BAREKENG: Journal of Mathematics and Its Applications
June 2024 Volume 18 Issue 2 Page 1009–1022
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss2pp1009-1022

TWOFOLD SUBAREA MODEL FOR ESTIMATING COMMUTER PROPORTION IN 10 METROPOLITAN AREAS

Yudi Fathul Amin[1], Indahwati[2], Anang Kurnia[3*]

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

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

Article History:

Received: 30th December 2023
Revised: 18th January 2024
Accepted: 19th March 2024
Published: 1st June 2024

Keywords:

Commuter;
Hierarchical Bayes;
Metropolitan Area;
Small Area Estimation;
Subarea Twofold.

ABSTRACT

The metropolitan area is a major contributor to national GDP. The metropolitan area is a center of attraction for many people who come to earn income as commuters. Commuters are people who carry out work activities in the center of the metropolitan area, which are carried out by residents who live in suburban areas around the center of the metropolitan area and commute regularly every day. The availability of commuter statistics from surveys for presentation level down to the smallest administrative level, such as regencies/municipalities, is unreliable. This happens because this level of presentation has poor precision due to insufficient samples due to the Statistics Indonesia survey design for making estimates at the national and provincial levels. It can be done using small area estimation (SAE) to meet increasing data needs, but existing SAE models can often estimate only at one level. To meet data requests more effectively, a model is needed that can estimate several small areas simultaneously. In SAE, one of the SAE models that can do this is the twofold subarea model. The twofold subarea model produces estimates of the proportion of commuters with good precision at the subarea level (regencies/municipalities) and area level (metropolitan area), with the RRMSE percentage value of the estimated proportion of commuters being below 25% for all regions. The results of this research can be used to present commuter data at the regencies/municipalities level and metropolitan area level where there is a lack of samples and become a new opportunity for Statistics Indonesia to increase statistical production in small areas, which is more effective compared to other SAE methods which have so far been used only to estimate one area level.

How to cite this article:

Y. F. Amin, Indahwati and A. Kurnia author., "TWOFOLD SUBAREA MODEL FOR ESTIMATING COMMUTER PROPORTION IN 10 METROPOLITAN AREAS," *BAREKENG: J. Math. & App.*, vol. 18, iss. 2, pp. 1009-1022, June, 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

The development of metropolitan areas has a positive impact on national economic growth. Metropolitan areas have grown from the beginning as small, separate cities and then expanding. Metropolitan areas then become integrated and interconnected in terms of structuring and unifying economic activities, spatial planning, and institutional forms into one urban form [1].

A metropolitan area is an urban area consisting of a stand-alone urban area or core urban area with surrounding cities that are functionally interconnected according to an infrastructure network system that connects areas with a population ranging from 1,000,000 (one million) people or more. [2].

Metropolitan area economic activity influences the level of regional economic recovery because it is one of the pillars of growth. To overcome the economic inequality between Java and outside Java, the government is trying to increase the role of metropolitan areas in Indonesia by optimizing their role and function. The contribution of metropolitan areas on Java Island to the Gross Regional Domestic Product (GRDP) is quite high compared to metropolitan areas outside Java Island. Overall, the metropolitan area on the island of Java contributes more than a third of Indonesia's GDP. It is known that Java Island has an economy that contributes 59% to the national GDP. The Jabodetabek region is the main contributor, contributing 17.31% of the national GDP [1]. Therefore, to strengthen development allocations, it is necessary to strengthen the role of metropolitan areas outside Java as growth centers.

The development of metropolitan areas in Indonesia is important for equalizing the Indonesian economy. The government has designated 10 metropolitan areas to focus on planning connectivity within integrated urban areas, connecting residential areas with strategic economic areas. The plans for 10 metropolitan areas that will be developed as stated in the 2020 - 2024 National Medium-Term Development Plan are as follows: Jabodetabekpunjur, Bandung Raya (West Java), Gerbangkertosusila (East Java), Kedungsepur (Central Java), Mebidangro (North Sumatra), Patungraya Agung (South Sumatra), Banjarbakula (South Kalimantan), Mamminasata (South Sulawesi), Sarbagita (Bali), Bimindo (North Sulawesi).

The metropolitan area has become a growing center of special attraction for many people who come here in search of income. However, house prices in the center of the metropolitan area tend to be unaffordable, so in the end many residents who work in the center of the metropolitan area choose to live on the outskirts of the metropolitan area, resulting in a process of suburbanization [3]. Working activities in the center of the metropolitan area carried out by residents who live in the peripheral areas around the center of the metropolitan area and who commute home every day routinely give rise to the commuting phenomenon and those who do this are called commuters.

Every day, commuter mobility flows rapidly from various suburban areas to the city center [4]. Most commuters commute to work [5]. The center of the metropolitan area is becoming a commuter destination due to accelerated development and industrial growth continuing to increase. Cities are core areas that act as service centers and development centers and can provide economic opportunities [6].

Sustainable urban areas as centers of economic growth and residential areas can support the achievement of the vision of a Golden Indonesia 2045. The growth of these areas not only creates opportunities but is accompanied by complex problems and challenges. This requires the provision of appropriate infrastructure to anticipate rapid urbanization, pressure on available resources, and degradation of environmental quality. Commuter data depicts the development of population mobility patterns between regencies/municipalities. It is hoped that this information will be useful for planners, policymakers, academics, and the general public to understand the impact of metropolitan area development on population mobility patterns.

Commuter data can come from several sources, one of which is the commuter survey, however, the availability of commuter statistics from the commuter survey is limited in two ways, namely that it is not carried out every year and is not carried out in all metropolitan areas in the year the survey is carried out. There is a lag in providing commuter data and it cannot be provided at once for 10 metropolitan areas. Statistics Indonesia conducts surveys designed to present data at the national and provincial administrative levels so that at certain administrative levels below the provinces, there will be a sample shortage.

Surveys have several advantages, such as saving costs, energy, time, and more accurate estimates. Surveys can collect more information [7]. However, if the sample used in the survey is inadequate, the estimation results will be inaccurate because they will cause high variance and standard errors [8]. Efforts to overcome this can be made by adding samples, but this requires quite a lot of money, so another alternative that can be done is to optimize the available data using Small Area Estimation (SAE). SAE is a statistical technique used to estimate population parameters with a small sample size. This estimation technique uses data from a large domain to estimate parameters in a smaller domain, whether regency/municipality, sub-district, or village.

The SAE method is a concept of indirect parameter estimation in a relatively small area in survey sampling where direct estimation cannot provide sufficient accuracy if the sample in a small area is small so that the resulting statistics will have a large variation or even estimates cannot be made because they are not represented in the survey [9]. Estimation in SAE based on small area models requires additional information about the relationship with observed variables which are also known as auxiliary variables. These auxiliary variables can be obtained from other surveys or censuses and must be correlated with the observed variables [10].

Commonly used SAE methods include Empirical Best Linear Unbiased Prediction (EBLUP), Empirical Bayesian (EB), and Hierarchical Bayes (HB) [11]. The EBLUP method with a linear mixed model (LMM) approach is generally used in SAE for continuous response variables and is known to perform well in SAE [12]. The EBLUP method is a parametric estimator that minimizes the mean square error (MSE) by replacing unknown variance components with a variance estimate using sample data. In this research, the SAE method is used because it can estimate at a lower level of aggregation, so the SAE method is considered the most suitable in this research. To increase efficiency, an efficient method is needed with the ability to carry out estimates simultaneously at two area levels. In SAE, one method that can do this is the twofold subarea model.

The twofold subarea model is a development of area-based models, especially the Fay Herriot (FH) model [13]. This model can be used to estimate two domains/areas simultaneously [14]. This model can be used to estimate the area mean (μ_i) and subarea mean (μ_{ij}) by borrowing the power of area and subarea random effects. Based on these references, an EBLUP estimator can be obtained for this model. Apart from that, based on previous references, we recommend using the HB estimator method for this model because it will produce accurate estimates [15].

Indeed, SAE estimation often encounters problems related to model complexity. The method of choice that can be used to overcome this problem is the SAE Hierarchical Bayes (HB) method. In the HB method, prior model parameters must be determined first. Based on the prior distribution, the posterior distribution of the estimated parameters is found. Estimation and inference of the parameters to be estimated are carried out from the posterior distribution to obtain accurate results [13]. This method can also handle data that is not normally distributed and is presented as a discrete data type [16]. Several studies use the HB method to estimate the percentage of poor people at the village level [17], [18]. The HB estimation method can be applied to the twofold subarea model. Several studies examine the twofold model [19], [20].

The need for data in small areas is increasing. Commuter data to understand the impact of metropolitan area development on population mobility patterns requires data that has a level of presentation down to the smallest administrative level such as regency/municipality, sub-district, or village. Statistics Indonesia only carries out designs to carry out estimates at the national and provincial levels. If estimates are still carried out at the administrative level, the results of the estimates carried out will not be reliable because they have poor precision due to insufficient samples. This problem is an obstacle for Statistics Indonesia in meeting data needs and providing commuter data to understand the impact of metropolitan area development. The SAE method can be used to overcome this problem, but several existing methods can only estimate at one area level. An effective method is needed to meet the increasing demand for data with the ability to carry out estimates at two area levels simultaneously. This method is the twofold subarea model.

A subarea twofold model can be used to estimate two area levels simultaneously, namely at the area and subarea levels [14]. In this model, the estimator obtained has better performance than the estimates produced by the SAE Fay-Herriot model [15]. Seeing this opportunity, researchers wish to apply this model to improve commuter outcomes at the regency/municipality level. However, the use of this model requires two levels of area estimation, namely subarea and area. In this research, the regency/municipality level is the same as the subarea level, and the metropolitan area level is the area level. This model will be able to produce better accuracy in estimating the proportion of commuters at the regency/municipality level compared to the

FH model. Apart from obtaining a better estimate of the proportion of commuters at the regency/municipality level, this method can produce a better estimate at the metropolitan area level.

2. RESEARCH METHODS

The data used in this research is secondary data collected by Statistics Indonesia, namely Sakernas (The National Labor Force Survey) and PODES (Village Potential Statistics of Indonesia). The data used is 2021 data in 10 metropolitan areas. List of metropolitan areas and regencies/municipalities with their codes:

1. Mebidangro: Karo Regency (1211), Deli Serdang Regency (1212), Medan Municipality (1275), Binjai Municipality (1276)
2. Patungraya Agung: Ogan Komering Ilir Regency (1602), Banyuasin Regency (1607), Ogan Ilir Regency (1610), Palembang Municipality (1671)
3. Jabodetabekpunjur: South Jakarta Municipality (3171), East Jakarta Municipality (3172), Central Jakarta Municipality (3173), West Jakarta Municipality (3174), North Jakarta Municipality (3175), Bogor Regency (3201), Cianjur Regency (3203), Bekasi Regency (3216), Bogor Municipality (3271), Bekasi Municipality (3275), Depok Municipality (3276), Tangerang Regency (3603), Tangerang Municipality (3671), South Tangerang Municipality (3674)
4. Bandung Raya: Bandung Regency (3204), Sumedang Regency (3211), West Bandung Regency (3217), Bandung Municipality (3273), Cimahi Municipality (3277)
5. Kedungsepur: Grobogan Regency (3315), Demak Regency (3321), Semarang Regency (3322), Kendal Regency (3324), Salatiga Municipality (3373), Semarang Municipality (3374)
6. Gerbangkertosusila: Sidoarjo Regency (3515), Mojokerto Regency (3516), Lamongan Regency (3524), Gresik Regency (3525), Bangkalan Regency (3526), Mojokerto Municipality (3576), Surabaya Municipality (3578)
7. Sarbagita: Tabanan Regency (5102), Badung Regency (5103), Gianyar Regency (5104), Denpasar Municipality (5171)
8. Banjar Bakula: Tanah Laut Regency (6301), Banjar Regency (6303), Barito Kuala Regency (6304), Banjarmasin Municipality (6371), Banjar Baru Municipality (6372)
9. Bimindo: Minahasa Regency (7102), North Minahasa Regency (7106), Manado Municipality (7171), Bitung Municipality (7172), Tomohon Municipality (7173)
10. Mamminasata: Takalar Regency (7305), Gowa Regency (7306), Maros Regency (7308), Makassar Municipality (7371)

In this study, researchers are interested in estimating the proportion of commuters at the regency/municipality and metropolitan area level using the twofold subarea model. In this research, the regency level is analogous to the subarea level, while the metropolitan area level is analogous to the area level. The variables used in this research are:

1. The response variable used is the proportion of commuters (Y) which is the proportion of commuters in regencies/municipalities and metropolitan areas. The proportion of commuters is the number of working commuters aged 15 years and over divided by the number of working residents aged 15 years and over.
2. The auxiliary variables which is variable X consists of:
   a. Average distance to economic support facilities in the village (X_1)
   b. Number of industrial centers (X_2)
   c. Percentage of villages that have lighting on the main road (X_3)
   d. The average family uses PLN electricity (X_4)
   e. Number of villages that have internet (X_5)
   f. Number of villages with 4G/LTE mobile internet signal (X_6)

2.1 Twofold Subarea Hierarchical Bayes Model

The twofold subarea model is a development of the Fay-Herriot model with two-area random effects. The twofold subarea model has the following specifications. Suppose there are m areas where each area is divided into n_i subarea [15]. We want to estimate parameters at the area level $\theta_{i.}$ and at the subarea level θ_{ij} where $i = 1, \dots, m$ and $j = 1, \dots, n_i$ so the linking model form for the subarea parameters can be written as follows:

$$\theta_{ij} = x_{ij}^T\beta + v_i + u_{ij} \quad (1)$$

where:
x_{ij} : The vector of auxiliary variables at the level of the *j*-th subarea in the *i*-th area is (*p* x 1) where (*m* > *p*)
β : vector of regression coefficients of size (*p* x 1)
v_i : area-level random effects that distribute identically and independently (*iid*) $v_i \sim N(0, \sigma_v^2)$
u_{ij} : subarea-level random effects that distribute identically and independently (*iid*) $u_{ij} \sim N(0, \sigma_u^2)$

If n_i subareas are sampled in each area, the form of the sampling model can be written as follows:

$$\hat{\theta}_{ij} = \theta_{ij} + e_{ij} \quad (2)$$

In this case $\hat{\theta}_{ij}$ is the result of direct estimation at the subarea level with sampling error e_{ij} and it is assumed that $e_{ij} \sim N(0, \psi_{ij})$ with ψ_{ij} is a known variance of direct estimator. If **Equation (1)** and **Equation (2)** are connected, a twofold subarea model will be produced which can be written as follows:

$$\hat{\theta}_{ij} = x_{ij}^T\beta + v_i + u_{ij} + e_{ij} \quad (3)$$

The subarea twofold Hierarchical Bayes model was adapted from research conducted [21]. The model can be written as follows:

$$\hat{\theta}_{ij} | \theta_{ij} \sim N(\theta_{ij}, \psi_{ij}) \quad (4)$$
$$\theta_{ij} = x_{ij}^T\beta + v_i + u_{ij} \text{ with } v_i | \sigma_v^2 \sim N(0, \sigma_v^2) \text{ and } u_{ij} | \sigma_u^2 \sim N(0, \sigma_u^2) \quad (5)$$
$$\beta \text{ with } \beta \sim N(\mu_{\beta[p]}, \tau_{\beta[p]}) \quad (6)$$
$$\sigma_u^2 \text{ with } \sigma_u^2 \sim IG(\tau_{ua}, \tau_{ub}) \ (prior) \quad (7)$$
$$\sigma_v^2 \text{ with } \sigma_v^2 \sim IG(\tau_{va}, \tau_{vb}) \quad (8)$$

where:
$\hat{\theta}_{ij}$: direct estimator at the level of the *j*-th subarea in the *i*-th area
ψ_{ij} : variance of direct estimation results at the *j*-th subarea level in the *i*-th area
x_{ij} : vector of auxiliary variables at the *j*-th subarea level in the *i*-th area
β : vector of regression coefficients
v_i : area-level random effects
u_{ij} : subarea-level random effects
σ_v^2 : random effect variance of area v_i
σ_u^2 : random effect variance of subarea u_{ij}
μ_β : parameters of the mean prior distribution of β
τ_β : variance parameter as a prior of β
τ_v : prior distribution parameters σ_v^2 with $\tau_{va} > 0$ and $\tau_{vb} > 0$
τ_u : prior distribution parameters σ_u^2 with $\tau_{ua} > 0$ and $\tau_{ub} > 0$

2.2 Accuracy

Generally, the goodness of the estimation results can be measured using the resulting sampling variance. Relative Root Mean Square Error (RRMSE) is the root mean square error normalized by the root mean square value where each residual is scaled to the actual value [22], [23]. RRMSE can be calculated with the following formula:

$$RRMSE\,(\hat{\theta}) = \frac{\sqrt{MSE(\hat{\theta})}}{\hat{\theta}} \times 100\% \qquad (9)$$

where:

$\hat{\theta}$: estimator value produced in a particular domain

$RRMSE(\hat{\theta})$: a measure of precision that can be calculated from the sampling variance as $\sqrt{MSE(\hat{\theta})}$

2.3 Data Analysis Procedures

Before carrying out the research stage, the data pre-processing stage is first carried out which will be processed as follows:

1. Preparing data on the proportion of commuters from Sakernas 2021 data
2. Preparing data for auxiliary variables (X_1, X_2, X_3, X_4, X_5, X_6) per regency/municipality in 10 metropolitan areas originating from PODES 2021

Furthermore, the research method and stages that will be carried out to achieve the research objectives are to estimate the proportion of commuters using the twofold subarea model approach as follows:

1. Estimating variance components
2. Estimating model parameters
3. Estimating the proportion of commuters for each regency/municipality and metropolitan area
4. Calculating RRMSE value

In this research, the software used is the R programming language with the saeHB.twofold package. In this package there is a function that is used to estimate small areas using a twofold subarea model.

3. RESULTS AND DISCUSSION

Commuter data to capture developments in population mobility patterns can be useful for planners, policymakers, academics, and the general public to understand the impact of metropolitan area development on population mobility patterns. This data is needed at the regional/municipal, sub-district, and village levels of presentation. However, surveys conducted by Statistics Indonesia are generally designed to present statistical data at the national and provincial levels. The presentation of statistical data at the regencies/municipalities level will result in a lack of samples, so statistical estimation results have poor precision with high relative standard error (RSE) values. The SAE method can be used to overcome this problem.

The SAE method used so far is limited to estimating a single area level. More effective methods are needed to carry out estimates at two small area levels simultaneously to meet the increasing need for statistical data in small areas. The model that can meet these needs is the SAE subarea twofold model. This model develops the Fay-Herriot model with random effects at the area and subarea levels. The EBLUP estimation method on this model with results showing that the resulting estimator is more efficient compared to the estimator produced from the SAE Fay-Herriot model [15]. The resulting estimator has better precision compared to direct estimation [21].

The use of the twofold subarea model can be a new opportunity for BPS to increase statistical production in small areas, which is more effective than other SAE methods that have so far been used to estimate only a single area level. At the parameter estimation process is carried out in the twofold subarea model using the Hierarchical Bayes method. Next, using the saeHB.twofold package, load data, namely data on response variables and auxiliary variables, determine the number of Markov Chains (chains), number of iterations, iterations for burn-in period and thin and carry out the estimation process using the MCMC method, the Gibbs Sampling algorithm with syntax in Rstudio format.

The number of chains used is 10 with the number of iterations for each chain being 250,000, with a burn-in period of 1000 and a thinning performance rate of 8. Determination of the number of chains, iterations, thinning, and burn-in is obtained from several experimental processes starting from a small number to convergence.

3.1 Selection of Auxiliary Variables

Based on the credible interval between 95% of the posterior distribution for the β_k parameter in **Table 1**, five auxiliary variables have a significant effect on the response variable for the proportion of commuters because they do not contain the value 0 in the interval. These significant variables are the average distance to economic support facilities in the village, the number of industrial centers, the percentage of villages that have lighting on the main road, the average number of families using PLN electricity, and the number of villages that there is internet.

Table 1. Summary of Estimation Results for Twofold Subarea Model

Par.	Mean	Standard Error	2.5%	Median	97.5%
β_0	0.2968	0.0263	0.2455	0.2964	0.3478
β_1	-00446	0.0098	-0.0635	-0.0446	-0.0253
β_2	-0.0275	0.0081	-0.0433	-0.0276	-0.0115
β_3	0.0255	0.0089	0.0080	0.0255	0.0427
β_4	0.0215	0.0102	0.0016	0.0215	0.0413
β_5	-0.0524	0.0114	-0.0751	-0.0525	-0.0303

***Data source**: the output of estimation parameter from R program*

Convergence checks are carried out visually for several parameters in **Figure 1** with a traceplot which shows that the graph of the Markov chain has a random pattern and is stable at a certain mean. The burn-in process on the trace plot has been completed so that the posterior distribution has converged and provides quite stable values because it does not form a particular pattern.

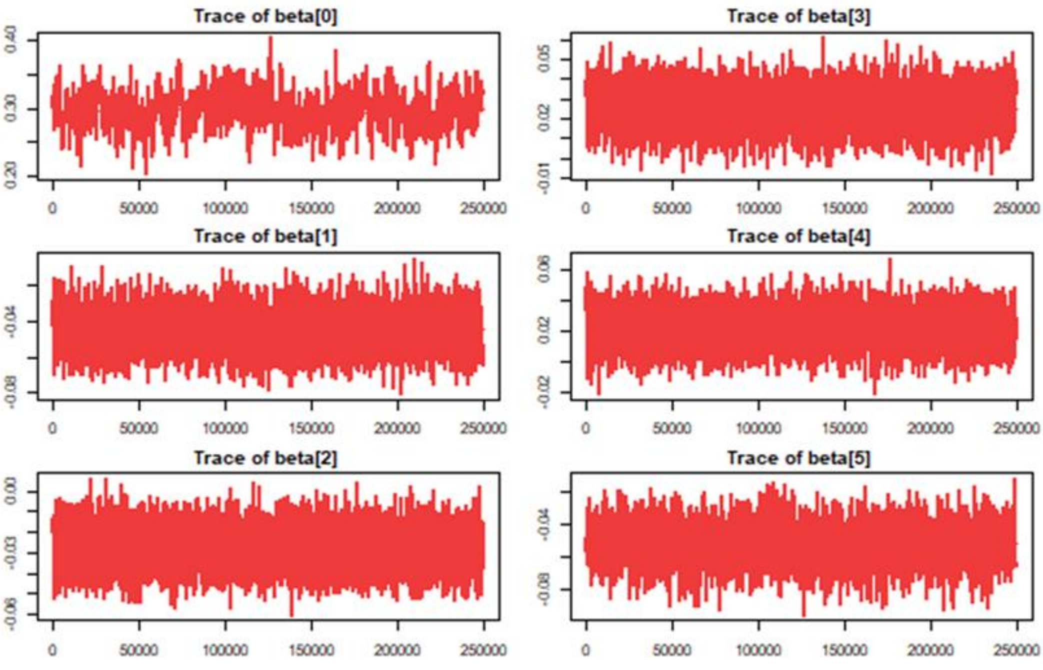

Figure 1. Trace Plot Parameters β_0 – β_5 Twofold Subarea Model

The autocorrelation plot shows a decreasing pattern as shown in Figure 2. The plot shows that the autocorrelation values approach one and then the values decrease towards zero so it can be said that in the chain there is a correlation in estimating the proportion of commuters between iterations. This correlation shows that the algorithm is already in the target distribution area.

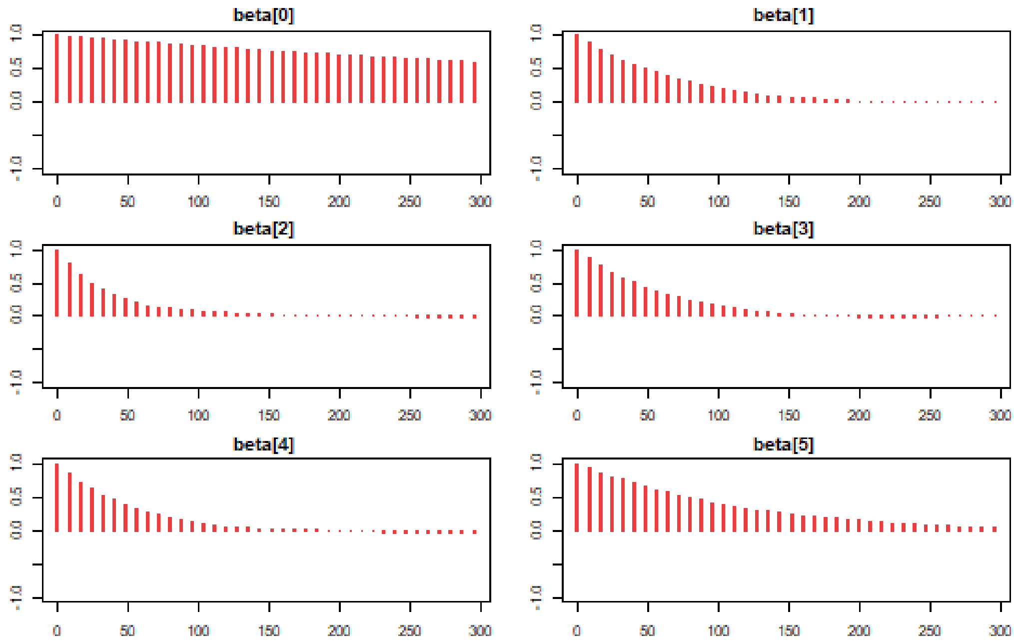

Figure 2. Autocorrelation Plot of Parameters $\beta_0 - \beta_5$ Twofold Subarea Model

The density plot in Figure 3 of the posterior parameters β0 – β5 in the twofold subarea model visually follows a normal distribution. The density plot shows that the distribution pattern of the observed parameter estimates tends to be symmetrical.

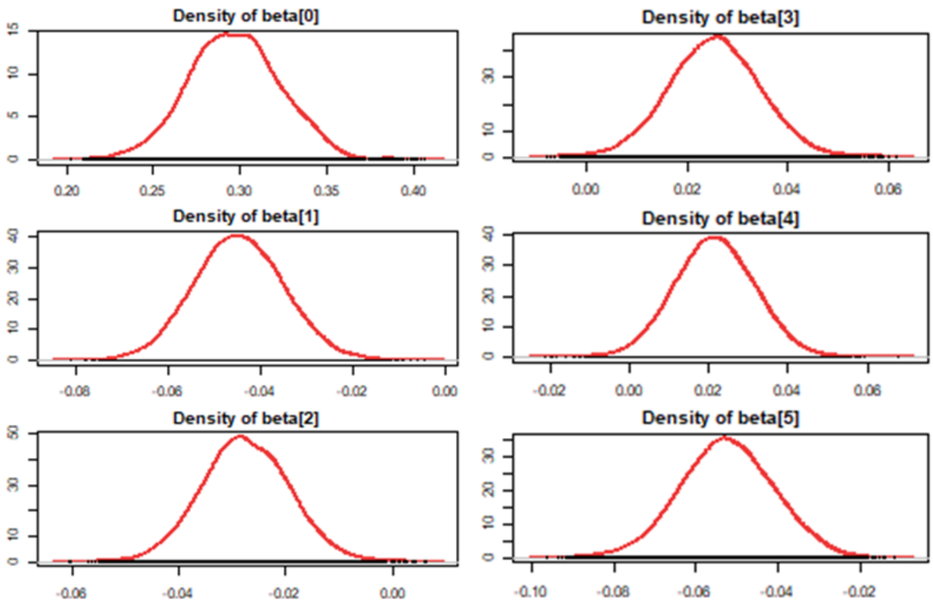

Figure 3. Density Plot of Posterior Distribution of Parameters $\beta_0 - \beta_5$ Twofold Subarea Model

3.2 Results of Estimating the Proportion of Commuters

In this study, the results of estimating the proportion of commuters at the regency/municipality (subarea) level using the twofold subarea model can be seen in Figure 4 and Figure 5.

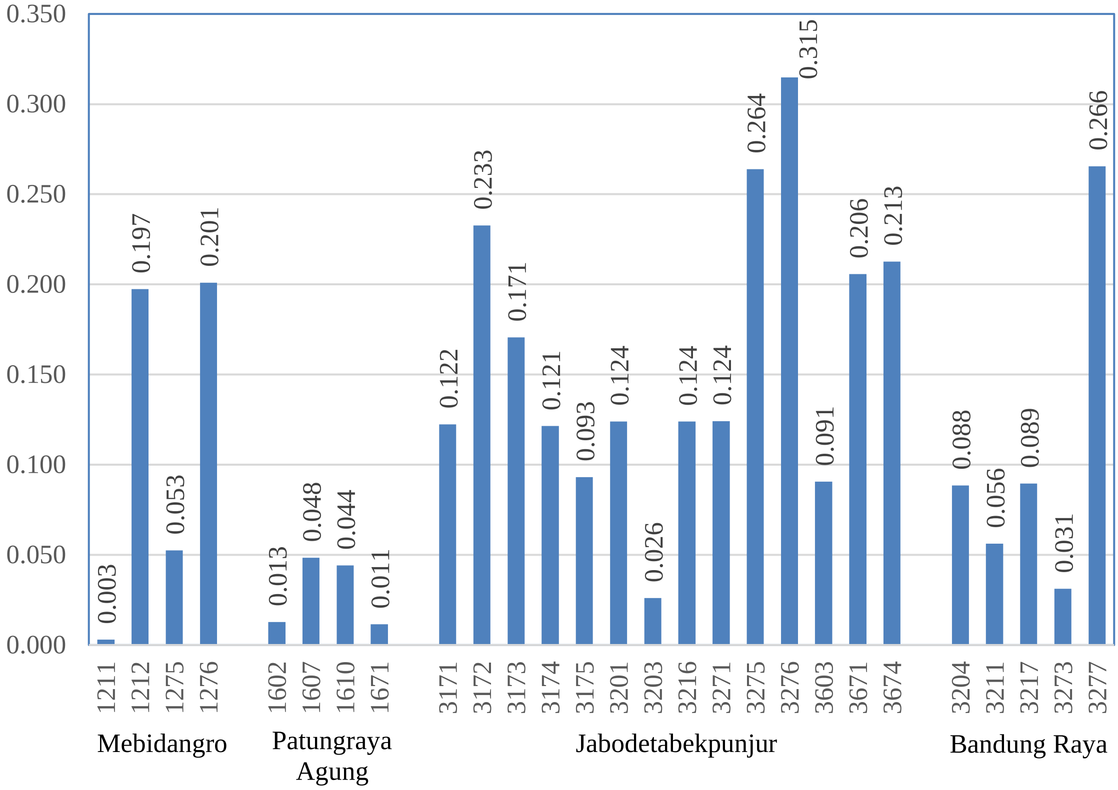

Figure 4. Results of Estimating the Proportion of Commuters at Regency/Municipality Level (Code 1211-3277) Using the Twofold Subarea Model

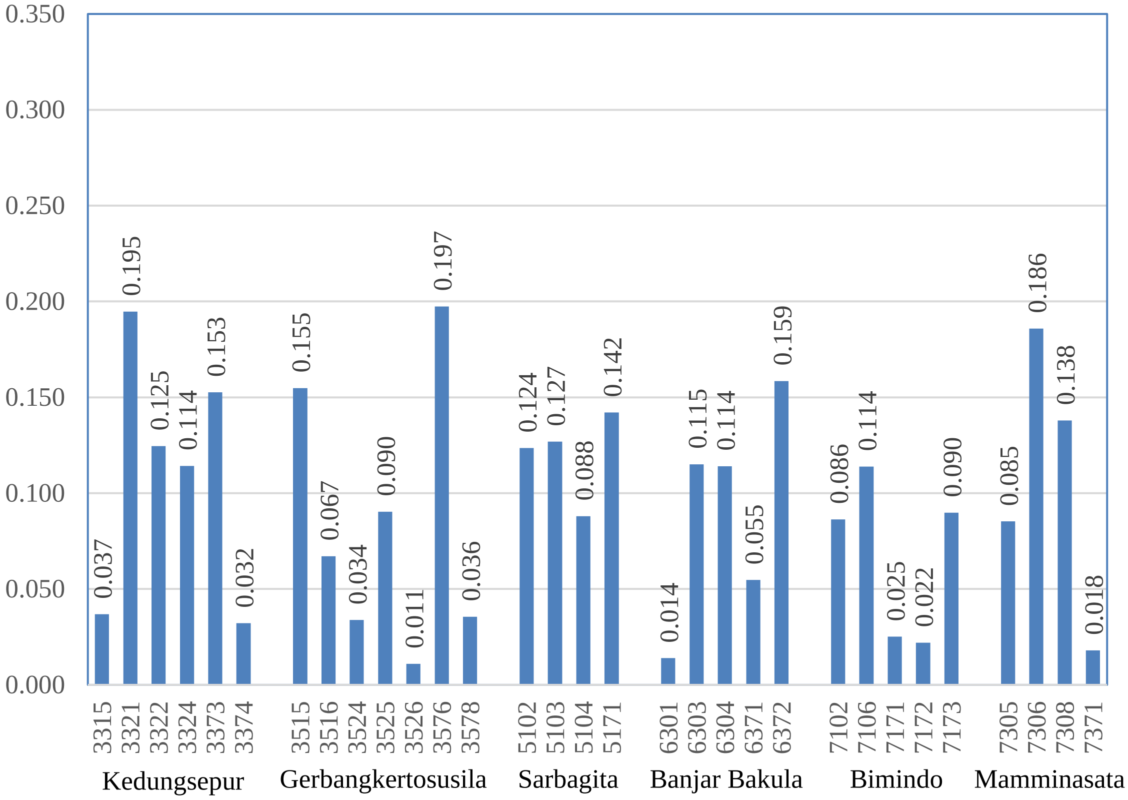

Figure 5. Results of Estimating the Proportion of Commuters at Regency/Municipality Level (Code 3315-7371) Using the Twofold Subarea Model

Based on **Figure 4** and **Figure 5**, the results of estimating the proportion of commuters at the regency/municipality level (subarea) using the twofold subarea model produce values of 0.003 to 0.315. The regencies/municipalities with the five highest proportions of commuters are Depok Municipality (3276) with a value of 0.315, then Cimahi Municipality (3277) with 0.266, Bekasi Municipality (3275) with 0.264, East Jakarta Municipality (3172) with 0.233, and South Tangerang Municipality (3674) with 0.213. The regencies/municipalities with the five highest proportions of commuters, four of which are in the Jabodetabekpunjur metropolitan area.

The regencies/municipalities with the five lowest proportions of commuters are Karo Regency (1211) with a commuter proportion of 0.003, then Bangkalan Regency (3526) with 0.011, Palembang Municipality (1671) with 0.011, Ogan Komering Ilir Regency (1602) with 0.013, and Tanah Laut Regency (6301) with 0.014. The regencies/municipalities with the five lowest proportions of commuters, four of which are regencies/municipalities located outside Java.

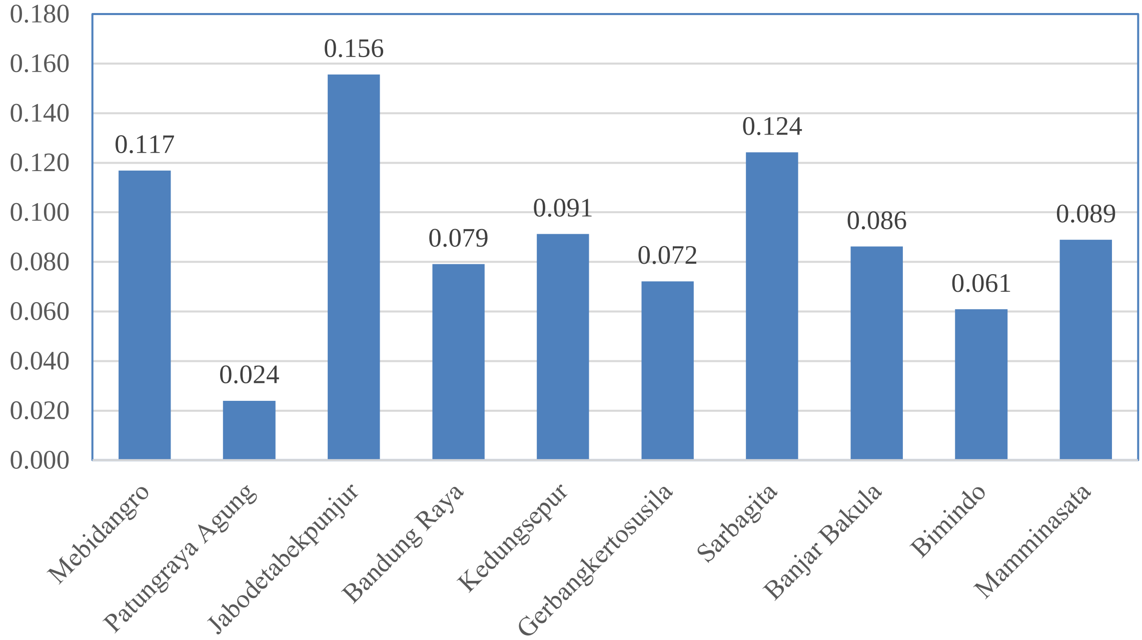

Figure 6. Results of Estimating the Proportion of Commuters at the Metropolitan Area Level Using the Twofold Subarea Model

The results of estimating the proportion of commuters at the metropolitan area (area) level can be seen in **Figure 6**. Based on the figure above, the metropolitan area with the highest proportion of commuters is Jabodetabekpunjur with a proportion of 0.156. Meanwhile, the metropolitan area with the lowest proportion of commuters is Patungraya Agung with a proportion of 0.024.

3.3 RRMSE Results of Estimating the Proportion of Commuters

The measure used to compare the goodness of the estimate of the proportion of commuters at the regency/municipality level (subarea) and the metropolitan area (area) level is to use the RRMSE percentage measure. The regency/municipality with the highest RRMSE percentage is Ogan Komering Ilir Regency (1602) with a value of 21.07. This regency is the only regency/municipality with an RRMSE percentage value above 20%, other regencies/municipalities have an RRMSE percentage value below 20%. Meanwhile, the regency/municipality with the lowest RRMSE percentage value is Tomohon Municipality (7173) with an RRMSE value of 2.35. Based on **Figure 7** and **Figure 8**, the RRMSE results from estimating the proportion of commuters for all regencies/municipalities are below the value of 25%. This shows that the use of the twofold subarea model can provide good precision for estimating the proportion of commuters at the regency/municipality level in all regions.

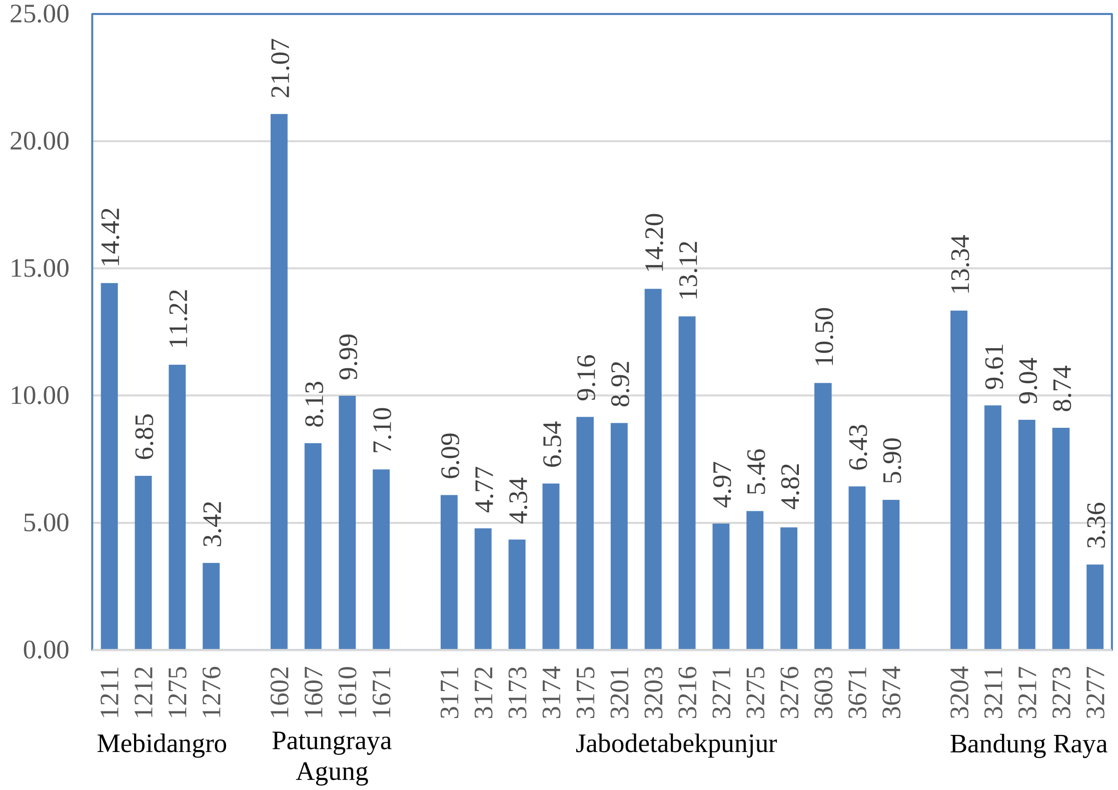

Figure 7. RRMSE Figures for Estimating the Proportion of Commuters at Regency/Municipality Level (Code 1211-3277) Using the Twofold Subarea Model

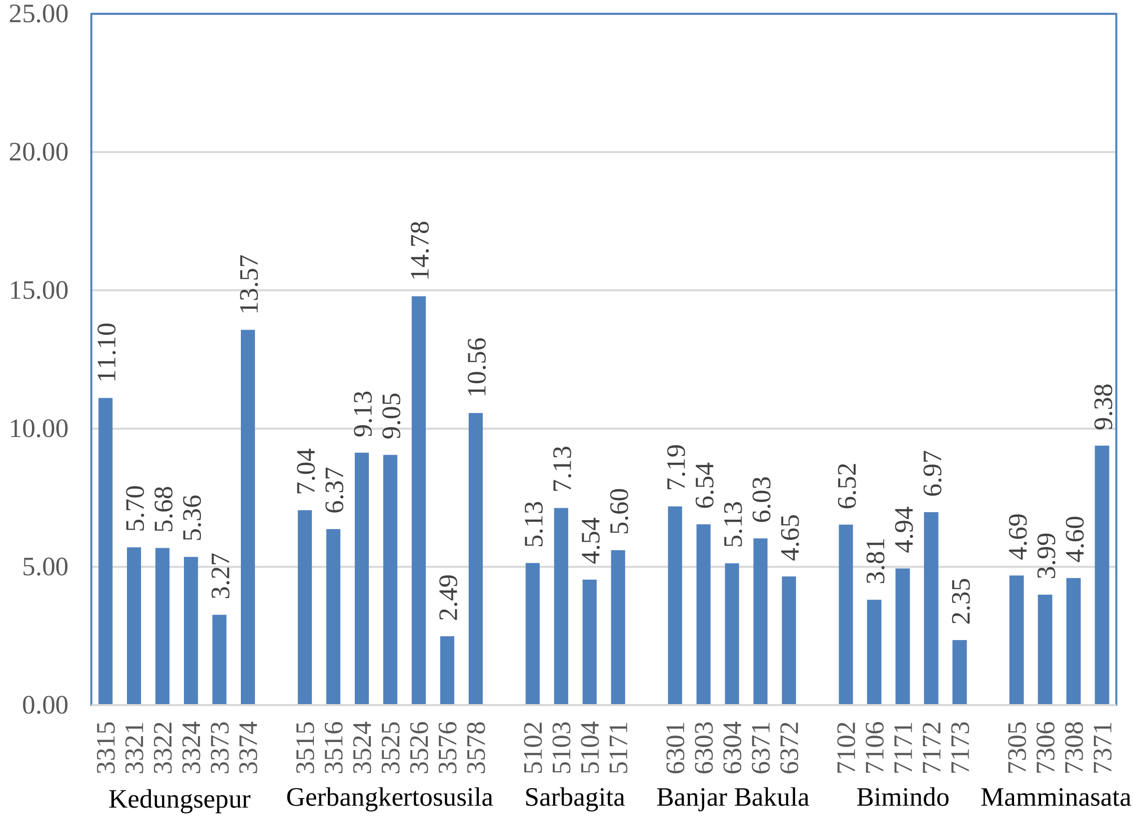

Figure 8. RRMSE Figure Estimating the Proportion of Commuters at Regency/Municipality Level (Code 3315-7371) Using the Twofold Subarea Model

Based on **Figure 9**, the RRMSE percentage value for estimating the proportion of commuters at the metropolitan area (area) level using the twofold subarea model produces a value of less than 25%. The metropolitan area with the highest RRMSE percentage value is Bandung Raya with 6.10. Meanwhile, the metropolitan area with the lowest RRMSE percentage is Jabodetabekpunjur with a value of 2.22.

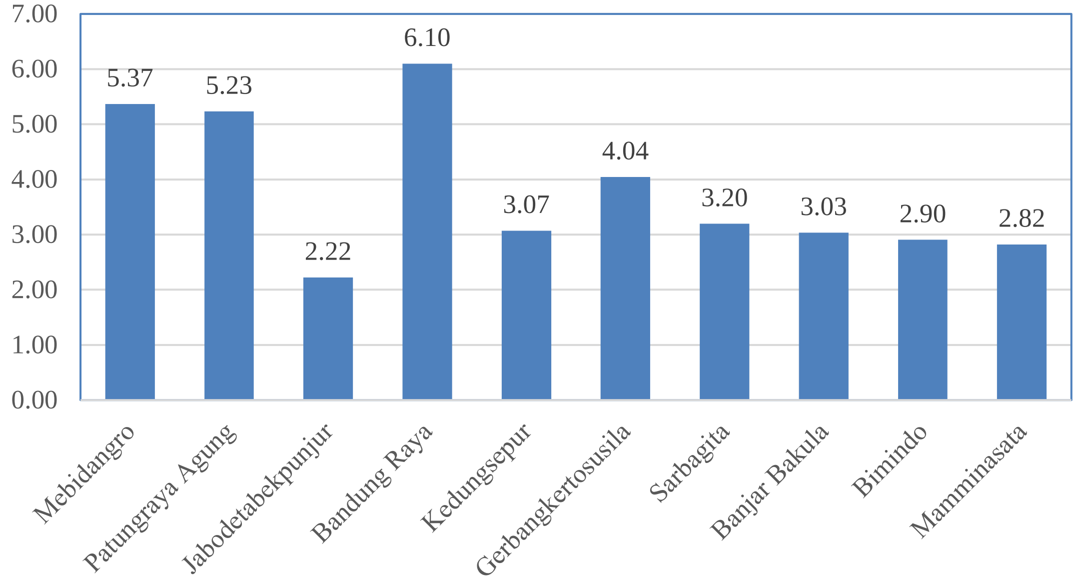

Figure 9. RRMSE Figures for Estimating the Proportion of Commuters at Metropolitan Area Level Using the Twofold Subarea Model

The final review is related to the results of estimating the variance of random effects. In the twofold subarea model, there are two random effect components, namely at the regency/municipality level (subarea) and the metropolitan area level (area). At the regency/municipality (subarea) level, the random effect variance estimate obtained is 0.0166. The estimation results are greater than zero, which indicates that the subarea has a random effect on the estimation results at the regency/municipality (subarea) level. In estimating the variance of random effects at the metropolitan area (area) level, the result obtained was 0.0302. This value is greater than zero, which indicates that there is a random effect on the estimation results at the metropolitan area (area) level. At both area levels, the estimated value of the random effect variance is more than zero. The twofold subarea model is appropriate to use to estimate the proportion of commuters at the regency/municipality and metropolitan area levels. Based on the results mentioned previously, the random influence information in the twofold subarea model is distributed at both area levels, namely the regency/municipality level and the metropolitan area level.

4. CONCLUSIONS

Based on the results and discussion, the conclusion that can be drawn is that the RRMSE percentage value for the estimated proportion of commuters is below 25% for all regions. The twofold subarea model strengthens the estimation results by including random effect information. This information is obtained from the random effect component at two different levels, namely at the regency/municipality level and the metropolitan area level. Based on these results, it can be concluded that the twofold subarea model produces estimates of the proportion of commuters with good precision at the subarea level (regency/municipality) and area level (metropolitan area). The results of this research can be used to present commuter data at the regencies/municipalities level and metropolitan area level where there is a lack of samples and become a new opportunity for Statistics Indonesia to increase statistical production in small areas which is more effective compared to other SAE methods which have so far been used only to estimate one area level.

REFERENCES

[1] [BPS] Badan Pusat Statistik, *Kajian Wilayah Metropolitan Patungraya Agung dan Banjarbakula Menggunakan MSA MPD*. Jakarta: Badan Pusat Statistik, 2022.

[2] I. Buchori, M. B. A. Yusuf, A. W. Sejati, dan A. Sugiri, “Rancang Bangun Sistem Informasi Keruangan Berbasis Web-Sistem Informasi Geografis (SIG) Untuk Pembangunan Metropolitan Berkelanjutan Studi Kasus: Wilayah Metropolitan Semarang,” dalam *Prosiding Seminar Nasional Tata Ruang dan Space# 2*, Universitas Hindu Indonesia, 2015.
[3] A. D. Irjayanti, D. W. Sari, dan I. Rosida, “Perilaku Pemilihan Moda Transportasi Pekerja Komuter: Studi Kasus Jabodetabek,” *Jurnal Ekonomi Dan Pembangunan Indonesia*, vol. 21, no. 2, hlm. 125–147, 2021.
[4] S. Khadijah, D. P. Lubis, dan R. A. Kinseng, “Komunikasi Kelompok Di Kalangan Komuter,” *Jurnal Komunikasi Pembangunan*, vol. 13, no. 2, 2015.
[5] F. Abdillah, “Fungsi Kereta Api dan Karakteristik Mobilitas Komuter Yogyakarta - Surakarta,” Universitas Gajah Mada, Yogyakarta, 2014.
[6] D. Noviyanti, A. E. Pravitasari, dan S. Sahara, “Analisis perkembangan wilayah provinsi Jawa Barat untuk arahan pembangunan berbasis wilayah pengembangan,” *Jurnal Geografi*, vol. 12, no. 1, hlm. 57–73, 2020.
[7] T. L. Jones, M. A. J. Baxter, dan V. Khanduja, “A quick guide to survey research,” *The Annals of The Royal College of Surgeons of England*, vol. 95, no. 1, hlm. 5–7, 2013.
[8] D. Firmansyah, “Teknik Pengambilan Sampel Umum dalam Metodologi Penelitian: Literature Review,” *Jurnal Ilmiah Pendidikan Holistik (JIPH)*, vol. 1, no. 2, hlm. 85–114, 2022.
[9] M. Y. Darsyah dan S. Iriyanto, “Small Area Estimation Untuk Pemetaan Kemiskinan di Kabupaten Demak,” dalam *PROSIDING SEMINAR NASIONAL & INTERNASIONAL*, 2014.
[10] A. H. Hakim dan N. Hajarisman, “Pendugaan Rata-rata Pengeluaran Per Kapita Menurut Kabupaten/Kota di Provinsi Jawa Barat Melalui Empirical Best Linear Unbiased Prediction dalam Pendugaan Area Kecil,” dalam *Bandung Conference Series: Statistics*, 2022, hlm. 474–481.
[11] N. Permatasari dan W. Larasati, “Perbandingan Metode SAE EBLUP dan SAE HB Pada Pendugaan Area Kecil (Studi Kasus Pendugaan Kemiskinan di Provinsi Jawa Timur),” *Jurnal Statistika Dan Aplikasinya*, vol. 6, no. 1, hlm. 96–108, 2022.
[12] F. Apriani, “Pemodelan pengeluaran per kapita menggunakan small area estimation dengan pendekatan semiparametrik penalized spline,” *Program Pascasarjana, Institut Teknologi Sepuluh Nopember, Surabaya*, 2017.
[13] R. Saadi dan A. Ubaidillah, “Pendugaan Area Kecil Angka Partisipasi Kasar Perguruan Tinggi di Indonesia Menggunakan Model Subarea Twofold dengan Pendekatan Hierarchical Bayes,” dalam *Seminar Nasional Official Statistics*, 2022, hlm. 123–136.
[14] J. N. K. Rao dan I. Molina, *Small area estimation*. John Wiley & Sons, 2015.
[15] M. Torabi dan J. Rao, “On small area estimation under a sub-area level model,” *J Multivar Anal*, vol. 127, hlm. 36–55, 2014.
[16] E. Sunandi, A. Kurnia, K. Sadik, dan K. A. Notodiputro, “A Bayesian Logit-Normal Model in Small Area Estimation,” dalam *Journal of Physics: Conference Series*, IOP Publishing, 2021, hlm. 012039.
[17] I. Y. Wulansari dan G. Pawitan, “Pendugaan area kecil terhadap proporsi rumah tangga miskin level kelurahan di Kabupaten Sampang menggunakan Hierarchical Bayes (HB) logit normal,” 2014.
[18] I. K. G. Sukarsa dan I. G. K. Gandhiadi, “Pendugaan Proporsi Rumah Tangga Miskin Tingkat Desa di Provinsi Bali dengan Metode Empirical Best Linear Unbiased Prediction dan Bayesian,” *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 15, no. 2, hlm. 215–222, 2021.
[19] B. Nandram, “Bayesian predictive inference of a proportion under a twofold small-area model,” *J Off Stat*, vol. 32, no. 1, hlm. 187–208, 2016.
[20] S. Cai, J. N. K. Rao, L. Dumitrescu, dan G. Chatrchi, “Effective transformation-based variable selection under two-fold subarea models in small area estimation,” *Statistics in transition new series*, vol. 21, no. 4, hlm. 68–83, 2020.
[21] A. L. Erciulescu, N. B. Cruze, dan B. Nandram, “Model-based county level crop estimates incorporating auxiliary sources of information,” *J R Stat Soc Ser A Stat Soc*, vol. 182, no. 1, hlm. 283–303, 2019.
[22] E. Eyduran, D. Zaborski, A. Waheed, S. Celik, K. Karadas, dan W. Grzesiak, “Comparison of the predictive capabilities of several data mining algorithms and multiple linear regression in the prediction of body weight by means of body measurements in the indigenous Beetal goat of Pakistan,” *Pak J Zool*, vol. 49, no. 1, 2017.
[23] P. Mahlangu *dkk.*, “Indirect estimation of structural parameters in South African forests using MISR-HR and LiDAR remote sensing data,” *Remote Sens (Basel)*, vol. 10, no. 10, hlm. 1537, 2018.