IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Zero Inflated Binomial Models in Small Area Estimation with Application to Unemployment Data in Indonesia

[1] Budi Hartono; [2] Anang Kurnia; [3] Indahwati

[1] Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

[2] Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

[3] Department of Statistics, Bogor Agricultural University
Bogor, Indonesia

Abstract - Binary response variables are commonly modeled by binomial models. The binomial overdispersion occurs if the variability is greater than the variance of assumed model. The overdispersion can be caused by excess zeros. The overdispersion may produce underestimated standard error which in turn will produce underestimated p-value. Therefore, Zero Inflated Binomial (ZIB) models are considered to overcome the excess zeros in binomial data. A simulation study is employed to evaluate the performance of models by using RRMSE and relative bias. The simulation showed that the proposed method SAE ZIB has better fit than SAE ZIB Synthetic in terms of the smaller RRMSE. The proposed SAE ZIB method applies to unemployment data to estimate proportion of unemployment in each district/regency during period of August 2016 In Jambi Province, Indonesia. The real data application showed that SAE ZIB method is better than the direct estimates method in terms of the smaller standard error.

Keywords – *Small Area Estimation, Zero Inflated Binomial, Unemployment.*

1. Introduction

Binary response variables (Y_i) are commonly modeled by binomial models, in which the variance is given by $var(Y_i)=n_i\pi_i(1-\pi_i)$. However if the variability is greater than the assumed model, $var(Y_i)>n_i\pi_i(1-\pi_i)$, then it is called overdispersed binomial [1]. The overdispersion can be caused by excess zeros [1]. The overdispersion may produce underestimated standard error which in turn will produce underestimated p-value [1]. This means that non-significant assosiation will appear to be significant, or the hypothesis testing tends to reject the null hypothesis. Hall [2] applied ZIB models to the case in which the response variable is an upper-bounded count opening the possibility to fit ZIB models. Based on these researches, ZIB models in which the explanatory variables consist of fixed and random effects are useful for modeling source of heterogenity and dependencies in zero inflated data [2]. Bodromurti [3] applied Small Area Estimation (SAE) using ZIB models for infant mortality data, it showed that SAE with ZIB model has better fit than SAE with binomial models to overcome the excess zeros in binomial data.

SAE deals with the producing reliable parameter estimates for small areas which have inadequate size samples [4]. However, direct estimators for small area, based on sample data from surveys are likely unacceptable due to its hace large standard errors. SAE can improve the direct estimators by assuming regression models that link all the sample data by borrowing the strength of relation between the set of auxilary variables and the variables of interest [5]. In recent years, the needs of small area statistics have greatly increased in various development sectors, such us the need for indicators of health, education, employment, and others at the level of smaller areas. SAE can be applied in National Labour Force Survey (Sakernas). This survey is specifically designed to collect information in labor force statistics in Indonesia. The first Sakernas was conducted in 1976 by BPS and since 2015 Sakernas has been conducted in semester period, i.e. February (semester I) and August

(semester II) [6]. Sakernas February to produces employment indicators up to province level, while Sakernas August to district/regency level. In 2016, Sakernas August not produced employment indicators in district/regency level because inadequate size samples.
Unemployment is a binary event with two possibilities, unemployment or working. Binary responses with probability of success, π, and the upper bound (n) generally follow the binomial (n, π) distribution, hereafter the number of unemployments are assumed to follow a binomial distribution. The number of unemployment data from Sakernas August 2016 occured overdispersed at the villages units level. So, it was assumed as overdispersed binomial data due to large number of zeros or zero inflated.

In this research, Zero Inflated Binomial (ZIB) models are considered to overcome the excess zeros in binomial data. This research applies the proposed SAE ZIB method to unemployment data at the village unit level and estimate proportion of unemployment in each district/regency during period August 2016 in Jambi Province, Indonesia.

2. Proposed Method

This research proposed SAE method using ZIB model to solve the problems of ZIB model in SAE. The method is combination as follows:

- ZIB model by Hall [2] for the probability in equation (1) is modeled using two model, those are *logit*(π) and *logit*(p) with $E(Y_i)=(1-p_i)n_i\pi_i$ and variance $Var(Y_i)=(1-p_i)\, n_i\pi_i[1-\pi_i\,(1-p_i n_i)]$.

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

- Binomial model in SAE by Chandra *et al.* [7] and Erciulescu and Fuller [8] in equation (2) for *logit*(π) model in SAE with ZIB.

$$logit\,(\pi_{ij}) = log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = x'_{nz,ij}\beta_{nz} + v_{nz,i} \quad (2)$$

- Zero inflated indicator in SAE for Zero Inflated data in SAE by Krieg *et al.* [9] in equation (3) for *logit*(p) model in SAE with ZIB.

$$logit\,(p_{ij}) = log\left(\frac{p_{ij}}{1-p_{ij}}\right) = x'_{z,ij}\beta_z + v_{z,i} \quad (3)$$

Based on those models, the Empirical Best Predictor (EBP) for π_{ij} and p_{ij},

$$\hat{\pi}_{ij} = \frac{\exp(x'_{nz,ij}\hat{\beta}_{nz} + \hat{v}_{nz,i})}{1 + \exp(x'_{nz,ij}\hat{\beta}_{nz} + \hat{v}_{nz,i})} \quad (4)$$

$$\hat{p}_{ij} = \frac{\exp(x'_{z,ij}\hat{\beta}_z + \hat{v}_{z,i})}{1 + \exp(x'_{z,ij}\hat{\beta}_z + \hat{v}_{z,i})} \quad (5)$$

The proposed EBP are combination from small area estimation of proportion in binary data [7], small area prediction of the mean of a binomial random variable [8], and small area estimation for zero inflated data [9]. Hereafter, 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 $E(\hat{y}_{ij}) = n_{ij}\hat{\pi}_{ij}(1-\hat{p}_{ij})$.

Bodromurti [3] proposed SAE method using ZIB model to solve the problems of ZIB model in SAE involved only from survey units or SAE method using ZIB synthetic model (SAE ZIB Synthetic). Therefore, this research is interested to develop small area estimation model proposes SAE method using ZIB models by considering unsurvey units. The improvement of this research is SAE method using ZIB model (SAE ZIB) by using population units (survey and unsurvey units) as the principle has been developed in the small area estimation.

The important things in EBP for GLMM with logistic link funtion for SAE are likelihood function which often involves high dimensional integrals and estimation of the MSE [7]. Estimation of the MSE is still a stunning problem in SAE, even in EBLUP [10]. Some approximation methods have beeen done by several authors such as bootstrap method to calculate the MSE in SAE with zero inflated data in papers of Chandra and Sud [10], Krieg et al. [9]. Bootstrap method is using random sampling with replacement to generate the samples with size as many as of the sample data in order to approximate statistics distribution which is empirical distribution function of the sample data.

The bootstrap technique can obtain the approximation of MSE while the close form of MSE is difficult to be obtained, in case of mixture model which is nonlinier with complex structure. GLMM usually tend to more complex situation for MSE [11]. Therefore, the approximation of MSE in SAE with ZIB data in this research used a bootstrap method which is computationally intensive in order to obtain standard error of the proposed method.

3. Methodology

3.1. Simulation

This research uses simulation for the proposed SAE ZIB to measure the fitness when compared to SAE ZIB Synthetic. The steps to generate population in the simulation is on the below.

1. Let number of areas as many as 11. It considers as the number of regencies/cities in Jambi Province, Indonesia.
2. Let number units in each area as many as 100. It consists 10 units are survey units and 90 units are unsurvey units. It considers as the number of villages in each district/regency.
3. Let parameter a0 and a1 as parameters model in equation (2) for non zero part which is binomial disributed.
4. Let parameter b0 and b1 as parameters model for averall data in equation (3).
5. Let variance parameter of variability between area, v (0.015).
6. Generate explanatory variable, $x1$ which follows normal distribution with mean 0.014 and standard deviation 0.011. It set by $x6$ auxilary variable condition.
7. Generate number of labor forces samples, n, that follows binomial distribution with probability 0.48 of 38.
8. Generate number of labor forces population in each unit, N, that follows normal distribution with mean 2.203 and standard deviation 2.025. It based on the labor force participation rate of Jambi Province, 2016.
9. Generate varible of variability between area, v, that follows normal distribution with mean 0 and variance 0.015.
10. Calculate the values of:
    a. $x_{ij}b\_nz = a0+a1*x1_{ij}$
    b. $x_{ij}b\_z = b0+b1*x1_{ij}$
    c. $logit(\pi_{ij}) = x_{ij}b\_nz + v_i$
    d. $logit(p_{ij}) = x_{ij}b\_z + v_i$
    e. $\pi_{ij} = (exp(logit(\pi_{ij}))) / (1 + exp(logit(\pi_{ij})))$
    f. $p_{ij} = (exp(logit(p_{ij}))) / (1 + exp(logit(p_{ij})))$
11. Generate delta variable, δ_{ij}, as the zero indicator which follows Bernoulli distribution with probability of zero is p, thus delta will equal to 1 with probability $(1 - p)$.
12. Generate $y_{ij}*$ variable as the non zero data that follows binomial ditribution with probability of be unemployed π of n labor forces.
13. Calculate the values of y_{ij} as the number of unemployment, $y_{ij} = y_{ij}* \text{ x } \delta_{ij}$.
14. Take sample using Simple Random Sampling as many as 10 observations in each survey units.
15. Initialization values using logistic model.
16. Estimate the parameter of $logit(\pi)$ and $logit(p)$ using ZIB Synthetic Method.
17. Calculate the estimation of SAE ZIB Synthetic:
    - $\hat{\pi}_{ij} = (\exp\left(logit(\hat{\pi}_{ij})\right))/(1 + \exp\left(logit(\hat{\pi}_{ij})\right))$
    - $\hat{p}_{ij} = (\exp\left(logit(\hat{p}_{ij})\right))/(1 + \exp\left(logit(\hat{p}_{ij})\right))$
    - $\hat{y}_{ij} = n_{ij} \text{ x } \hat{\pi}_{ij} \text{ x } (1 - \hat{p}_{ij})$
    - Proportion of unemployment ($\hat{y}$), $\widehat{proportion}_i = \sum_{\forall j} \hat{y}_{ij} / \sum_{\forall j} n_{ij}$
18. Calculate the estimation of SAE ZIB:
    a. Number of unemployment in each survey units (j) :
    - $\hat{\pi}_{ij} = (\exp\left(logit(\hat{\pi}_{ij})\right))/(1 + \exp\left(logit(\hat{\pi}_{ij})\right))$
    - $\hat{p}_{ij} = (\exp\left(logit(\hat{p}_{ij})\right))/(1 + \exp\left(logit(\hat{p}_{ij})\right))$
    - $\hat{y}_{ij} = N_{ij} \text{ x } \hat{\pi}_{ij} \text{ x } (1 - \hat{p}_{ij})$

    Number of unemployment in each unsurvey units (j^*):
    - $\hat{\pi}_{ij^*} = (\exp\left(logit(\hat{\pi}_{ij^*})\right))/(1 + \exp\left(logit(\hat{\pi}_{ij^*})\right))$
    - $\hat{p}_{ij^*} = (\exp\left(logit(\hat{p}_{ij^*})\right))/(1 + \exp\left(logit(\hat{p}_{ij^*})\right))$
    - $\hat{y}_{ij^*} = N_{ij^*} \text{ x } \hat{\pi}_{ij^*} \text{ x } (1 - \hat{p}_{ij^*})$

    b. Proportion of unemployment
    $\widehat{proportion}_i = \frac{1}{N_i}\left(\sum_{\forall j} \hat{y}_{ij} + \sum_{\forall j^*} \hat{y}_{ij^*}\right)$
19. Do step 14 until 18 repeatedly in 1000 times. Measure the RRMSE and relative bias of $\widehat{proportion}$ is on the below.

$$RMSE_i = \sqrt{\sum_{q=1}^{R} (\widehat{proportion}_{i,q} - proportion_i^{mean})^2 / R}$$

$$RRMSE_i(\%) = \left(\frac{RMSE_i}{proportion_i^{mean}}\right) x100\%$$

$$Bias_i = \sum_{q=1}^{R} (proportion_i^{mean} - \widehat{proportion}_{i,q}) / R$$

$$Relative\ Bias_i(\%) = \left(\frac{Bias_i}{proportion_i^{mean}}\right) x100\%$$

where,
$RMSE_i$ = Root Mean Square Error of area i
$RRMSE_i$ = Relative Root Mean Square Error of area i
$\widehat{proportion}_{i,q}$ = Proportion estimate of area i from SAE ZIB method in qth simulation
$proportion_i^{mean}$ = Proportion parameter of area i
R = Number of simulations

3.2. Application

This research applies the proposed SAE ZIB method to unemployment data in order to overcome excess zeros in binomial data at the village level and estimate proportion of unemployment in each district/regency. To obtain the estimates at the district/regency level, the resulting estimates could be agregated up to the district/regency. Respon variable data is from Sakernas August 2016 and some variables for auxiliary data are from Village Potential data (Podes 2014). The data used in small area estimation is using village as unit level in 11 regencies/cities (i = 1, ..., 11) in Jambi Province, Indonesia. The number of units is 1562 villages, it consists of 99 villages as survey units (j = 1, ..., 99) and 1463 villages as unsurvey units (j*= 1, ... , 1463). The variables used is presented in Table 1.

Table 1: Variables used and the information

Variable	Information	Data Source
Y	Number of unemployment	Sakernas August 2016
X1	% of male population	Podes 2014
X2	% of agiculture family	Podes 2014
X3	Number of poor certificates (SKM) issued by the village government	Podes 2014
X4	Number of senior high schools and collages	Podes 2014
X5	Number of industries	Podes 2014
X6	Number of markets and shops	Podes 2014
X7	Number of hotels/inns	Podes 2014

These are the steps of the aplication SAE ZIB to unemployment data.

1. Weighting the response variable with w_{ij}.

$$w_{ij} = \left(\frac{HH_{ij}}{hh_{ij}}\right) \Big/ \left(\sum_{\forall j} \frac{HH_{ij}}{hh_{ij}}\right) \quad (6)$$

   where,
   w_{ij} = weighting of district/regency i and village j
   hh_{ij} = number of households population of district/regency i and village j
   HH_{ij} = number of households sample of district/regency i and village j
2. Modelling ZIB model for unemployment data in SAE.
   - Select the variables into $logit(\pi)$ and $logit(p)$ models.
   - Initialization values using ZIB model.
   - modeling the proposed ZIB model in SAE using the initial values.
3. Estimate the parameter of $logit(\pi)$ and $logit(p)$ using SAE ZIB method.
4. Calculate the estimates:
   a. Number of unemployment in each survey units (j):
      - $\hat{\pi}_{ij} = (\exp\left(logit(\hat{\pi}_{ij})\right))/(1 + \exp\left(logit(\hat{\pi}_{ij})\right))$
      - $\hat{p}_{ij} = (\exp\left(logit(\hat{p}_{ij})\right))/(1 + \exp\left(logit(\hat{p}_{ij})\right))$
      - $\hat{y}_{ij} = N_{ij} \text{ x } \hat{\pi}_{ij} \text{ x } (1 - \hat{p}_{ij})$

   b. Number of unemployment in each unsurvey units (j*):
      - $\hat{\pi}_{ij^*} = (\exp\left(logit(\hat{\pi}_{ij^*})\right))/(1 + \exp\left(logit(\hat{\pi}_{ij^*})\right))$
      - $\hat{p}_{ij^*} = (\exp\left(logit(\hat{p}_{ij^*})\right))/(1 + \exp\left(logit(\hat{p}_{ij^*})\right))$
      - $\hat{y}_{ij^*} = N_{ij^*} \text{x } \hat{\pi}_{ij^*} \text{ x } (1 - \hat{p}_{ij^*})$

   c. Proportion of unemployment and standard error:
      Proportion of $\hat{y}$,

$$\widehat{proportion}_i = \frac{1}{N_i}\left(\sum_{\forall j} \hat{y}_{ij} + \sum_{\forall j^*} \hat{y}_{ij^*}\right)$$

      Hall [2] :

$$Var(y_i) = (1 - p_i)n_i\pi_i[1 - \pi_i(1 - p_i n_i)]$$

      The approximation of MSE using standard error,

$$Var\left(\frac{Y_i}{N_i}\right) = \frac{1}{N_i^2}\,\mathrm{Var}(Y_i)$$

$$Var\left(\frac{Y_i}{N_i}\right) = \frac{1}{N_i^2}\left((1 - \hat{p}_i) \text{ x } N_i \text{ x } \hat{\pi}_i \ [1 - \hat{\pi}_i \text{ x } (1 - \hat{p}_i \text{ x } N_i)]\right)$$

      Standard error of proportion,

$$\widehat{SE}_{proportion,i} = \sqrt{\frac{1}{N_i^2}\left((1 - \hat{p}_i) \text{ x } N_i \text{ x } \hat{\pi}_i \ [1 - \hat{\pi}_i \text{ x } (1 - \hat{p}_i \text{ x } N_i)]\right)}$$

5. Estimate proportion of unemployment and standard error using direct estimation.

4. Result

4.1. Simulation

Small area model used in the simulation is unit level model in 11 areas and 100 units in each area. This units consists of 10 survey units and 90 unsurvey units. Overdispersed binomial occurs at the unit level and then proportion estimates is aggregated to the area level. Simulation is designed to calculate RRMSE and relative bias of the estimates between SAE ZIB and SAE ZIB Synthetic method. The proposed SAE ZIB and SAE ZIB

IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

Synthetic used initialization from logistic model to prevent failures of the algorithms in the simulation.

Table 2: Summary of the simulation using SAE ZIB and SAE ZIB Synthetic

Area	RRMSE (%)			Relative Bias (%)	
	SAE ZIB	SAE ZIB Synthetic	Difference	SAE ZIB	SAE ZIB Synthetic
1	13.161	13.367	-0.206	1.332	1.305
2	23.753	24.264	-0.511	19.679	19.650
3	13.911	14.418	-0.507	3.739	3.712
4	15.665	14.964	0.701	-9.263	-9.289
5	16.979	17.388	-0.409	9.429	9.401
6	13.134	13.936	-0.802	-4.429	-4.456
7	21.247	20.744	0.503	-17.787	-17.813
8	18.006	17.404	0.602	-13.285	-13.312
9	13.835	14.141	-0.307	3.541	3.514
10	14.762	14.362	0.400	-8.026	-8.053
11	12.830	13.534	-0.704	-2.050	-2.077
Average	16.117	16.229	-0.113	-1.556	-1.583

Table 2 showed RRMSE and relative bias of the proportion estimates from simulation result of the SAE ZIB and SAE ZIB Synthetic method. The simulation result showed that the proposed method produces smaller RRMSE (average = 16.117 percent) when compared to SAE ZIB Synthetic (RRMSE = 16.229 percent). Figure 1 showed that SAE ZIB produces smaller RRMSE than SAE ZIB Synthetic method. It also can be seen in Figure 2, mostly difference between SAE ZIB and SAE ZIB Synthetic method is under zero value. Table 2 showed that the relative bias of SAE ZIB method is mostly closer to zero value (average = -1.556 percent) than SAE ZIB Synthetic method (average = -1.583 percent). It means the proposed SAE ZIB method has better fit than SAE ZIB Synthetic method to overcome overdispersed binomial data.

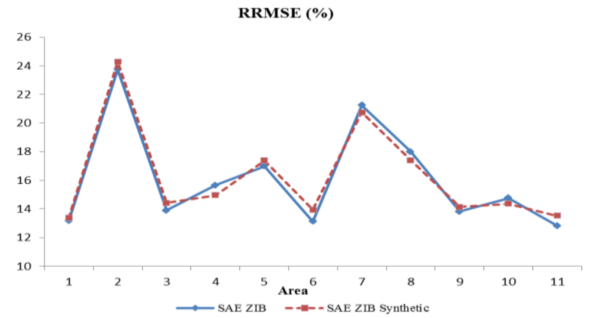

Fig. 1 Line chart of RRMSE (%) of the proportion estimate between SAE ZIB and SAE ZIB Synthetic method.

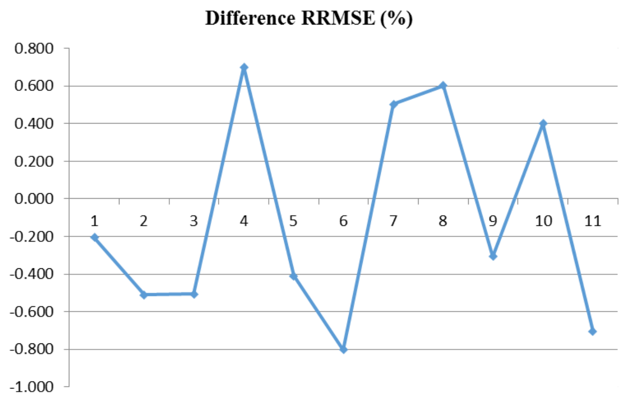

Fig. 2 Line chart of RRMSE difference (%) between SAE ZIB and SAE ZIB Synthetic method.

4.2. Application on Unemployment Data

Labor force is persons of 15 years old and over who, in the previous week, were working temporarily absent from work out but having jobs, and those who did not have work and were looking for work. Unemployment which consists of person without work but looking for work, person without work who have established a new business/firm, person without work who were not looking for work because they do not expect to find work, or person who have made arrangements to start work on a date subsequent to the reference period (future starts). Unemployment rate is percentage of unemployment to the number of labor force [6].

The current publication [12] showed that unemployment rate in Jambi Province during period of August 2016 is 4.00 percent. It means that there are 4 unemployments of 100 labor forces. Based on Sakernas August 2016 data, in 11 regencies/cities in Jambi Province there are 67 unemployments of 1812 labor forces (3.70 percent). The number of unemployment occured in 39 villages of 99

IJCSN - International Journal of Computer Science and Network, Volume 6, Issue 6, December 2017
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

villages surveyed (39.39 percent), it means that there are 60.61 percent zero values in data at the village level so those unemployment data has excess zeros in binomial data.

In simulation data is fit to describe the unemployment data in terms of overdispersed binomial data in SAE at the unit level. Based on the simulation result, the proposed SAE ZIB method has better fit than SAE ZIB Synthetic method in terms of smaller RRMSE. Hereafter, this research applies the proposed SAE ZIB method to the unemployment data. The number of unemployment data is modeled by ZIB model in SAE in order to overcome overdispersed binomial data at the village level. ZIB model in SAE from survey units data is used to estimate the number of uneployments in unsurvey units. This research is using initialitation values from ZIB model without SAE to prevent failures of algoritms in SAE ZIB method. The appropriate choice of initialization values and the appropriate choice of variables used are important things to overcome failures of the algoritms. The results in Table 3 showed that X4 = number of senior high schools and collages and X6 = number of markets and shops are auxilary variables used in models which can overcome failures of the algoritms.

Table 3: Parameter estimate on the SAE ZIB method

Logit	Parameter	Estimate	Standard Error	DF	t-Value	Pr > |t|
π	X4	-2471.20	3481.50	97	-0.71	0.48
	X6	-446.60	124.00	97	-3.60	0.00
p	X4	-808.86	678.10	97	-1.19	0.23
	X6	-17.78	13.37	97	-1.33	0.18

These are the selected variables for the logit models based on equations (2) and (3). The first model is non zero part sample data via logit link function,

$$logit\,(\hat{\pi}_{ij}) = log\left(\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}}\right) = \hat{\alpha}_4 X4 + \hat{\alpha}_6 X6 + v_{nz,i} \quad (7)$$

where i = 1, 2, ..., 11; dan j = 1, 2, ..., N_i. N_i is number of villages in area i and $v_{nz,i} \sim N(0, \sigma^2_{v,nz})$. The second model is whole sample data that describes probability of non zero that is $(1 - p_{ij}) = P(\delta_{ij} = 1) = P(y_{ij} \neq 0)$, the model is,

$$logit\,(\hat{p}_{ij}) = log\left(\frac{\hat{p}_{ij}}{1-\hat{p}_{ij}}\right) = \hat{\beta}_4 X4 + \hat{\beta}_6 X6 + v_{z,i} \quad (8)$$

Table 4: Proportion estimate of unemployment between SAE ZIB and direct estimation

District/Regency		SAE ZIB		Direct Estimation	
		Estimate	SE	Estimate	SE
1501	Kerinci	0.0506	0.0163	0.0541	0.0191
1502	Merangin	0.0436	0.0162	0.0495	0.0216
1503	Sarolangun	0.0389	0.0162	0.0425	0.0198
1504	Batang Hari	0.0494	0.0187	0.0330	0.0121
1505	Muaro Jambi	0.0303	0.0160	0.0604	0.0198
1506	Tanjung Jabung Timur	0.0334	0.0163	0.0106	0.0098
1507	Tanjung Jabung Barat	0.0362	0.0135	0.0519	0.0186
1508	Tebo	0.0335	0.0180	0.0175	0.0114
1509	Bungo	0.0469	0.0142	0.0452	0.0156
1571	Kota Jambi	0.0515	0.0154	0.0425	0.0138
1572	Kota Sungai Penuh	0.0290	0.0184	0.0371	0.0186
Jambi Province		0.0417	0.0163	0.0403	0.0164

Table 4 showed that the proportion estimate of unemployments to the number of labor forces result of SAE ZIB and direct estimation method. Figure 3 represents the proportion estimate of unemployments result of SAE ZIB and direct estimation method in 11 regencies/cities in Jambi Province. The proportion estimate of unemployment result of ZAE ZIB method in Jambi Province is 0.0417 (unemployment rate = 4.17 percent). The proportion estimate of unemployment result of direct estimation method is 0.0403 (unemployment rate = 4.03 percent). The current publication of BPS during period of August 2016 showed that unemployment rate = 4.00 percent (proportion = 0.4000). SAE ZIB estimates, direct estimation, and current publication have similar conclusion that 4 person unemployments to the 100 of labor forces.

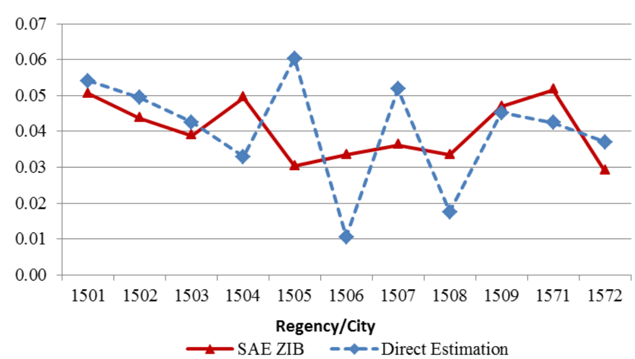

Fig. 3 Line chart of proportion estimate between SAE ZIB and direct estimation method.

Table 4 also showed that in Jambi Province standard error of SAE ZIB method (0.0163) less than direct estimates method (0.0164). It means the proposed SAE ZIB method is better than the direct estimates method.

4. Conclusions

Small area estimation with zero inflated binomial model showed that the proposed method SAE ZIB (survey and unsurvey units) has better fit than SAE ZIB Synthetic (survey units only) in terms of the smaller RRMSE. The appropriate initialization values and arrangement of explanatory variables included in the models are important things to overcome the failures in algorithms. This model can be used to estimate proportion of unemployments for each area as well as for each district/regency in Jambi Province, Indonesia. The application to unemployment data results that the proportion estimate of SAE ZIB method is similar with the current publication (4 unemployments of 100 labor forces). Standard error of SAE ZIB method (0.0163) in Jambi Province less than direct estimates method (0.0164). It means the proposed SAE ZIB method is better than the direct estimates method when applied to unemployment data.

References

[1] J. Hinde, and CGB. Demetrio, Overdispersion: Models and Estimation A short Course for SINAPE 1998, Sao Paulo : Associacio Brasileira de Estatisca, 2007.

[2] DB. Hall, "Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study", Biometrics 56, 2000, pp. 1030-1039.

[3] W. Bodromurti, "Zero Inflated Binomial Models in Small Area Estimation with Application to Infant Mortality Data in Indonesia" Thesis, Bogor Agriculture University, Bogor, Indonesia, 2016.

[4] JNK. Rao, Small Area Estimation, New York: John Wiley and Sons, 2003.

[5] A. Kurnia, "Prediksi Terbaik Empirik untuk Model Transformasi Logaritma di dalam Pendugaan Area Kecil dengan Penerapan pada Data Susenas" Disertasi, Bogor Agriculture University, Bogor, Indonesia, 2009.

[6] BPS, Pedoman Pencacahan Sakernas Semesteran 2016, Jakarta: BPS, 2016.

[7] H. Chandra, R. Chambers, and N. Salvati, "Small Area Estimation of Proportions in Business Surveys [Internet]", Centre for Statistical and Survey Methodology, University of Wollongong, 2009, Working Paper: 15-09 22p. [accesed 2017 March 27], Available from: http://ro.uow.edu.au/cssmwp/35/.

[8] AL. Erciulescu, and WA. Fuller, "Small Area Prediction of the Mean of a Binomial Random Variable", Survey Research Methods Section, JSM Proceedings, Biometrics 56, 2013, pp. 1030-1039.

[9] S. Krieg, HJ. Boonstra, and M. Smeets, "Small Area Estimation with Zero-Inflated Data – A Simulation Study", Statistics Netherlands, Discussion Paper 1, 2015, pp. 1- 45.

[10] H. Chandra, and C. Sud, "Small Area Estimation for Zero-Inflated Data", Communications in Statistics– Simulation and Computation 41(5), 2012, pp. 632-643.

[11] G. Mantaega, MJ. Lambordia, I. Molina, D. Morales, and L. Santamaria, "Estimation of the mean squared error of predoctors of small area linier parameters under a logistic mixed model", Computational Statistics and Data Analysis 51, 2008, pp. 2720-2733.

[12] BPS Provinsi Jambi, Indikator Kesejahteraan Rakyat Provinsi Jambi 2016, Jambi: BPS Provinsi Jambi, 2017.

Authors –

B. Hartono master student in Department of Statistics, Bogor Agricultural University. His main interests is on small area estimation.

Dr. A. Kurnia ccurrently worked as a senior lecture in Department of Statistics, Bogor Agricultural University. He is head in Department of Statistics, Bogor Agricultural University. His main interests is on statistical modellng and small area estimation.

Dr. Indahwati ccurrently worked as a senior lecture in Department of Statistics, Bogor Agricultural University. His main interests is on statistical modellng, sampling design, and methodology.