BAREKENG: Journal of Mathematics and Its Applications
March 2024 Volume 18 Issue 1 Page 0493–0506
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss1pp0493-0506

SMALL AREA ESTIMATION WITH HIERARCHICAL BAYES FOR CROSS-SECTIONAL AND TIME SERIES SKEWED DATA

Titin Yuniarty[1*], Indahwati[2], Aji Hamim Wigena[3]

[1,2,3]*Department of Statistics, Faculty of Mathematics and Natural Sciences, IPB University*
Dramaga Campus., Bogor, West Java, 16680,, Indonesia

*Corresponding author's e-mail: * **yuniarty_titin@apps.ipb.ac.id***

Article History:

Received: 16th October 2023
Revised: 2nd January 2024
Accepted: 23rd January 2024

Keywords:

Cross-sectional and Time Series;
Hierarchical Bayes;
SAE;
Skew-normal.

ABSTRACT

Small Area Estimation (SAE) is a method based on modeling for estimating small area parameters, that applies Linear Mixed Model (LMM) as its basic. It is conventionally solved with Empirical Best Linear Unbiased Prediction (EBLUP). The main requirement for LMM to produce high precision estimates is normally distributed. The observation unit is food crop farmer households from Sulawesi Tenggara Province to estimate food and non-food per capita expenditure at the district/city level using SAE that has been positively skewed. Applying EBLUP for positively skewed data will result less accurate estimates. Meanwhile, transformation will be potentially result biased estimates. Therefore, the problem of skewed data and small area level in this research was completed by Hierarchical Bayes (HB) on combination cross-sectional and time series under skew-normal distribution assumption. The results obtained were skew-normal SAE HB model was significantly reducing Relative Root Mean Squared Error (RRMSE) than the direct estimation. It indicates that SAE modeling is able to provide a shrinkage effect on the direct estimation results. However, there is a slight difference in interpretating between direct estimation and skew-normal SAE HB. It is possible because the modeling used the assumption that the autocorrelation coefficient is equal to 1 or known as the random walk effect. However, in reality, Susenas is not a panel data, so unit of observation for each time period may be different. Therefore, further research should be compared with the skew-normal or another skewed distribution that assumes the autocorrelation coefficient is unknown and should be estimated in the model.

How to cite this article:

T. Yuniarty, Indahwati and A. H. Wigena., "SMALL AREA ESTIMATION WITH HIERARCHICAL BAYES FOR CROSS-SECTIONAL AND TIME SERIES SKEWED DATA," *BAREKENG: J. Math. & App.*, vol. 18, iss. 1, pp. 0493-0506, March, 2024.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Small Area Estimation (SAE) is a method of estimating parameters in a small area or domain, where the sample size is not sufficient if a direct estimation carried out. This method is known as 'indirect estimation' that is borrowing information strength from other adjacent areas by formatting the appropriate models. The resulting estimates are more efficient than direct estimates because they utilize the strength of the relationship between auxiliary variables and interest variables to increase the effectiveness of the sample size [1]. A good auxiliary variable is able to explain variation between small areas and is not susceptible to small sample sizes. These auxiliary variables can be obtained from census data or administrative compilations [2].

SAE studies mostly review social survey analysis that generally have skewed data distributions [3]. As a model-based estimation method, SAE applies a linear mixed model (LMM) to the basic model which has two sources of diversity, namely diversity from fixed effects and diversity from random effects. It is conventionally solved by using Empirical Best Linear Unbiased Prediction (EBLUP). SAE is grouped into two types of basic models based on the availability of auxiliary variables, namely area-level models which was first introduced by Fay-Herriot on 1979 and unit-level models by Battese, Harter and Fuller on 1988. The EBLUP basic assumption is the normality distribution of the area random effect to get precise estimation results. The skewed data causes a violation of the normality assumption. Several data transformation studies are applied to normal LMM, including logarithmic transformation [4] or Box-Cox and Dual Power method [5]. However, estimating the scale of the original data requires back transformation which has the potential to be a source of bias and the difficulty of obtaining the mean square error (MSE) value [3]–[5]. Another approach that can be used because of its regarding normality assumptions flexibility is Hierarchical Bayes (HB) approach. The main principle of HB is to estimate small-area parameters based on the posterior distribution. The posterior distributions of the parameters are not always obtained in closed forms. Therefore, it is necessary to use numerical approximation by using the Markov Chain Monte Carlo (MCMC) [6].

Per capita expenditure is one of the socio-economic data that tends to be skewed the right [7]. This data is important to measure on a micro-scale because it can be an indicator of achieving economic prosperity. It was obtained from estimates from the annual National Socio-Economic Survey (Susenas) in March at the district/city level and divided into expenditure on food and non-food commodities. Improving the economic welfare of society is a global goal of sustainable development goals (SDGs). This is also the case in the agricultural sector, namely to achieve food self-sufficiency so that improving farmers' welfare is the main target that must be achieved in agricultural development. The farmers welfare can be seen by their ability to satisfy household basic needs such as clothing, food, shelter, education, and health [8]. Meanwhile, so far, farmers' welfare has been measured by the Farmer's Exchange Rate (NTP) which is calculated from the ratio of the price index received by farmers for the results of their farming (It) and the price index paid by farmers (Ib) for the production and consumption expenditure of the farmer's household. [9]. However, NTP does not represent the farmers welfare because it is assumed that production is constant and only prices change. On the other hand, There is an increasing necessity for farmer welfare measures that are not only theoretically and empirically valid but also practical and cheap to implement, especially for regular monitoring instrument and evaluation of agricultural development performance [10], [11]. Therefore, this research makes an analogy of farmer's welfare from per capita expenditure data for farmer households. This research uses Sulawesi Tenggara Province per capita expenditure for food crop farmer households by districts/cities level, which is based on data pre-processing results tends skewed to the right. This province was chosen because it has a stable trend below 100 for food crop farmer exchange rate during 2009-2021 which is interpreted as a low purchasing power of farmers for farming income or as an indication of a low level of farmer welfare in a region. Through the average per capita expenditure for food and non-food commodities, the level of welfare of the population view in a region can be obtained. If the composition of household expenditure is still dominated by food commodities, the population welfare will still be relatively low [12].

SAE modeling on skewed data that considers the original scale of data was carried out by [13]. Two extensions of the Fay–Herriot small area level model are proposed, i.e. it allows for non-symmetrically distributed sampling errors and proposed to jointly model the direct survey estimator and its variance. However, the agricultural sector is closely related to the continuity of the production process on a regular basis, so it would be very valuable if the SAE estimator was obtained from several points in time. That is why this research uses a combination of cross-section and time series data for the 2018-2021 time period by considering data availability, apart from the fact that applying the area level in this research will produce a small number of observations, namely 17 districts/cities in Sulawesi Tenggara Province. In surveys carried out periodically such as Susenas, the efficiency of estimates can be increased by combining cross-sectional

and time series data [14], [15]. Therefore, this research applied the Rao-Yu small area level model for combination of cross-sectional and time series as extension of the Fay-Heriot. Meanwhile, SAE modeling on combined cross-sectional and time series data which is assumed to follow a skew-normal distribution was carried out in [16] using HB approach. The research considered domain and time random effects on the intercept and slope model, unlike Rao-Yu model, which is reasonable assessed to suppose influence of lag random effects when working with economic skewed data. By the several proposed model, the model with random walk effects is better than the others, which is almost the same as the Rao-Yu model.

Thus, to obtain view of welfare conditions for the food crop farmer household at the district/city level in Sulawesi Tenggara Province more effectively and efficiently, this research aims to estimate small area parameters, namely the mean of per capita expenditure for food and non-food commodities respectively. This is because changes in the ratio of the two types of expenditure can be a benchmark for the level of social welfare [12]**.** Due to limitations in data processing, the two types of expenditure is assumed to be a single variables in this research.

2. RESEARCH METHODS

This research will compare the effectiveness of direct estimation results and SAE HB with combined cross-sectional and time series data which are assumed to follow a skew-normal distribution. Data processing uses R version 4.1.3 and RStudio version 2022.07.1+554 software with the help of 'BRugs' packages through the following stages: (1) Carrying out direct estimates, (2) Preparing auxiliary variables for SAE modeling, namely those closely correlated with the interest variable from direct estimation results, (3) Forming a skew-normal SAE HB model, (4) Comparing the direct estimation and skew-normal SAE HB, (5) Calculating the ratio of per capita expenditure on food crop farmer household based on the best method.

2.1 Data

The data used empirical data from BPS-Statistics Indonesia for 2018-2021, namely March Susenas, Village Potential Data Collection (Podes) for Sulawesi Tenggara Province along with the Sulawesi Tenggara Publication in Figures (DDA). The data is divided to the type of variable, namely interest variable data and auxiliary variable data. The interest variable (Y) is the SAE interest variable, which consists of the food per capita expenditure (Y_1) and non-food per capita expenditure (Y_2) variables expressed in a thousand rupiahs. Variable Y comes from the March Susenas with the observation unit in the form of food crop farmer household at the district/city level in Sulawesi Tenggara Province. Auxiliary variables (X) are variables that provide additional information for SAE modeling, obtained from Podes and DDA. Determination of auxiliary variables refers to the study [17], where the farmer welfare model is compiled from production indicators of farming, including climate, infrastructure, social-economics, policies, institutions, and production techniques. There are 13 candidate of auxiliary variables, namely:

X_1 : Amount of rainfall per year (thousand mm)

X_2 : Proportion of food crop centers villages/sub-districts that have irrigation canals/dams/reservoirs/embung for irrigating agricultural land

X_3 : Proportion of food crop centers villages/sub-districts that have good condition farming roads

X_4 : Population dependency ratio

X_5 : Ratio of natural disasters in food crop centers villages/sub-districts

X_6 : Ratio of active village unit cooperatives (KUD) in food crop centers villages/sub-districts

X_7 : Ratio of farmer groups in food crop centers villages/sub-districts

X_8 : Ratio of Indigent Certificates (SKTM) from government of the food crop centers villages/sub-districts per 100 populations

X_9 : Ratio of residents suffering from malnutrition in food crop centers villages/sub-districts

X_{10} : Ratio of slum settlements families in food crop centers villages/sub-districts

X_{11} : Ratio of Elementary School (SD/MI) in food crop centers villages/sub-districts per 100 populations

X_{12} : Ratio of Junior High School (SMP/MTS) in food crop centers villages/sub-districts per 100 populations

X_{13} : Ratio of health workers in food crop centers villages/sub-districts per 100 populations

2.2 Direct Estimation

Direct estimation is a classic method of estimating small area parameters based on a specific survey design model [18]. In this study, the average per capita expenditure on food crop farmers household at the district/city level was estimated using weighted probability sampling (WPS) because it followed the Susenas sampling technique, namely two stage one phase.

$$\hat{\bar{Y}}_{it}^{\ WPS} = \frac{1}{\sum_{j=1}^{n_{it}} w_{(0)ijt}} \sum_{j=1}^{n_{it}} w_{(0)ijt}\, y_{ijt} \quad (1)$$

where, y_{ijt} is food/non-food per capita expenditure for unit *j*-th in district/city *i*-th year *t*-th, n_{it} is number of sample in district/city *i*-th year *t*-th, $w_{(0)ijt}$ is sample weighting for unit *j*-th in district/city *i*-th year *t*-th.

The goodness of parameter estimates can be determined by calculating the Mean Squared Error (MSE) value of the direct WPS estimator obtained, namely:

$$MSE\left(\hat{\bar{Y}}_{it}^{\ WPS}\right) = \frac{s_{it}^2}{\left(\sum_{j=1}^{n_{it}} w_{(0)ijt}\right)^2}\left(\sum_{j=1}^{n_i} {w_{(0)ijt}}^2\right) \quad (2)$$

$i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n_{it}$, $s_{it}^2 = \frac{1}{n_{it}-1}\sum_{j=1}^{n_{it}}\left(\hat{\bar{Y}}_{\mathrm{it}}^{\ WPS} - y_{ijt}\right)^2$, $RSE\left(\hat{\bar{Y}}_{it}^{\ WPS}\right) = \frac{\sqrt{MSE\left(\hat{\bar{Y}}_{it}^{\ WPS}\right)}}{\hat{\bar{Y}}_{it}^{\ WPS}} * 100\%$.

2.3 SAE Cross-sectional and Time Series Model

Estimating small area parameters from periodic survey data, such as Susenas, the estimation efficiency can be increased by including the random effects of area and time in small area specific variance component. This model was popularly introduced by Rao and Yu [19], as a development of the area level model developed by Fay and Herriot [18], which consists of a sampling error model.

$$\hat{Y}_{it} = \theta_{it} + e_{it};\ t = 1, \ldots, T;\ i = 1, \ldots, m \quad (4)$$

and the small area model:

$$\theta_{it} = \boldsymbol{x_{it}^T\beta} + v_i + u_{it} \quad (5)$$

If model (4) and (5) are combined, then will be:

$$\hat{Y}_{it} = \boldsymbol{x_{it}^T\beta} + v_i + u_{it} + e_{it} \quad (6)$$

where $\boldsymbol{\beta}$ is parameter coefficient from variable $\boldsymbol{x_{it}^T}$ which is assumed to be constant, $\boldsymbol{x_{it}^T}$ is vector of fixed auxiliary variables for area *i*-th on year *t*-th, $\hat{Y}_{it}$ is direct estimation for area *i*-th on year *t*-th, θ_{it} is mean function for small area *i*-th on year *t*-th, and e_{it} is sampling error in normal distribution with expected value 0 and variance covariance diagonal matrices $\boldsymbol{\psi_i} = blockdiag(\boldsymbol{\psi_1}, \ldots, \boldsymbol{\psi_m})$ that might be changed by time, $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$ and u_{it} follow this following process:

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it},\ |\rho| < 1 \text{ and } \varepsilon_{it} \overset{iid}{\sim} N(0, \sigma_u^2) \quad (7)$$

with $\{e_{it}\}$, $\{v_i\}$, and $\{\varepsilon_{it}\}$ assumed independently. θ_{it} is area effect $v_{i,}$ and u_{it} is area- time effect.

If the variance of random area effect σ_v^2, the variance of random area-time effect σ_u^2, and the autocorrelation coefficient ρ are unknown and substituted with its estimator respectively $\hat{\sigma}_v^2$, $\hat{\sigma}_u^2$, and $\hat{\rho}$, then the estimator obtained is called EBLUP. The EBLUP estimator for θ_{it} is:

$$\hat{\theta}_{it}^{\ EBLUP} = w_{iT}^*\hat{Y}_{it} + (1 - w_{iT}^*)\boldsymbol{x_{it}^T\hat{\beta}} + \sum_{j=1}^{T-1} w_{ij}^*\left(\hat{Y}_{it} - \boldsymbol{x_{it}^T\hat{\beta}}\right) \quad (8)$$

where $(w_{i1}^*, \dots, w_{i2}^*) = (\hat{\sigma}_v^2 \mathbf{1}_T + \hat{\sigma}_u^2 \boldsymbol{\gamma}_T)^T V_i^{-1}$; γ_T is the T-th row of $\boldsymbol{\Gamma}$; $\boldsymbol{\Gamma}$ is a TxT matrix with elements $\hat{\rho}^{|i-j|}/(1-\hat{\rho}^2)$; $\boldsymbol{V} = block\ diag_i(\boldsymbol{\psi_i} + \hat{\sigma}_u^2 \boldsymbol{\Gamma} + \hat{\sigma}_v^2 \boldsymbol{J}_T) = block\ diag_i(\mathrm{V}_i)$; $\boldsymbol{J}_T = \boldsymbol{Z}\boldsymbol{Z}^T$; $\boldsymbol{Z} = \boldsymbol{I}_m \otimes \mathbf{1}_T$; $\mathbf{1}_T$ is a vector of 1's; $\boldsymbol{I}_m$ is the identity matrix of order m, and $\otimes$ denotes the direct product.

The MSE value of the EBLUP estimator is used as a measure of the goodness of the resulting EBLUP estimator. An exact expression for $MSE\left(\hat{\theta}_{it}^{EBLUP}\right)$ is given by:

$$MSE\left(\hat{\theta}_{it}^{EBLUP}\right) = g_{1iT}(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) + g_{2iT}(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) \quad (9)$$

where

$g_{1iT}(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) = \hat{\sigma}_v^2 + \frac{\hat{\sigma}_u{}^2}{1-\hat{\rho}^2} - (\hat{\sigma}_v^2 \mathbf{1}_T + \hat{\sigma}_u^2 \boldsymbol{\gamma}_T)^T V_i^{-1} (\hat{\sigma}_v^2 \mathbf{1}_T + \hat{\sigma}_u^2 \boldsymbol{\gamma}_T)$, and

$g_{2iT}(\hat{\sigma}_u^2, \hat{\sigma}_v^2, \hat{\rho}) = \left\{\boldsymbol{x}_{iT} - \boldsymbol{X_i^T V_i^{-1}} (\hat{\sigma}_v^2 \mathbf{1}_T + \hat{\sigma}_u^2 \boldsymbol{\gamma}_T)\right\}^T \left(\boldsymbol{X}^T \boldsymbol{V^{-1}} \boldsymbol{X}\right)^{-1} x \left\{\boldsymbol{x}_{iT} - \boldsymbol{X_i^T V_i^{-1}} (\hat{\sigma}_v^2 \mathbf{1}_T + \hat{\sigma}_u^2 \boldsymbol{\gamma}_T)\right\}$

So, $RRMSE\left(\hat{\theta}_{it}^{EBLUP}\right) = \frac{\sqrt{MSE\left(\hat{\theta}_{it}^{EBLUP}\right)}}{\hat{\theta}_{it}^{EBLUP}} * 100\%$ is a measure of the model goodness-of-fit.

2.4 SAE HB for Skew-normal Distribution

The research uses a cross-sectional and time series model assuming $\rho = 1$, then $u_{it} = u_{i,t-1} + \varepsilon_{it}$. It is the so-called random walk effect which assumes that from one period to the next, the original time series merely takes a random step away from its last recorded position. Random walk effects have the advantage of being simpler and faster to fit **[15]**. SAE modeling with HB approach assumes that all unknown parameters are considered as variables and have a distribution **[20].** Estimation of small area parameters is carried out on the posterior distribution, which is the result of multiplying prior distribution and likelihood function of the observed data $f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$.

Y is the SAE interest variable which is assumed to follow skewness distribution. Although the direct estimate of Y is a linear combination of individual observations that can be assumed to be normally distributed for large samples, but not for small samples **[3]**. For this reason, in this study the direct estimate of the average Y is assumed to follow a skew-normal distribution, with the following probability density function:

$$Y|\mu \sim SN(\mu, \sigma, \lambda) \Leftrightarrow f_Y(y) = \frac{2}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\lambda \frac{y-\mu}{\sigma}\right) \quad (10)$$

where $\Phi(.)$ is the cumulative distribution function, $\phi(.)$ is the probability density function from standardized normal distribution, and the parameters μ, σ, and λ respectively are the location, scale, and skewness parameters. The mean and variance of the skew-normal distribution are given by:

$$E(Y) = \mu + \delta\sqrt{\frac{2\sigma}{\pi}} \text{ and } V(Y) = \sigma\{1 - 2\pi^{-1}\delta^2\} \text{ where } \delta = \lambda/\sqrt{1+\lambda^2} \quad (11)$$

The implementation of HB for $\hat{\bar{\boldsymbol{Y}}}_{\boldsymbol{it}}$ ~ skew-normal, is structured in the hierarchical framework:

Level 1: $\hat{\boldsymbol{Y}}_{\boldsymbol{it}}|\boldsymbol{\theta}_{\boldsymbol{it}}, \lambda, n_{it}, \phi_i \sim SN\left(\boldsymbol{\theta}_{\boldsymbol{it}}, \sqrt{\psi_i}, \lambda/\sqrt{n_{it}}\right)$

$\psi_{it}|n_{it}, \phi_i \sim G\left(\frac{1}{2}(n_{it}-1), \frac{1}{2}(n_{it}-1)\phi_i{}^{-1}\right) \rightarrow$ sampling variance

$\phi_i{}^{-1}|a_\phi, b_\phi \sim G(a_\phi, b_\phi) \rightarrow$ parameter of scale

$\lambda \sim N(0; 0.01) \rightarrow$ parameter of skewness

Level 2: $\boldsymbol{\theta}_{\boldsymbol{it}}|\boldsymbol{\beta}, u_{it}, \tau_v \sim N\left(\boldsymbol{x}_{\boldsymbol{it}}^{\boldsymbol{T}}\boldsymbol{\beta} + u_{it}, \tau_v\right)$

Level 3: $u_{it}|u_{i,t-1}, \tau_u \sim N\left(u_{i,t-1}, \tau_u\right)$

Level 4: $f(\boldsymbol{\beta}, \tau_v, \tau_u) = f(\boldsymbol{\beta}) f(\tau_v)\, f(\tau_u)$

$$f(\boldsymbol{\beta}) \sim N\left(\hat{\beta}_k, \frac{1}{\left(se(\hat{\beta}_k)\right)^2}\right), \tau_v \sim G(a_v, b_v),\ \tau_u \sim G(a_u, b_u) \quad (12)$$

where, $a_\phi \sim G(0.01, 0.01)$ and $b_\phi \sim G(0.01, 0.01)$, $\tau_v = 1/\sigma_v^2$ and $\tau_u = 1/\sigma_u^2$, while prior parameter value for τ_v is $\tau_v \sim G(0.01; 0.01)$ and $\tau_u \sim G(0.01; 0.01)$, n_{it} is sample size for area i-th and year t-th.

The mean posterior of $(\boldsymbol{\theta}|\boldsymbol{y})$ in the HB approach is used as an estimate of the positional point and variance of $V(\boldsymbol{\theta}|\boldsymbol{y})$ as measure of diversity. We define the MCMC sample as $\left\{\left(\beta^{(s)}, \theta^{(s)}, {\sigma_v^2}^{(s)}, {\sigma_u^2}^{(s)}\right), s = h+1, \dots, h+H\right\}$ with posterior mean from close form exploration when σ_v^2, ${\sigma_u}^2$, and ρ are known:

$$\hat{\theta}_{it}^{\ HB-SN} = \frac{1}{H}\sum_{s=h+1}^{h+H} \hat{\theta}_{it}^{\ B}\left(\sigma_v^{2(s)}, \sigma_u^{2(s)}, \rho^{(s)}\right) = \hat{\theta}_{it}^{\ B}(.) \quad (13)$$

and the variance posterior:

$$MSE\left(\hat{\theta}_{it}^{\ HB-SN}\right) = \hat{V}\left(\hat{\theta}_{it}^{\ HB-SN}|\hat{\boldsymbol{Y}}_{\boldsymbol{it}}\right) = \frac{1}{H}\sum_{s=h+1}^{h+H}\left[g_{1iT}\left(\hat{\sigma}_u^{\ 2}, \hat{\sigma}_v^2, \hat{\rho}\right) + g_{2iT}\left(\hat{\sigma}_u^{\ 2}, \hat{\sigma}_v^2, \hat{\rho}\right)\right] + \frac{1}{H-1}\sum_{s=h+1}^{h+H}\left[\hat{\theta}_{it}^{\ B}\left(\sigma_v^{2(s)}, \sigma_u^{2(s)}, \rho^{(s)}\right) - \hat{\theta}_{it}^{\ B}(.)\right] \quad (14)$$

3. RESULTS AND DISCUSSION

3.1 Overview of Direct Estimation 2018-2021

The average sub-sample size of food crop farmer household in Sulawesi Tenggara Province by Susenas data in **Table 1** is only 1 percent of the estimated sub-population from food crop farmer household in Sulawesi Tenggara Province. Meanwhile, the average when compared with Susenas sample size in each time period is aproximately 12 percent. On average, the sub-sample size of the Susenas sample is relatively small. The term small in small area is not just seen from the relative size of the sub-sample to the sample in a survey, more than that, when sub-samples are repeatedly sampled from sub-populations it will produce different direct estimates. If these direct estimates are averaged, they will produce very large variations. That is why small area estimates are needed because Susenas is not designed to estimate the characteristics of food crop farmer household, especially at the district/city level when the average sub-sample percentage of the sub-population is only 1 percent which is very small.

Table 1. Summary of Direct Estimation, 2018-2021

Summary	Food Expenditure				Non-food Expenditure			
	2018	2019	2020	2021	2018	2019	2020	2021
Statistics Central Measure (Thousand Rupiahs)								
Minimum	51.32	59.14	123.43	103.53	45.32	58.08	32.57	25.46
Average	397.68	403.96	453.04	436.76	356.19	359.47	360.87	376.78
Maximum	2262.86	1956.77	1975.71	2622.86	7045.00	2785.58	5421.67	7377.36
Standard of deviation	240.93	252.65	243.40	236.94	367.42	297.40	370.17	463.65
Others Statistic Measure								
Sub-sample size	746	1065	1232	1285	746	1065	1232	1285
Sample size	6141	8710	9164	9216	6141	8710	9164	9216
Sub-population size	72879	69677	77774	89743	72879	69677	77774	89743

Further observation found that the data distribution, per capita expenditure on food crop farmer household in Sulawesi Tenggara Province, both food and non-food during 2018-2021, is not distributed normally. This can be clearly seen from **Figure 1**. The skewness of the data distribution makes it difficult to fulfill the normality assumption for the variance component while still using conventional SAE methods, such as EBLUP Rao-Yu. Therefore, an alternative approach to SAE modeling used in this research is HB because it adaptively determines the distribution based on the data.

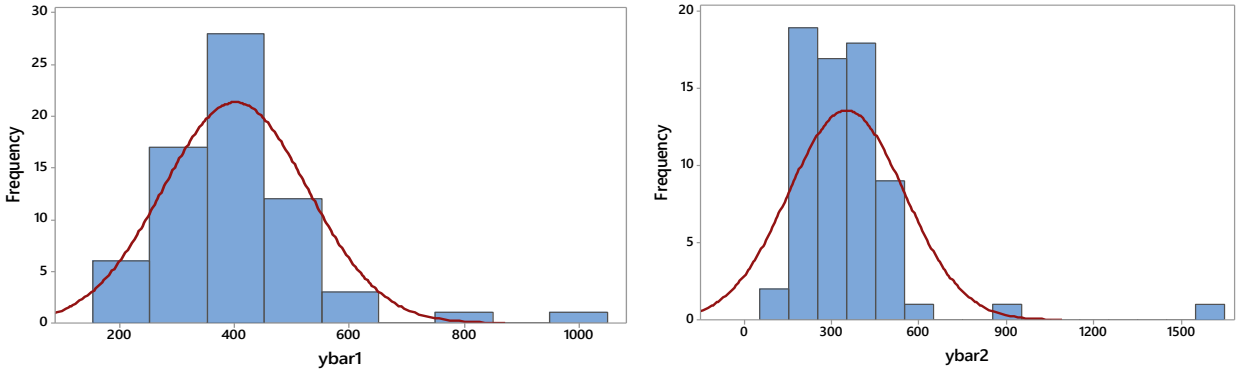


Figure 1. Histogram of Direct Estimates of Per Capita Expenditure on Food Crop Farmer Household at the District/City Level in Sulawesi Tenggara Province 2018-2021 for (a) food and (b) non-food commodities

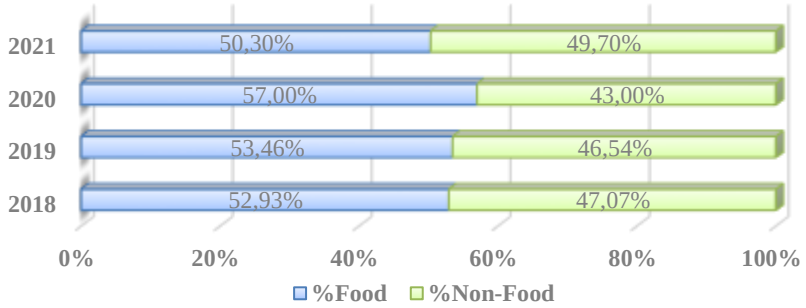


Figure 2. Comparison of Food and Non-food Expenditure Ratio (%) for Food Crop Farmer Household at Sulawesi Tenggara Province, 2018-2021

If we observe the average per capita expenditure on food crop farmer household for 2018-2021 based on direct estimation results, the ratio of food per capita expenditure is still greater than non-food. This is shown in **Figure 2** where the ratio of food per capita expenditure was stable at above 50 percent during 2018-2021. In addition, the significant increase in the ratio of per capita food expenditure in 2020 indicates that the food crops farmer household in Sulawesi Tenggara Province is a households group that has been significantly impacted by the Covid-19 pandemic as income decreases.

3.2 Auxiliary Variable Selection

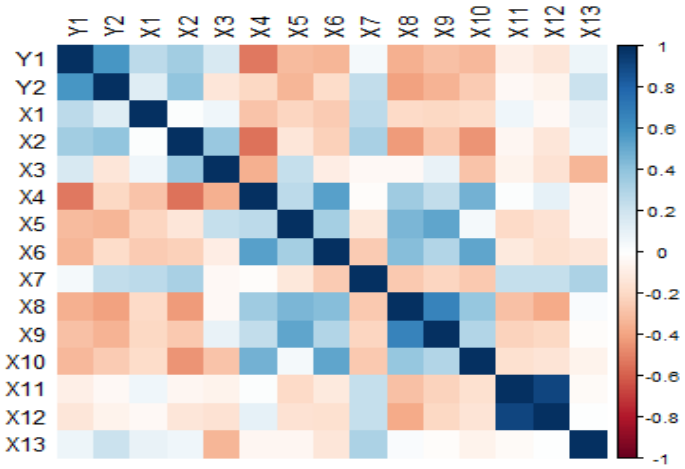


Figure 3. Visualization of Pearson Correlation between Research Variable

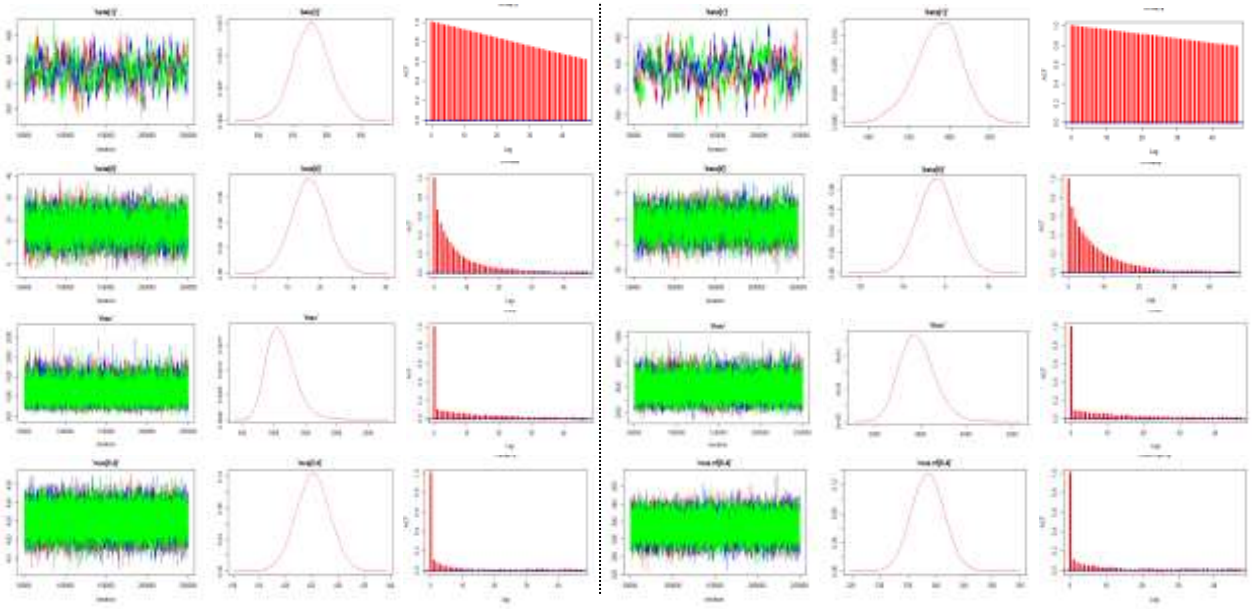
Before estimate the mean of food and non-food per capita expenditure for food crop farmer household using the SAE method, the auxiliary variables must first be selected to be included in the SAE modeling. The process of selecting auxiliary variables is carried out by calculating the Pearson correlation between each of 13 candidate auxiliary variables and the interest variable. The results of the correlation analysis are visualized in **Figure 3** by observing at the color gradations, where for variable Y_1 there are 10 variable that are quite

closely correlated, namely X_1, X_2, X_3, X_4, X_5, X_6, X_8, X_9, X_{10}, and X_{12}. Meanwhile, for the variable Y_2 there are 10 variables, namely X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10} and X_{13}. Apart from observing the respective correlations with the interest variables Y_1 and Y_2, it is very important to examine the correlation between the auxiliary variables. The Pearson correlation results show that the variables X_{11} and X_{12} has an almost perfect linear relationship, which is close to 1. Therefore, to avoid increasing error in the regression parameter estimate, only one of the two variables was chosen to be included in the SAE model. Variable X_{12} was chosen because it has a greater Pearson correlation value to the interest variable. Thus, the auxiliary variables that will be included in the SAE model are X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{12}, and X_{13}, both for interest variable Y_1 and Y_2.

3.3 Skew-normal SAE HB Model

The estimated parameters in the model equation (6) are solved using HB approach with the assumption that the estimated parameters directly follow a skew-normal distribution as described hierarchically in equation (10). Because all parameters in the model are considered as variables and have a distribution, to solve the integral complexity in the likelihood function and the posterior distribution of each parameter, the MCMC approach with the Gibbs Sampling algorithm is used. In building a model, for both food and non-food commodities, the number of total iterations, iterations for burn-in period, and thin must first be determined by trial and error, starting from a small value until algorithm convergence is achieved for all model parameters. After going through several simulation stages, the best model combination was obtained when the number of Markov chain (numChains) was 3 with the number of iterations for each chain is being 200,000, of which 50,000 iterations for the burn-in period and 150,000 iterations for the posterior analysis. This chain is then thinned by taking every 10th sample value to reduce autocorrelation between the sample generated. Convergent conditions can be seen from **Figure 4** by the trace plot, the density plot and the autocorrelation plot for the following parameters.

Figure 4 shows a visualization of parameter convergence after going through the MCMC process. The green plot is a trace plot when convergence is achieved, where the mean value is quite stable and does not form a particular pattern. Likewise, visualization from density plot achieved when it tends to be symmetrical, so it resembles a normal distribution curve. Meanwhile, from an autocorrelation plot, convergence obtained when initial value starts close to one but slowly decreases towards zero as the lag increases and it shows that the algorithm is already in target distribution area. Different visualizations of convergence are seen in β_1, especially in the trace plot and the autocorrelation plot. This happens because the MC error in β_1 is relatively larger than other parameters, so the convergence process is slow. However, these parameters have converged because there is relatively no special pattern that forms as the iterations increase in the trace plot, whereas in the autocorrelation plot there appears to be a decrease with each increase in lag, although it tends to be slow.



(a) (b)

Figure 4. Visualization of Convergence with Trace Plot, Density Plot, and Autocorrelation Plot for Expenditure of: (a) Food and (b) Non-food

After MCMC simulation process with the Gibbs Sampling algorithm succeeded in obtaining convergent parameters, the next step was to analyze the posterior distribution for the fixed effect parameter β_k from 12 selected variables included in the modeling process. Based on a 95 percent confidence interval that does not contain zero, there are 10 auxiliary variables that have a significant effect on the average food per capita expenditure for food crop farmer household at district/city level in Sulawesi Tenggara Province. Variables that have a positive influence are the amount of rainfall (X_1), the proportion of villages/subdistricts that have irrigation channels/dams/reservoirs/embung for irrigating agricultural land (X_2), natural disasters ratio (X_5), slum settlements families ratio (X_{10}), ratio SMP/MTS per 100 population (X_{12}), and ratio of health worker per 100 population (X_{13}); while villages/subdistricts with good condition farming roads proportion (X_3), ratio of population dependency (X_4), farmer group ratio (X_7), and ratio of people suffering from malnutrition (X_9) has a negative influence on the average food per capita expenditure for food crop farmer household at district/city level in Sulawesi Tenggara Province. The equation of the skew-normal SAE HB model for food commodities is as follows:

$$\hat{Y}_{1.it} = \mathbf{577.30} + \mathbf{17.01}X_1 + \mathbf{87.78}X_2 - \mathbf{38.79}X_3 - \mathbf{594.60}X_4 + \mathbf{16.69}X_5 - 17.67X_6 - \mathbf{10.87}X_7 - 6.31X_8 - \mathbf{34.21}X_9 + \mathbf{27.63}X_{10} + \mathbf{618.50}X_{12} + \mathbf{170.10}X_{13} + v_i + u_{it}$$

which $\hat{\sigma}_v^2 = 1114$, $\hat{\sigma}_u^2 = 0.9928$, and time autocorrelation $\hat{\rho} = 0.9772$.

In the skew-normal SAE HB model for non-food commodities, there are 10 auxiliary variables that have a significant effect, based on 95 percent confidence interval. Among these variables, there are 5 variables that have a positive influence, namely ratio of village unit cooperatives (KUD) with active status (X_6), ratio of residents suffering from malnutrition (X_9), ratio of families in slum settlements (X_{10}), ratio of SMP/MTS per 100 population (X_{12}), and health worker per 100 population ratio (X_{13}). Meanwhile, amount of rainfall (X_1), proportion of villages/subdistricts that have irrigation channels/dams/reservoirs/embung for irrigating agricultural land (X_2), population dependency ratio (X_4), farmer groups ratio (X_7), and SKTM ratio per 100 population (X_8) is a variable that has a negative effect on average non-food per capita expenditure for food crop farmer household at the district/city level in Sulawesi Tenggara Province. The skew-normal SAE HB model for non-food commodities is expressed in the following equation:

$$\hat{Y}_{2.it} = \mathbf{586.20} - \mathbf{32.87}X_1 - \mathbf{132.70}X_2 + 23.34X_3 - \mathbf{357.50}X_4 - 2.07X_5 + \mathbf{695.30}X_6 - \mathbf{23.11}X_7 - \mathbf{141.80}X_8 + \mathbf{43.43}X_9 + \mathbf{20.57}X_{10} + \mathbf{252.90}X_{12} + \mathbf{145.10}X_{13} + v_i + u_{it}$$

with $\hat{\sigma}_v^2 = 2936$, $\hat{\sigma}_u^2 = 1.935$, and time autocorrelation $\hat{\rho} = 0.9785$.

3.4 Comparison between Direct Estimation and Skew-normal SAE HB Model

Table 2. Comparison Between The Method of Direct Estimation and SAE HB Skew-normal for Per Capita Expenditure of Food Crop Farmer Household in Sulawesi Tenggara Province, 2018-2021

Num.	Value for all Small Area	Method of Direct Estimation			Method of SAE HB skew-normal		
		$\hat{\theta}$	MSE	RRMSE	$\hat{\theta}^{HB-SN}$	MSE	RRMSE
	Food Expenditure						
1	Minimum	172.47	12.03	1.07	212.80	6.06	0.66
2	First Quartile	328.29	43.79	1.66	313.90	11.51	0.95
3	Mean	399.80	9 051.51	8.37	380.57	240.43	2.33
4	Median	392.78	101.18	2.39	363.90	21.72	1.32
5	Third Quartile	449.41	648.65	7.49	444.93	79.81	2.89
6	Maximum	983.35	221 298.23	99.71	601.50	6 416.01	15.47
	Non-food Expenditure						
1	Minimum	135.87	5.18	1.25	106.70	2.60	0.72
2	First Quartile	227.94	28.06	1.89	233.10	9.66	1.06
3	Mean	349.82	14 255.61	8.80	306.60	312.88	2.95
4	Median	336.41	96.79	2.96	299.00	21.23	1.43
5	Third Quartile	403.49	523.46	7.54	375.85	64.41	3.24
6	Maximum	1 579.49	412 750.64	99.71	559.60	8 361.27	18.29

The results of applying method of direct estimation and SAE HB to per capita expenditure on food crop farmer household at district/city level in Sulawesi Tenggara Province 2018-2021 which is assumed to follow skew-normal distribution are presented in **Table 2**. $\hat{\theta}$ is direct estimate of the average per capita expenditure on food crop farmer household at district/city level based on Susenas sampling method, while $\hat{\theta}^{HB-SN}$ is an estimate of per capita expenditure on food crop farmer household (in thousand rupiahs) at the district/city level based on the model SAE HB approach uses a skew-normal distribution for 2018-2021 Susenas data. RRMSE describes the magnitude of the relative error rate of the estimated results, which is standardized to remove the unit factor or relative value of the mean root mean squared error **[21]**.

Based on **Table 2** and **Figure 5**, it can be seen that overall the RRMSE value in the skew-normal SAE HB model is smaller than the direct estimation method. On average, the direct estimator RRMSE was 8.37, decreasing to 72.15 percent for food commodities and for non-food commodities, namely 8.80 or decreasing to 66.46 percent in the skew-normal SAE HB method. This shows that the area random effect and the area-time random effect function to calibrate the results of direct estimates based only on survey data. The decrease in RRMSE is a result of the decomposition of the variance components contained in the SAE model, which consists of sampling variance (σ_e^2), area random variance (σ_v^2), and area and time random variance (σ_u^2). If observed from the lowest to highest value range, the RRMSE in the skew-normal SAE HB method was recorded less than 25 percent, which means that it satisfied to be considered as the accurate statistics.

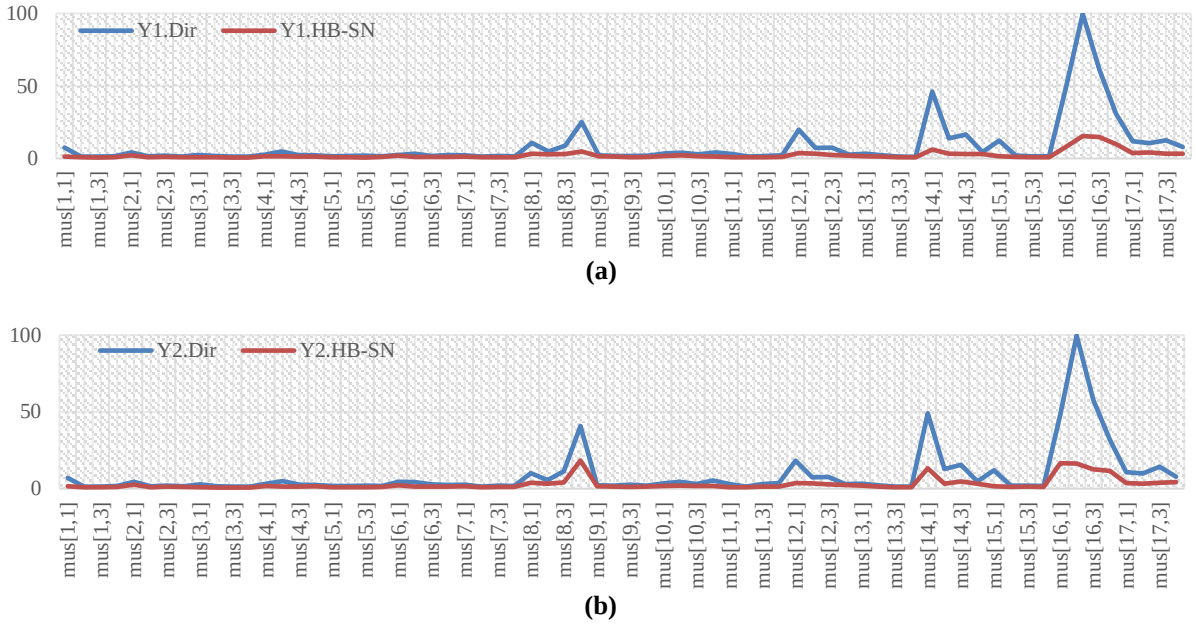


Figure 5. RRMSE Plot of Direct Estimate and SAE-HB Skew-normal for (a) Food and (b) Non-food Commodities of Food Crop Farmer Household in Sulawesi Tenggara Province, 2018-2021

The magnitude of the decrease in MSE and RRMSE of the skew-normal SAE HB modeling relative to the direct estimation method indicates that SAE modeling is able to provide a shrinkage effect on the direct estimation results. This finding is in line with the results of almost previous studies, especially for the skewed data, among others **[3]**, **[6]**, **[13]**, **[16]**.

3.5 Overview of Skew-normal SAE HB Model as the Best Estimator

The estimated results of the skew-normal SAE HB model as the best model is presented through a boxplot visualization in **Figure 6**. This figure is shown that the median of per capita expenditure on food commodities in 2018-2021 period is higher than the median per capita expenditure on non-food commodities. This means that more than 50 percent of per capita expenditure on food crop farmer household in Sulawesi Tenggara Province is used for food consumption. According to Dealton and Muellbauer (1980), the greater ratio of expenditure on food in a household, so the lower welfare of society in an area **[12]**.

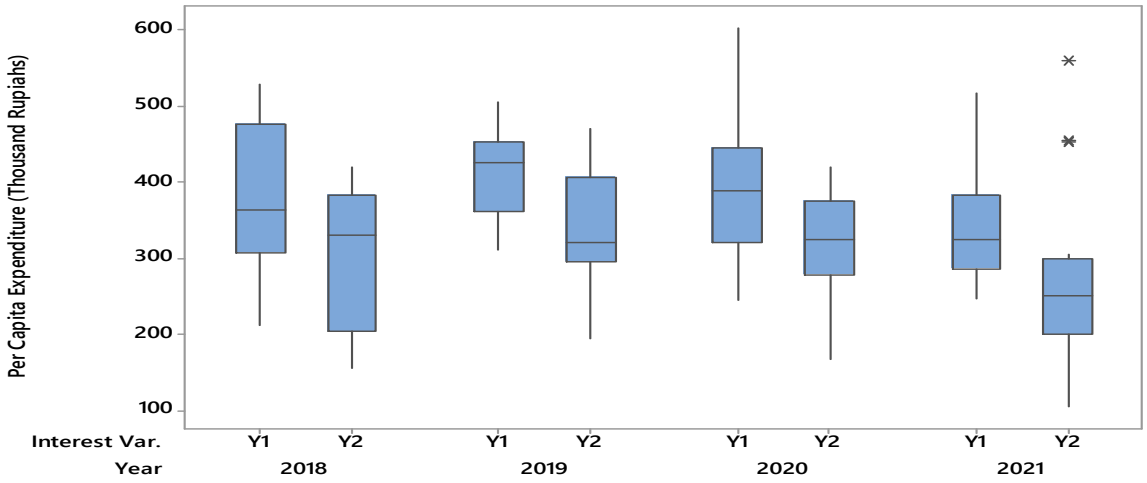


Figure 6. Comparison of Per Capita Expenditure on Food (Y1) and Non-food (Y2) Commodities of Food Crop Farmer Household in Sulawesi Tenggara Province 2018-2021, based on Skew-normal SAE HB

Table 3. Comparison of Food Commodities Per Capita Expenditure Ratio to Total Household Expenditure between The Generally Households based on Susenas Publication and The Food Crop Farmer Households at District/City Level in Sulawesi Tenggara Province, 2018-2021

Code	District/City	The Generally Households*				The Food Crop Farmer Households			
		2018	2019	2020	2021	2018	2019	2020	2021
7401	Buton	0.5196	0.5223	0.4915	0.5602	0.5989	0.5223	0.5809	0.6004
7402	Muna	0.5002	0.4620	0.4958	0.5438	0.5966	0.5568	0.5944	0.6984
7403	Konawe	0.5157	0.4695	0.5072	0.4891	0.6391	0.5588	0.5443	0.6041
7404	Kolaka	0.4807	0.4866	0.5203	0.4817	0.6150	0.4228	0.4990	0.6751
7405	Konawe Selatan	0.5326	0.4854	0.5394	0.4901	0.5762	0.4731	0.5791	0.6496
7406	Bombana	0.5101	0.4599	0.5121	0.4990	0.5588	0.4983	0.5890	0.5494
7407	Wakatobi	0.5119	0.5211	0.5318	0.5429	0.5886	0.4836	0.4888	0.5359
7408	Kolaka Utara	0.5196	0.5416	0.5261	0.4901	0.5786	0.5578	0.4975	0.6006
7409	Buton Utara	0.5143	0.5076	0.5008	0.5076	0.5348	0.5518	0.5567	0.5943
7410	Konawe Utara	0.5002	0.5255	0.5379	0.4946	0.5252	0.5821	0.5317	0.6012
7411	Kolaka Timur	0.4981	0.4936	0.5202	0.5006	0.5402	0.5785	0.4949	0.4806
7412	Konawe Kepulauan	0.4917	0.4998	0.5619	0.5279	0.5385	0.6001	0.5358	0.4703
7413	Muna Barat	0.5006	0.5058	0.5730	0.5354	0.5451	0.5868	0.5561	0.4785
7414	Buton Tengah	0.3985	0.5260	0.6180	0.5932	0.5442	0.6268	0.5441	0.5236
7415	Buton Selatan	0.5567	0.5167	0.5718	0.5575	0.5842	0.6039	0.6078	0.4710
7471	Kota Kendari	0.4007	0.4159	0.4029	0.4093	0.5684	0.6145	0.5457	0.5602
7472	Kota Bau Bau	0.3939	0.3738	0.4167	0.3943	0.5200	0.5518	0.6455	0.5867
	Sulawesi Tenggara	0.4718	0.4653	0.4883	0.4762	0.5631	0.5476	0.5504	0.5550

Note: *) obtained from the 2019-2021 DDA Publication of Susenas results
ratio of per capita expenditure on food commodities <0.5

On average, the ratio between per capita food expenditure to total expenditure of food crop farmer housholds during 2018-2021 was recorded at 0.56; 0.55; 0.55; and 0.56, which is slightly different with the direct expected results. However, this ratio was recorded to be higher than the per capita food expenditure ratio for all households in Sulawesi Tenggara Province during 2018-2021, namely 0.47; 0.46; 0.49; and 0.48 **[22]**. The high ratio of per capita food expenditure for food crop farmer household relative to the regional average in Sulawesi Tenggara Province indicates that the level of welfare for the food crop farmer households is lower than for the generally household group. In addition, increasing in the ratio of per capita food expenditure in 2020 and 2021 indicates that for food crop farmer households in Sulawesi Tenggara Province is a household group that has been significantly impacted by the Covid-19 pandemic as income decreases **[23]**. According to Maxwell, households with the lowest quantile spend almost 60 percent of their total expenditure on food needs. Dealton and Muellbauer stated that the greater the food expenditure ratio in a household, the lower the welfare of the people in an area **[12]**.

4. CONCLUSIONS

SAE modeling using HB approach on data that is assumed to follow skew-normal distribution in this research has proven to be more efficient in estimating per capita expenditure on food crop farmer household at the district/city level in Sulawesi Tenggara Province on positively skewed data. The magnitude decreasing in MSE and RRMSE of the skew-normal SAE HB modeling relative to the direct estimation method indicates that SAE modeling is able to provide a shrinkage effect on the direct estimation results. From the ratio of the direct estimates and the skew-normal SAE HB estimate, it can be seen that there is different interpretations of per capita expenditure patterns of food crop farmer households at district/city level in Sulawesi Tenggara Province during the period 2018-2021. It is possible because the modeling used the assumption that coefficient of autocorrelation in area-time effect is equal to 1 or known as the random walk effect. However, in reality, Susenas data is not a panel data. The unit of observation for each time period may be different. Therefore, further researches should be compared with a skew-normal SAE HB model or another skewed distribution that assumes the autocorrelation coefficient is unknown and should be estimated in the model.

ACKNOWLEDGMENT

The authors presented the highest appreciation to BPS-Statistics Indonesia for the financial assistance in this research.

REFERENCES

[1] N. Tzavidis, L. Zhang, and A. Luna, "From start to finish : a framework for the production of small area official statistics," *J.R.Statististical Soc.*, vol. 4, no. 181, pp. 927–979, 2018.
[2] [ADB] Asian Development Bank, *Introduction to Small Area Estimation Techniques: A Practical Guide for National Statistics Offices*, no. May. Manila: Asian Development Bank, 2020.
[3] E. Fabrizi, M. R. Ferrante, and C. Trivisano, "Bayesian small area estimation for skewed business survey variables," *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 67, no. 4, pp. 861–879, Aug. 2018, doi: 10.1111/rssc.12254.
[4] D. B. do N. S. Silva, A. F. A. Neves, and S. C. Onel, "Small Domain Estimation for a Brazilian Service Sector Survey," *59 ISI World Stat. Congr.*, no. December, p. 6, 2013.
[5] D. K. Bodro, K. Sadik, and B. Sartono, "Kajian peningkatan kualitas pendugaan area kecil melalui transformasi peubah target [tesis]," Bogor: Institut Pertanian Bogor, 2019.
[6] F. A. S. Moura, A. F. Neves, and D. B. do N. Silva, "Small area models for skewed Brazilian business survey data," *J. R. Stat. Soc. Ser. A Stat. Soc.*, vol. 180, no. 4, pp. 1039–1055, 2017, doi: 10.1111/rssa.12301.
[7] A. R. Nulkarim and I. Y. Wulansari, "M-quantile chambers-dunstan untuk pendugaan area kecil: studi kasus data pengeluaran rumah tangga per kapita di Yogyakarta tahun 2018," in *Seminar Nasional Official Statistics 2021*, 2021, vol. 1, pp. 80–89.
[8] Martina and R. Yuristia, "Analisis pendapatan dan pengeluaran rumah tangga petani padi sawah di kecamatan Sawang kabupaten Aceh Utara," *Agrica Ekstensia*, vol. 15, no. 1, pp. 56–63, 2021.
[9] [BPS] Badan Pusat Statistik, *Statistik Nilai Tukar Petani Provinsi Sulawesi Tenggara 2021*. Kendari: BPS Provinsi Sulawesi Tenggara, 2022.
[10] M. Rachmat, "Nilai tukar petani: konsep, pengukuran dan relevansinya sebagai indikator kesejahteraan petani," *J. Agro Ekon.*, vol. 31, no. 2, pp. 111–122, 2013.
[11] P. Simatupang, M. Rahmat, Supriyati, and M. Maulana, "Kajian isu-isu aktual kebijakan pembangunan pertanian: review dan perumusan indikator kesejahteraan petani," 2016.
[12] [BPS] Badan Pusat Statistik, *Pengeluaran untuk Konsumsi Penduduk Indonesia Per Provinsi: Berdasarkan Hasil Susenas September 2021*. Jakarta: Badan Pusat Statistik, 2022.
[13] V. R. S. Ferraz and F. A. S. Moura, "Small area estimation using skew normal models," *Comput. Stat. Data Anal.*, vol. 56, no. 10, pp. 2864–2874, 2012.
[14] F. E. Supriatin, B. Susetyo, and K. Sadik, "EBLUP method of time series and cross-section data for estimating education index in district Purwakarta," *Indones. J. Stat.*, vol. 20, no. 7, pp. 34–38, 2015.
[15] H. J. Boonstra, "Time-series small area estimation for unemployment based on a rotating panel survey," *Stat. Netherlands*, vol. 17, no. June, pp. 1–39, 2014.
[16] A. Neves, D. Britz, and F. Ant, "Skew normal small area time models for the Brazilian annual service sector survey," *Stat. Transit.*, vol. 21, no. 4, pp. 84–102, 2020, doi: 10.21307/stattrans-2020-032.
[17] A. Paranata, Wahyunadi, and A. Daeng, "Mengurai model kesejahteraan petani," *Jejak*, vol. 5, no. 1, pp. 90–102, 2012.
[18] J. N. . Rao and I. Molina, *Small Area Estimation*. New Jersey: John Wiley & Sons, 2015.
[19] S. Muchlisoh, A. Kurnia, K. A. Notodiputro, and I. W. Mangku, "Small area estimation of unemployment rate based on unit level model with first order autoregressive time effects," *Appl. Probabilty Stat.*, vol. 12, no. 2, pp. 51–63, 2017.
[20] G. Chen and S. Luo, "Bayesian hierarchical joint modeling using skew-normal independent distributions," *Commun Stat Simul Comput*, vol. 47, no. 5, pp. 1420–1438, 2019.
[21] N. P. Istiqomah and I. Y. Wulansari, "Estimasi angka partisipasi kasar perguruan tinggi level kabupaten/kota di pulau Kalimantan tahun 2020 dengan small area estimation hierarchical bayes beta-logistic," in *Seminar Nasional Official Statistics 2022 Statististic 2022*, 2020, pp. 137–146.
[22] [BPS] Badan Pusat Statistik, *Provinsi Sulawesi Tenggara Dalam Angka 2022*. Kendari: BPS Provinsi Sulawesi Tenggara, 2022.
[23] R. A. Trianto, "Perubahan pola pengeluaran makanan masyarakat Indonesia akibat pandemi covid-19," *J. Ecogen*, vol. 4, no. 4, p. 471, 2021, doi: 10.24036/jmpe.v4i4.12093.