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

Small Area Estimation in Estimating Unemployment Rate in Bogor District of Sampled and Non-Sampled Areas Using A Calibration Modeling Approach

[1] Siti Aprizkiyandari; [2]Anang Kurnia; [3] Indahwati

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

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

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

Abstract – The main problem in Indonesia is unemployment. There are some various government policies to resolve unemployment, such as the availability of statistical data in unemployment. The National Labor Survey conducted by the Statistics Indonesia (BPS) only generates estimates at the national levels, whereas to carry out various government policies requires the availability of unemployment information to smaller levels. The Small Area Estimation (SAE) method is one of the solutions to estimate small area without adds sampling units. The method is borrowing strength from nearby observation sample areas. The study focused on estimating unemployment rate in Bogor sub-district level using Generalized Linear Mixed Models (GLMM) method with calibration approach. The results of the proposed method can produce the same result as published by BPS and are able to generate the result to sub-district level..

Keywords – *Generalized Linear Mixed Models (GLMM), Calibration modeling approach, Clustering analysis.*

1. Introduction

The unemployment issue is a very important issue to be resolved by the government because one of the milestones in the process of the movement of the economy. the Country that tends to be forward will have a lot of unemployment. A wide range of Government policy ever done in tackling unemployment one of them that is available related statistics of unemployment. Statistics Indonesia (BPS) only presents data for the unemployment level area is higher e.g. national, provinces or districts, so as to lower level areas such as sub-district or village than any sample unit number small. If a direct prediction is used then it will produce a low precision. Therefore, one solution to overcome these problems is prediction method for a small area (Small Area Estimation). According to [6] defines the small area estimation as estimating an area of relatively small sample size by utilizing information from outside the area, information from within the area itself and from outside the survey. The response module will be associated with fixed-match variables and the small, random-area specific diversity. Through the estimation of small areas, the available samples will be strengthened in their role to be effective. The increase in effectiveness will increase the precision in the estimation because the default error will be small.

Research on unemployment of small area estimation method has been done by many previous researchers. Study on applying of empirical best linear unbiased prediction method to small area estimation in estimating of the unemployment rate in Bogor city, the result showed by using EBLUP method generates smaller RRMSE when compared with direct estimation method by Harsanti [4]. While a study of a small area estimation method for estimating unemployment using the Bayes approach with education level, this results obtained suggest that indirect estimation is better than direct estimates by Hanike [3].

The unemployment rate at Bogor district had been increased in different ways. In 2011 until 2013 the unemployment rate in Bogor regency decreased from 10.73% to 7.87, but in 2015 the unemployment rate in Bogor District experienced a significant change that is increased to 10.01%. Therefore, the researchers are interested to observe the unemployment rate incident that occurred in Bogor regency. The difference of this study with previous research is that this research will be conducted at the village level and assumes the influence of the number of sample units in the village into the model. The study focused on 38 of surveyed sub-districts and 2 other of not surveyed sub-districts. Unemployment incidence in Bogor district is assumed to have Poisson distribution because unemployment incidence has small incidence number. The assumption that is paid for Poisson distribution is equidispersion. The estimation of a small area that will be used in this research using Generalized Linear Mixed Models (GLMM) method with calibration modeling approach.

2. Research Method

2.1 Data

The data source for this research obtained from statistics national (BPS). It classified into two parts, such as national labor force survey (SAKERNAS) and potential of the village (Podes). Sakernas data is used as a response variable in calculating the number of unemployed. The number of unemployed in this study refers to the definition of open unemployment according to BPS are individuals aged 15 years and over, not working more than a week, looking for work, preparing for business, desperate / feeling impossible to get a job, already have a job but not work now. The second data source used this research is a potential of the village (Podes) as an indicator of unemployment.

Table 1: Unemployment indicator from podes data

No	Name of variable	Status	Explanation
1	Central government funds (x_1)	Concomitant variable	The amount of funds sourced by the central government
2	Number of users of state electricity companies (x_2)	Concomitant variable	The amount of families using electricity in each village

2.2 Data Analysis Procedure

The analysis method will be done as follows:

1. Direct estimation
2. Indirect estimation
   a) Calculating the correlation of each variable
   b) Checking the dispersion parameters
   c) Establishing the best unit-level model with the assumption of number of unemployment events distributed poisson through modeling of General Linear Mixed Models (GLMM) by considering the effect of number of sample units as follows:
   $E(Y_{ij}|U_i) = \mu_{ij} = e^{\beta_0 + x_{ij}^T\beta + u_{i+\ln(n_{ij})}}; u_i \sim N(0, \sigma_u^2);$
   with $u_i \sim N(0, \sigma_u^2)$
   d) Performing parameter estimation using Retriscted Maximum Likelihood Method (REML)
   e) In the sub-districts surveyed, it will count the number of unemployment in villages of surveyed and not surveyed
   - In the surveyed villages, the average unemployed through equations
   $$\hat{\mu}_{ij} = e^{\hat{\beta}_0 + \bar{x}_{1ij}\hat{\beta}_1 + \bar{x}_{2ij}\hat{\beta}_2 + \hat{u}_i + \ln(n_{ij})}$$
   and number of unemployed through equations
   $$\hat{Y}_{ij} = \frac{\hat{\mu}_{ij}}{n_{ij}} * N_{ij}$$
   where n_{ij} is the sample village unit surveyed and N_{ij} is the resident population of the sub-district where i is sub-district and j is village
   - In the not surveyed villages, the average unemployment is obtained through synthetic estimators which assume that the surveyed and not surveyed villages in a sub-district have the same character ($\hat{\mu}_{ij} = \hat{\mu}_{ij*} = \hat{\mu}_i$) as follows
   $$\hat{\mu}_{ij*} = e^{\hat{\beta}_0 + \bar{x}_{1ij*}\hat{\beta}_1 + \bar{x}_{2ij*}\hat{\beta}_2 + \hat{u}_i + \ln(n_{ij*})}$$
   and number of unemployed through equations
   $$\hat{Y}_{ij*} = \frac{\hat{\mu}_{ij*}}{\hat{n}_{ij*}} * N_{ij*}$$
   where is an estimate of the number of sample village units not surveyed obtained through the average number of sample village units surveyed
   f) Calculating number of umployment as follows
   $$\hat{Y}_i = \hat{Y}_{ij} + \hat{Y}_{ij*}$$
3. Indirect estimation of non-sample areas
   a) Performing aggregation of average coefficient variable of sub-district
   b) Performing clustering based on the variable
   c) Identify an unpollected subdistrict area and then enter it based on cluster
   d) Calculating a random average of each cluster through $\hat{\bar{u}}_r = \frac{1}{m_r}\sum_{i=1}^{m_r}\hat{u}_i$ and the number of sample units through $\hat{\bar{u}}_r = \frac{1}{m_r}\sum_{i=1}^{m_r}\hat{u}_i$ where m_r is the number of sample areas in the r cluster.
   e) Doing predictions of number of unemployed in non sample-areas.

4) Calibration modeling approach.
   a) Calculating the calibration modeling approach with various combinations.
   b) Determining the best calibration approach through weighted averages as follows
   $$\hat{\bar{\mu}}_i = \frac{\sum(N_i.\hat{\mu}_{il_t})}{\sum N_i}$$

with t is the calibration approach index

3. Result and Discussion

3.1 Direct estimation

The results of direct estimation of the unemployment rate were obtained by dividing the incidence of the number of unemployment populations against the total population of the surveyed. The average unemployment rate using weighted average in Bogor regency is 0.038 or 3.8%. Table 2 represents the result of direct estimation of unemployment rate of each sub-district in Bogor district

Table 2: Unemployment rate for each sub-district is based on direct estimates

Sub-district	Unemployment rate	Sub-district	Unemployment rate
Babakan madang	0.00	Jasinga	5.88
Bojong gede	6.98	Jonggol	10.00
Caringin	0.00	Kelapa nunggal	0.00
Cariu	0.00	Kemang	10.26
Ciampea	0.00	Leuwiliang	0.00
Ciawi	0.00	Leuwisadeng	0.00
Cibinong	0.00	Megamendung	0.00
Cibungbulang	2.27	Nanggung	16.67
Cigombong	18.75	Pamijahan	0.00
Cigudeg	4.35	Parung	0.00
Cijeruk	5.41	Parung panjang	11.11
Cileungsi	8.00	Rumpin	0.00
Ciomas	0.00	Sukajaya	0.00
Cisarua	0.00	Sukamakmur	0.00
Ciseeng	0.00	Sukaraja	4.00
Citeureup	0.00	Tajur halang	0.00
Dramaga	16.67	Tamansari	5.56
Gunung putri	1.67	Tanjungsari	15.38
Gunung sindur	11.76	Tenjo	8.33

Based on the descriptive data, the maximum value of unemployment rate in Bogor district is Cigombong with 18.75%. In other words, almost one-fifth of the population in Cigombong is said to be unemployment. Furthermore, the second highest level of unemployment was Dramaga and Nanggung with the percentage of unemployment rate are 16.67%. Table 2 represents that there are 20 sub-districts that have direct estimates of unemployment rate are zero. It is not necessarily that in this sub-district has no unemployment, but the possibility that the population as unemployed is not selected as a sample unit or other possibility is that the number of small sample area units is very small.

3.2 Indirect estimation of sampled area and non-sampled area using the calibration modeling approach

The sample size of direct estimation in each sub-district in Bogor district has a small sample unit number while the population in each sub-district is large. Therefore, if the direct estimation is used as the estimated value of the unemployment rate then the error will be large. To be able to improve the results of direct estimation, its introduced the method of Small Area Estimation (SAE). Based on the considerations made, the researchers only use 2 variables in the SAE prediction model. The result of the correlation between two variables showed that both did not indicate the occurrence of multicollinearity due to the second correlation value of 0.263.

A good predictive result is obtained when the modeling is appropriate and in accordance with the shape of the data. Modeling errors are common in Poisson distribution when there is an overdispersion or underdispersion problem. In this research, we will see the form of the deviation of data dissemination through 2 ways that are by plotting between residual against average predicted [7] and through the ratio between deviance value to its degrees of freedom.

Figure 2 shows that the result of the plot of the remainder against the average value of the assumption. In Figure 2 it does not indicate that the remainder forms certain patterns or homogeneous remains so explorations do not indicate the occurrence of underdispersion or overdispersion problems.

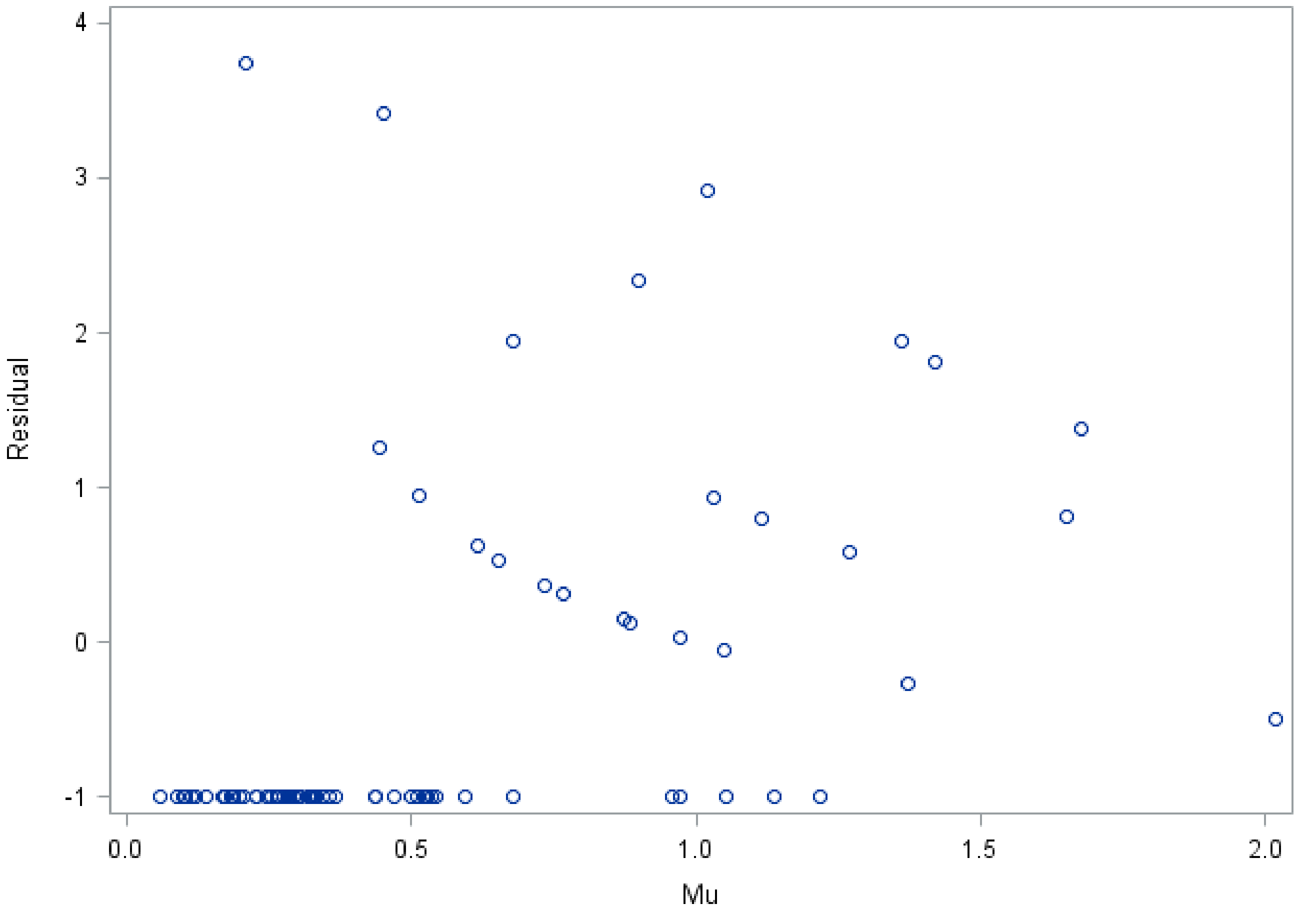

Fig. 1 Scatterplot residual against average estimated value

The second way to detect the occurrence of overdispersion or underdispersion is to divide the value of deviance to degrees of freedom. The ratio value is known as the dispersion parameter. If the dispersion parameter is greater than 1, it indicates an overdispersion problem occurring in the model and if the dispersion parameter less than 1 indicates an underdispersion problem. The good prediction model is when the dispersion parameter will be 1.

Table 3: Fit statistics of indirect estimation

Fit Statistics	
-2 res log pseudo-likelihood	319.79
Generalized chi-square	78.2
Gener. Chi-square / df	0.98

Based on the results of checking the dispersion parameters in table 4 shows that the modeling did not indicate the occurrence of overdispersion or underdispersion so that the model has been the good model for estimation. This can be seen from the value of dispersion parameters approaching 1. The confidence level used in modeling is 95% and it is seen that the significant coefficients at the 95% confidence interval are X_1 and intercept with p-value <0.05.

Table 4 : Solutions for fixed effect

Effect	Estimate	P-value
Intercept	-3.0282	<.0001
X_1	-0.2132	0.0392
X_2	2.5442	0.1996

The model in the study is a unit level model by assuming the influence of number of sample units (n) and random factor area is the sub-district area as follows

$$\hat{\mu}_{ij} = e^{-3.0282-0.2132X_{1j}+2.5442X_{2j}+u_i+\ln(n_{ij})} \qquad (1)$$

The model shown in equation (1) is an estimated form of a small village-level area of the surveyed village, furthermore for not surveyed villages are used, synthetic estimators. According to [2], the synthetic predictors are unbiased estimators derived from sample surveys (large

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

estimates) for large areas, these synthetic predictions are used for a sub-area (smaller area) assuming that the sub-area has the same characteristics to the large area. In this study, synthetic estimates are used to estimate unemployment in not surveyed villages by assuming surveyed and not surveyed villages in a sub-district have the same characteristics.

In this research, there are 2 sub-districts that are not taken as sample units. According to [1] cluster analysis of small area estimation is used in estimating small areas if there are unexpected areas as sample units. The analysis of the clustering in the study used the hierarchical method through the Wald-test to obtain the following gains.

Table 5: Forming group based on cluster analysis

Cluster	Number of clusters	Member of cluster
1	4	Gunung putri,Cileungsi,Bojong gede,Cibinong
2	11	Babakan madang,Cisarua, Ciampea, Gunung sindur, Jonggol, Taman sari, Tajur halang, Citeureup, Sukaraja, Ciomas, Parung
3	25	Ranca bungur,Tenjo laya,Caringin,Cariu,Ciawi, Cibungbulang,Cigombong,Cijeruk,Cigudeg,Ciseeng,Dramaga,Jasinga, Klapa nunggal, Kemang, Leuwiliang, Leuwisadeng, Megamendung, Nanggung, Pamijahan, Parung panjang,Rumpin,Sukajaya, Suka makmur, Tenjo, Tanjung sari

According to [5], the calibration modeling is used to estimate the predicted value which is then converted to a different measuring value (still within in the model) so that the results of the predict calibration modeling can produce an accurate model and are expected to represent an important aspect of the actual situation. Proportional weighting is considered to result in two consequences, such as measurements using proportional weighted measurements are more accurate than weightless measurements and although they have different observations, the sum of the weighted values will be equal to 1.

Table 6: The approximation formula and the result of calibration modeling approach with Q= 10.01%

No	Calibration formula	Unemployment rate
1	$\hat{\mu}_{il_1} = \frac{\frac{\hat{Y}_i}{N_i} * 100}{\frac{\sum_{i=1}^{40} \hat{Y}_i}{\sum_{i=1}^{40} N_i} * 100} * Q$	10.26%
2	$\hat{\mu}_{il_2} = \frac{\frac{\hat{Y}_i}{N_i} * 100}{\sum_{i=1}^{40} \left(\frac{\hat{Y}_i}{N_i} * 100\right)} * (Q * C)$	8.81%
3	$\hat{\mu}_{il_3} = \frac{\hat{Y}_i}{\sum_{i=1}^{40} \hat{Y}_i} * (Q * C)$	11.40%

The calibration modeling approach used in this study using proportional weighting with Q is the publication value of unemployment rate from BPS, C is the number of sub-district in Bogor district, $\hat{Y}_i$ is the estimated of number unemployment rate for each sub-district and N_i is the resident population for each sub-district. Based on the comparison using 3 calibration modeling approach shows that the best calibration model approach for this research data is obtained from formula no 1, this is indicated from the estimated average unemployment rate is 10.26%, the result is close to the estimation result of BPS publication in 2015 which is 10.01%. The results of the estimated unemployment rate in Bogor District using the calibration model approach shown in Table 7.

Table 7: Unemployment rate for each sub-district is based on indirect estimates with calibration modeling approach

Sub-district	Unemployment rate	Sub-district	Unemployment rate
Babakan Madang	6.83	Jonggol	17.06
Bojong Gede	11.69	Kelapa Nunggal	6.95
Caringin	4.90	Kemang	19.59
Cariu	6.90	Leuwiliang	7.29
Ciampea	6.20	Leuwisadeng	8.94
Ciawi	8.98	Megamendung	8.54
Cibinong	4.81	Nanggung	22.83
Cibungbulang	8.69	Pamijahan	8.60
Cigombong	27.22	Parung	5.43
Cigudeg	13.45	Parung Panjang	19.18
Cijeruk	15.73	Rumpin	5.81
Cileungsi	12.40	Sukajaya	3.77
Ciomas	6.61	Sukamakmur	6.67
Cisarua	7.99	Sukaraja	8.94
Ciseeng	6.49	Tajur Halang	7.74
Citeureup	5.59	Tamansari	13.22
Dramaga	30.66	Tanjungsari	18.61
Gunung Putri	6.67	Tenjo	12.43
Gunung Sindur	17.36	Ranca Bungur	13.50
Jasinga	7.24	Tenjolaya	11.16

4. Conclusions

Overall, the best calibration modeling for the unemployment rate is proportional to sub-district level and the unemployment rate at a district level. The result shows the calibration modeling is 10.26% and the result is BPS result that is 10.01%. Using the method can also generate predictions for the sub-district level.

References

[1] Annisa R, "Kajian Pengaruh Penambahan Informasi Gerombol terhadap Hasil Prediksi Area Nircontoh (Studi Kasus Pengeluaran per Kapita Kecamatan di Kota dan Kabupaten Bogor)", Tesis, Bogor Agriculture University.

[2] Gonzalez ME. "Use and Evaluation of Synthetic Estimators", in *Proceedings of the Social Statistics Section*. 1973, 33-36.

[3] Hanike Y, "Post Stratification sampling in Small Area Estimation (SAE) Model for Unemployment Rate Estimation by Bayes Approach, in .*Proceedings of The 7th SEAMS UGM International Conference of Mathematics and Its Application 2015*. 2015, doi 10.1063/1.4940876.

[4] Harsanti R, "Penerapan Metode Empirical Best Linear Unbiased Prediction pada Model Small Area Estimation dalam Pendugaan Tingkat Pengangguran di Kota Bogor", Skripsi, Bogor Agriculture University.

[5] Hill,CM. Methods and Guidlines for Effective Model for Calibration. Colorado: US. Geological Survey, 1998.

[6] Longford NT. "Missing Data and Small Area Estimation: Modern Analytical Equipment for the Survey Statistician." Springer. New York, 2005.

[7] Stroup WW. "Generalized Linear Mixed Models: Modern Concepts,Methods and Applications". Chapman and Hall. New York, 2013.

S. Aprizkiyandari graduate student in Department of Statistics, Bogor Agricultural University. Her main interest is on statistical modeling.

Dr. A. Kurnia Currently worked as a senior lecturer and Head of Department of Statistics, Bogor Agricultural University. His main interest are on small area estimation and statistical modeling.

Indahwati Currently worked as a lecturer and Head of Program of Applied Statistics, Bogor Agricultural University. Her main interests are on small area estimation and statistical modeling.