Small Area Estimation using Data Integration from Two Survey with Different Data Strength

Dr. S. Islam[1], Dr. H. Chandra[2] & Dr. P. R. Ojasvi[3]

ICAR-Indian Institute of Soil and Water Conservation[1,3], Dehradun
ICAR-Indian Agricultural Statistics Research Institute[2], New Delhi
Email Id.:sadikul.islamiasri@gmail.com, hchandra12@gmail.com & projasvi@yahoo.com

Introduction

- Survey samples: Cost ↗ Non-response rate ↗
- More demand for
  - ✓ Faster statistics
  - ✓ Detailed information at Small area (or domain) level
- Availability of different survey data source on same ↗ population
  - Leverage advantages of each source
  - More and cheaper information

Data integration of different (two) survey data in small area estimation : Precision ↗ (Islam and Chandra, 2019)

Data Scenarios

Study variable: Y
Auxiliary variables: X, Z
Sample Size very small
Survey 1

Absent: Y & X
Available: Z & variable L linearly related to Z
Sample Size larger
Survey 2

Population

Available: L, X and Population aggregate: X & Z
Census/Other

Method used

Data integration model

Linear mixed model

$$y_{ij} = \begin{bmatrix} 1 & x_{ij} & \mathbf{z}_{ij} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_x \\ \beta_z \end{bmatrix} + u_i + e_{ij}$$

Survey 1
Y, X, Z

Model: Impute Y

Census/Others
L, X

Model: Impute X

Survey 2
Z, L

Estimators Developed:
EP(b)3 & EP(b)3

Performance Evaluation

Model-based & Design-based simulation

Key Results

Model-based Simulation

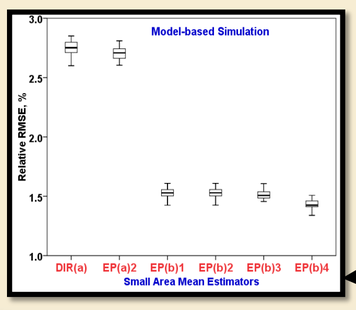
Area-specific sample size pair	Areas	PREDICTOR	Area effect=10% RB	Area effect=10% RRMSE	Area effect=10% RE	Area effect=20% RB	Area effect=20% RRMSE	Area effect=20% RE
Survey1=3 Survey2=60	Sampled	DIR(a)	0.001	2.55	100	0.001	2.55	100
		EP(a)2	0.008	2.50	102	0.009	2.53	101
		EP(b)1	0.001	1.29	198	0.001	1.63	157
		EP(b)2	0.000	1.29	198	0.001	1.63	157
		EP(b)3	0.019	1.23	207	0.019	1.52	168
		EP(b)4	0.010	1.17	218	0.012	1.47	174
	Non-sampled	M-EP(b)1	-0.007	1.35	-	-0.018	1.90	-
		M-EP(b)2	-0.007	1.35	-	-0.018	1.90	-
		M-EP(b)3	-0.024	1.31	-	-0.035	1.88	-
		M-EP(b)4	-0.015	1.26	-	-0.025	1.82	-
Survey1=3 Survey2=70	Sampled	DIR(a)	-0.002	2.55	100	-0.002	2.55	100
		EP(a)2	-0.009	2.49	103	-0.010	2.52	101
		EP(b)1	-0.004	1.30	197	-0.004	1.64	156
		EP(b)2	-0.005	1.30	197	-0.005	1.64	156
		EP(b)3	-0.011	1.24	206	-0.011	1.52	168
		EP(b)4	-0.007	1.18	217	-0.009	1.47	174
	Non-sampled	M-EP(b)1	0.023	1.31	-	0.036	1.86	-
		M-EP(b)2	0.023	1.31	-	0.036	1.86	-
		M-EP(b)3	0.027	1.29	-	0.040	1.84	-
		M-EP(b)4	0.025	1.22	-	0.037	1.76	-

Small Area Mean Estimators

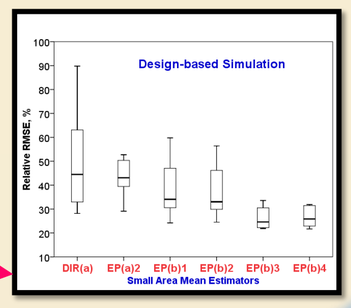
Predictor	Description	Data used
DIR(a)	Direct estimator	Survey1 data
EP(a)2	Chandra *et al.*(2015)	Survey1 data
EP(b)1	Islam &Chandra(2019)	Combined data
EP(b)2	Islam &Chandra(2019)	Combined data
EP(b)3	Developed estimator1	Combined data
EP(b)4	Developed estimator2	Combined data
M-EP(b)3	Modified EP(b)3	Non-sampled area
M-EP(b)4	Modified EP(b)4	Non-sampled area

Design-based Simulation

Predictor	Survey1: n_1=60 Survey2: n_2=759 RB	RRMSE	RE
DIR(a)	1.36	57.35	100
EP(a)2	0.89	56.28	102
EP(b)1	7.41	40.43	142
EP(b)2	6.65	39.22	146
EP(b)3	-0.44	31.73	181
EP(b)4	0.22	29.54	194



Box-plot: Area-specific estimates



Brief Discussion

- EP(b)3 and EP(b)4 out perform EP(b)1 and EP(b)2 estimator in terms of RB, RRMSE and RE
- Use of additional auxiliary variable of Survey 1 along with common auxiliary variables in data integration achieved significant gains in efficiency than utilising only common auxiliary variables
- M-EP(b)3 and M-EP(b)3 outperform M-EP(b)1 and M-EP(b)2 for non-sampled areas
- The performance of the developed estimators are consistent over change in sample size as well as area effect value

Conclusion

- The EP(b)3 and EP(b)4 found to be outperformer
- EP(b)4 slightly outperformer than EP(b)3 due to utilize of population level auxiliary variable
- EP(b)4 is proposed if population level auxiliary variables are available, otherwise EP(b)3

References

Chandra, H., Sud, U.C. and Gharde, Y., 2015. Small area estimation using estimated population level auxiliary data, Communications in Statistics-Simulation and Computation, 44(5), 1197-1209.

Islam, S. and Chandra, H. (2019). Small area estimation combining data from two independent surveys, *Communications in Statistics-Simulation and computation*. doi.org/10.1080/03610918.2019.1588308

Islam, S. (2018). PAT report entitled "Integration of Survey Data for Small Domain Inference" submitted to ICAR-IASRI, New Delhi & NAARM, Hyderabad.

Acknowledgement

Director ICAR-Indian Institute of Soil & Water Conservation, Dehradun

ICAS VIII 2019 – NEW DELHI