it skyrocketed when the government opened the market for globalisation. Researchers termed this era as "Billionaire Raj", which is more unequal than the colonial times. These studies provide valuable high-level estimates that help in understanding inequality patterns in the Indian economy. However, due to the lack of datasets at disaggregated levels, it is difficult to replicate studies like Chancel and Piketty (2019) and Bharti et al. (2024). In this case, increasing sample size is the only solution, but this becomes a more costly and time-consuming process. Hence, small area estimation (SAE) is an effective and efficient alternative technique.

The SAE is used to precisely generate estimates for a small domain by combining information from numerous sources. It relates the interest variable to auxiliary information through different statistical models. Basically, it is categorised into two models - the unit level model and the area level model. The unit-level model combines individual-level data with unit-specific auxiliary information (Battese et al., 1988). Generally, area-level models are widely applied in studies, as unit-level or individual-level datasets are difficult to obtain. The area level model was first proposed by Fay and Herriot in 1979 while calculating the per capita income at the county level using census and administrative data. Several studies in India have extensively used SAE methods across various fields such as economics, agriculture, and health to generate micro-level estimates (e.g., Pramanik et al., 2015; Chandra et al., 2018; Srivastava et al., 2021; Guha and Chandra, 2021, 2022a). For instance, Pramanik et al. (2015) estimated vaccination rates for states not covered under the Annual Health Survey (AHS-I) in 2011. Chandra et al. (2011) estimated the proportion of indebted households in Uttar Pradesh by integrating data from the Debt-Investment Survey, the Population Census, and the Agriculture Census using a Logistic Linear Mixed Model (LLMM). Their results demonstrated the effectiveness of the methodology in generating estimates for domains with small sample sizes. Using the same model, Chandra et al. (2018) assessed the poverty incidence in Bihar based on per capita consumption expenditure from the 2011-12 household consumption expenditure survey and census data. The results indicated a higher poverty incidence in the western districts- such as Buxar, Rohtas, Pashchim

Champaran, and Bhojpur - and a lower incidence in districts like Supaul, Begusarai, Araria, Saharsa, and Madhubani.

In another study, Guha and Chandra (2022a) used data from the 2018-19 Periodic Labour Force Survey (PLFS) with the Multivariate Fay-Herriot (MFH) model to examine earning inequality across districts in rural and urban Uttar Pradesh. Their spatial analysis revealed substantial disparities in income distribution, with rural areas exhibiting significantly lower levels of earning inequality compared to urban areas. Similarly, Guha and Chandra (2022b) explored earning inequality in Bihar by computing the Theil Index to compare rural and urban sectors. Srivastava et al. (2021) produced district-level estimates of childhood stunting using National Family Health Survey (NFHS-2, 1998-99) data and analysed the trends in child health outcomes with NFHS-4 (2015-16).

However, there is a lack of studies in the literature that explore income distribution at the micro level. In 2022a, Guha and Chandra attempted to estimate earning inequality in the rural and urban areas of Uttar Pradesh for the year 2018. They combined PLFS (2018-19) and Census (2011) datasets using the MFH model and provided district-level estimates with low standard errors. Although the methodology is significant and can be extended to other states to reveal their true income patterns. But the problem is PLFS dataset, which covers a large amount of information on labour income but does not include non-labour earnings. Ignoring non-labour income can lead to serious consequences when measuring income inequality (Somanchi, 2019). In this context, it becomes relevant to re-estimate income using datasets that incorporate a broader range of income sources to obtain more accurate estimates with minimal variation. Therefore, in this study, we attempt to examine the income estimates for Chhattisgarh state using an alternative data source. The rationale for choosing this specific state is discussed in a later section. The key parameter of interest is average household income at the district level. Usually, to obtain the small area means, studies consider the empirical-based unbiased predictor (EBLUP), which is based on a frequentist approach (Srivastava et al., 2007; Sud et al., 2012; Chandra, 2013; Desiyant et al., 2022). Depending on the nature of the target parameter (e.g., mean, total, count), different models have been used, such as the logistic linear mixed model (Chandra et al., 2011), generalized linear mixed models (Chandra & Verma, 2022), the multivariate FH model (Guha and Chandra, 2022a, 2022b), and the spatial Fay-Herriot model (Porter et al., 2014). Although widely applied, the frequentist approach relies on certain approximations to derive estimates. In contrast, the Hierarchical Bayesian (HB) approach is based on the Bayesian approach, which considers the probability distribution of the dataset (Anjoy & Chandra, 2019). Anjoy et al. (2020) used the HB model to study poverty incidence in rural districts of Chhattisgarh and found a high concentration of poverty in the northern and southern regions, particularly among Scheduled Tribes. Several studies have shown that the HB model outperforms the FH model and provides more precise and robust estimates (You & Chapman, 2004; Anjoy & Chandra, 2019).

1.1 Study Areas

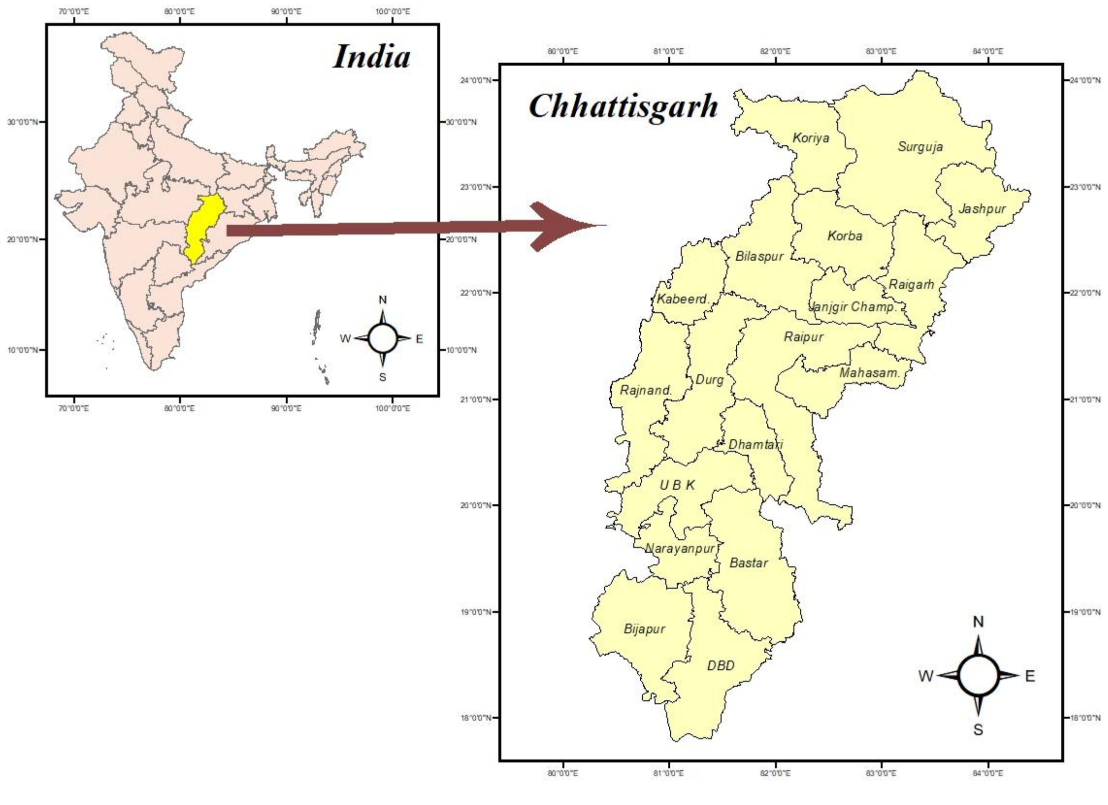
Chhattisgarh is the 10th largest and 18th most populous state located in central India, covering 1,35,192 sq. km of area. The state was carved out of Madhya Pradesh in November 2000 to ensure better governance and decentralised power for the region's overall development. At present, the state consists of 33 districts under five divisions[6] which are given below –

- Surguja - Balrampur-Ramanujganj, Jashpur, Koriya, Manendragarh-Chirmiri-Bharatpur, Surajpur, Surguja
- Bilaspur - Bilaspur, Gaurella-Pendra-Marwahi, Janjgir Champ. (Janjgir-Champa), Korba, Mungeli, Raigarh, Sakti, Sarangarh-Bilaigarh
- Durg - Balod, Bemetara, Durg, Kabirdham, Khairagarh-Chhuikhadan-Gandai, Mohla-Manpur-Ambagarh Chowki, Rajnand. (Rajnandgaon)
- Raipur - Raipur, Baloda Bazar, Dhamtari, Gariaband, Mahasam. (Mahasamund)
- Bastar - Bastar, Bijapur, DBD (Dakshin Bastar Dantewada/Dantewada), UBK (Uttar Bastar kanker/ kanker), Kondagaon, Narayanpur, Sukma

[6] https://cgstate.gov.in/en

This study is confined to only 18 districts as per the Census 2011, as shown in Figure 1. There is a significant tribal population with diverse languages, socio-cultural, livelihood and development levels in the state. It is the home of 42 tribes, with the Gind tribe being dominant. With the vast forest coverage, cultivation is the state's dominant occupation mode. About 80% of farmers fall under the small and marginal categories. The primary cultivation of Kharif crops includes paddy, soyabean, urad and arhar, while chickpea and lathyrus are sown in the Rabi season. However, 70 % of the state's agriculture is impacted by bad monsoons, affecting rural income. The state is also famous for its industrial development in rice mills and cement and steel plants, which are mainly situated in the Durg, Raipur, Korba, and Bilaspur districts.

Fig 1: Geographical map of Chhattisgarh



Source: ArcGIS

Janjgir Champ- Janjgir Champa,

Kabeerd.-Kabeerdham/Kawardha

Mahasam.- Mahasamund

Rajnand.- Rajnandgaon

UBK- Uttar Bastar kanker/ kanker

DBK- Dakshin Bastar Dantewada/Dantewada

Besides the geographical pattern, the socio-economic picture mirrors a different image of the state. Out of the total population (2,55,45,198), almost 76.76 % of the population lived in rural areas, while 23.24 % resided in urban areas (Population Census, 2011). The poverty rate is highest among other states, where 10 million people live below the poverty line (World Bank, 2016). In 2020, the state had a 22 % birth rate and the highest death rate of 7.9 % (Chhattisgarh Economic Survey, 2023-24). In addition, the sex ratio is 991 females per 1000 males, which is lower than the national average of 943 (Census, 2011). The per capita income of a state remains lower than that of major states like Gujarat, Maharashtra, Haryana, Kerala, and Tamil Nadu. This has further exacerbated the inter-state disparity over the last decade (Fourteenth Finance Commission Memorandum, 2014). In 2021, Chandrashekhar et al. showed that the median monthly per capita household earnings (MPCHE) of Chhattisgarh are lower among other states, especially in rural areas (2021). The study also highlighted that it has the highest Gini coefficient of 0.47 amongst the eight poorer states. As a large section of the state's population resides in villages and remains underdeveloped despite government interventions, there is a need to uncover the actual income distribution in the state and identify the most backward regions. This would enable efficient allocation of funds and targeted policy interventions.

The remaining paper is arranged as follows. Section 2 presents data and methodology. Section 3 discusses results by comparing the direct, FH and HB estimates, along with their diagnostic tests. Finally, Section 4 summarises the study, discusses the key findings and concludes by emphasising the importance of producing precise estimates at the subpopulation level.

2. Data and Methodology

The household level income dataset is obtained from the India Human Development Survey, IHDS-I (2004-05) and II (2011-12). The survey provides a wide range of information on various indicators, including health, education, employment, infrastructure, economic status, wage levels, etc. The first round was conducted in 2004-05, which involved a total of 41,554 households, with 26,734 rural and 14,820 urban households. Then in 2011-12, 83 per cent of households were re-interviewed, with an additional 2,134 households and provided the information on 42,152 households, with 27,579 in rural and 14,573 in urban areas, using the stratified random sampling and proportional population (PPP) for rural and urban areas, respectively. The income is provided for 8 different sources, including agriculture, business, non-agricultural wage, farm income, salary, government transfer benefits, remittances and income from estates. The auxiliary variables are extracted from the Population Census (2001) and (2011) and linked with survey datasets. According to the Population Census (2001), the state had 16 districts; later, two districts were carved out from Bastar and DBD in 2011, leading the total to 18. However, IHDS collected samples only from 15 districts in both rounds. The sample size for rural districts ranges between 20 to 148 in IHDS-I, while it ranges from 20 to 165 in IHDS-II. The average income of districts is computed using a design-based direct estimator (i.e., Horvitz-Thompson)[7] from survey data. These direct estimates will then be linked with the auxiliary variables from the population census by applying separate SAE modelling to produce new income estimates for two time periods, making the analysis cross-sectional. Since the census data provide auxiliary information as an aggregate of districts, we are using an area-level model in this analysis. Following Anjoy and Chandra (2021), we describe the FH and HB methods of small area estimation below.

2.1 Standard FH method

Let y_d be the direct estimator of true population parameter Y_d in domain d $(1,2,\ldots,D)$ and independent sampling error e_d with $E(e_d|y_d) = 0$ and $var(e_d|y_d) = \sigma_e^2$ in area d. Let x_d are the p-vector of area-level auxiliary variables that are related to the population parameter. The FH model suggested two-stage model is described as

$$y_d = Y_d + e_d \qquad \text{and} \qquad Y_d = x_d^T\beta + u_d \qquad (1)$$

[7] For detailed explanation see D. Morales et al., 2021.

The first stage expresses the sampling variability of survey estimates y_d and the second stage accounts for linking the direct estimates Y_d with fixed effect paramters x_d , regression coefficients β and random effects u_d which captures the variability not explained by auxiliary variables. The random errors are independently-identically distributed (iid) with the $E(u_d) = 0$ and $var(u_d) = \sigma_u^2$. By combining the sampling and linking models FH model can be expressed in the form of linear mixed model

$$y_d = x_d^T \beta + u_d + e_d \qquad (2)$$

The model assumes both error terms are independent of each other within and across domains.[8] Generally, sampling variances are known and random error variances are unknown and must be estimated from the data. The methods for estimating σ_u^2 include maximum likelihood (ML) and restricted maximum likelihood (REML) methods under the normality assumption. We have used REML method to obtain model parameters. For out-of sampled districts, income estimates are computed with a synthetic estimator as given below

$$y_d = x_d^T \beta \qquad (3)$$

2.2 HB model for SAE

The HB model assigns a prior distribution to all unknown parameters and hyperparameters and then makes the inference based on the posterior distribution. It offers an advantage for handling the different types of target variables (i.e., continuous, categorical) using the Gibbs sampling method (V. Gomez-Rubio´ et al., 2010). The equation is expressed as a sampling and linking model given below

$$y_d | Y_d \sim N(Y_d, \sigma_e^2) \qquad (4)$$

$$Y_d | \beta, \sigma_u^2 \sim N(x_d^T \beta, \sigma_u^2) \qquad (5)$$

The choice of prior distribution is a crucial step in the Bayesian approach, as further inference depends on it. In the literature, usually the prior of β is set to be $N(0, 10^6)$, where the variance of the coefficient is taken very large, reflecting the uncertainty or lack of prior knowledge of the

[8] For a detailed explanation, follow Chandra et al. (2013), Srivastava and Chandra (2021), Sud et al. (2012), Halbmeier et al. (2019).

distribution. Here σ_u^2 follows the inverse gamma distribution (i.e., $IG(a_0, b_0)$) and is taken to be very small, which reflects the vague knowledge about the variance of random effect. In this analysis, we choose a prior for β is $N(0, 10^6)$ and $IG(0.001, 0.001)$ to generate a target variable's mean and standard deviation. The model uses the Monte Carlo Markov Chain (MCMC) simulation technique to deal with the high computational complexities of the HB method (You and Rao, 2002; Anjoy et al., 2019; Anjoy & Chandra, 2019; Anjoy & Chandra, 2021).

3. Results and Analysis

This section presents the results obtained from SAE modelling using two approaches and provides a detailed discussion on the reliability of the estimates and the variation in income patterns across the state from 2004 to 2011. The modelling process begins with the selection of appropriate covariates, which is a crucial step in the SAE technique, as the inclusion of weak or irrelevant variables can lead to biased estimates. While selecting auxiliary variables, it is important to assess both their statistical and practical significance to ensure that the model remains parsimonious (Asian Development Bank, 2020).

Various techniques have been proposed in the literature for variable selection, including stepwise regression, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Principal Component Analysis (PCA), Least Absolute Shrinkage and Selection Operator (LASSO), decision trees and random forests, cross-validation methods, and expert judgment (Ren et al., 2022). In this study, we employed the LASSO technique, which effectively addresses issues of overfitting and multicollinearity among predictors (Masaki et al., 2022; Edochie et al., 2025). By adding a penalty term to the regression model, LASSO shrinks the coefficients of non-influential variables to zero, thereby retaining only the most relevant variables and enhancing model interpretability (Lykou & Ntzoufras, 2013).

Compared to traditional methods such as stepwise regression, which rely on forward or backwards selection, LASSO offers a more stable and computationally efficient approach, particularly in high-dimensional settings. Based on a review of the literature and relevance to income estimation, we initially considered a wide range of variables, including demographic characteristics (e.g., age structure, sex ratio, population size, household size), literacy rate, economic indicators (e.g., social group composition, workforce participation of main and marginal workers), and access to household amenities (e.g., television, two-wheelers, cars, phones, banking services).

After applying the LASSO technique, three variables emerged as the most influential in the model: the percentage of Scheduled Tribe population (ST_pop), access to banking services (Bank_access), and mobile phone accessibility (Phone). Given the large tribal population in the state, inclusion of ST_pop is contextually relevant. Moreover, financial and mobile accessibility are significant determinants of income, particularly in rural areas. Therefore, these variables were retained in the final model. The regression results, including significance levels and the model's R-squared value, are reported in Table 1.

Table 1: Regression Analysis for the target variable in 2004 and 2011

Parameters	**2004**	**2011**
ST_pop	-161.2049** (67.47322)	-562.1734** (202.2763)
Bank_access	709.4341** (263.9237)	1246.974*** (295.9558)
Phone	-7064.702** (2446.676)	-39534.96** (13292.41)
Intercept	28638.05 (5587.465)	49123.66 (19351.99)
R-squared	0.60	0.71

Standard errors are reported in parentheses.

*,**,*** indicates significance at 10%, 5% and 1% level, respectively

The selected covariates were then incorporated to obtain income estimates using both the FH and HB models. In the case of small domains, data often follows non-linearity on the original scale; therefore, log-transformation of the dataset is recommended to make it assumption-neutral (Chandra et al., 2018). Accordingly, we have log-transformed the direct estimates (i.e., household average income) and used the log-linear model in this study. The income estimates of direct, FH and HB models, along with their coefficients of variation (CV), are presented in Tables 2 and 3. The CV estimates suggest that the HB model yields lower variability and thus provides more precise estimates compared to the FH model. In 2004-05, the FH model estimated average district-level income to range between Rs 20,817 and Rs 40,366, whereas the HB model estimates ranged from Rs 19,884 to Rs 40,039. Raigarh and Kabeerd. were found to have the lowest and highest income levels, respectively. By 2011-12, income estimates showed an upward trend. The FH estimates ranged from Rs 27,172 to Rs 1,14,103, while HB model estimates ranged between Rs 32,916 and Rs 1,20,412. Korba and Kabeerd. were identified as the lowest and highest income districts, respectively. This increase in income levels from 2004 to 2011 also highlights the presence of income disparities across districts within the state.

It is important to note that one district (DBD) remained non-sampled in the IHDS-I survey, and three districts - Narayanpur, DBD, and Bijapur in IHDS-II. Nevertheless, the income estimates for these non-sampled districts were generated using the same covariates in both the FH and HB models, demonstrating the strength of SAE methods in generating estimates when sample size is zero.

Districts	sample size	Direct estimates		FH estimates		HB estimates	
		Income (Rs)	CV (%)	Income (Rs)	CV (%)	Income (Rs)	CV (%)
Koriya	57	28351	13.11	31707	10.38	31607	4.35
Surguja	71	26565	11.61	27143	9.05	25858	3.84
Jashpur	30	35794	12.14	28673	9.40	23951	4.63
Raigarh	77	21463	8.89	20817	8.47	19884	4.14
Korba	20	20729	9.82	21980	8.21	21831	4.07
Janjgir Champ.	35	28056	5.82	29141	5.90	31548	3.84
Bilaspur	58	30888	13.77	31104	9.23	30943	3.60
Kabeerd.	49	47847	16.48	40366	10.92	40039	4.57
Rajnand.	126	28228	6.88	28529	6.78	29539	4.02
Durg	148	26848	6.98	26692	6.84	27273	3.91
Raipur	55	35004	11.51	32491	8.81	32555	4.07
Mahasam.	20	23525	15.77	25752	10.80	25097	3.61
Dhamtari	51	29872	10.41	29569	8.30	29028	3.55
UBK	38	25378	15.34	25792	10.12	25334	4.05
Bastar	69	20137	9.60	21236	8.50	20549	4.13
DBD	0	-	-	22958	12.93	21344	3.51

Table 2: Direct, FH and HB income estimates and their coefficient of variation in 2004

Districts	sample size	Direct estimates	FH estimates	HB estimates

		Income (Rs)	CV (%)	Income (Rs)	CV(%)	Income (Rs)	CV(%)
Koriya	45	49434	18.98	60649	16.23	77467	9.99
Surguja	92	81096	41.50	60895	18.74	66281	8.55
Jashpur	34	47330	23.92	42657	16.17	43941	7.84
Raigarh	90	34380	13.09	33924	12.19	32916	7.83
Korba	22	22233	15.41	27172	13.26	33697	9.48
Janjgir Champ.	39	45954	17.45	49255	15.37	49564	7.15
Bilaspur	73	48478	15.58	51083	13.38	53401	7.47
Kabeerd.	53	127054	27.28	114103	20.13	120412	7.75
Rajnand.	152	54039	6.17	54864	6.29	64641	8.91
Durg	165	53820	9.28	55125	9.04	59616	7.98
Raipur	43	67975	19.63	63149	14.92	61222	7.75
Mahasam.	20	84533	23.45	76481	16.44	78044	7.86
Dhamtari	59	67114	18.54	56581	14.75	52881	9.13
UBK	42	50419	19.68	43855	14.88	43359	8.46
Bastar	77	37993	8.73	37400	8.74	37776	7.92
Narayanpur	0	-	-	25687	23.92	26169	7.56
DBD	0	-	-	23135	23.95	23700	7.38
Bijapur	0	-	-	22188	24.23	22453	7.53

Table 3: Direct, FH and HB income estimates and their coefficient of variation in 2011

3.1 Coefficient of Variation

The validity and reliability of the FH and HB estimates are supported by their respective coefficients of variation (CVs). The CV measures relative error and serves as an indicator of the precision of estimates, especially when comparing direct and SAE methods (Baffour et al., 2019). Although there is no universally accepted threshold for CV, lower values are generally preferred. The UK Office for National Statistics considers estimates with a CV below 20% to be acceptable, while the National Centre for Health Statistics (USA) accepts estimates with a CV under 30% for national health statistics (Srivastava et al., 2021). A comparison of the CVs across all three estimators clearly indicates that direct estimates exhibit higher variability than both FH and HB estimates, as reported in Table 2. In 2004-05, the CVs of direct estimates for sampled districts ranged from 5.82% to 16.47%, with an average of 11.20%. In contrast, FH estimates had a narrower range of 5.90% to 10.92%, averaging 8.78%. The HB estimates demonstrated the least variability, with CVs ranging from 3.54% to 4.63% across all districts. In 2011-12, the

CVs of direct estimates were higher and more dispersed, ranging from 6.17% to 41.49%. FH estimates in the same year ranged from 6.29% to 20.12%, with an average of 14.03%. HB estimates again showed greater precision, with CVs between 7.14% and 9.99%. These low CV values for the HB model suggest that the income estimates are more stable and closely distributed compared to direct estimates. For non-sampled districts, both FH and HB models provided low CVs, indicating reliable predictions. Figures 2(a) and 2(b) visually illustrate the CVs for direct, FH, and HB estimates in 2004 and 2011, respectively. The graphs confirm that the HB model consistently produces estimates with the lowest CVs, followed by FH, and then the direct method.

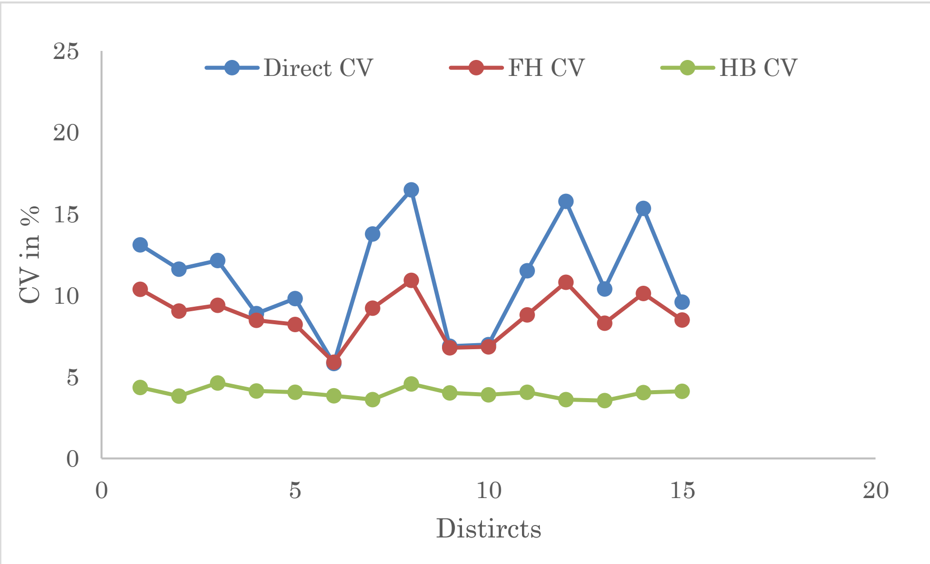


Fig 2(a): CV plot of direct, FH and HB estimates in 2004

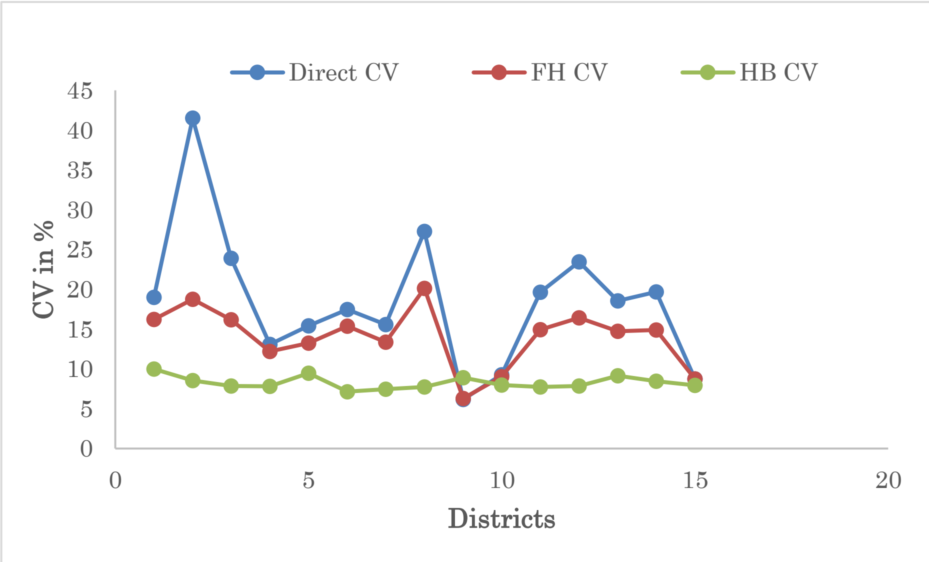


Fig 2(b): CV plot of direct, FH and HB estimates in 2011

3.2 Diagnostic Test

Following Guha and Chandra (2022a), we performed bias diagnostic tests to assess the validity of the SAE estimates. These tests suggest that the regression of direct estimates on the population values should align with the 45-degree line, as direct estimates are assumed to represent the actual population values. If the SAE estimates closely approximate the population values, the regression of direct estimates on SAE estimates is also expected to follow a similar linear relationship. The regression of direct estimates on HB estimates is shown in Figure 3. The graph indicates that the SAE estimates are consistent with the direct estimates, as the regression line closely overlaps the linearity line.

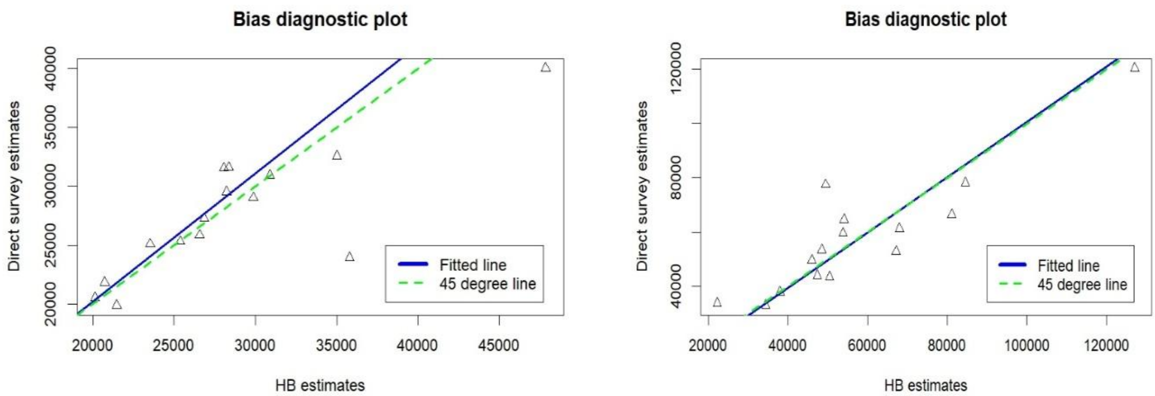


Fig 3: Bias diagnostic Plot of direct and HB estimates in 2004 (left) and 20011(right)

4. Income Distribution and Inequality

The income distribution across the state becomes more apparent through the choropleth maps presented in Figure 4. These maps highlight a significant shift in income levels over time. In 2004, the estimated income range varied between Rs 19,000 and Rs 40,000. By 2011, this range expanded considerably, from Rs 20,000 to Rs 1,30,000. Although the lower end of the income spectrum remained nearly unchanged, particularly in the southern districts (Bastar and DBD), which latter are subdivided into additional districts (Bastar, Narayanpur, Bijapur, DBD). These divisions were intended to facilitate targeted governance and development interventions. However, the persistent low-income levels in these newly formed districts reflect a lack of significant improvement, suggesting the limited effectiveness of government policies and local governance in addressing regional disparities.

This observed spatial disparity is consistent with findings from previous literature. For instance, Mohanty et al. (2016) analysed intra- and inter-state disparities in India by creating an index and ranking districts between 2009 and 2012. Their study ranked Jashpur, Koriya, and Bilaspur among the lowest in terms of poverty, while Bastar, Narayanpur, and Bijapur were among the highest, which aligns with our findings. These high-poverty districts continue to face challenges such as inadequate infrastructure, poor education and health services, limited connectivity to developed regions, and the persistent presence of Naxalite activity, all of which contribute to their underdevelopment (Roy, 2017). The concentration of forests and low literacy rates in these regions are also cited as major factors behind the continued Naxalite influence (Sikdar and Singh, 2019). In contrast, the northern and central districts, such as Surguja, Koriya, Jashpur, Korba, Raigarh, Janjgir Champ., Bilaspur, Raipur, Mahasam., Durg, Rajnand., and Dhamtari, fall into the middle to high-income categories in both time periods. Notably, Kabeerd. has shown the highest income level among all districts.

Given the significant tribal population in the state, caste-based inequality has also contributed to the socio-economic backwardness observed in certain regions. Despite the various government welfare schemes, marginalised communities continue to experience substantial disparities in education, health, infrastructure, and income, largely due to structural barriers and inefficiencies in policy implementation (Sonwari, 2024). Therefore, it is essential to address such gaps for effectively targeting the most deprived groups and ensuring inclusive development across districts.

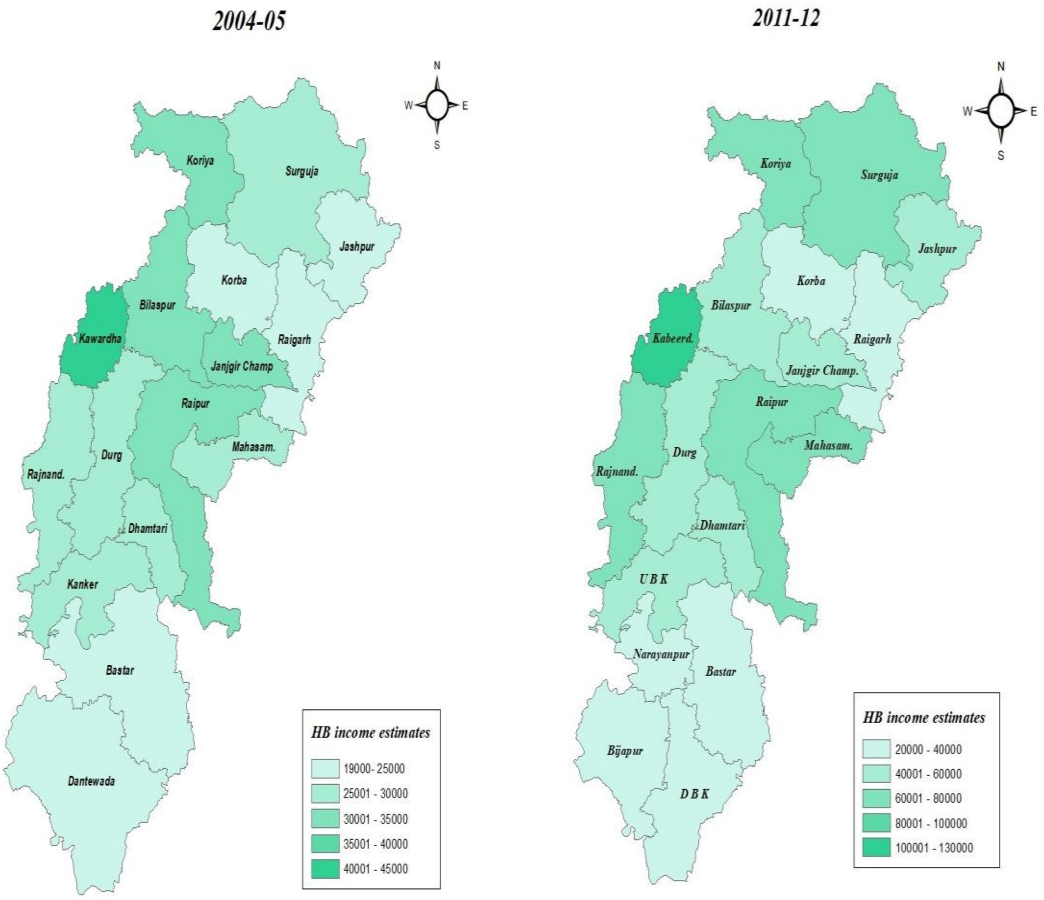


Fig 4: Choropleth map of income distribution in the state in 2004 (Left) and 2011 (right)

5. Conclusion and Discussion

The recent Poverty and Equity Brief: April 2025 highlights that the extreme poverty rate and consumption inequality in India declined between 2011-12 and 2022-23. On the other hand, several studies have shown that the income and wealth share of the top 1% and 10% has continued to rise after liberalisation (Chancel and Piketty, 2019; Somanchi, 2023; Bharti et al., 2024). At the state and district levels, the situation is often more concerning, while some areas have developed over time, others still struggle to attain even the minimum subsistence level. Due to the lack of reliable income data and discrepancies in existing datasets, many studies have combined survey and administrative sources to capture a more accurate picture of income distribution. However, such approaches are not directly applicable at the district level due to the absence of a dataset at the disaggregated level. Moreover, survey data alone are not adequate because of high sampling variability at small geographic levels. In this context, SAE technique can be used to link survey and census data to produce reliable estimates for such levels through mixed models. In this study, we employ two area-level models - FH and HB to assess their applicability in the real-world context of a developing country like India, where data scarcity is a major challenge. We generated income estimates for rural districts of Chhattisgarh and compared the survey based, FH, and HB estimates.

To assess the reliability of these estimates, we computed the CV for all three approaches. The results highlight the superior performance of the HB estimates, which exhibited a lower range of CV values, indicating lower standard errors and greater reliability. Further, the bias diagnostic test validated the HB model closely followed the survey estimates while effectively shrinking extreme values towards the mean. Choropleth maps of income distribution revealed that low-income districts are concentrated in the southern and a few northern parts of the state, which are highly vulnerable due to inadequate access to health, education, and infrastructure, and further affected by Naxalism. In contrast, high-income districts are mainly located in the northern region, including Koriya, Korba, Raigarh, and Raipur. These findings clearly indicate persistent intra-state disparities.

Based on the analysis, the following recommendations are proposed:

- This study underscores the importance of SAE methodology as a reliable and cost-effective solution for generating precise estimates in contexts of small sample sizes. The Bayesian approach, by incorporating posterior distributions, further improves precision and reduces sampling variability. Such income estimates enable the government to frame and monitor area-specific policies to address uneven development within the state.
- It also helps the government in better resource allocation so that they can be effectively used in schemes which are specifically meant for poverty and disparity reduction, such as Mahatma Gandhi National Rural Employment Guarantee Scheme (MGNREGA), Deen Dayal Antyodaya Yojana (DAY) for rural and urban livelihood, Vibrant Village Programme, Umbrella Programme for Development of Minorities.[9]
- The uneven development is largely due to poor access to basic facilities like education, health, and infrastructure in southern districts like Bastar, Narayanpur, and DBD, which requires a strategic plan. Schemes such as the mid-day meal programme, maternal benefit schemes, MGNREGA, and state-level programmes like Mahila Samakhya and Mukhya Mantri Swavalamban for women's empowerment should be intensified in these regions (Oxford Poverty and Human Development Initiative, 2020). The government should also incentivise private investors to direct resources towards these backwards regions to stimulate economic growth (Roy, 2017).
- Persistent policy ineffectiveness remains a major concern, requiring stronger interventions, comprehensive approaches, stricter implementation, and more targeted programmes to address structural barriers (Sonwani, 2024). Regular monitoring of backwards regions, such as through the Aspirational District/Block Programmes, can help the state government to

[9] SDG India Index 2023-24. https://www.niti.gov.in/sites/default/files/2024-07/SDG_India_Index_2023-24.pdf

track progress on various indicators and implement area-specific programs to uplift the backwards regions.

Finally, to conclude, the continuous monitoring of SDGs by NITI Aayog shows that Chhattisgarh has a low index score in the case of Poverty (40), Industry, Innovation and Infrastructure (39), and Climate Action (47), all below the national average, requiring urgent attention. However, the state has made progress in recent years, scoring above 65 in nine other SDG goals and slowly improving the state's performance. According to the Fiscal Health Index (2025) report, the state has also improved in demographic indicators, public expenditure, revenue, and fiscal stability. This shows its potential that can accelerate growth with appropriate government intervention and governance. Moreover, the application of SAE can be extended to estimate income distribution in other districts, enabling large-scale comparability. The study has some limitations that future research can address. First, although the IHDS dataset provides valuable information, incorporating temporal models could further improve estimates. Second, spatial correlatedness is also an important factor that can be incorporated to take into account the information from the clustering of similar types of areas.

References:

Anjoy, P., & Chandra, H. (2019). Hierarchical Bayes aggregated level spatial model for crop yield estimation. http://isas.org.in/jsp/volume/vol73/issue2/9-Priyanka.pdf

Anjoy, P., & Chandra, H. (2020). Disaggregate-level disparity in the incidence of poverty in Chhattisgarh, India. *Agricultural Economics Research Review*, *33*(1). DOI: 10.22004/ag.econ.304153

Anjoy, P., & Chandra, H. (2023). Hierarchical Bayes estimation of small area means under a spatial nonstationary Fay–Herriot model. *Communications in Statistics-Simulation and Computation*, *52*(7), 3043-3061. https://doi.org/10.1080/03610918.2021.1926501

Banerjee, A., & Piketty, T. (2005). Top indian incomes, 1922–2000. *The World Bank Economic Review*, *19*(1), 1-20. https://doi.org/10.1093/wber/lhi001

Battese, G. E., Harter, R. M., & Fuller, W. A. (1988). An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, *83*(401), 28–36. https://doi.org/10.1080/01621459.1988.10478561

Bharti, N. K., Chancel, L., Piketty, T., & Somanchi, A. (2024). Income and Wealth Inequality in India, 1922-2023: The Rise of the Billionaire Raj. https://shs.hal.science/halshs-04563836

Chancel, L., & Piketty, T. (2019). Indian income inequality, 1922-2015: from british raj to billionaire raj?. *Review of Income and Wealth*, *65*, S33-S62. https://doi.org/10.1111/roiw.12439

Chandra, H., & Verma, B. (2022). Small area estimation of food insecurity prevalence for the state of uttar pradesh in India. *Model Assisted Statistics and Applications*, *17*(2), 73-85. https://doi.org/10.3233/MAS-220011

Chandra, H., Aditya, K., & Sud, U. C. (2018). Localised estimates and spatial mapping of poverty incidence in the state of Bihar in India—An application of small area estimation techniques. *Plos one*, *13*(6), e0198502. https://doi.org/10.1371/journal.pone.0198502

Chandra, H., Sud, U. C., & Gupta, V. K. (2013). Small area estimation under area level model using R software. *New Delhi: Indian Agricultural Statistics Research Institute*.

Chandra, H., Sud, U. C., & Salvati, N. (2011). Estimation of district level poor households in the state of Uttar Pradesh in India by combining NSSO Survey and Census Data. *Journal of the Indian Society of Agricultural Statistics*, *65*(1), 1-8.

Chandrasekhar, S., Naraparaju, K., & Sharma, A. (2021). *Spatial Disparities in Household Earnings in India: Role Of urbanization, Sectoral Inequalities, and Rural-urban Differences*. Indira Gandhi Institute of Development Research.

Chhattisgarh poverty, growth, and inequality (English). India state briefs Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/166551468194958356

Desiyanti, A., Toharudin, T., & Suparman, Y. (2022). The implementation of empirical best linear unbiased prediction-Fay Herriot (EBLUP-FH) on the estimation of average per capita expenditure at district level in West Sumatra Province in 2019. *J. Math. Comput. Sci.*, *12*, Article-ID. DOI: 10.28919/jmcs/7233

Edochie, I., Newhouse, D., Tzavidis, N., Schmid, T., Foster, E., Hernandez, A.L., Ouedraogo, A., Sanoh, A. and Savadogo, A. (2025). Small area estimation of poverty in four west African countries by integrating survey and geospatial data. *Journal of Official Statistics*, *41*(1), 96-124. https://doi.org/10.1177/0282423X241284890

Fay, R.E., and R. Herriot. 1979. "Estimates of income for small places: an application of James stein procedures to census data". Journal of the American Statistical Association 74: 269–277.DOI: https://doi.org/10.2307/2286322.

Fiscal Health Index 2025. https://www.niti.gov.in/sites/default/files/2025-01/Fiscal_Health_Index_24012025_Final.pdf

Gómez-Rubio, V., Best, N., Richardson, S., Li, G., & Clarke, P. (2010). Bayesian statistics small area estimation. https://eprints.ncrm.ac.uk/id/eprint/1686

Guha, S., & Chandra, H. (2021). Measuring and mapping disaggregate level disparities in food consumption and nutritional status via multivariate small area modelling. *Social Indicators Research*, *154*(2), 623-646. https://doi.org/10.1007/s11205-020-02573-8

Guha, S., & Chandra, H. (2022a). Multivariate small area modelling for measuring micro level earning inequality: Evidence from Periodic Labour Force Survey of India. *Social Indicators Research*, *162*(2), 643-663. https://doi.org/10.1007/s11205-021-02857-7

Guha, S., & Chandra, H. (2022b). Measuring and mapping micro level earning inequality towards addressing the sustainable development goals–a multivariate small area modelling approach. *Journal of Official Statistics*, *38*(3), 823-845. DOI: 10.2478/JOS-2022-0036

Halbmeier, C., Kreutzmann, A. K., Schmid, T., & Schröder, C. (2019). The fayherriot command for estimating small-area indicators. *The Stata Journal*, *19*(3), 626-644. https://doi.org/10.1177/1536867X19874238

Hukum Chandra (2013) Exploring spatial dependence in area-level random effect model for disaggregate-level crop yield estimation, Journal of Applied Statistics, 40:4, 823-842, DOI: 10.1080/02664763.2012.756858

India Poverty and Equity Brief : April 2025 (English). Poverty and Equity Brief Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/099722104222534584

Introduction to Small Area Estimation Techniques (2020). Asian Development Bank. http://dx.doi.org/10.22617/TIM200160-2

Lykou, A., & Ntzoufras, I. (2013). On Bayesian lasso variable selection and the specification of the shrinkage parameter. *Statistics and Computing*, *23*, 361-390. DOI 10.1007/s11222-012-9316-x

Masaki, T., Newhouse, D., Silwal, A. R., Bedada, A., & Engstrom, R. (2022). Small area estimation of non-monetary poverty with geospatial data. *Statistical Journal of the IAOS*, *38*(3), 1035-1051. https://doi.org/10.3233/SJI-210902

Mohanty, S. K., Govil, D., Chauhan, R. K., Kim, R., & Subramanian, S. V. (2016). Estimates of poverty and inequality in the districts of India, 2011–2012. *Journal of Development Policy and Practice*, *1*(2), 142-202. DOI: 10.1177/2455133316642338

Morales, D., Esteban, M. D., Pérez, A., & Hobza, T. (2021). A course on small area estimation and mixed models. *Methods, theory and applications in R*.

Porter, A. T., Holan, S. H., Wikle, C. K., & Cressie, N. (2014). Spatial Fay–Herriot models for small area estimation with functional covariates. *Spatial Statistics*, *10*, 27-42. https://doi.org/10.1016/j.spasta.2014.07.001

Poverty, O., & Human Development Initiative. (2020). Multidimensional poverty in Chhattisgarh: a measure for action. https://ophi.org.uk/multidimensional-poverty-in-chhattisgarh-a-measure-for-action/

Pramanik, S., Muthusamy, N., Gera, R., & Laxminarayan, R. (2015). Vaccination coverage in India: a small area estimation approach. *Vaccine*, *33*(14), 1731-1738. https://doi.org/10.1016/j.vaccine.2015.01.083

Priyanka Anjoy & Hukum Chandra (2021): Hierarchical Bayes estimation of small area means under a spatial nonstationary Fay-Herriot model. *Communications in Statistics Simulation and Computation*. DOI: 10.1080/03610918.2021.1926501

Ren, W., Li, J., Erciulescu, A., Krenzke, T., & Mohadjer, L. (2022). A variable selection method for small area estimation modeling of the proficiency of adult competency. *Stats, 5*(3), 689. doi:https://doi.org/10.3390/stats5030041

Roy, E. (2017). Regional imbalance in chhattisgarh's economy: An intra-state analysis. *Asian Journal of Research in Business Economics and Management*, *7*(9), 67-74. DOI: 10.5958/2249-7307.2017.00162.1

SDG India Index 2023-24. TOWARDS VIKSIT BHARAT SUSTAINABLE PROGRESS, INCLUSIVE GROWTH.

Sikdar, A. K., & Singh, V. (2019). Causes of Naxalism and its impact on Tourism in Bastar region of Chhattisgarh. *Research Journal of Humanities and Social Sciences*, *10*(4), 1135-1143. DOI: 10.5958/2321-5828.2019.00187.6

Somanchi, A. (2023). *Income Inequality in India, 2000-2020* (Doctoral dissertation, MSc thesis, Paris School of Economics).

Sonwani, F. K. (2024). A Study of Caste-Based Social and Economic Inequality in Chhattisgarh State. *Research Journal of Humanities and Social Sciences*, *15*(3), 180-182. DOI: 10.52711/2321-5828.2024.00027

Srivastava, A. K. (2007). Small area estimation–A perspective and some applications. *J. Ind. Soc. Agril. Statist*, *61*(3), 295-309.

Srivastava, A. K., & Chandra, H. (2021). Small Area Estimation–Some Applications in NSSO Surveys. *Statistics and Applications* {ISSN 2454-7395 (online)} Volume 19, No. 1, 2021 (New Series) 149-172. http://krishi.icar.gov.in/jspui/handle/123456789/46425

Srivastava, A. K., & Chandra, H. Small Area Estimation–Some Applications in NSSO Surveys. *Statistics and Applications*, Volume 19, No. 1, 2021 (New Series), 149-172

Subramanian, S., & Jayaraj, D. (2013). The evolution of consumption and wealth inequality in India: A quantitative assessment. *Journal of Globalization and Development*, *4*(2), 253-281. DOI: 10.1515/jgd-2013-0018

Sud, U. C., Hukum Chandra, H. C., & Srivastava, A. K. (2012). Crop yield estimation at district level using improvement of crop statistics scheme data-an application of small area estimation technique. *Journal of the Indian Society of Agricultural Statistics*. 66(2) 2012 321-326. http://www.isas.org.in/jisas

Taunk, A., & Nimbalkar, V. (2021). ECONOMIC GROWTH THROUGH SUSTAINABLE DEVELOPMENT GOAL-8 DECENT WORK AND ECONOMIC GROWTH IN INDIA. *NeuroQuantology*, *19*(12), 604.

Tyagi, A. (2023). Income Inequality In Indian States. *Economic and Political Weekly*, 52(12).

You, Y., & Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, *32*(1), 97.

You, Y., & Rao, J. N. K. (2002). Small area estimation using unmatched sampling and linking models. *Canadian Journal of Statistics*, *30*(1), 3-15. https://doi.org/10.2307/3315862