Estimation of Total Fertility Rate (TFR) Using Small Area Estimation (SAE) in Nusa Tenggara Timur (NTT) Province

Mellinda[1], **C Sumarni**[1]

[1]STIS Polytechnic of Statistics, Otto Iskandardinata Road 64C, Jakarta. 13330.

*Corresponding author's e-mail: 211709829@stis.ac.id

Abstract. The large population in Indonesia has an impact on providing basic services for population which is not optimal so the condition and distribution of the population in a country must be addressed through fertility control methods. Total Fertility Rate (TFR) is one of fertility measures used in Indonesia. The estimation of TFR at the district level is very important, especially for the Nusa Tenggara Timur (NTT) Province as the province with the highest TFR in Indonesia. The availability of TFR data up to the district level is difficult to obtain every year due to data limitations. This study uses the National Socio-Economic Survey to address these problems. TFR estimation through survey data (direct estimation) generally results in a large Relative Standard Error (RSE) value, so it is necessary to estimate using an indirect estimate in the form of Small Area Estimation (SAE). By using SAERestricted Maximum Likelihood (REML) procedure, TFR with an RSE that is lower than the direct estimate is obtained. There are 5 district that have a medium-high TFR, namely: Sumba Barat Daya, Sumba Tengah, Sabu Raijua, Sumba Barat, and Manggarai Barat. The government is recommended to focus more on that 5 districts to suppress the high TFR in NTT.

1. Introduction

The population is one of the important conditions for the formation of a country. In various aspects, the population is the party that takes the role, including advancing the welfare of a country. Based on population projection year 2015-2045, Indonesia's population in the year 2021 is 272.248 million [1]. In fact, in 2021 Indonesia would be in the fourth position of the country with the largest population in the world [2]. Huge population and a limited state budget have an impact on the provision of basic services to the population that is not optimal in a given country. Because of these problems, the condition and distribution of the population in a country must be considered. To support the success of development, as well as to deal with population problems including the number, composition, and distribution of the population, it is necessary to control the population through fertility control [3]. The fertility control in Indonesia carried out by the National Population and Family Planning Board (BKKBN) is an effort to suppress the population with specific steps in the form of the Keluarga Berencana (KB) program. The KB program is implemented to reduce the Total Fertility Rate (TFR) to a value of 2.1 in 2025.

TFR is the average number of children born to a woman during her childbearing age (WUS) obtained through the accumulation of the Age-Specific Fertility Rate (ASFR) [4]. TFR is useful as an indicator to compare success between regions in implementing socio-economic development, showing the level of success of family planning programs, helping development program planners to increase the average age of marriage, improving health service programs related to maternal and child care services, and developing birth rate reduction program [4]. Based on Indonesia Demographic and Health Survey (IDHS) 2017, the values TFR Indonesia in 2017 is 2.4 which means that Indonesian women aged 15-49 years on average have 2 to 3 children during her fertile age. Based on IDHS 2017, the province with the highest TFR is Nusa Tenggara Timur (NTT) (3.4) and the province with the lowest TFR is Jawa Timur (2.1) and Yogyakarta (2.1). Thus, the handling of the high TFR in Indonesia can be focused on the province with the highest TFR, namely NTT. If look at the condition of TFR more specifically, then the data required TFR at the districts level. Estimation of TFR which was published by IDHS and the National Socio-Economic Survey (SUSENAS) only national and provincial level. The availability of TFR data at the provincial level is still difficult to obtain every year. In several surveys with the level of presentation in the form of districts/cities, TFR estimates are also rarely published. With their decentralized systems and data needs, required estimation of TFR to a smaller area by geographic (district level).

TFR estimation can be done using direct and indirect estimation methods. The direct estimation method is very suitable if applied to data that is supported by a good registration system so that the estimation results are accurate. However, the use of direct estimation methods sometimes results in large standard errors [5]. Estimation TFR at the district level also can be done through surveys (direct estimation) who have publications county level, but the indirect estimation methods can still be used to lower the standard error. Indirect estimation methods are widely applied in countries with poor vital registration systems such as Indonesia [6]. The TFR in 2013 was calculated using the indirect estimation method of the Own Child Method (OCM) [7]. This method turns out to have a weakness namely, it does not have a Relative Standard Error (RSE) precision measure, so it cannot be known whether the results of the estimations carried out are more reliable than direct estimates. This method also requires census data as a component of calculating TFR, even though this data is difficult to obtain every year. This study applies a different indirect method, namely Small Area Estimation (SAE). This SAE method has the advantage that it can provide more reliable estimation results with the strength of the loan from the accompanying variables used with comprehensive coverage [8]. SAE Restricted Maximum Likelihood (REML) method was used in this research to reduce the standard error of the direct estimation method.

Research related to the SAE method to estimate ASFR and TFR was conducted at the district level in Portugal [9]. In that study, SAE was carried out without the use of auxiliary variables because no suitable auxiliary variables were found. Also in that study, SAE was applied to estimate the fertility rate of 28 districts in Portugal referring to 7 age groups of women of childbearing age, namely 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49 in 2009. Research on fertility rates was also conducted to estimate the percentage of women of childbearing age with high fertility in Mamuju and Mamuju Tengah Districts using the SAE method of Spatial Empirical Best Linear Unbiased Predictor (SEBLUP) [10]. With a significance level of 20%, of the 9 proposed variables, there are 4 significant accompanying variables in this case. The concomitant significant variable is the ratio of school per 1000 women was significant, at $\alpha = 10\%$, the industry ratio per 100 women was significant at $\alpha = 5\%$, and the average distance to health facilities which was significant at $\alpha = 20\%$. The results of this study, the SEBLUP method SAE with the Restricted Maximum Likelihood (REML) procedure resulted in the estimated value of the percentage of women of childbearing age with high fertility at village level is better than the direct estimation results and SEBLUP Maximum Likelihood (ML) procedure.

In this study, the estimated variable is TFR. This study aims to estimate TFR at the district level so that it is useful for the government especially local government in suppressing the high TFR. The SAE EBLUP method is considered appropriate to use on continuous data [11]. TFR is a variable of the continuous type so that the estimation of TFR using SAE would be more appropriate if using the EBLUP Fay-Herriot (FH) with the REML method. Therefore, the purpose of this study is to examine the estimation of the district level TFR in NTT in 2019 through the Small Area Estimation. This study uses a method that produces a precision measure in the form of RSE so that the goodness of the estimation results can be known. Previous research has calculated the TFR using the Own Children Method but this method does not have a precision measure that guarantees the goodness of the estimation results [6]. In terms of data use, this study uses data that can be obtained every year such as SUSENAS so that data availability is guaranteed and estimations can be carried out continuously for every year, in contrast to the previous research which requires census data in its estimation [7].

2. Methodology

The data used in this study is secondary data sourced from BPS, namely the March 2019 SUSENAS data, *Village Potential Statistics* (PODES) in 2018, and the 2010 Population Census (SP) in Nusa Tenggara Timur Province. This study discusses the direct and indirect estimation of TFR in NTT in 2019. NTT consists of 21 districts and 1 city. A direct estimate of TFR is obtained using the following formula.

$$\hat{Y}_{gh} = \sum_{g=1}^{G} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_i} w_{ghij} y_{ghij} \tag{1}$$

$$\hat{X}_{gh} = \sum_{g=1}^{G} \sum_{h=1}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_i} w_{ghij} x_{ghij} \tag{2}$$

$$\widehat{ASFR}_{gh} = \hat{R}_{gh} = \frac{\hat{Y}_{gh}}{\hat{X}_{gh}} \tag{3}$$

$$\widehat{TFR}_g = 5 \sum_{h=1}^{H} \widehat{ASFR}_{gh} \tag{4}$$

with w_{ghij} is weighted at the *g*th district/city, the *h*th age group in the *i*th census block (BS), and *j*th households. Then, y_{ghij} is the number of children born in the district/city to-*g*, the age group of to-*h* in the census block (BS) to-*i*, and households to-*j* and is the number of women of childbearing age in the district/city to-*g*, age group *h* in the *i*-th census block, and the *j*-th household. The estimation of the direct estimation variance is based on the accumulated ASFR variance. Assuming that the ASFRs are independent, the following is the formula for the variance [12].

$$var(\hat{R}_{gh}) = \frac{\sum_{g=1}^{G} \sum_{h=i}^{H} \sum_{i=1}^{n_h} \sum_{j=1}^{m_i} w_{ghij}^2 \left(y_{ghij} - \hat{R}_{gh} x_{ghij}\right)^2}{\left(\sum_{i=1}^{n_h} \sum_{j=1}^{m_i} w_{ghij} x_{ghij}\right)^2} \tag{5}$$

$$var(\widehat{TFR}_g) = 25 \times \sum_{g=1}^{G} \sum_{h=1}^{H} var(\hat{R}_{gh}) \tag{6}$$

2.1. *Small Area Estimation*

Small area modeling is a model with random area-specific effects that calculates the variation between areas that is not explained by the auxiliary variable. Indirect estimation based on a small area model is called model-based estimation. Area-level-based model is a model based on the availability of auxiliary variable data that only exists for a certain area level through a linear model as follows [13]:

$$\hat{\theta}_i = \boldsymbol{x}_i^T \boldsymbol{\beta} + b_i v_i + e_i \qquad i = 1, \dots, m \tag{7}$$

Where $\hat{\theta}_i$ is a direct estimate of the parameters θ_i for each of the *i*th areas. Then, x_i is the vector size $p \times 1$ of concomitant variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is a fixed effect model coefficients, v_i are random effects area with $v_i \sim \text{iid } N(0, \sigma_{vi}^2)$ and e_i is sampling error with $e_i \sim N(0, \Psi_i)$ the variance (Ψ_i) obtained from direct estimation. It is assumed that v_i and e_i are mutually independent with *i* being the index for the area. Meanwhile, *b* is a positive constant which is known, in which the Fay-Herriot model for the level of the basic area b is 1. Estimates of indirect estimators Best Linear Unbiased Predictor (BLUP) for θ_i the following:

$$\tilde{\theta}_i^{BLUP} = \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}} + \gamma_i (\hat{\theta}_i - \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}}) \tag{8}$$

$$\tilde{\theta}_i^{BLUP} = \gamma_i \hat{\theta}_i + (1 - \gamma_i) \boldsymbol{x}_i^T \tilde{\boldsymbol{\beta}} \tag{9}$$

$$\gamma_i = \frac{\sigma_v^2}{\Psi_i + \sigma_v^2} \tag{10}$$

σ_v^2 is the variance of the random effects area, Ψ_i is the variance of sampling error of $\tilde{\theta}_i$ for the area to- i, and $\widetilde{\boldsymbol{\beta}}$ is the regression coefficient BLUP. $\widetilde{\boldsymbol{\beta}}$ is the Best Linear Unbiased Estimator (BLUE) which is stated as follows.

$$\widetilde{\boldsymbol{\beta}} = \widetilde{\boldsymbol{\beta}}(\sigma_v^2) = \left[\sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^T/(\Psi_i + \sigma_v^2 b_i^2)\right]^{-1}\left[\sum_{i=1}^{m} \boldsymbol{x}_i\tilde{\theta}_i/(\Psi_i + \sigma_v^2 b_i^2)\right] \tag{11}$$

The BLUP estimator requires components σ_v^2 in its estimation, which in practice the component is σ_v^2 not known for its value. The components σ_v^2 need to be estimated using the Restricted Maximum Likelihood (REML) method. By replacing the value σ_v^2 with the value of the estimator($\hat{\sigma}_v^2$), the BLUP estimator would turn into an Empirical Best Linear Unbiased Predictor (EBLUP) estimator which can be written as follows:

$$\hat{\theta}^{EBLUP} = \hat{\gamma}_i\hat{\theta}_i + (1-\hat{\gamma}_i)\,\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}} \tag{12}$$

$$\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\Psi_i + \hat{\sigma}_v^2} \tag{13}$$

$$\hat{v}_i = \hat{\gamma}_i(\hat{\theta}_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}) \tag{14}$$

Equation (12) shows that the EBLUP estimator can be written as a weighted average of a direct estimate and a synthetic estimate ($\boldsymbol{x^T}\widehat{\boldsymbol{\beta}}$) where the weight is γ with a value $0 \le \gamma \le 1$. The *gamma* value measures the uncertainty in the modeling parameters. If the gamma value is small, then the weight attached to the synthetic estimate is greater.

2.2. MSE of EBLUP Fay-Herriot Estimator

To measure how good EBLUP is, the *Mean Square Error* (MSE) value is used which can be written as follows.

$$MSE(\hat{\theta}^{EBLUP}) \approx g_{1i}(\hat{\sigma}_v^2) + g_{2i}(\hat{\sigma}_v^2) + 2g_{3i}(\hat{\sigma}_v^2) \tag{15}$$

$$g_{1i}(\hat{\sigma}_v^2) = \frac{\hat{\sigma}_v^2 b_i^2 \Psi_i}{\Psi_i + \hat{\sigma}_v^2 b_i^2} = \hat{\gamma}_i\Psi_i \tag{16}$$

$$g_{2i}(\hat{\sigma}_v^2) = (1-\hat{\gamma}_i)^2\boldsymbol{x}_i^T\left[\sum_{i=1}^{m} \boldsymbol{x}_i\boldsymbol{x}_i^T/(\Psi_i + \hat{\sigma}_v^2 b_i^2)\right]^{-1}\boldsymbol{x}_i \tag{17}$$

$$g_{3i}(\hat{\sigma}_v^2) = {\Psi_i}^2(\Psi_i + \hat{\sigma}_v^2 b_i^2)^{-3}\bar{V}(\hat{\sigma}_v^2) \tag{18}$$

$$\bar{V}(\hat{\sigma}_v^2) = 2\left[\sum_{i=1}^{m} b_i^4/(\Psi_i + \hat{\sigma}_v^2 b_i^2)^2\right]^{-1} \tag{19}$$

$\bar{V}(\hat{\sigma}_v^2)$ = asymptotic diversity of $\hat{\sigma}_v^2$

2.3. TFR Indirect Estimation Stage

1. Prepare the results of the direct estimation (TFR and variance) and the candidate auxiliary variables used. The estimation results are obtained by using Equations (1) and (6).
2. Select candidate auxiliary variables by considering the significance of the correlation between candidate auxiliary variables and TFR. Candidate variables in full can be found inAppendix D.
3. Select variable using the backward elimination method, which is removing the variable that has the largest p-value. Akaike Information Criterion (AIC) value is also used to see the goodness of the model obtained. After that, re-modeling was carried out with the remaining auxiliary variables until the best model was obtained. AIC is obtained through the following formula:

$$AIC = -2\, l(\widehat{\boldsymbol{\beta}}, \hat{\sigma}_v^2) + 2p$$
$$l(\boldsymbol{\beta}, \sigma_v^2) = const - \frac{1}{2}\left[log|\boldsymbol{V_e}| + (\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zv})^{\boldsymbol{T}}\boldsymbol{V_e}^{-\boldsymbol{1}}(\boldsymbol{y} - \boldsymbol{X\beta} - \boldsymbol{Zv})\right] \quad (20)$$

$\hat{\sigma}_v^2$ =*random effects variance*
$\hat{\beta}$ =regression coefficients
p = dimension of $\hat{\beta}$
m = number of observation
$\boldsymbol{X}$ is the vector size $m \times p$ from auxiliary variables, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^T$ is a fixed effect model coefficients with size $p \times 1$, $\boldsymbol{v}$ is a *random effects area* with size $m \times 1$, $\boldsymbol{V_e}$ is a matrix variance of direct estimation with $\mathbf{V_e} = \text{diag}[\boldsymbol{V_{ei}}]$ with size $m \times m$ where *i* for each of the *i*th areas, and $\mathbf{Z}$ is a positive constant matrix that is worth 1 (identity matrix) with size $m \times m$.

4. Estimate the TFR value with the EBLUP FH Model (REML), MSE, and Relative Standard Error (RSE) of each region through the RStudio application with the "sae" package. MSE is obtained from Equation (15). RSE is obtained through the following formula.

$$RSE(\hat{\theta}_i) = \frac{\sqrt{MSE(\hat{\theta}_i)}}{\hat{\theta}_i} \times 100\% \quad (21)$$

5. Estimate the random effects area and gamma. Estimates are made using Equations((13) and (14).
6. Select the best model of indirect estimation then check the RSE, the results of the random effect area normality test and model error, and the gamma value. The normality test used is the Shapiro-Wilk test.
7. Compare direct and indirect methods through model evaluation and comparing the estimation results by considering the residual value. The residual formula used is as follows:

$$\hat{e}_i = \hat{\theta}_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}$$
$$standardized: \hat{e}_i = \frac{\hat{\theta}_i - \boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}}{\sqrt{\Psi_i + \hat{\sigma}_v^2}} \quad (22)$$

8. Check the efficiency of the model.
9. Present TFR visualization through the thematic map of the indirect estimation TFR distribution (EBLUP FH REML model) and compares it with the direct estimation TFR thematic map.

3. Result and Discussion

A general description is needed to determine the condition of the area by paying attention to aspects related to the fertility component. Table 1 presents an overview of the percentage of WUS and births in NTT Province based on 2019 SUSENAS data.

Table 1. Estimated Percentage of WUS and Births by Age Group

Age Group	Percentage of WUS	Percentage of Births
15-19	19.295%	4.699%
20-24	15.397%	20.638%
25-29	15.684%	27.810%
30-34	13.717%	22.965%
35-39	13.018%	16.349%
40-44	11.887%	6.475%
45-49	11.002%	1.064%
Total	100% (1349441)	100% (124886)

From Table 1, it is known that most of the WUS in NTT is 15-19 years old (19.295%) while the least WUS is 45-49 years old (11.002%). This indicates that NTT is dominated by WUS who have a long reach in their fertile period so that WUS aged 15-19 years have a great opportunity to give birth when they are in other age groups. This has an impact on increasing fertility in NTT in the future. Meanwhile, the largest number of births in NTT was in WUS in the 25-29 year age group, which was marked by the highest percentage (27.810%). In the 45-49 year age group, it is known that the lowest percentage of births (very low) is 1.064%.

3.1. Direct Estimation of TFR

After knowing the general description of NTT Province in 2019, the analysis continued by making a direct estimate of the TFR in NTT Province at the district level.

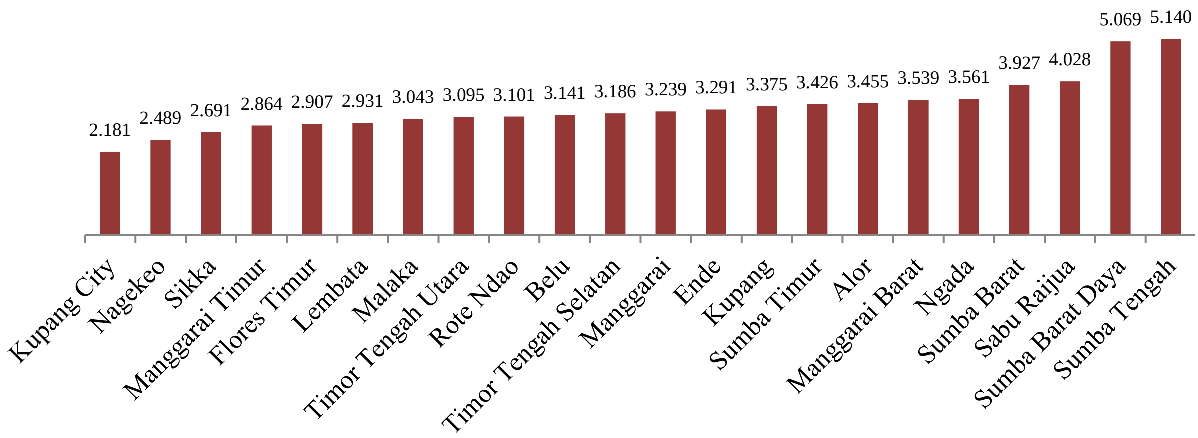

Figure 1. Direct estimation of TFR by district/city of NTT Province in 2019

Based on Figure 1, it can be seen that the districts of Sumba Tengah, Sumba Barat Daya, and Sabu Raijua are the 3 districts with the highest TFR in the province of NTT. United Nations (2015) states that the TFR is over 5 into the category of high TFR (high) so that the TFR is owned by Southwest Sumba (5.069) and Sumba Barat (5.140) fall into the category of high TFR. Meanwhile, the only city in NTT, namely Kupang City, has the lowest TFR compared to 21 other regencies. The lowest TFR at the district level in 22 districts/cities in NTT Province is 2.181 and the largest is 5.140. This indicates that in the districts with the lowest TFR in NTT in 2019, women aged 15-49 years on average had 2 to 3 children during their fertile period, while in the districts with the largest TFR they had an average of 5 to 6 children during their fertile period. After making a direct estimate of the TFR, the analysis is continued by calculating the RSE from the direct estimate. This RSE is used to see the goodness of direct estimates made. The RSE generated by direct estimation is quite good because it is below 25% [14]. However, the indirect estimation of TFR was continued to get a better estimation precision. Thus, the resulting TFR is more accurate which is characterized by a lower RSE value.

3.2. Indirect Estimation of TFR

In building an indirect estimation model using SAE EBLUP FH, an auxiliary variable that does not contain errors is needed. After checking the significance of the correlation, the candidate auxiliary variables that can be used are levels of Public and Private Academy/University (x_5), the percentage of families using electricity (x_7), the closest distance to the Community Health Centers (Puskesmas)without hospitalization (x_{11}) and the number of doctors' offices (x_{27}). The variables that have been obtained are then used to estimate TFR. This step begins with finding the estimated value of the regression coefficient ($\hat{\beta}$) and the area-level random effect using the Restricted Maximum Likelihood (REML) estimation methods. The estimation of the regression coefficient and the area-level random effect is done by eliminating the variable that has the highest p-value (backward elimination). The model chosen from this series is the model that has the smallest AIC value. Table 2 contains a summary of the results of the process of forming the best model.

Table 2. EBLUP-REML Model Selection Process

Model	Variables	AIC
1	x_5, x_7, x_{11}, x_{27}	38,932
2	x_7, x_{11}, x_{27}	37,043
3	x_7, x_{11}	34,855

From Table 2, it is known that the 3rd model are the best models for the EBLUP indirect estimation method. The estimation of regression coefficients and coefficients of random variance in the selected models is shown in Table 3.

Table 3. Estimation of Regression Coefficients and Random Diversity EBLUP-REML Methods

Variables	$\hat{\beta}$	p-value	std.error
Intercept	4,855	0,000	0,539
x_7	-0,026	0,000	0,007
x_{11}	0,013	0,013	0,005
$\hat{\sigma}_v^2$	0,115	-	-

From Table 3, it is known that all variables are significant at a significance level of 0.05. Table presents the complete estimation results using the EBLUP REML method.

Table 4. Results of the 2019 NTT District Level TFR Estimation

District/City	TFR	RSE (%)
Sumba Barat	3.863	7.262
Sumba Timur	3.452	7.004
Kupang	3.306	7.531
Timor Tengah Selatan	3.448	7.717
Timor Tengah Utara	3.160	8.417
Belu	3.020	8.458
Alor	3.291	8.157
Lembata	3.197	8.825
Flores Timur	2.721	9.494
Sikka	2.756	8.256
Ende	3.232	8.587
Ngada	3.227	8.242
Manggarai	3.150	8.485
Rote Ndao	3.233	8.104
Manggarai Barat	3.543	7.314
Sumba Tengah	4.554	6.713
Sumba Barat Daya	4.602	6.272
Nagekeo	2.686	8.643
Manggarai Timur	3.222	8.004
Sabu Raijua	4.145	7.504

District/City	TFR	RSE (%)
Malaka	3.234	8.658
Kupang City	2.196	10.027

Based on Table 4, it is known that of the 22 districts in NTT Province, the lowest district level TFRs are 2.196 (EBLUP REML) namely Kupang City. Meanwhile, the largest district-level TFRs are 4.602 (EBLUP REML) which are both located in Sumba Barat Daya Regency. Through the smallest TFR, it can be interpreted that the Kupang City in 2019 women aged 15-49 years on average had 2 to 3 children during their childbearing age, while in the district that had the largest TFR (Sumba Barat Daya) on average had 4 up to 5 children during their childbearing years. The average RSE of EBLUP REML is 8.076%. The highest RSE is in Kupang City with an RSE of 10.027%. Meanwhile, the lowest RSE was in Southwest Sumba Regency with an RSE of 6.272%.

3.3. Model Evaluation

In addition to using the RSE value, checking the best model is also done by assessing whether the model meets the assumptions of random effect area normality and model error. Furthermore, the gamma value is also checked. The results of testing the assumption of normal random effect area through the Shapiro-Wilk test prove that with a significance level of 0.05, there is not enough evidence to support that the random effect area is not normally distributed. This is indicated by p-values of 0.789 (EBLUP REML). In the error model, with a significance level of 0.05, there is not enough evidence to support that the model error is not normally distributed. This is indicated by a p-value of 0.407 (error model EBLUP REML). Based on the model formed, it was found that the average gamma values obtained were 0.524. This indicates that the estimated value is highly dependent on the accompanying variables used in the model.

The evaluation of the model was carried out to see the goodness of the best model that was found, namely the SAE EBLUP FH REML indirect estimation method. It is assumed that the equally *unbiased* direct estimate fluctuates around the model-based estimate. The plot distribution is distributed around the bisector distribution (black line). The regression line is marked by a red line.

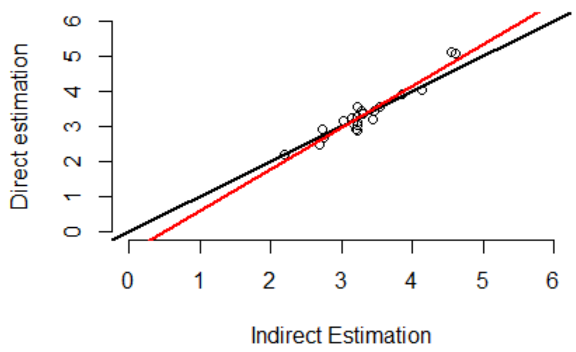

Figure 2. The relationship between direct estimation and indirect estimation of TFR

Based on Figure 2, it is known that the direct estimator and the indirect estimator have the same results. This is evident from the overlap of the bisector line and the regression line. This indicates that the resulting model leads to unbiased results. This is evident from the overlap of the bisector line and the regression line which indicates that the resulting model leads to unbiased results because the indirect estimates produced are similar to the direct estimates [15].

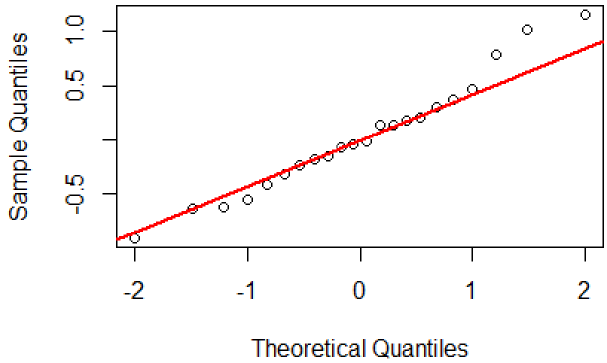

Figure 3. Q-Q Plot Standardized Residual

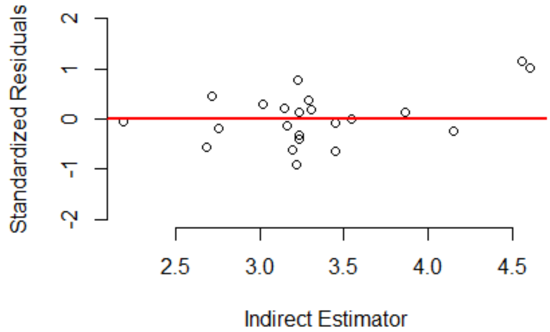

Figure 4. Residual plot with model estimation results

The Normal Quantile-Quantile plot of the standardized residual in Figure 3 shows a slight deviation from the normal distribution. In other words, the standardized residual is normally distributed. Based on Figure, it can be seen that the residuals of the indirect estimator are spread around the standardized residual = 0 so that visually it can be said that the sampling error tends to have a constant variance. From the evaluation of this model, it can be said that the model formed is quite good.

3.4. Comparison of Direct and Indirect Estimation Results

After deciding that the SAE EBLUP REML model is the best model for estimating TFR in NTT Province in 2019, the next step is to compare the results of the direct and indirect estimates to see the differences in the estimation results form.

Table 5. Results of the 2019 NTT District Level TFR Estimation

District/City	Direct Estimation		Indirect Estimation	
	TFR	RSE (%)	TFR	RSE (%)
Sumba Barat	3.927	9.287	3.863	7.262
Sumba Timur	3.426	8.255	3.452	7.004
Kupang	3.375	9.026	3.306	7.531
Timor Tengah Selatan	3.186	9.854	3.448	7.717
Timor Tengah Utara	3.095	11.123	3.160	8.417
Belu	3.141	9.905	3.020	8.458
Alor	3.455	10.152	3.291	8.157
Lembata	2.931	11.045	3.197	8.825
Flores Timur	2.907	10.501	2.721	9.494
Sikka	2.691	9.679	2.756	8.256
Ende	3.291	10.061	3.232	8.587
Ngada	3.561	9.448	3.227	8.242
Manggarai	3.239	10.638	3.150	8.485
Rote Ndao	3.101	10.730	3.233	8.104
Manggarai Barat	3.539	9.104	3.543	7.314
Sumba Tengah	5.140	7.900	4.554	6.713
Sumba Barat Daya	5.069	6.963	4.602	6.272
Nagekeo	2.489	10.841	2.686	8.643
Manggarai Timur	2.864	10.476	3.222	8.004
Sabu Raijua	4.028	9.755	4.145	7.504
Malaka	3.043	12.281	3.234	8.658
Kupang City	2.181	10.800	2.196	10.027

From Table 5, it can be seen that through direct and indirect estimation methods the lowest TFR are both in Kupang City. Through the lowest TFR, it can be interpreted that in Kupang City in 2019 women aged 15-49 years on average have 2 to 3 children during their childbearing age. Meanwhile, the highest TFR through direct estimation is located in Sumba Tengah with a value of 5.140 while the highest TFR by an indirect method is 4.602 located in Sumba Barat Daya Regency. Through the highest TFR, it can be interpreted that in 2019 women aged 15-49 years on average had 5 to 6 children (based on direct estimates) or 4 to 5 children (based on indirect estimates) during their childbearing years.

3.5. Comparison of Relative Standard Error (RSE)

To find out that the indirect estimation method gives better estimation results, it is necessary to compare the estimation results between the direct method and the indirect method. The comparison of this method can be visualized through the comparison of RSE values.

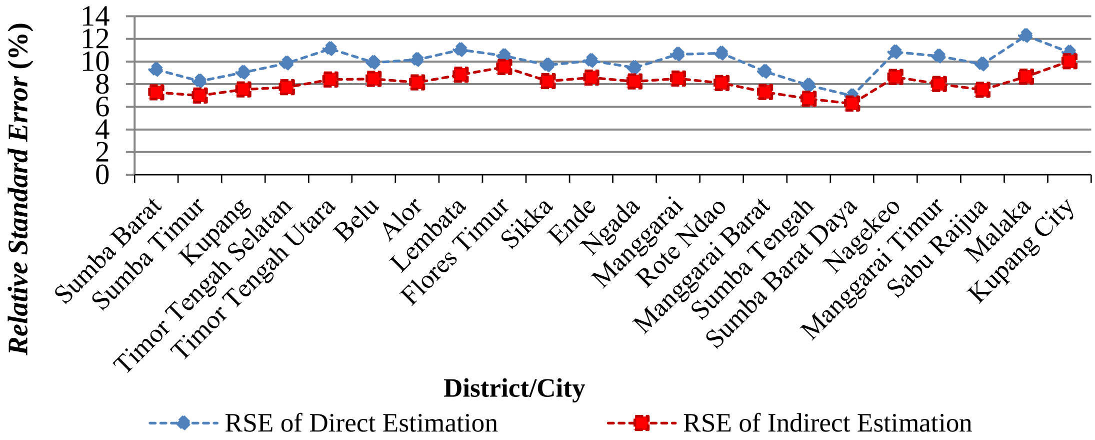

Figure 51. RSE direct and indirect estimation of NTT Province TFR 2019

From Figure 5, it can be seen that the results of estimating TFR through the indirect method give a smaller RSE than the results of the estimation by the direct method. Based on descriptive statistics, with a number of observations of 22 districts/cities, the RSE of the direct estimator ranged from 6.963% to 12.281% with a variance of 1.408. Meanwhile, the RSE of the indirect estimator has a value ranging from 6.272% to 10.027% with a variance of 0.765. The average RSE of the direct estimator was 9.901% while the average RSE of the indirect estimator was 8.076%.

3.6. Model Efficiency

To see the efficiency gains that can be achieved by the application of indirect estimation, a graph of the efficiency of the model is shown by comparing the MSE of the indirect estimator and the variance of the direct estimator in Figure 6. With the observations as the horizontal axis and the ratio between the MSE EBLUP REML and the variance of the direct estimator as the axis Based on Figure 6, it can be obtained that the ratio value between MSE and the direct estimation variance is less than 1. This indicates that the MSE value is smaller than the direct estimation variance so that the SAE EBLUP REML model provides efficiency advantages in estimating TFR.

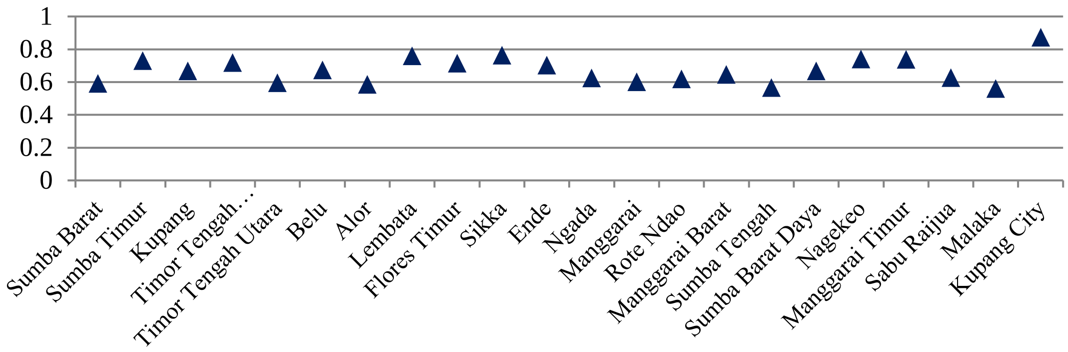

Figure 6. Model efficiency

3.7. Visualization of TFR Indirect Estimation Results

The fertility rate is classified into 5 levels, namely very low for a TFR of 1.5 or less, low for a TFR of less than 2.1 but more than 1.5, medium for a TFR of 2.1 to 3.5, medium-high for TFR 3.5 to 5, and high for TFR more than 5 [16]. Visualization of TFR indirect estimation through thematic maps can be compared with TFR direct estimation as shown in Figure 7 and Figure 8 below.

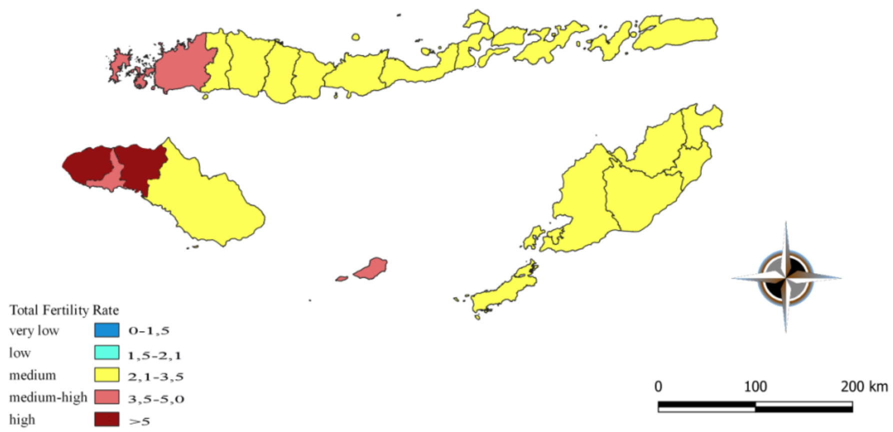

Figure 7. District level TFR in NTT 2019 based on direct estimation

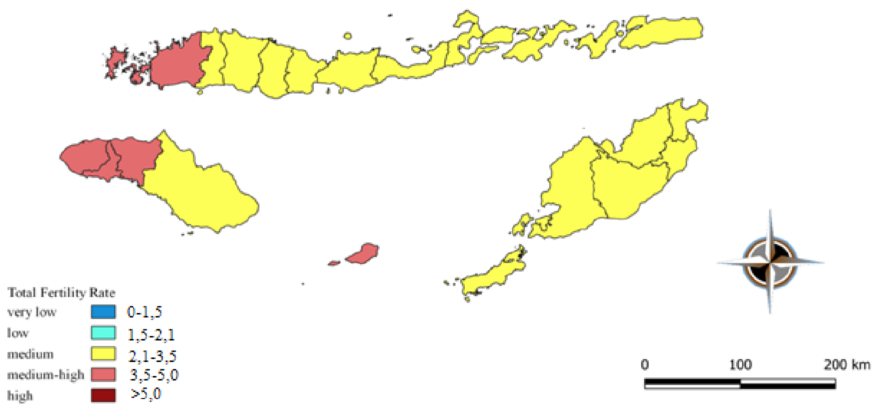

Figure 8. District level TFR in NTT 2019 based on indirect estimation

Based on the comparison between Figure 7 and Figure 8, it can be seen that there are 2 districts, namely Southwest Sumba and Central Sumba which have different TFR categories. Through direct estimation, these 2 districts fall into the high TFR category, while through indirect estimation these 2 districts fall into the medium-high TFR category. In Figure 8 there are 2 classifications obtained from TFR, namely medium and medium-high. There are no areas with very low, low, or high TFR. From Figure 8, it can also be seen that the TFR per district/city in the province of NTT is mostly medium TFR because 17 of the 22 districts (77.273%) fall into this classification. Meanwhile, there are five areas that have TFR medium-high that Sumba Barat Daya, Sumba Tengah, Sabu Raijua, Sumba Barat, and Manggarai Barat. Therefore, the high concentration of TFR handling in NTT Province can be more specifically emphasized in these 5 districts.

4. Conclusions and Suggestions

4.1. Conclusions

Based on the results of the discussions that have been carried out in this study, the conclusions and suggestions that can be obtained regarding the district/city level TFR in NTT Province in 2019 are as follows.

- Through the general description of the data, most of the WUS in NTT Province are in the age group 1 (15-19 years).
- The results of direct TFR estimation show that there are 2 districts in NTT that have high fertility rates, namely Sumba Barat Daya (5.069) and Sumba Tengah (5.140). This direct estimation produces a fairly good RSE, which is less than 25%.
- Auxilliary variables from PODES and SP data that can be used to estimate TFR in NTT arethe percentage of families using electricity and the closest distance to the Community Health Centers (Puskesmas) without hospitalization.
- Model evaluation shows that the results of the indirect estimator (EBLUP FH REML) have a satisfies the assumption of normality of the random effect area and the error model, variance that tends to be constant, and the RSE is lower than the direct estimator. In addition, the efficiency of the model proves that the indirect estimator is more efficient than the direct estimator.
- Based on the UN classification, the TFR thematic maps for the medium-high category are owned by the districts of Sumba Barat Daya, Sumba Tengah, Sabu Raijua, Sumba Barat, and Manggarai Barat. Meanwhile, other districts are included in the medium category.

4.2. Suggestions

- By utilizing the availability of data, further research can estimate a small area of TFR at the sub-district level.
- The government is recommended to increase the focus on handling the high TFR in districts with the medium-high category (the result of indirect estimation).
- To prevent high TFR in the future, the government is recommended to provide outreach activities to WUS in the 15-19 age group or in general, SMA/SMK age equivalent related to efforts to avoid early marriage which can be inserted through school activities, social media, and public service advertisements.
- The government can reduce the number of cases of short birth spacing through socialization and activation of family planning programs by assisting in the form of free family planning programs.

Appendix

Appendix D. Candidate Auxiliary Variables

Variables	Description	Source
x_1	The number of people who have Community Health insurance (Jamkesmas)	PODES 2018
x_2	The number of people who haveSurat Keterangan Tidak Mampu (SKTM)	PODES 2018
x_3	The number of education levels of SMP / MTS public and private	PODES 2018
x_4	The number of education levels of SMA / MA / SMK public and private	PODES 2018
x_5	The number of levels of education in the State and Private Colleges	PODES 2018
x_6	Number of processing industries	PODES 2018
x_7	percentage of the number of electric user families	PODES 2018
x_8	the closest distance to the hospital	PODES 2018
x_9	The closest distance to the maternity hospital	PODES 2018
x_{10}	The closest distance to the health center with hospitalization	PODES 2018
x_{11}	The closest distance to the health center without hospitalization	PODES 2018
x_{12}	The closest distance to the Puskesmas Pembantu	PODES 2018
x_{13}	The closest distance to the polyclinic / treatment center	PODES 2018
x_{14}	The closest distance to the doctor's practice	PODES 2018
x_{15}	The closest distance to the maternity house	PODES 2018
x_{16}	the closest distance to the midwife practice	PODES 2018
x_{17}	The closest distance to Poskesdes (village health post)	PODES 2018
x_{18}	The closest distance to Polindes (village maternity cottage)	PODES 2018
x_{19}	the closest distance to the pharmacy	PODES 2018
x_{20}	The closest distance to a drug specialty / herbal medicine	PODES 2018
x_{21}	Number of Hospitals	PODES 2018
x_{22}	Number of Maternity Hospitals	PODES 2018
x_{23}	Number of Puskesmas with hospitalization	PODES 2018
x_{24}	Number of Puskesmas without hospitalization	PODES 2018
x_{25}	Number of Puskesmas Pembantu	PODES 2018
x_{26}	Number of Polyclinic / Medical Balai	PODES 2018
x_{27}	Number of the doctor's practice	PODES 2018
x_{28}	Number of maternity houses	PODES 2018
x_{29}	Number of Midwife Practice	PODES 2018
x_{30}	Number of Poskesdes (Village Health Post)	PODES 2018
x_{31}	Number of Polindes (Village Maternity Cottage)	PODES 2018
x_{32}	Number of the pharmacy	PODES 2018
x_{33}	Number of specialized / medicinal stores	PODES 2018
x_{34}	WUS percentage with education more than high school equivalent	SP 2010

References

[1] Statistics Indonesia, "Proyeksi Penduduk Indonesia 2015-2045 Hasil SUPAS 2015," Jakarta : Indonesia, 2018.

[2] Worldometer, "Indonesia Population (LIVE)," 2021. [Online]. Available: https://www.worldometers.info/world-population/indonesia-population/. [Accessed 13 Juni 2021].

[3] A. Mahendra, "Analisis faktor-faktor yang mempengaruhi fertilitas di Indonesia," *JRAK*, pp.

3(02), 223-242, 2017.
[4] Statistics Indonesia, "Total Fertility Rate (TFR)," 2021. [Online]. Available: https://sirusa.bps.go.id/sirusa/index.php/indikator/49. [Accessed 14 Desember 2020].
[5] M. A. C. Hakim, "Penerapan Metode Fay-Herriot pada Estimasi Prevalensi Stunting Level Kecamatan di Nusa Tenggara Barat Tahun 2017[Skripsi]," Jakarta, 2019.
[6] National Population and Family Planning Board, "Penyajian TFR Kabupaten dan Kota : Data Susenas 2010," 2013.
[7] Indrawati, et al., "Perhitungan Fertilitas Menggunakan Metode Anak Kandung : Analisis Lanjut Data RISKESDAS 2013," *Populasi*, pp. 67-79, 2016.
[8] Asian Development Bank, "Introduction to Small Area Estimation Techniques," Metro Manila: Filipina, 2020.
[9] E. A. Castro, Z. Zhang, A. Bhattacharjee, J. M. Martins and T. Maiti, "Regional fertility data analysis: A small area Bayesian approach," *Current Trends in Bayesian Methodology with Applications*, pp. 203-224, 2015.
[10] A. Setyawan, “Small Area Estimation metode spatial empirical best linear unbiased predictor untuk estimasi persentase wanita usia subur dengan fertilitas tinggi di Kabupaten Mamuju dan Mamuju Tengah,” Surabaya, 2016.
[11] A. Bukhari, “Pendugaan Area Kecil Komponen Indeks Pendidikan Dalam IPM di Kabupaten Indramayu dengan Metode Hierarchical Bayes Berbasis Spasial,” Bandung, 2015.
[12] H. Lu dan A. Gelman, “method for estimating design-based sampling variance for surveys with weighting, poststratification, and raking,” *Journal of Official Statistics*, p. 133, 2003.
[13] Rao & Molina, Small Area Estimation Second Edition, New York: John Willey and Sons, Inc, 2015.
[14] M. Despotovic, V. Nedic, D. Despotovic and S. Cvetanovic, "Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation.," *Renewable and Sustainable Energy Reviews* , pp. 246-260, 2016.
[15] Eurostat, Guidelines on small area estimation for city statistics and other functional geographies, European, 2019.
[16] United Nations, “World Fertility Report 2015: Highlights,” New York, 2017.