© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET19613

Parametric Bootstrap for Estimating Mean Square Error of Proportion in Small Area Estimation

Anggun Permatasari*, Khairil Anwar Notodiputro*, Erfiani

Department of Statistics, Bogor Agricultural University, Bogor, Indonesia

ABSTRACT

Small area estimation (SAE) is an important alternative method to obtain information in a small area when the sample size is small. In this paper, we proposed a parametric bootstrap method to estimate mean square error (MSE) of proportion based on area unit levels. The purpose of this research has been focused on applying the parametric bootstrap method to estimate MSE in SAE for zero inflated binomial models (SAE ZIB). The results showed that the bootstrap method produced a smaller MSE than the direct estimation, implying that the SAE ZIB performs better when compared to the direct estimation.

Keywords : Bootstrap, Mean Square Error, Parametric, Small Area Estimation

I. INTRODUCTION

A method of collecting data by taking samples from population describing the population is called a survey. The purpose of the survey is to provide complete, fast and continuous statistical data. In general, surveys are designed to predict parameters in a larger area. If you want to obtain information for a smaller area, there will be problems in the survey. Smaller areas have a small survey sample size so that the statistics obtained will have a large variety and cannot even be estimated when the area is not selected as a sample unit [1]. So, a method has been developed to overcome this problems which is called small area estimation (SAE). SAE is a technique in small area that utilize information from the area, outside the area, and survey results. Estimates directly on small area will produce a large variety because of the small sample size [2]. One solution used is to do indirect estimation by adding auxiliary variables in estimating parameters.

The estimation of mean squared error (MSE) for small area models is complicated, in case of mixture model which is nonlinear with complex structure [3]. The computation of reliable mean squared error estimator in small area estimation problems is a complicated process. This is because the models used and the small sample sizes within the areas require the accounting for the contribution to the error resulting from estimating the model parameters. Some approximation methods have been done by several authors such as bootstrap method to calculate the mean square error in small area estimation with zero inflated data in papers of Chandra and Sud (2012), Krieg et al (2015).

In this paper we proposed parametric bootstrap methods to estimate MSE in SAE with zero inflated binomial models (SAE ZIB). The techniques are at least as accurate as existing ones, although valid in substantially more general settings, and do not require derivation of analytical expansions. Bootstrap method is using random sampling with

replacement to generate the samples with size as many as the size of the sample data in order to approximate statistics distributions which is empirical distribution function of the sample data [4]. The parametric method consists of generating parametrically a large number of area bootstrap samples from the model fitted to the original data, re-estimating the model parameters for each bootstrap sample and then estimating the separate component of the mean square error. A parametric bootstrap procedure is proposed for the mean squared error of the predictor based on a unit level area.

Therefore, the approximation of MSE in ZIB data in this research adopted a bootstrap method which is computationally intensive in order to obtain standard error of the proposed method. The purpose of this study was to apply parametric bootstrap method to estimate MSE in SAE ZIB models and to compare both methods.

II. DATA AND METHODS

A. Data

This research used real data and simulation data. From the two data, the standard error would be calculated and compare. The real data used in this study was secondary data obtained from National Labour Force Survey (Sakernas) August 2015 conducted by Central Bureau of Statistics (BPS). Observations on Sakernas data were 38 districts in Bogor regency and 6 districts in Bogor city, Indonesia.

Simulation data describes population of the data that has overdispersed binomial distribution or ZIB data and the sample is obtained by taking samples from simulation data repeatedly in order to see the characteristics of the data which is assumed. The sampling design used in simulation is simple random sampling. This research used simulation to determine standard error in SAE ZIB model. The steps to generate population in the simulation data were on the below.

1. Let number of areas as many as 44.
2. Let number units in each area as many as 20.
3. Let parameter $\alpha_1 = -2.1$ and $\alpha_7 = 0.7$ as parameters model:
$$logit(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x'_{nz,ij}\beta_{nz} + v_{nz,i} \quad (1)$$
4. Let parameter $\beta_3 = -0.7$ and $\beta_7 = 1.1$ as parameters model:
$$logit(p_{ij}) = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = x'_{z,ij}\beta_z + v_{z,i} \quad (2)$$
5. Let variance parameter of variability between area, v (0.005)
6. Generate explanatory variable, x_{1i} distributing normally with mean 95 and variance 30.
7. Generate explanatory variable, x_{3i} distributing normally with mean 1 and variance 2.
8. Generate explanatory variable, x_{7i} distributing normally with mean 2 and variance 90.
9. Generate number of labor forces samples, n, distributing binomially with probability 0.15 of 40.
10. Generate variable of variability between area, v_i, distributing normally with mean 0 and variance 0.05.
11. By using parameters specified in points 1 to 8, calculate the value
    a. $xb_i^{nz} = \alpha_1 x_{1i} + \alpha_7 x_{7i}$
    b. $xb_i^{z} = \beta_3 x_{3i} + \beta_7 x_{7i}$
    c. $logit(\pi_i) = xb_i^{nz} + v_i$
    d. $logit(p_i) = xb_i^{z} + v_i$
    e. $\pi_i = \frac{exp(logit(\pi_i))}{1+(exp(logit(\pi_i)))}$
    f. $p_i = \frac{exp(logit(p_i))}{1+(exp(logit(p_i)))}$
12. Generate delta variable, δ_{ij} as the zero indicator following Bernoulli distibution with $P(\delta_{ij} = 0) = p$ dan $P(\delta_{ij} = 1) = 1 - p$.

13. Generate y_{ij}^{*} variable as the non zero data distributing binomially with probability of unemployed π_i of n_i labor forces.
14. Calculate of the values of y_{ij} as the number of unemployment, $y_{ij} = y_{ij}^{*}$ x δ_{ij}

B. Methods

This study proposed small area estimation method using ZIB model. The method is combination as follows:

1. ZIB model by Hall (2000) for probability in (3) is modeled using two model, those are logit (π) and logit (p).

$$Y_i \sim \begin{cases} 0, with\ probability\ p_i + (1-p_i)(1-\pi_i)^{\,n_i} \\ k, with\ probability\ (1-p_i)\binom{n_i}{k}\pi_i^{k}(1-\pi_i)^{\,n_i-k} \end{cases} \quad (3)$$

where $i = 1,2,\dots,m$ and $y_i = 0,1,2,\dots,n_i$ with ecpectation value $E(y_i) = (1-p_i)n_i\pi_i$ and variance $var(y_i) = (1-p_i)n_i\pi_i[1-\pi_i(1-p_in_i)]$.

2. Binomial model in small area estimation by Chandra *et al.* (2009) and Erciulescu and Fuller (2013) in (4) for logit (π)

$$logit(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x'_{nz,ij}\beta_{nz} + \nu_{nz,i} \quad (4)$$

where $i = 1,2,\dots,m$ and $y_i = 0,1,2,\dots,n_i$.

3. Zero inflated indicator in SAE for Zero Inflated data in SAE calculated by Krieg *et al.* (2015) in (5) for logit (p)

$$logit(p_{ij}) = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = x'_{z,ij}\beta_z + \nu_{z,i} \quad (5)$$

where $i = 1,2,\dots,m$ and $y_i = 0,1,2,\dots,n_i$. Based on those models, the expectation value for ZIB model in SAE is $E(\hat{y}_{ij}) = P(\delta_{ij}=1) = E(y_{ij}|\delta_{ij}=1) = (1-\hat{p}_{ij})\hat{\mu}_{ij}$ where $\hat{\mu}_{ij} = n_i\hat{\pi}_{ij}$ thus $E(\hat{y}_{ij}) = n_i\hat{\pi}_{ij}(1-\hat{p}_{ij})$.

Technically the above model can be implemented as follows:

a. Take a sample using simple random sampling as many as 10 observations in each area.
b. Estimate proportion (π_i) using SAE binomial method in each unit
c. Initialization values using logistic model
d. Estimate the parameters of logit (π) and logit (p) models using SAE ZIB method and calculate:
   1. $\hat{\pi}_{ij} = \frac{exp(logit(\hat{\pi}_{ij}))}{1+exp(logit(\hat{\pi}_{ij}))}$
   2. $\hat{p}_{ij} = \frac{exp(logit(\hat{p}_{ij}))}{1+exp(logit(\hat{p}_{ij}))}$
   3. $\hat{y}_{ij} = n_{ij}$ x $\hat{\pi}_{ij}$ x $(1-\hat{p}_{ij})$
   4. Proportion of $\hat{y}_i$, $\widehat{prop}_i = \sum_{\forall j}\hat{y}_{ij}/n_{ij}$
e. Do step a until d repeatedly in 1000 times.
f. Calculate standar error of proportion, $SE\widehat{prop}_i$:

$$SE\widehat{prop}_{i,q} = \sqrt{\frac{\sum_{q=1}^{R}\left(\widehat{prop}_{i,q} - prop_i^{\,mean}\right)^2}{R}}$$

where,

$SE\widehat{prop}_{i,q}$ = standard error estimate of area *i* from SAE ZIB method in *q*th simulation

$\widehat{prop}_{i,q}$ = proportion estimate of area *i* from SAE ZIB method in *q*th simulation

$prop_i^{\,mean}$ = proportion parameter of area *i*

R = number of simulations

III. RESULTS AND DISCUSSION

Bootstrap method uses random sampling with replacement to generate the samples with size as many as the size of the sample data. The bootstrap method can obtain the approximation of MSE while the close form of MSE is difficult to be obtained, especially in case of mixture model which is nonlinear with complex structure [3]. Therefore, the approximation of MSE in SAE with ZIB data in this study used a bootstrap method computationally intensive in order to obtain standard error of SAE ZIB method.

Figure 1 and 2 show the standard error values of proportion for each district in Bogor regency and Bogor city. From Figure 1, it can be seen that the standard error values with bootstrap method tent to be more stable and lower than direct estimators.

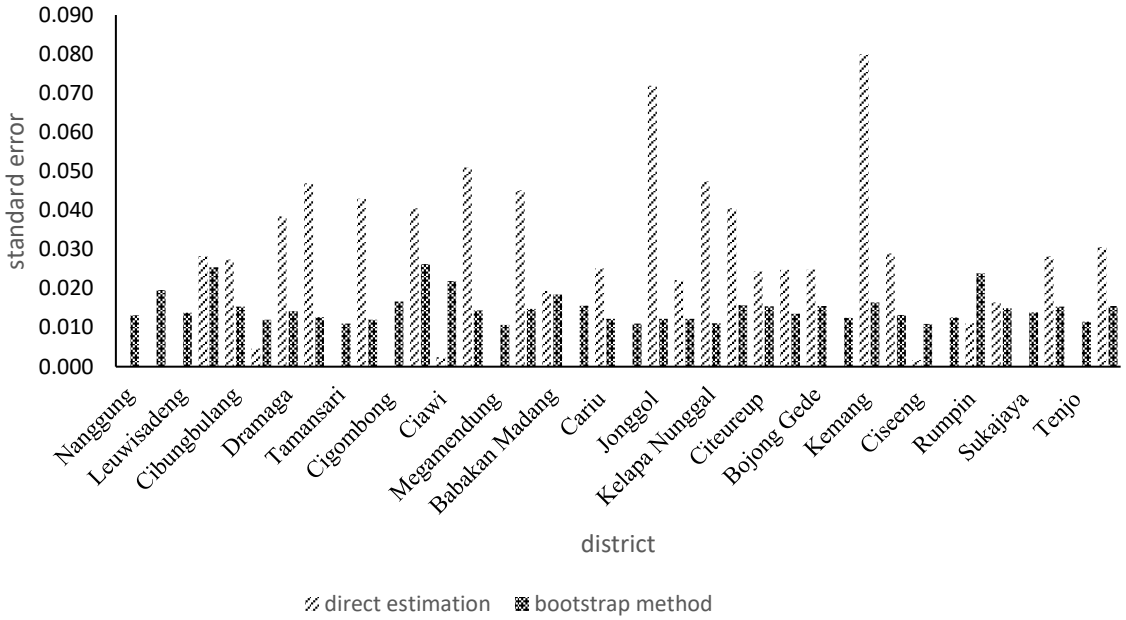

Figure 1 : Graph of standard error in Bogor regency

However, there were some districts in Bogor regency having very small standard error values that close to zero. The districts were Nanggung, Leuwisadeng, Cigombong, Megamendung and Tanjungsari. There were also districts having no standard error value because the estimator had zero proportions. That were Leuwiliang, Tamansari, Sukamakmur, Tajurhalang and Gunungsindur. For the proportion estimates which are equal to zero using direct estimation, the standard error was not relevant because of normal approximation method requiring the proportion estimates did not close to zero or one.

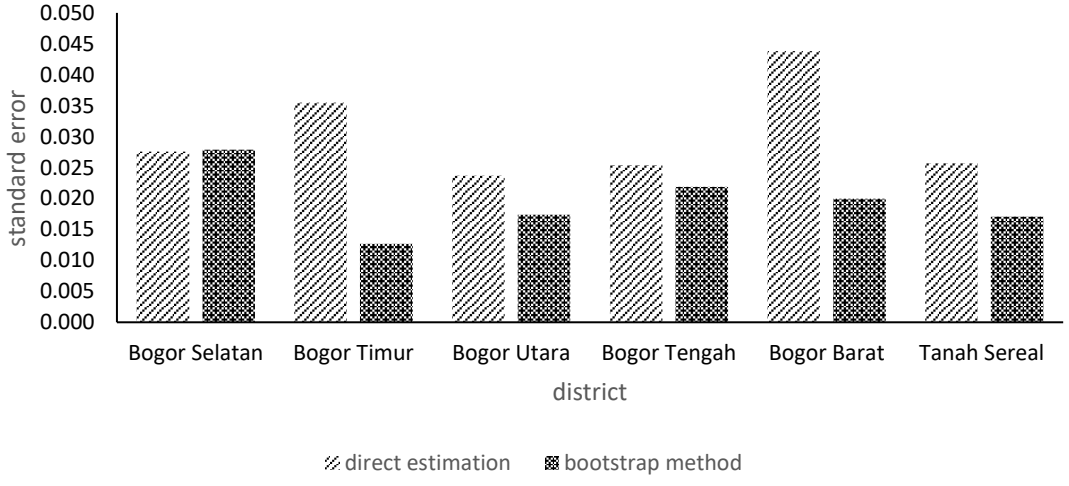

Figure 2 : Graph of standard error in Bogor city

The value of standard error in Bogor city in Figure 2 shows that, the estimation of standard error with the bootstrap method tent to have a lower value than the direct estimation. But, there was one district in Bogor city where the standard error value was almost same between the bootstrap method and direct estimation, i.e. Bogor Selatan district. The estimation of standard error with bootstrap used SAE ZIB method. Based on the standard value, SAE ZIB method revealed an estimate with precision that was better than direct estimation.

IV.CONCLUSION

This research can provided an overview of the use of bootstrap methods in small area estimation to estimate mean square, especially zero inflated binomial data. The estimation of MSE for small area models is complicated. Therefore, the approximation of MSE in SAE ZIB model in this study used standard errors. Estimation value of standard errors using bootstrap method tent to produce smaller value than direct estimates. Based on the estimated standard error of the bootstrap method and direct estimator, the estimation of proportions with SAE ZIB model had better precision so as to improve the direct estimation results.

V. REFERENCES

[1]. Sadik K. 2009. Metode prediksi tak-bias linier terbaik dan bayes berhirarki untuk pendugaan area kecil berdasarkan model state space [disertasi]. Bogor (ID): Institut Pertanian Bogor.

[2]. Benavent R, Morales D. 2016. Multivariate Fay-Herriot models for small area estimation. Computational Statistics & Data Analysis. 94: 372-390.

[3]. Manteiga G, Lambordia MJ, Molina I, Morales D, Santamaria L. 2007. Estimation of the mean squared error of predictors of small area linear parameters under a logistic mixed model. Computational Statistics & Data Analysis. 51: 2720-2733.

[4]. Bodromurti W. 2017. Zero inflated binomial models in small area estimation with application to infant mortality data in Indonesia [tesis]. Bogor (ID): Institut Pertanian Bogor.

[5]. Chandra H, Sud UC. 2012 small area estimation for zero-inflated data. Communications in Statistics – Simulation and Computation. 41(5): 632 – 643.

[6]. Krieg S, Booontra HJ, Smeets M. 2015. Small area estimation with zero-inflated data – a simulation study. Statisitics Netherlands, Discussion paper. 1: 1 – 45.

Cite this article as :

Anggun Permatasari, Khairil Anwar Notodiputro, Erfiani, "Parametric Bootstrap for Estimating Mean Square Error of Proportion in Small Area Estimation", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 311-315, March-April 2019. Available at doi :
https://doi.org/10.32628/IJSRSET19613
Journal URL : http://ijsrset.com/IJSRSET19613