Journal of Biostatistics and Epidemiology

J Biostat Epidemiol. 2022;8(3):258-268

Original Article

Sampling Variance Estimation Method and Precision of Small Area Estimation in the Exponential Spatial Structure

Yadollah Mehrabi[1], Amir Kavousi[2], Mojtaba Soltani-Kermanshahi[3*]

[1]Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[2]Workplace Health Promotion Research Center and Department of Epidemiology, School of Public Health and Safety, Shahid Beheshti University of Medical Sciences, Tehran, Iran.
[3]Social Determinants of Health Research Center, School of Medicine, Semnan University of Medical Sciences, Semnan, Iran.

ARTICLE INFO

Received 27.02.2022
Revised 15.03.2022
Accepted 09.04.2022
Published 15.10.2022

Key words:
Spatiality;
Small area estimation;
Exponential structure;
Household food expenditure;
Sampling variance;
Simulation.

ABSTRACT

Introduction: In various practical applications, neighbouring small area data have spatial correlation. More recently, an extension of the Fay–Herriot model through the spatial (exponential) has been considered. This spatial area-level model like the fundamental area-level model (was first suggested by Fay III and Herriot) has a powerful assumption of known sampling variance. Several methods have been suggested for smoothing of sampling variance and there is no unique method for sampling variance estimation, more studies need.
Methods: This research examines four techniques for sampling variance estimates including of Direct, Probability Distribution, Bayes and Bootstrap methods. We used households' food expenditures (HFE) data 2013 and other socio-economic ancillary data to fit the read model and at last conduct a simulation study based on this data to compare the effects of four variance estimation methods on precision of small area estimates.
Results: The best model on real data showed that the lowest and the highest HFE belonged to Pishva district (in Tehran province) with 26,707 thousand rials (TRs) and Omidiyeh (in Khouzestan province) with 101,961 TRs, respectively. Accordingly on simulation study, the probability distribution and direct methods, respectively and approximately had the smallest and the highest Root Average Mean Square Errors (RAMSE) for all conditions.
Conclusion: The results showed the best fitting with direct method in real data and best precision with Probability Distribution method in simulation study.

*.Corresponding Author: msoltani@semums.ac.ir

Introduction

Sample surveys have long been used as a cost-effective alternative for data collection. Small area estimation (SAE) methods are gaining popularity in survey sampling due to the increasing demand for reliable estimates for small areas. If the domain-specific sample size is not large enough to provide a direct and reliable estimate, a domain or an area is referred to as "small".[1-5]

Direct estimation of the small area parameter (s) depends entirely on samples from that area, which is usually unreliable. However, much of the popularity and usefulness of SAE techniques can be attributed to model-based techniques that are now widely used and have been increasingly applied in the last three decades. These approaches draw strength from adjacent areas and use appropriate techniques that link direct estimates from small areas. We can easily use these models when there are area-level summary of covariate variables. Fay and Herriot (1979) proposed a fundamental model at area-level and others followed.[1-7]

In various practical applications, neighbouring small areas data have spatial correlation. In these cases, unless adequate ancillary variables are available, between-area correlations must be somehow represented in the covariance structure of the model.[1, 8-9]

Mehrabi et. al. (2020) proposed the exponential model in the Mat`ern function with:

$$\rho(d)=\exp(-d) \qquad \text{Eq. 1.1}$$

where d and ρ(.) are the distance between two areas and correlation function, respectively.[1,7, 10] This spatial area-level model like the fundamental area-level model (was first suggested by Fay III and Herriot[6] has a powerful assumption of known sampling variance (SV), Where SV is the variance of sampling errors in the Fay and Herriot model. Usually the SVs are estimated from the drawn sample by the model, which may not be reliable.[11]

Thus, with spatial area-level model, we need to estimate the sampling variance. Several methods have been suggested for smoothing of SV, each for an appropriate situation[3,6,12, 13, 14, 15] For example, Dick (1995) proposed to use generalized variance function (GVF), constructing a regression model based on direct estimates of SVs along with some external covariates.[16] Bell (2008) with some SV estimation methods concluded that most critical possible problems from use of direct SV estimates in small area models come from critical underestimation of the SV when it is large. He also emphasized that more investigation on this subject is required.[17]

An application of generalized design effects for smoothing sampling covariance matrices had been suggested by Singh et al., at 2005.[18] Also, the model of common design effects to smooth the SVs had been implicated by you at 2008.[19]

Soltani-Kermanshahi et al. at 2017 proposed the Bayesian technique for SV estimation and in a simulation study showed that the probability distribution (PD) method had approximately the best relative root mean square errors (RRMSE) in Fay III and Herriot model.[5]

In our best knowledge, there is not any study for sampling variance estimation method on the precision of spatial small area estimation and since there is no unique method for SV estimation, further study is required.

In the first part of this research, we applied spatial small area model with the exponential structure in the Mat`ern function with four methods of SV estimation to predict households' food expenditures (HFE) in urban areas of Iran. Data were collected by the Statistics Center of Iran in 2013. Although the sample size was optimal at the provincial level, it was not optimal at the district level.[20] In the secondary part, we performed a simulation study based on HFE data to compare the effects of four methods of variance estimation on the accuracy of small area estimates.

Methods

Suppose that we have M areas in the population and only m areas are sampled. The linear form of SAE model with spatial dependence is:

$$Y = X\beta + u + e \qquad \text{Eq. 2.1}$$

where Y is the vector of design-unbiased direct estimator available for each of the small areas (from 1 to m), X the vector of area-level auxiliary covariates, e the vector of independent sampling errors; and u the vector by using an exponential model (Eq. 1.1) in the Mat`ern function as follows:

$$u = (u_1,\dots,u_m) \sim MvN(0,V),\ e_i \sim (0,\sigma_i^2),\quad v_{ij} = \sigma_u^2 \rho(d_{ij}),\ i, j = 1, 2, \dots m \qquad \text{Eq. 2.2}$$

where V is a variance-covariance matrix with elements v_{ij} and correlation function ρ(.); an isotropic correlation function that decays as the Euclidean distance $d_{ij}=\|s_i-s_j\|$ between two individuals increases. In the exponential model ρ(.) has the form describe by Eq. 1.1. Also, u_i's and e_i's are assumed to be independent.

In this research, we estimated σ_u^2 and β's using the Restricted Maximum Likelihood (REML) and Spatial Empirical Best Linear Unbiased Prediction (SEBLUP), respectively.

For the estimation of sampling variance (σ_i^2's), we focused on four methods namely Direct, Probability Distribution, Bootstrap and Bayes methods.

In direct method, we estimated SV at area i by

$$S_i^2 = \frac{\sum(y_{ij}-\bar{y}_i)^2}{n_i-1} \qquad i = 1, \dots, m\ ;\ j = 1, \dots, n_i \qquad \text{Eq. 2.3}$$

Where n_i and $\bar{y}_i$ are the sample size and mean at area i, respectively.

In probability distribution (PD) method, with normality assumption of interested parameter, we estimated SV based on:

$$(n_i-1)s_i^2 \sim \sigma_i^2\, \chi^2_{(n_i-1)} \qquad i = 1, \dots, m \qquad \text{Eq. 2.4}$$

Where σ_i^2 is the SV at area i. In the Bootstrap method, we 5000 resampled data sets at each area and estimated the SV using direct method explained beforehand (Eq. 2.3). Finally, in the Bayes method, we assumed the distribution of HFE at area i to be $N(\bar{y}_i, \sigma_i^2)$ Where $\bar{y}_i$ assumed to be known but $1/\sigma_i^2$ as random with prior distribution

$$r_i = \frac{1}{\sigma_i^2} \sim \text{Gamma}(1, 1/S_i^2) \qquad \text{Eq. 2.5}$$

In which S_i^2 is the direct variance in area i.

Furthermore we calculated the posterior inverse variance as below.[5]

$$r_i|\underline{y} \sim Gamma\left(1+\frac{n_i}{2}, \left(S_i^2+\frac{1}{2}\sum\left(y_{ij}-\bar{y}_i\right)^2\right)^{-1}\right) \qquad \text{Eq. 2.6}$$

$$\xrightarrow{yields} \hat{\sigma}_i^2 = \frac{S_i^2+\frac{1}{2}\sum\left(y_{ij}-\bar{y}_i\right)^2}{1+\frac{n_i}{2}} \qquad \text{Eq. 2.7}$$

Real Data Analysis

The Iranian Rural and Urban Households' Expenditures and Income Survey (IRUHEIS), is carried out annually by Statistical Center of Iran (SCI). In this study we analysed urban data of IRUHEIS 2013 (collected between 21 April 2013 and 20 April 2014).[20]

A total of 387 districts were selected by IRUHEIS out of 429 districts of urban areas of Iran in 2013. Of the 18,876 households that participated in IRUHEIS 2013, we analyzed complete data consisting of 18,850 households. The HFE includes all payments made to purchase the essential food items and needed nutritious including dairy products, meat, bread and flour, cereal and bean, oil and butter, biscuits and cakes, nuts, fruits and vegetables, sweets and sugar, additives and dressings, as well as cigarette and tobacco.[20]

We considered district-level variables emanating from Iran's census of 2011that included the average number of households (ANH); the proportion of households headed by a male (PMH); the average number of rooms of each household; the proportion of the active population employed; the proportion of population of the following age groups: >65 years, 25-64 years, 15-24 years and <15 years; sex ratio and the proportion of higher education.[21] We also considered the gross domestic product (GDP); the distance from province capital; the proportion of households that had joined a charity organization; the migration rate and the per capita income for municipalities. This data was prepared by the Ministry of Interior in a project to identify less developed areas of the country. We also used geographic information such as latitude and longitude of the capital of each area to calculate exponential correlations.[21]

To achieve the highest correlation with dependent variables, the appropriate conversion of independent variables (power for continuous variables and logarithm for proportion variables) was applied. We also used variance inflation factors (VIFs) to assess the correlation between the independent variables, with VIF> 5, or coefficients of multiple determination with respect to other independent variables greater than 0.8, indicating serious multilinearity for the predictor.[22] The forward selection method was used to create the final model. In the forward method, independent variables (respectively based on their correlation with the dependent variable) are serially inserted in the model, with only significant variables remaining.

According to VIF values, the variable proportion of population at 25 to 64 years was omitted (Appendix s1).

Table 1 shows the results of the final models obtained using four types of sampling variance estimation methods.

Accordingly, only the variables including those of ANH, PMH, logarithm proportion of higher education (LHE) and proportion of population at 15 to 24 years (P15-24) had a significant effect on the HFE. It is shown that P15-24 had reversion effect on HFE and on the other hand by increase of one percent on proportion of male-headed households, the HFE increased about 0.5 million rials on average.

According to AIC, direct method (13619) had best fit among four methods while probability distribution method (13755) resulted in the poorest fit.

Figure 1 shows the zoning of the HFE in the urban areas of Iran by using spatial (exponential) small area model with probability distribution

Table1. Results of spatial (exponential) small area model with four types of sampling variance estimation on households Food Expenditures (1000 Rials)

	Direct	Probability Distribution	Bootstrap	Bayes
	Beta (SE[2])	Beta (SE)	Beta (SE)	Beta (SE)
ANH[3]	4.59×10^{3} (1.98×10^{3})*	8.85×10^{3} (2.38×10^{3})**	4.65×10^{3} (1.98×10^{3})*	4.63×10^{3} (1.99×10^{3})*
PMH[4]	5.45×10^{4} (9.36×10^{3})**	6.71×10^{4} (1.14×10^{4})**	5.43×10^{4} (9.36×10^{3})**	5.43×10^{4} (9.37×10^{3})**
LHE[5]	6.20×10^{3} (1.48×10^{3})**	$1.0^{4}\times10^{4}$ ($1.4^{5}\times10^{3}$)**	6.22×10^{3} (1.48×103)**	6.20×10^{3} (1.48×10^{3})**
P15-24[6]	-	-9.14×10^{4} (3.19×10^{4})*	-	-
σ_u^2	1.31×10^{8}	2.72×10^{8}	1.36×10^{8}	1.34×10^{8}
AIC[7]	13619	13755	13621	13620

*P-value<0.05 ,
**P-value<0.001
[2]Standard error
[3]Average number of households
[4]Proportion of male headed households
[5]Logarithm proportion of higher education
[6]Proportion of population at 15 to 24 years
[7]Akaike Information Criterion
[8]Probability distribution

method of SV estimation. The highest and the lowest HFE belonged to Omidiyeh (in Khouzestan province) with 101,961 Thousand Rials (TRs) and Pishva district (in Tehran province) with 26,707 TRs, respectively.

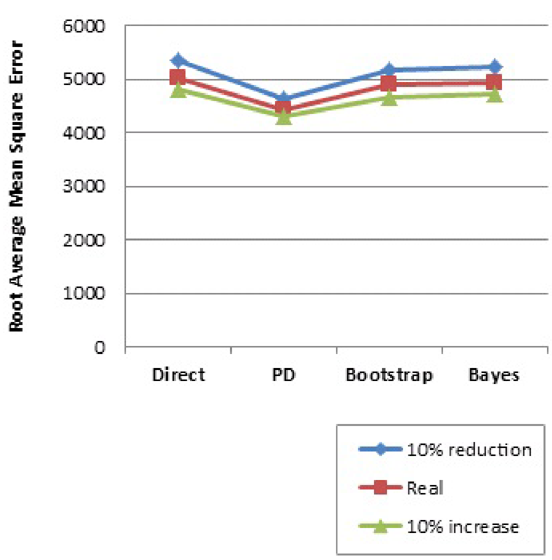

Figure 1. The zoning of Annual Household Food Expenditures at the district-level in Iran estimated with spatial (exponential) small area model and sampling variance estimation with Probability Distribution

Simulation with HFE Data

To investigate the effects of four methods of SV estimation on precision estimations, a simulation study was carried out. We used data based on HFE in Iran with 387 selected districts out of 429 districts of urban areas of Iran from IRUHEIS2013. Because only four auxiliary variables (ANH, PMH, LHE and P15-24) found significant in real data analysis, we only use them in our spatial model.

We also added latitude and longitude of the capital of each area to calculate exponential correlations (21). We generated data for each of the 387 areas by the following approach:

1. Multivariate normal distribution (from Eq. 2.1 and Eq. 2.2) with mean and co-variance of the correspondent real data was used, i.e. for the direct method of SV, σ_u^2 was 1.31×10^{8} and d_{ij} calculated from the following equation:

$$d_{ij} = \sqrt{(l_i - l_j)^2 + (t_i - t_j)^2} \qquad \text{Eq. 4.1}$$

Where l and t are the latitude and longitude of each pair areas i , j.

2. The sample size in each area was accurately imitated from real data to make the simulation as realistic as possible;
3. Data generation was repeated 1000 times in each of 378 small areas and mean and variance of each area were calculated;
4. To achieve small area estimations, we used the auxiliary variables included in the continues variables of ANH, PMH, LHE and P15-24 from the 2011 Iranian census, because they were significant in analysing real data.
5. We fitted the exponential structure spatial model to each of the 1000 datasets generated and stored the estimates for each area (1).
6. To assess the precision of SAEs, we employed Root Average Mean Square Errors (RAMSEs) for small area estimates:

$$RAMSE = \sqrt{\sum_{i=1}^{378} \mathrm{MSE_i} /378} \qquad \text{i. Eq. 4.2}$$

$$\mathrm{MSE_i} = \frac{1}{1000}\sum_{q=1}^{1000}(\hat{g}_{i,q} - \hat{g}_{i,M})^2 \qquad \text{Eq. 4.3}$$

Where $\hat{g}_{i,M}$ is the estimate of g_i based on real data with the four sampling variance estimation methods (M) and $\hat{g}_{i,q}$ shows the estimated g_i at run q (4-5).

For better assessment we used sample size sensitive analysis using the real sample as well as 10% above and below of it. All calculations used utilized the R package SAE (23) and the codes given in Appendix (S2 and S3).

Simulation results

The results of RAMSE with sample size sensitivity are shown in Table 2 and Figure 2. Accordingly, the probability distribution and direct methods, respectively and approximately had the smallest and the highest RAMSE for all conditions. Also, with increasing sample size, RAMSEs almost decreased.

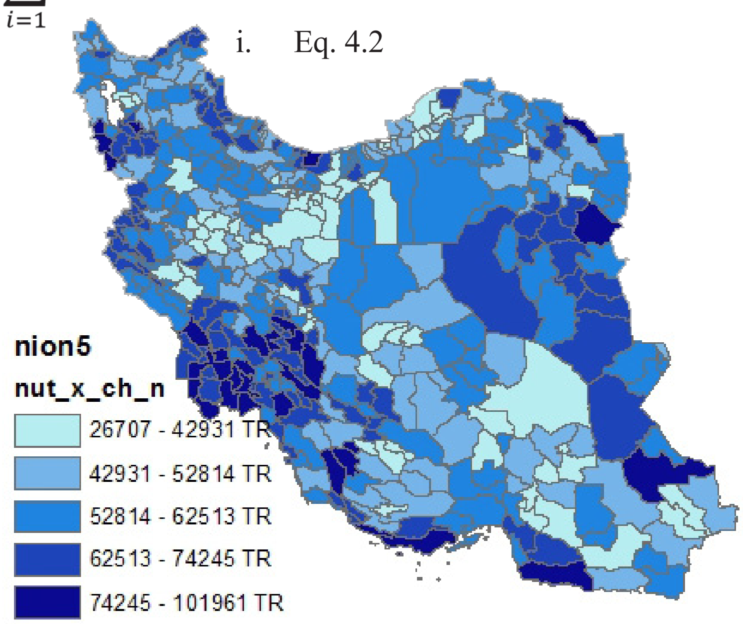

Figure 2. Root Average Mean Square Error of small area estimates with area sample size changes in four types of sampling variance estimation method in spatial (exponential) small area model by simulation[1] (PD, probability distribution)

Table 2. Root Average Mean Square Error of small area estimates with area sample size changes in four types of sampling variance estimation method in spatial (exponential) small area model by simulation

Method	Sample Size		
	10% increase	Real	10% reduction
Direct	5356	5030	4799
PD*	4623	4440	4301
Bootstrap	5186	4900	4667
Bayes	5247	4940	4715

*Probability distribution

Discussion

We demonstrated precision of spatial (exponential) SAE with various types of SVs estimation methods in the HFE urban data and a simulation based on real data.

From a purely statistical point of view, the position of the small area is related to the modelling of its parameters, and further improvement in the EBLUP estimator can be achieved by including possible spatial interaction between random area effects as previously discussed.[25, 26] Indeed, the inclusion of ancillary variables to obtain the spatial effects may be beneficial even when the strength of the spatial link is weak.[27]

Our study showed that four variables had a significant effect on HFE. The Probability Distribution method had bigger AIC and parameters standard errors than other three methods, approximately. Accordingly, the Probability Distribution method was not preferred but in simulation study, Probability Distribution method had approximately the best RAMSE.

Although with increasing sample size other methods such as Bayes and Boot strap had the better decrease in RAMSE.

Bell (2008) with some SV estimation methods concluded that most critical possible problems from use of direct SV estimates in small area models come from critical underestimation of the SV when it is large.[17]

In examining the use of administrative records in small area estimation, Erciulescu et al. Point out the challenging issue of SV estimation in the application of area level model. They used state-level GVF - GVF with random effects and simpler approaches that make assumptions about uniformity over larger aggregation levels to obtain "smooth" sampling variance estimates for small areas.[28] In a paper to expand on previous work on variable selection for area-level models, Cai S and Rao obtained sampling variances by smoothing out SV estimators, using the generalized variance function (GVF) method, and then they considered smooth estimators as sample variance.[29] Ciginas also proposed a new method for estimating the mean squared error (MSE) in the Fay Herriot model. He evaluated SVs from external sources or by smoothing out their direct estimates.[30]

Our findings approximately showed outclass of modelling relative to other smoothing methods and other studies had similar results.[11, 19, 17, 24, 5]

One reason of SAE method using is the small sample in subdomain. Due to the fact that we had less RAMSE in larger samples, so as the sample size increases, the accuracy increases.[5, 1]

On the other hand, when the sample size increasing, RAMSE of other methods decreased faster than Probability Distribution method, approximately. Roughly, RAMSE of Probability Distribution method have stable relative to sample size changing, too.

We did not find any works to compare the SV in spatial small area models but Soltani-Kermanshahi et al. at 2017 in a simulation study compared some SV methods and showed that the probability distribution (PD) method had approximately the best relative root mean square errors (RRMSE) in Fay III and Herriot model.[5]

In real data process, there were some areas without sample data. For such areas where the values of auxiliary variables are available at the area-level from any other data source, possible estimators are $\hat{Y}_i = X_i\ \beta_i$.[9] HFE zoning showed that border areas, especially those on the western border, had higher HFEs, approximately. The findings showed the convergence of HFE and suggest using spatial models. However, we expected to have the highest HFE in the districts of the capital province, but our findings showed the opposite. One of the reasons for this could be Iran's economic development.

The biggest difference between our study and similar studies was the inadequacy of households in each district, but the rate of weakness was reduced using the SAE method and we were finally able to find an acceptable model for predicting HFE in urban areas of Iran.

Furthermore, one of the basic limitations of the SAE was using the variables having high correlation with dependent variables. In this research, we have tried to find those variables and access them.

Conclusion

The results showed the best fitting with Direct method in real data and best precision with PD method in simulation study.

Acknowledgment

The authors are thankful to SCI, Statistical Research and Training Center and Mr. Yuosef Bashiri for providing data for this project. This research has the ethical code 1390-P-M.A.D-518 from Shahid Beheshti University of medical sciences, Tehran, Iran.

Authors' Contribution

Study concept and design: Dr. Mehrabi and Dr. Soltani-Kermanshahi, Acquisition of data: Dr. Soltani-Kermanshahi, Analysis and interpretation of data: Dr. Mehrabi and Dr. Soltani-Kermanshahi, Drafting of the manuscript: Dr.Soltani-Kermanshahi, Critical revision of the manuscript for important intellectual content: Dr. Mehrabi and Dr. Kavousi, Statistical analysis: Dr. Mehrabi, Dr. Kavousi and Dr. Soltani-Kermanshahi, Study supervision: Dr Mehrabi.

Conflict of interest

There is no conflict of interest.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

1. Mehrabi Y, Kavousi A, Baghestani AR, Soltani-Kermanshahi M. Spatiality in

small area estimation: A new structure with a simulation study. Geospatial Health. 2020 Dec 29;15(2).

2. Rao JN, Molina 2nd I. Small Area Estimation, A John Wiley & Sons. Inc, New Jersey. 2003.

3. Rao JN, Wu W. Sample Surveys: Inference and Analysis, vol. 29B of Handbook of statistics, chap. Empirical Likelihood Methods. 2009:189-208.

4. Soltani-Kermanshahi M, Mehrabi Y, Kavousi A, Baghestani AR, Mohammadi-Nasrabadi F, Aliakbari-Saba R. Households' food expenditures in urban areas of Iran: An application of small area estimation. Epidemiology, Biostatistics and Public Health. 2017 Dec 21;14(4).

5. Soltani-Kermanshahi M, Mehrabi Y, Kavousi A, Baghestani AR, Mohammadi-Nasrabadi F. EFFECTS OF SAMPLING VARIANCE ESTIMATION METHODS ON PRECISION OF SMALL AREA ESTIMATION. JP JOURNAL OF BIOSTATISTICS. 2017 Dec 1;14(2):93-106.

6. Fay III RE, Herriot RA. Estimates of income for small places: an application of James- Stein procedures to census data. Journal of the American Statistical Association. 1979;74(366a):269-77.

7. Waller LA, Gotway CA. Applied spatial statistics for public health data. John Wiley & Sons; 2004 Jul 29.

8. Prasad NN, Rao JN. The estimation of the mean squared error of small-area estimators. Journal of the American statistical association. 1990 Mar 1;85(409):163-71.

9. Molina I, Salvati N, Pratesi M. Bootstrap for estimating the MSE of the Spatial EBLUP. Computational Statistics. 2009 Aug 1;24(3):441-58.

10. Li Y, Tang H, Lin X. Spatial linear mixed models with covariate measurement errors. Statistica Sinica. 2009;19(3):1077.

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

12. Rivest LP, Vandal N. Mean squared error estimation for small areas when the small area variances are estimated. InProceedings of the International Conference on Recent Advances in Survey Sampling 2003. Laboratory for Research in Statistics and Probability, Carleton University Ottawa, ON, Canada.

13. Wang J, Fuller WA. The mean squared error of small area predictors constructed with estimated area variances. Journal of the American Statistical Association. 2003 Sep 1;98(463):716-23.

14. Lahiri P. On the impact of bootstrap in survey sampling and small-area estimation. Statistical Science. 2003 May 1:199-210.

15. Herrador M, Morales D, Esteban MD, Sánchez Á, Santamaría Arana L, Marhuenda Y, Pérez A. Sampling design variance estimation of small area estimators in the Spanish Labour

Force Survey. SORT. 2008;32(2):177-98.

16. Dick P. Modelling net undercoverage in the 1991 Canadian Census. Survey Methodology. 1995 Jun;21(1):45-54.

17. Bell WR. Examining sensitivity of small area inferences to uncertainty about sampling error variances. InProceedings of the American Statistical Association, Survey Research Methods Section 2008 Aug 4 (Vol. 327, p. 334).

18. Singh AC, Folsom Jr RE, Vaish AK. Small area modeling for survey data with smoothed error covariance structure via generalized design effects. InFederal Committee on Statistical Methods Conference Proceedings 2005.

19. You Y. Small area estimation using area level models with model checking and applications. InProceedings of the Survey Methods Section, Statistical Society of Canada 2008.

20. Statistical Center of Iran (SCI), 2013. [Results of Urban Households' Expenditures and Income Survey.] Available from: http://amar.org.ir/Portals/0/Files/fulltext/1392/n_ahdkhsh_92.pdf Accessed on: Jan 2021. [Website in Persian].

21. Statistical Center of Iran (SCI), 2011. [National Census of Population and Housing.] Available from: https://nnt.sci.org.ir/_LAYOUTS/15/PsaLogins/sciLogin.aspx?ReturnUrl=%2fsites%2fnnt%2f_layouts%2f15%2fAuthenticate.aspx%3fSource%3d%252Fsites%252Fnnt%252FSitePages%252Freport%255F90%252Fgozaresh%255Fostani%255Fshahrestani%252Easpx&Source=%2Fsites%2Fnnt%2FSitePages%2Freport%5F90%2Fgozaresh%5Fostani%5Fshahrestani%Easpx Accessed on: Jan 2021. [Website in Persian].

22. Iranian Noojum, 2018. [Longitude and latitude of the Iran's districts.] Available from: http://www.noojum.com/other/astronomy-tools/187-online-tools/6266-longitudelatitude.html Accessed on: Jan 2021. [Website in Persian].

23. Belsley DA, Kuh E, Welsch RE. Regression diagnostics: Identifying influential data and sources of collinearity. rdii. 1980.

24. Molina I, Marhuenda Y. sae: An R Package for Small Area Estimation. R J.. 2015 Jun 1;7(1):81.

25. Maples J, Bell W, Huang ET. Small area variance modeling with application to county poverty estimates from the American community survey. InProceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association 2009 (pp. 5056-5067).

26. Petrucci A, Salvati N, 2004. Small area estimation using spatial information. The rathbun lake watershed case study. Dipartimento di Statistica "G. Parenti" viale morgagani, pp.59-50134. Available at: https://www.academia.edu/1376417/Small_Area_Estimation_Using_Spatial_Information_The_Rathbun_Lake_Watershed_Case_Study [Accessed: September 1, 2020].

27. Pratesi M, Salvati N. Small area estimation in the presence of correlated random area effects. Journal of Official Statistics. 2009 Jan 1;25(1):37.

28. Petrucci A, Pratesi M, Salvati N. Geographic information in small area estimation: small area models and spatially correlated random area effects. Statistics in Transition. 2005;7(3):609-23.

29. Erciulescu AL, Franco C, Lahiri P. Use of administrative records in small area estimation. Administrative Records for Survey Methodology. 2021 Apr 6:231-67.

30. Cai S, Rao JN. Selection of Auxiliary Variables for Three-Fold Linking Models in Small Area Estimation: A Simple and Effective Method. Stats. 2022 Feb 5;5(1):128-38.

31. Ciginas A. ON DESIGN MEAN SQUARE ERROR ESTIMATION FOR MODEL-BASED SMALL AREA ESTIMATORS. SUMMER SCHOOL ON SURVEY STATISTICS 2021. 2021:48.