Small Area Estimation Under A Pseudo-Unit Level Model with First-Order Autoregressive Time Effects and Measurement Error

Anang Kurnia, Dian Surida, Siti Muchlisoh, Ita Wulandari, Dian Handayani

Abstract—**Statistics Indonesia (BPS) provides employment data through the National Labor Force Survey (SAKERNAS). SAKERNAS data collection is carried out quarterly under a rotating panel survey with a sample unit of the rotation group. Our study focuses on the Rao-Yu unit level model and its modifications. The modified Rao-Yu unit level model is required to handle pseudo panel units i.e., the units are replaced by the summary of units in the rotation group. In this context, these units are referred to as pseudo units. A modification was made by adding a measurement error pseudo unit into the Rao-Yu unit level model. The simulation study showed that the greater variance between areas, the better the capability of Rao-Yu pseudo unit level model compared with the Rao-Yu unit level model in estimating the parameters of interest. Our method proposal was applied to data on the quarterly unemployment rate at the district/city level in one of Indonesia's provinces according to the 2011-2014 SAKERNAS panel rotation data.**

Index Terms—**Empirical best linear unbiased prediction, Rotation group, Pseudo panel, Pseudo unit.**

I. Introduction

Statistics Indonesia (BPS) provides employment data through the National Labor Force Survey (SAKERNAS). The SAKERNAS aims to obtain information on the unemployment rate and its changes from time to time at the national, provincial, or district/city levels [1]. The 2011-2014 SAKERNAS is a survey with a quarterly sampling method. The SAKERNAS data was designed in panel rotation by the rotation group as a sample unit. In the rotation panel framework, the total sample of census blocks is divided into four sample packages. Each package in each census block is formed by four household groups. Quarterly, the sample size consists of four groups from different packages. The SAKERNAS rotation panel is designed quarterly by maintaining three-quarter groups of the previous quarter and adding one-fourth of new groups of the current quarter [2].

However, some problems occur in the labor estimation at the district/city level which uses the quarterly SAKERNAS data with inadequate sample size because the quarterly SAKERNAS is only designed to estimate the parameter at the provincial and national levels. If the estimation of employment at the district/city level is done with direct estimation, then it can produce a large standard error due to the small sample size. To resolve these issues, it is necessary to study how to estimate parameters at the district/city level according to the rotation panel survey when the sample size is insufficient. The small area estimation (SAE) model is an alternative that can be used to estimate the parameters in an area when the sample size in the area is too small to obtain an adequate level of accuracy in addition to direct estimation [3].

The Rao-Yu model modified by Muchlisoh [2] is suitable for unit-level analysis to estimate the unemployment rate using the 2011-2014 SAKERNAS data. The study results showed that the estimators were commensurate with the Rao-Yu model estimators and better than the direct estimators; therefore, the model can be used as an alternative in the SAE when the variances of sampling errors are unavailable. However, to apply the model in the SAKERNAS rotating panel data, Muchlisoh [2] eliminated information on rotation groups, as if the observed units (rotation groups) are the same all the time. While in practice, these units are different from time to time. One way to solve the problem is to use the pseudo panel method [4].

Deaton was the first to introduce the pseudo panel method [5]. In his paper, Deaton revealed that, in a survey, the same individual in the sample cannot be observed over time. Hence, individual groups are divided into sub-populations based on certain characteristic similarities, commonly referred to as "cohorts" which can be observed [4]. The principle of the pseudo panel is to follow cohorts (i.e. groups of individuals sharing a set of characteristics that are improved from time to time), rather than individuals over time. Individual values are replaced by their intra-cohort means. Hence, the value of the observed

Manuscript received May 13, 2023; revised December 19, 2023.
A. Kurnia is an associate professor in Statistics and Data Science at Department of Statistics, IPB University, 16680, Indonesia. (e-mail: anangk@apps.ipb.ac.id).
D. Surida is a statistician at Ministry of Women's Empowerment and Child Protection, Indonesia (e-mail: dsurida@gmail.com)
S. Muchlisoh is a lecturer at Politeknik Statistika STIS, Indonesia (e-mail: sitim@bps.go.id).
I Wulandari is a doctoral student at IPB University. She is a lecturer at Politeknik Statistika STIS, Indonesia (e-mail: ita.wulandari@stis.ac.id).
D. Handayani is an associate professor in Statistics at State University of Jakarta, Indonesia (e-mail: dianh@unj.ac.id).

variables in the cohort still has a measurement error. If not handled properly, the measurement error will result in biased estimators.

The cohorts in the pseudo panel model are assumed as units if applied to the SAE-unit level. Individuals in this unit are different from time to time, so the units are then referred to as pseudo units. Individual values in the units are replaced by their intra-unit means. Referring to Muchlisoh, this pseudo unit is the rotation groups that are assumed the same over time, then the measurement error that arises from the pseudo unit is ignored. Therefore, in this article, we would improve the model modified by Muchlisoh [2], by applying the pseudo panel model to the model by paying attention to a measurement error arising from the pseudo units. A simulation study was conducted and the quarterly rate of unemployment at the district level was also estimated.

Handayani et al. [6] compared the three approximation methods to obtain maximum likelihood estimates in GLMM. The comparison is based on a simulation study. Furthermore, these methods are applied to analyze the salamander data. Handayani et al. [7] reviewed the Empirical Based Predictor (EBP) to estimate small area linear parameters as well as nonlinear parameters whenever variable interest has skewed distribution. They assume that the variable of interest would follow log normal distribution after taking logarithm transformation. Then, the EBP would be derived under SAE unit level model (nested error regression model) using this transformed variable of interest.

The remainder of the paper is arranged as follows. Section 2 describes the Rao-Yu model and its modification; Section 3 describes the pseudo panel method; Section 4 describes the proposed model; Section 5 describes our simulation and application studies; and Section 6 describes the conclusion.

II. Methods

A. Rao-Yu Model and Its Modification

The Rao-Yu model is the development of the basic SAE which is Fay-Herriot model by adding a random area-time effect following the first autoregressive process [8]. The model is as follows:

$$\hat{\theta}_{it} = \theta_{it} + e_{it} \tag{1}$$
$$\theta_{it} = x_{it}^T\beta + v_i + u_{it},\ \ u_{it} = \rho\, u_{i,t-1} + \varepsilon_{it},\ |\rho| < 1 \tag{2}$$

θ_{it} is the parameter of interest on small area i-th and time t ($i = 1, \dots, m;\ t = 1, \dots, T$) also $\hat{\theta}_{it}$ is the direct unbiased estimator of θ_{it}. The $x_{it} = (x_{ij1}, \dots, x_{ijp})^T$ is area-time auxiliary variable size $p\ x\ 1$ which is assumed to be available in the i-th small areas and time t. The $\beta = (\beta_1, \dots, \beta_p)^T$ is a column vector of the $p\ x\ 1$ regression coefficient. Also, the sampling error e_{it} is independent of that normally distributed given θ_{it}, with $e_{it} | \theta_{it} \sim iidN(0, \sigma^2_{it})$. Furthermore, v_i is a random effect for i-th small area with $v_i \sim iid\ N(0, \sigma^2_v)$ and u_{it} is a random effect for i-th small area at time point t with $\varepsilon_{it} \sim iid\ N(0, \sigma^2_\varepsilon)$, that is assumed to follow the first-order autoregressive process within each area i. The constant ρ is the autoregressive coefficient. The errors $\{e_{it}\}$, $\{v_i\}$, and $\{u_{it}\}$ are assumed to be independent. The combination of the sampling error model (1) and the linking model (2) will produce the following model:

$$\hat{\theta}_{it} = x_{it}^T\beta + v_i + u_{it} + e_{it},$$
$$u_{it} = \rho\, u_{i,t-1} + \varepsilon_{it},\ |\rho| < 1 \tag{3}$$

The random effect v_i, u_{it} and e_{it} are assumed mutually independent. The condition $|\rho| < 1$ ensures the stationarity of the series defined by (3) to obtain the first-order autoregressive process.

Muchlisoh [2] modified the Rao-Yu model area level to Rao-Yu model unit level. The modification aims to make the Rao-Yu Model suitable for analysis at the unit level when the sampling error is unavailable. Let y_{itj} is j-th sample unit of i-th small area at time t and assume that unit specific auxiliary variable x_{itj} is available for each population element j in i-th small area at time t. The model is defined as,

$$y_{itj} = x_{itj}^T\beta + v_i + u_{it} + e_{itj},$$
$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it},\ |\rho| < 1 \tag{4}$$

Then we called this model as Model-1. While $i = 1, 2, \dots, m$ is the index for a small area, $t = 1, 2, \dots, T$ is the index for time, and $j = 1, 2, \dots, \eta_{it}$ with η_{it} is the number of sample units in a small area i at time t. The $\beta = (\beta_1, \dots, \beta_p)^T$ is a column vector of the $p\ x\ 1$ regression coefficient. Assume that $v_i \sim iid\ N(0, \sigma^2_v)$ is a random effect for i-th small area at time point t. The u_{it}'s is assumed to follow the first-order autoregressive process within each area i, with $\varepsilon_{it} \sim iid\ N(0, \sigma^2_\varepsilon)$. The error e_{itj} is assumed $e_{itj} \sim iid\ N(0, \sigma^2_e)$ and the v_i, u_{it}, and e_{itj} are mutually independent.

B. Pseudo Panel Method

The Pseudo panel method is an alternative to using panel data for estimating fixed effects models when only independent repeated cross-sectional data are available. They are widely used to estimate price or income elasticities and carry out life-cycle analyses, for which long-term data are required, but panel data have limits in terms of availability over times and attrition [4]. Deaton [5], the first suggestion to using panel methods of repeated cross-sectional data, reveals that when the samples are withdrawn, the same individual cannot be observed over times. So that, the group of individuals divided into sub-populations based on certain characteristic similarities, commonly referred to as "cohorts", can be observed. The principle of the pseudo panel is to observe cohorts, i.e., stable groups of individuals, rather than individuals over time. Individual variables are replaced by their intra-cohort means. The pseudo panel model is obtained by approaching the linear model as follows:

$$y_{it} = x_{it}^T \beta + \alpha_i + \delta_{it} \tag{5}$$

where $i = 1,2,\ldots,N$ is the index for the individual in the population, $t = 1,2,\ldots,T$ is the index for time. The y_{it} is variable of interest for the *i*-th individual at *t*-th time, x_{it} is the explanatory variable for the *i*-th individual and the *t*-th time. α_i is the individual fixed effect of the variable of interest that is fixed over time. The δ_{it} is a sampling error. Then, by substituting the individual using the observed intra-cohort means, model (5) satisfies the following conditions:

$$y_{ct}^* = x_{ct}^* \beta + \alpha_c^* + \delta_{ct}^* \tag{6}$$

where $c = 1,2,\ldots,C$ is indexed for a cohort. The y_{ct}^* and x_{ct}^* are the means of the population in the cohorts for the variable of interest y and the explanatory variable x, respectively. However, in practice, the true values of y_{ct}^* and x_{ct}^* are unknown. Thus, an approximation is made using individual variables that are replaced by their intra-cohort means. In practice, the estimated model is

$$\bar{y}_{ct} = \bar{x}_{ct} \beta + \bar{\alpha}_c + \bar{\delta}_{ct}, \tag{7}$$

where $\bar{y}_{ct} = \frac{1}{n} \sum_{i \epsilon c, t} y_{it}$ and $\bar{x}_{ct} = \frac{1}{n} \sum_{i \epsilon c, t} x_{it}$ are the means of observed values of the individuals of the sample in the cohorts. The $\bar{\alpha}_c$ is the mean of a fixed effect for cohort c which appears in a survey and is assumed to be fixed over time. The $\bar{\alpha}_c$ has a linear relation with $\bar{x}_{ct}$.

The estimation method in equation (6) does not take the fact that the true intra-cohort means (y_{ct}^* and x_{ct}^*) should be measured with error using the means calculated for the sample ($\bar{y}_{ct}$ and $\bar{x}_{ct}$). This measurement error poses two problems: an error in the explanatory variables results in biased estimators and an error in the variable of interest and the variability of the cohort effect over time reduces the precision of the estimators. To solve this problem, Deaton adapted a theory by Fuller [4] in the pseudo panel estimation. Suppose that u_{ct} and v_{ct} are the measurement error in $\bar{y}_{ct}$ and $\bar{x}_{ct}$, are

$$\bar{y}_{ct} = y_{ct}^* + u_{ct} \text{ and } \bar{x}_{ct} = x_{ct}^* + v_{ct} \tag{8}$$

Then, by combining the model (6) and (8) we obtained

$$\bar{y}_{ct} = \bar{x}_{ct} \beta + \alpha_c + \tilde{\delta}_{ct} \tag{9}$$

as the pseudo panel model, where $\tilde{\delta}_{ct} = \delta_{ct}^* + u_{ct} - v_{ct} \beta$ is the residual value that correlated to $\bar{x}_{ct}$.

III. The Proposed Model

The proposed model is taken by combining the pseudo panel into Model (1). In this proposed model, we treat the unit as a cohort, so that the individual value of the units is not the same over time. Therefore, these units are referred to as pseudo units. The following is the pseudo unit model:

$$y_{itj}^* = y_{itj} + \delta_{ij} \tag{10}$$

where $i = 1,2,\ldots,m$ is the index for a small area, $t = 1,2,\ldots,T$ is the index for time, and $j = 1,2,\ldots,\eta_{it}$ with η_{it} is the number of sample units in *i*-th small area at *t*-th time. The y_{itj}^* is the pseudo unit sample for *i*-th small area at *t*-th time. The y_{itj} is population *j*-th units in *i*-th area at *t*-th time and δ_{ij} is measurement error for *j*-th pseudo unit in *i*-th area, where $\delta_{ij} \sim N\left(0, \sigma_\delta^2\right)$. So, based on Model-1 on (4) and pseudo unit on (10), therefore:

$$\begin{aligned} y_{itj}^* &= \boldsymbol{x}_{itj}^T \boldsymbol{\beta} + v_i + u_{it} + \delta_{ij} + e_{itj}, \\ u_{it} &= \rho u_{i,t-1} + \varepsilon_{it}, \ |\rho| < 1, \end{aligned} \tag{11}$$

Then we call (11) as Model-2. The $\boldsymbol{\beta} = \left(\beta_1, \ldots, \beta_q\right)^T$ is a column vector of the p x 1 regression coefficient. Assume $v_i \sim iid\ N\left(0, \sigma_v^2\right)$ is a random effect for *i*-th small area at time point *t*. The u_{it}'s are assumed to follow a first order autoregressive process within each area *i*, with $\varepsilon_{it} \sim iid\ N\left(0, \sigma_\varepsilon^2\right)$. The sampling error e_{itj} is assumed $e_{itj} \sim iid\ N\left(0, \sigma_e^2\right)$. The v_i, u_{it}, e_{itj}, and δ_{ij} are assumed mutually independent.

Based on Muchlisoh [2], for every small area *i*-th, model (11) can be denoted into the following matrix form

$$\begin{aligned} \boldsymbol{y}_i^* = \boldsymbol{X}_i \boldsymbol{\beta} + v_i \boldsymbol{1}_{\eta_i} + \left(blockdiag_{1 \le t \le T} \boldsymbol{1}_{\eta_{it}}\right) \boldsymbol{u}_i + \\ \left(\boldsymbol{1}_T \otimes \boldsymbol{I}_{\eta_{it}}\right) \boldsymbol{\delta}_i + \boldsymbol{e}_i \end{aligned} \tag{12}$$

where $\boldsymbol{y}_i^* = \left(\boldsymbol{y}_{i1}^*, \boldsymbol{y}_{i2}^*, \ldots, \boldsymbol{y}_{iT}^*\right)'$, $\boldsymbol{y}_{it}^* = \left(y_{it1}^*, y_{it2}^*, \ldots, y_{it\eta_{it}}^*\right)'$, $\boldsymbol{X}_i = \left(\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{iT}\right)'$ with $\boldsymbol{X}_{it} = \left(\boldsymbol{X}_{it1}, \boldsymbol{X}_{it2}, \ldots, \boldsymbol{X}_{it\eta_{it}}\right)'$ and $\boldsymbol{X}_{itj} = \left(1, x_{1_{it1}}, \ldots, x_{q_{it\eta_{it}}}\right)'$, $\boldsymbol{u}_i = \left(\boldsymbol{u}_{i1}, \boldsymbol{u}_{i2}, \ldots, \boldsymbol{u}_{iT}\right)'$, $\boldsymbol{e}_i = \left(\boldsymbol{e}_{i1}, \boldsymbol{e}_{i2}, \ldots, \boldsymbol{e}_{iT}\right)'$ with $\boldsymbol{e}_{itj} = \left(\boldsymbol{e}_{it1}, \boldsymbol{e}_{it2}, \ldots, \boldsymbol{e}_{it\eta_{it}}\right)'$, $\boldsymbol{\delta}_i = \left(\boldsymbol{\delta}_{i1}, \boldsymbol{\delta}_{i2}, \ldots, \boldsymbol{\delta}_{i\eta_{it}}\right)'$ and $\eta_i = \sum_{t=1}^{T} \eta_{it}$. The $\boldsymbol{1}_{\eta_i}$ and $\boldsymbol{1}_{\eta_{it}}$ are vectors size η_i and η_{it} with all the elements 1, respectively. $\boldsymbol{I}_{\eta_{it}}$ is identity matrix size η_{it} x η_{it}.

For balanced data, $\eta_{i1} = \eta_{i2} = \ldots = \eta_{iT}$, model (12) can be written as,

$$\boldsymbol{y}_i^* = \boldsymbol{X}_i \boldsymbol{\beta} + v_i \boldsymbol{1}_{\eta_i} + \left(\boldsymbol{I}_T \otimes \boldsymbol{1}_{\eta_{it}}\right) \boldsymbol{u}_i +$$

$$\left(1_T \otimes I_{\eta_{it}}\right)\delta_i + e_i \tag{13}$$

where I_T is an identity matrix of order T and $\otimes$ is the Kronecker product. Defined $Z_{1i} = 1_{\eta_i}$, $Z_{2i} = \left(I_T \otimes 1_{\eta_{it}}\right)$, and $Z_{3i} = \left(1_T \otimes I_{\eta_{it}}\right)$, so the model may be written as

$$y_i^* = X_i\beta + Z_{1i}v_i + Z_{2i}u_i + Z_{3i}\delta_i + e_i \tag{14}$$

and the model for all of the areas be

$$y^* = X\beta + Z_1 v + Z_2 u + Z_3\delta + e \tag{15}$$

where $Z_1 = I_m \otimes Z_{1i}$, $Z_2 = I_m \otimes Z_{2i}$, and $Z_3 = I_m \otimes Z_{3i}$, $y^* = \left(y_1^*, y_2^*, \dots, y_m^*\right)'$, $X = \left(X_1, X_2, \dots, X_m\right)'$, $v = \left(v_1, v_2, \dots, v_m\right)'$, $u = \left(u_1, u_2, \dots, u_m\right)'$, $e = \left(e_1, e_2, \dots, e_m\right)'$, $\delta = \left(\delta_1, \delta_2, \dots, \delta_m\right)'$ and I_m is identity matrix size m x m. Assumed that u_{it} is a stationer, so the expected value and covariance matrix u_i, are

$$E\left(u_i\right) = 0 \text{ and } Cov\left(u_i\right) = G_{2i} = \sigma_\varepsilon^2\Gamma$$

where Γ, is a symmetric matrix defined by $\frac{\rho^{|t-t'|}}{1-\rho^2}$, where $t = 1,2,\dots,T$ and $t' = 1,2,\dots,T$. The independence assumption of vector e_i leads to $E\left(e_i\right) = 0$ and $Cov\left(e_i\right) = R_{1i} = \sigma_e^2 I_{\eta_i}$, and the independence assumption of vector δ_i leads to $E\left(\delta_i\right) = 0$ and $Cov\left(\delta_i\right) = G_{3i} = \sigma_{\delta ij}^2$.

If $G_{1i} = Cov\left(v_i\right) = \sigma_v^2$ with the random effect v_i, u_{it}, e_{itj}, and δ_{ij} are assumed mutually independent. So the covariance matrix of y_i^* on the model is

$$\begin{aligned} Cov\left(y_i^*\right) = {} & Cov\left(Z_{1i}v_i\right) + Cov\left(Z_{2i}u_i\right) + \\ & Cov\left(Z_{3i}\delta_i\right) + Cov\left(e_i\right) \end{aligned}$$

So that,

$$\begin{aligned} V_i &= Z_{1i}G_{1i}Z'_{1i} + Z_{2i}G_{2i}Z'_{2i} + Z_{3i}G_{3i}Z'_{3i} + R_{1i} \\ &= Z_{1i}\sigma_v^2 Z'_{1i} + Z_{2i}\sigma_\varepsilon^2\Gamma Z'_{2i} + Z_{3i}\sigma_{\delta ij}^2 Z'_{3i} + \sigma_e^2 I_{\eta_i} \\ &= \sigma_v^2 1_{\eta_i}1'_{\eta_i} + \sigma_\varepsilon^2\left(I_T \otimes 1_{\eta_{it}}\right)\Gamma\left(I_T \otimes 1_{\eta_{it}}\right)' + \\ &\quad \sigma_{\delta ij}^2\left(1_T \otimes I_{\eta_{it}}\right)\left(1_T \otimes I_{\eta_{it}}\right)' + \sigma_e^2 I_{\eta_i} \\ &= \sigma_v^2 1_{\eta_i}1'_{\eta_i} + \sigma_\varepsilon^2\left(I_T\Gamma I'_T \otimes 1_{\eta_{it}}1'_{\eta_{it}}\right) + \\ &\quad \sigma_{\delta ij}^2\left(J_T \otimes I_{\eta_{it}}\right) + \sigma_e^2 I_{\eta_i} \\ &= \sigma_v^2 J_{\eta_i} + \sigma_\varepsilon^2\left(\Gamma \otimes J_{\eta_{it}}\right) + \sigma_{\delta ij}^2\left(J_T \otimes I_{\eta_{it}}\right) + \sigma_e^2 I_{\eta_i} \end{aligned} \tag{16}$$

The J_{η_i} and $J_{\eta_{it}}$ are matrices of order η_i x η_i and η_{it} x η_{it} with elements 1, respectively. For all areas, the structure of the covariance matrix is $V = blockdiag_{1 \le i \le m}\left(V_i\right)$, and if $\eta_1 = \eta_2 = \dots = \eta_m$, the structure of the covariance matrix can be written as,

$$\begin{aligned} V &= I_m \otimes V_i \\ &= \left[I_m \otimes \left(\sigma_v^2 J_{\eta_i} + \sigma_\varepsilon^2\left(\Gamma \otimes J_{\eta_{it}}\right) + \sigma_{\delta i}^2\left(J_T \otimes I_{\eta_{it}}\right) + \sigma_e^2 I_{\eta_i}\right)\right] \end{aligned} \tag{17}$$

Henderson [7] obtained the general form of BLUP (Best Linear Unbiased Prediction) of w for the general linear mixed model,

$$\tilde{w} = \left(\tilde{v}_i, \tilde{u}'_i\right)' = G_i Z'_i V_i^{-1}\left(\tilde{y}_i^* - X_i\tilde{\beta}\right)$$

where $G_i = \begin{pmatrix} G_{1i} & 0 \\ 0 & G_{2i} \end{pmatrix}$,

so

$$\begin{aligned} \left(\tilde{v}_i, \tilde{u}'_i\right)' &= \begin{pmatrix} G_{1i} & 0 \\ 0 & G_{2i} \end{pmatrix}\left(Z_{1i}, Z_{2i}\right)' V_i^{-1}\left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \\ &= \left(G_{1i}Z'_{1i}, G_{2i}Z'_{2i}\right) V_i^{-1}\left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \end{aligned}$$

or

$$\begin{aligned} \tilde{v}_i &= G_{1i}Z'_{1i}V_i^{-1}\left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \\ &= \sigma_v^2 1'_{\eta_i}\left[\sigma_v^2 J_{\eta_i} + \sigma_\varepsilon^2\left(\Gamma \otimes J_{\eta_{it}}\right) + \sigma_{\delta ij}^2\left(J_T \otimes I_{\eta_{it}}\right) + \sigma_e^2 I_{\eta_i}\right]^{-1} \\ &\quad \times \left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \end{aligned} \tag{18}$$

and

$$\begin{aligned} \tilde{u}_i &= G_{2i}Z'_{2i}V_i^{-1}\left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \\ &= \sigma_\varepsilon^2\Gamma\left(I_T \otimes 1_{\eta_{it}}\right)'\left[\sigma_v^2 J_{\eta_i} + \sigma_\varepsilon^2\left(\Gamma \otimes J_{\eta_{it}}\right) + \sigma_{\delta ij}^2\left(J_T \otimes I_{\eta_{it}}\right) + \sigma_e^2 I_{\eta_i}\right]^{-1} \times \\ &\quad \left(\tilde{y}_i^* - X_i\tilde{\beta}\right) \end{aligned} \tag{19}$$

The EBLUP (empirical best linear unbiased prediction) of $\tilde{v}_i$ and $\tilde{u}_i$ are obtained by replacing the component of variances $\rho, \sigma_v^2, \sigma_\varepsilon^2, \sigma_e^2$, and σ_δ^2, with their estimators, say $\hat{\rho}, \hat{\sigma}_v^2, \hat{\sigma}_\varepsilon^2, \hat{\sigma}_e^2$, and $\hat{\sigma}_\delta^2$, respectively. So, the EBLUP of v_i and u_i are

$$\begin{aligned} \hat{v}_i = {} & \hat{\sigma}_v^2 1'_{\eta_i}\left[\hat{\sigma}_v^2 J_{\eta_i} + \hat{\sigma}_\varepsilon^2\left(\hat{\Gamma} \otimes J_{\eta_{it}}\right) + \hat{\sigma}_e^2 I_{\eta_i} + \hat{\sigma}_\delta^2 I_{\eta_i}\right]^{-1} \times \\ & \left(\tilde{y}_i^* - X_i\hat{\beta}\right) \end{aligned} \tag{20}$$

and

$$\begin{aligned} \hat{u}_i = {} & \hat{\sigma}_\varepsilon^2\hat{\Gamma}\left(I_T \otimes 1_{\eta_{it}}\right)'\left[\hat{\sigma}_v^2 J_{\eta_i} + \hat{\sigma}_\varepsilon^2\left(\hat{\Gamma} \otimes J_{\eta_{it}}\right) + \hat{\sigma}_e^2 I_{\eta_i} + \hat{\sigma}_\delta^2 I_{\eta_i}\right]^{-1} \times \\ & \left(\tilde{y}_i^* - X_i\hat{\beta}\right) \end{aligned} \tag{21}$$

where $\hat{\beta} = \left[X'\hat{V}^{-1}X\right]^{-1}\left[X'\hat{V}^{-1}\tilde{y}^*\right]$. The component of variances $\hat{\sigma}_v^2$, $\hat{\sigma}_\varepsilon^2$, $\hat{\sigma}_e^2$, and $\hat{\sigma}_\delta^2$ are computed by using Restricted Maximum Likelihood (REML).

Based on Muchlisoh [2], for a finite population model, the *i*-th small area at *t*-th time, θ_{it} is

$$\begin{aligned} \theta_{it} &= \frac{1}{N_{it}}\left[\sum_{j=1}^{n_{it}} y_{itj} + \sum_{j=n_{it}+1}^{N_{it}} y_{itj}\right] \\ &= \left(\frac{n_{it}}{N_{it}}\right)\bar{y}_{it}^s + \left(1 - \frac{n_{it}}{N_{it}}\right)\bar{y}_{it}^r \\ &= f_{it}\bar{y}_{it}^s + \left(1 - f_{it}\right)\bar{y}_{it}^r \end{aligned}$$

where $\tilde{\bar{y}}_{it}^{r}=\frac{1}{N_{it}-n_{it}}\left[\left(N_{it}\left(\bar{X}_{it}^{p}\right)'-n_{it}\left(\bar{X}_{it}^{s}\right)'\right)\tilde{\beta}\right]+\left(\tilde{v}_{i}+\tilde{u}_{it}\right)$, N_{it} is the number of population units of i-th small area at t-th time, $f_{it}=\frac{n_{it}}{N_{it}}$ is a sampling fraction. The upper script "*s*" denoted the sampled units (observed units) and "*r*" denoted the non-sampled units (unobserved units). So the BLUP for θ_{it} is

$$\begin{aligned}\tilde{\theta}_{it}&=f_{it}\bar{y}_{it}^{s}+\left(1-f_{it}\right)\times\\&\quad\left[\frac{1}{N_{it}-n_{it}}\left(N_{it}\left(\bar{X}_{it}^{p}\right)'-n_{it}\left(\bar{X}_{it}^{s}\right)'\right)\tilde{\beta}+\left(\tilde{v}_{i}+\tilde{u}_{it}\right)\right]\\&=f_{it}\bar{y}_{it}^{s}+\left(\left(\bar{X}_{it}^{p}\right)'-f_{it}\left(\bar{X}_{it}^{s}\right)'\right)\tilde{\beta}+\left(1-f_{it}\right)\left(\tilde{v}_{i}+\tilde{u}_{it}\right)\end{aligned}$$

and when the f_{it} close to 0, the BLUP for θ_{it} becomes

$$\tilde{\theta}_{it}=\left(\bar{X}_{it}^{p}\right)'\tilde{\beta}+\left(\tilde{v}_{i}+\tilde{u}_{it}\right).$$

The EBLUP of θ_{it} when ρ is known, $\theta_{it(\rho)}$ is obtained by replacing $\tilde{\beta}$, $\tilde{v}_{i}$, and $\tilde{u}_{it}$ with $\hat{\beta}$, $\hat{v}_{i(\rho)}$, and $\hat{u}_{it(\rho)}$ respectively. As follow

$$\begin{aligned}\hat{\theta}_{it(\rho)}&=f_{it}\bar{y}_{it}^{s}+\left(\left(\bar{X}_{it}^{p}\right)'-f_{it}\left(\bar{X}_{it}^{s}\right)'\right)\hat{\beta}+\left(1-f_{it}\right)\times\\&\quad\left(\hat{v}_{i(\rho)}+\hat{u}_{it(\rho)}\right)\end{aligned}\qquad(22)$$

where

$$\hat{v}_{i(\rho)}=\hat{\sigma}_{v}^{2}1'_{\eta_i}\left[\hat{\sigma}_{v}^{2}J_{\eta_i}+\hat{\sigma}_{\varepsilon}^{2}\left(I_{T}\otimes J_{\eta_{it}}\right)+\hat{\sigma}_{e}^{2}I_{\eta_i}+\sigma_{\delta ij}^{2}\left(J_{T}\otimes I_{\eta_{it}}\right)\right]^{-1}\times\left(y_{i}^{*}-X_{i}\hat{\beta}\right)$$

and

$$\hat{u}_{it(\rho)}=\hat{\sigma}_{u}^{2}I_{T}\left(I_{T}\otimes 1_{\eta_{it}}\right)'\left[\hat{\sigma}_{v}^{2}J_{\eta_i}+\hat{\sigma}_{u}^{2}\left(I_{T}\otimes J_{\eta_{it}}\right)+\hat{\sigma}_{e}^{2}I_{\eta_i}+\sigma_{\delta ij}^{2}\left(J_{T}\otimes I_{\eta_{it}}\right)\right]^{-1}\times\left(y_{i}^{*}-X_{i}\hat{\beta}\right).$$

IV. Results And Discussion

A. The Simulation Study

In this simulation study, the determination of parameter values referred to quarterly unemployment data at the district/city level in West Java based on the SAKERNAS rotating panel data within 2011-2014 assumed to follow Model-2 (11). The simulation started by generating finite population data which consisted of 26 areas (M = 26), the observation time was 13 (T = 13), and each area consisted of 4 pseudo units, with the size of each unit being 300 households. We used $\beta_0=8$, $\beta_1=15$, $\sigma_{\varepsilon}^{2}=0.5$, $\sigma_{e}^{2}=19.5$ and ρ = 0.9. This value is the best value based on the study by Muchlisoh [2], $\sigma_{\delta ij}^{2}$ computed from real data. Then we determined three categories of variances of random effects of the area, i.e., $\sigma_{v}^{2}=2.5$, $\sigma_{v}^{2}=5$, and $\sigma_{v}^{2}=10$. The determination of the three categories was used to evaluate the effect of σ_{v}^{2}.

The parameter of the i-*th* small area at t-*th* time is defined by $\theta_{it}=\frac{1}{n_{it}}\frac{1}{N_{itj}}\sum_{j=1}^{n_{it}}\sum_{k=1}^{N_{itj}}y_{itjk}$ and the direct estimation is defined by $\bar{y}_{it}=\frac{1}{n_{it}}\frac{1}{n_{itj}}\sum_{j=1}^{n_{it}}\sum_{k=1}^{n_{itj}}y_{itjk}$. In this simulation, we selected n_{itj} about 10 to 30 samples. This sampling was done with as many as 1000 replications (R=1000). To study the efficiency of estimation of the proposed model, we evaluated the values of bias and MSE of direct estimator and Model-2 that were computed as follows:

$$Bias=\frac{1}{MT}\sum_{i=1}^{M}\sum_{t=1}^{T}\left(\frac{1}{R}\sum_{j=1}^{R}\left(\hat{\theta}_{itl}-\theta_{it}\right)\right)$$

$$MSE=\frac{1}{MT}\sum_{i=1}^{M}\sum_{t=1}^{T}\left(\frac{1}{R}\sum_{j=1}^{R}\left(\hat{\theta}_{itl}-\theta_{it}\right)^{2}\right)$$

$$CV=\frac{1}{MT}\sum_{i=1}^{M}\sum_{t=1}^{T}\left(\frac{\sqrt{\frac{1}{R}\sum_{j=1}^{R}\left(\hat{\theta}_{itl}-\theta_{it}\right)^{2}}}{\theta_{it}}\right)x100\%$$

TABLE I
COMPARISON OF BIAS AND MEAN SQUARE ERROR (MSE) BASED ON DIRECT ESTIMATION, MODEL-1 AND MODEL-2 AT ρ = 0.9 WITH $\sigma_{v}^{2}=2.5$, $\sigma_{v}^{2}=5$, $\sigma_{v}^{2}=10$.

VRE	Model	Bias (%)	MSE	CV (%)	MSE (%)	RT (hour)
2.5	A	0.118	11.48	7.18	5545.8	6.20
	B	0.079	0.38	.41	102.9	6.20
	C	0.079	0.36	1.38	100	5.38
5	A	0.118	38.19	12.53	14822.5	6.21
	B	0.079	0.58	1.67	105.2	6.21
	C	0.079	0.53	1.611	100	5.38
10	A	0.118	147.08	23.50	26633.8	6.25
	B	0.078	1.42	2.58	110.1	6.25
	C	0.078	1.25	2.43	100	6.10

VRE; Variance of Random Effect; A: direct estimate:B: Model-1; C: Model-2

The results of the computation are presented in Table I. Table I shows that Model-2 gives smaller values of Bias, MSE, and CV than the direct estimation does at any condition of variances of random effects of the area. Besides, Table I presents that the MSE ratio between the direct estimation with Model-2 and the ratio between Model-1 and Model-2 increases as the variance of the random effect of area increases. This means that the greater the variance between areas, the Model-2 is better than Model-1 in estimating the parameters of interest. Table 1 also shows that the running time of Model-2 is faster than Model-1. Therefore, it can be concluded that Model-2 has a capability better than Model-1 to estimate the parameters of interest.

B. The SAKERNAS Data Application in West Java, Indonesia

In this section, we applied Model-1 and Model-2 to estimate the quarterly rate of unemployment at the district/city level using the same data. The data used in this study are quartered at the district/city level in West Java based on the SAKERNAS rotating panel data from 2011-2014. The auxiliary variables in this study were taken from the village potential census of West Java Province in 2011.

Fig. 1 shows that direct estimation generates more fluctuating and unstable estimated values compared with Model-1 and Model-2. Furthermore, the addition of measurement error in the *i*-th area and the *j*-th unit of Model-1 found in Model-2 make the estimated value in Model-2 more stable than in Model-1.

Based on Table II, the MSE of the direct estimation of 4.32 drops to 2.31 and 2.18 on the estimation method with Model-1 and Model-2, respectively. This suggests that the addition of random effects serves to calibrate the results of direct estimation based solely on survey data. The decrease in MSE is due to the decomposition of the variance components that are present in Model-1 and Model-2. In addition, although the estimated value of MSE produced by Model-1 and Model-2 is not much different, Model-1 of 2.31 drops to 2.18. This suggests that adding the measurement error in the *i*-th and *j*-th units into Model-1 can improve the efficiency of Model-1.

TABLE II
COMPARATIVE STATISTICS OF THE MEAN OF MSE OF QUARTERLY UNEMPLOYMENT RATE AT DISTRICT/CITY LEVEL IN WEST JAVA PROVINCE USING DIRECT ESTIMATION, MODEL-1 AND MODEL-2

Estimation Method	Mean of MSE
Direct estimation	4.32
Model-1	2.31
Model-2	2.18

To see more clearly the MSE comparative results based on direct estimation, Model-1, and Model-2 can be seen in Fig. 2.

Fig. 2 shows that direct estimation provides the greatest estimated MSE values compared to Model-1 and Model-2. Overall, there is a significant reduction in MSE values when using Model-1 and Model-2. The estimated MSE results in Model-2 tend to have a smaller value than in Model-1. This indicates that Model-2 is better than Model-1 in lowering MSE.

V. Conclusion

The simulation study shows that the addition of the measurement error in the *i*-th area and the *j*-th unit of Model-1 that found in Model-2 make the estimated value in Model-2 more stable than Model-1. Besides, the ratio of MSE and the running time indicate that the greater the variance between areas, the Model-2 is better than Model-1 in estimating the parameters of interest. The application study points out that the estimation of the quarterly rate of unemployment at the district/city level in West Java Province in Indonesia using Model-2 provides the same conclusions as the simulation study.

References

[1] Badan Pusat Statistik (Statistics Indonesia), *Pedoman pengawas survey angkatan kerja nasional (Sakernas) triwulanan.*, Jakarta, 2015.

[2] Muchlisoh A, Kurnia A, Notodiputro KA, and Mangku IW, Small Area Estimation of Unemployment Rate Based on Unit Level Model with First Order Autoregressive Time Effects, *Journal of Applied Probability and Statistics*, 12(2), 65-77, 2017.

[3] Rao JNK, *Small Area Estimation*, Second Edition, Wiley, New York, 2003.

[4] Guillerm M, Pseudo-panel methods and an example of application to household wealth data, *Economie et Statistique*, 109-130. doi: 10.24187/ecostat.2017.491d.1908.

[5] Deaton A, Panel data from time series of cross-section, *Journal of Econometrics*, 30(1-2):109-126. doi: 10.1016/0304-4076(85)90134-4.

[6] Handayani D, Notodiputro KA, Sadik K, and Kurnia A, A Comparative Study of Approximation Methods for Maximum Likelihood Estimation in Generalized Linear Mixed Models (GLMM), *AIP Conf. Proc.* 1827, 020033. 2017. doi:10.1063/1.4979449.

[7] Handayani D, Notodiputro KA, Saefuddin A, Mangku IW and Kurnia A, Empirical Best Predictor for Nested Error Regression Small Area Models. *IOP Conf. Ser.: Earth Environ. Sci. 187012036.* 2018. doi: 10.1088/1755-1315/187/1/012036.

[8] Fay RE and Herriot RA, Estimates of income for small places an application of James-Stein procedures to census data, *Journal of the American Statistical Association*, vol.74, pp.269- 277, 1979.

[9] Henderson CR, Best linear unbiased estimation and prediction under selection model, *Biometrics*, 31, 423-447, 1975.

[10] Pfeffermann D, Terryn B and Fernando M, Small area estimation under a two part random effects model to estimation if literacy in developing countries, *Survey Methodology*, 34, 67-72, 2008.

[11] Friedman EM, Jang D, and William TV, Combined estimates from four quarterly survey data sets, *Proceeding of the Section on Survey Research Methods*, *American Statistical Association,* 1064-1069, 2002.

[12] Ghosh M and Rao JNK, Small area estimation: an appraisal, *Statistical Sciences*, 9, 55-93, 1994.

[13] Hidiroglou M, Small-area estimation: theory and practice. *Section on Survey Research Methods,* 3445–3456, 2007.

[14] Kurnia A and Notodiputro KA, Generalized additive mixed model for small area estimation. *2nd. International Conference on Mathematical Sciences 2007*, Universiti Teknologi Malaysia, 2007.

[15] Levy PS and Lemeshow S, *Sampling of populations, methods and applications*, John Wiley and Sons, Inc, New York, 1999.

[16] Longford NT, On standard errors of model-based small-area estimators, *Survey Methodology*, 33, 69-79, 2007.

[17] Rao CR, *Linear Statistical Inference and its applications*, John Wiley and Sons, Inc, New York, 1973.

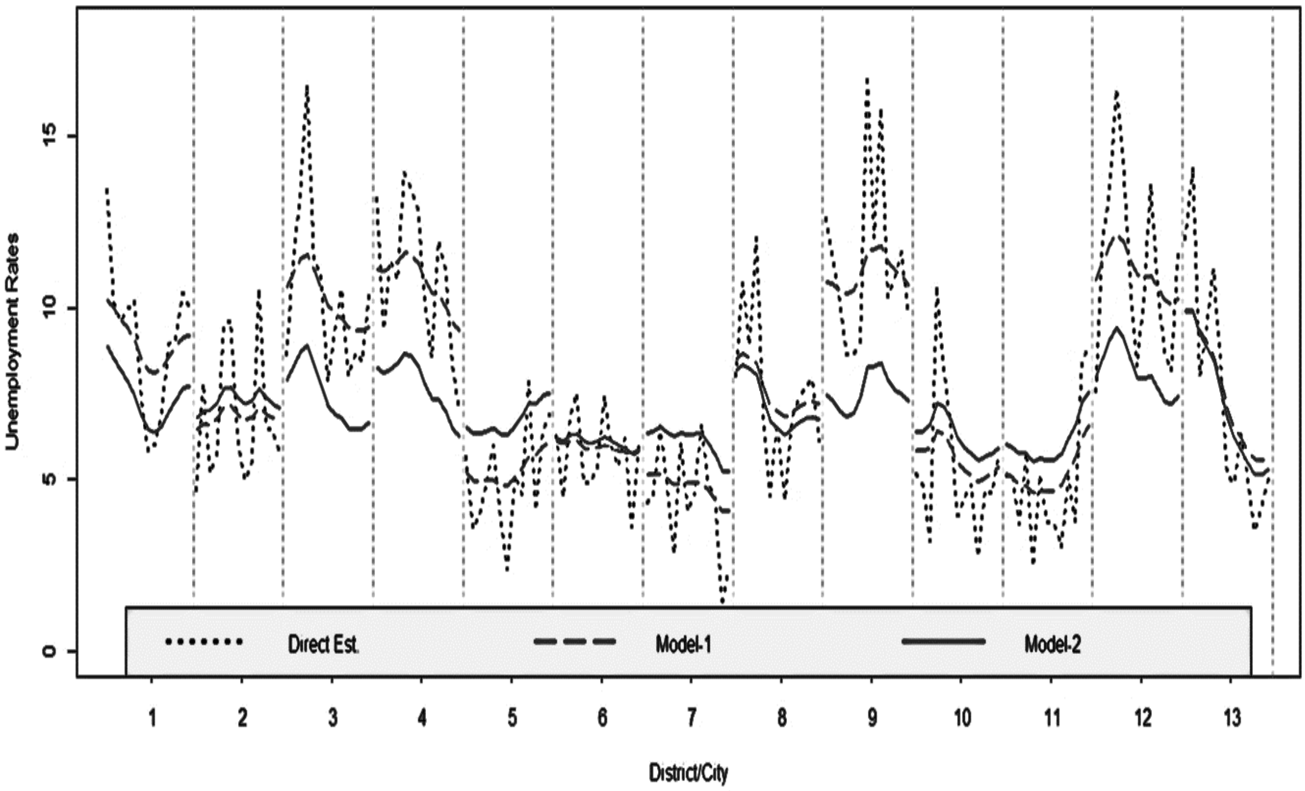

Fig. 1(A). The estimation of the quarterly unemployment rate in West Java Province using direct estimation, Model-1 and Model-2 for the district/city: 1-13

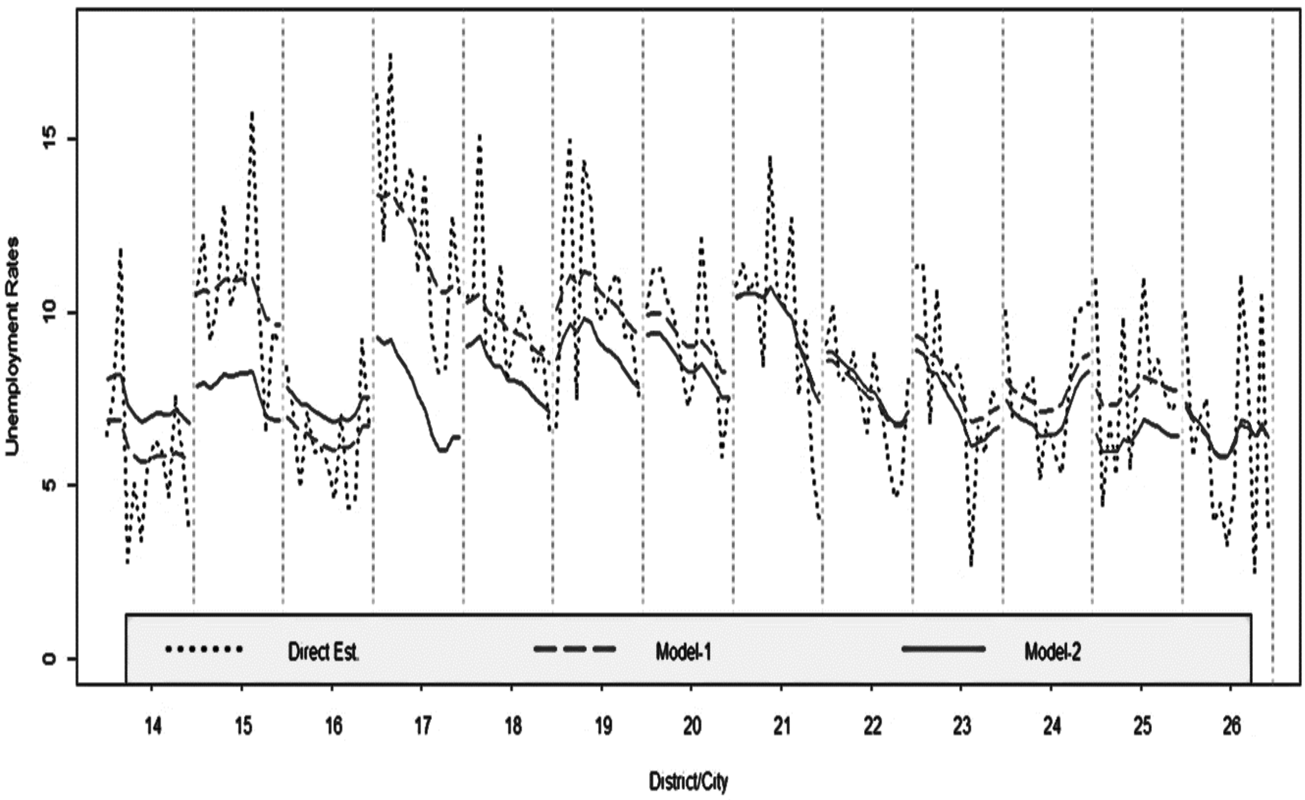

Fig. 1(B). The estimation of the quarterly unemployment rate in West Java Province using estimation, EBLUP of Model-1 and Model-2 for the district/city:

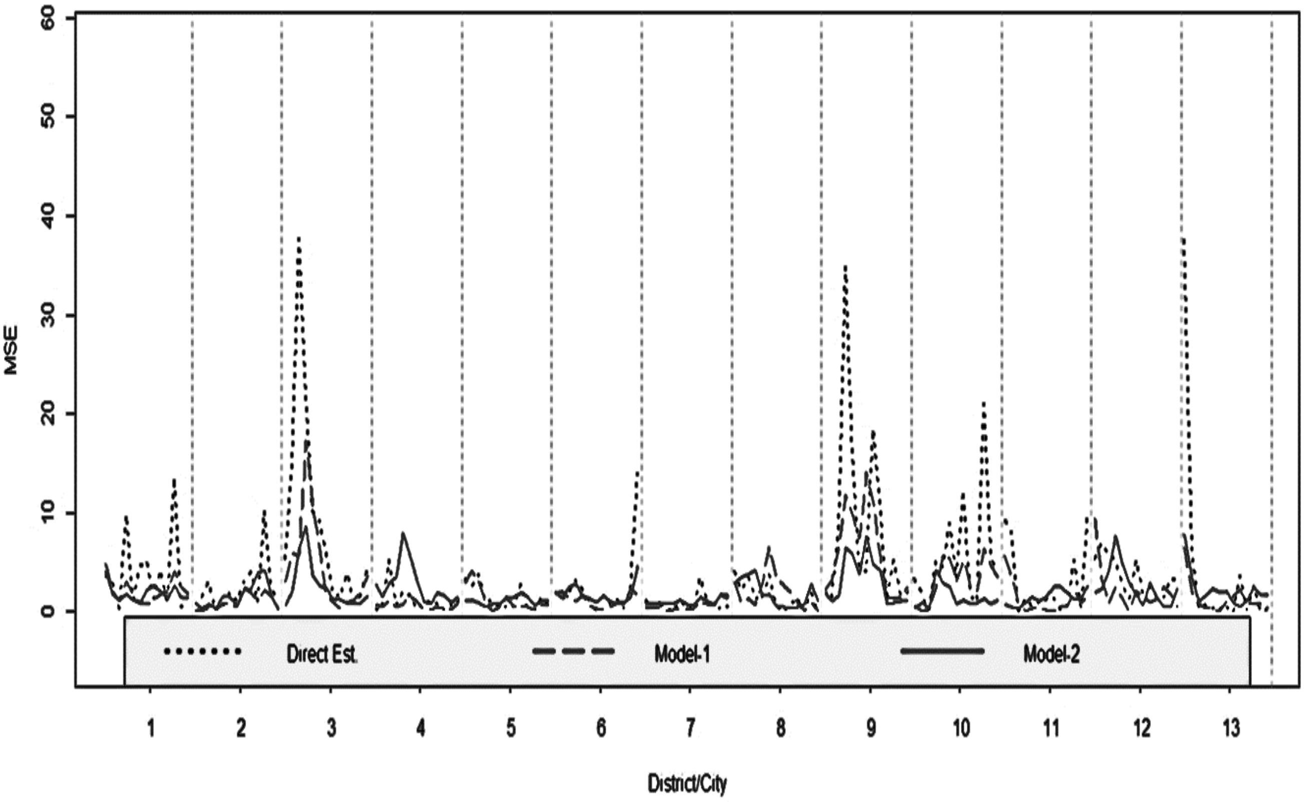

Fig. 2(A). The comparison of MSE estimation of the quarterly unemployment rate in West Java Province using direct estimation, Model-1, and Model-2 for district/city:1-13.

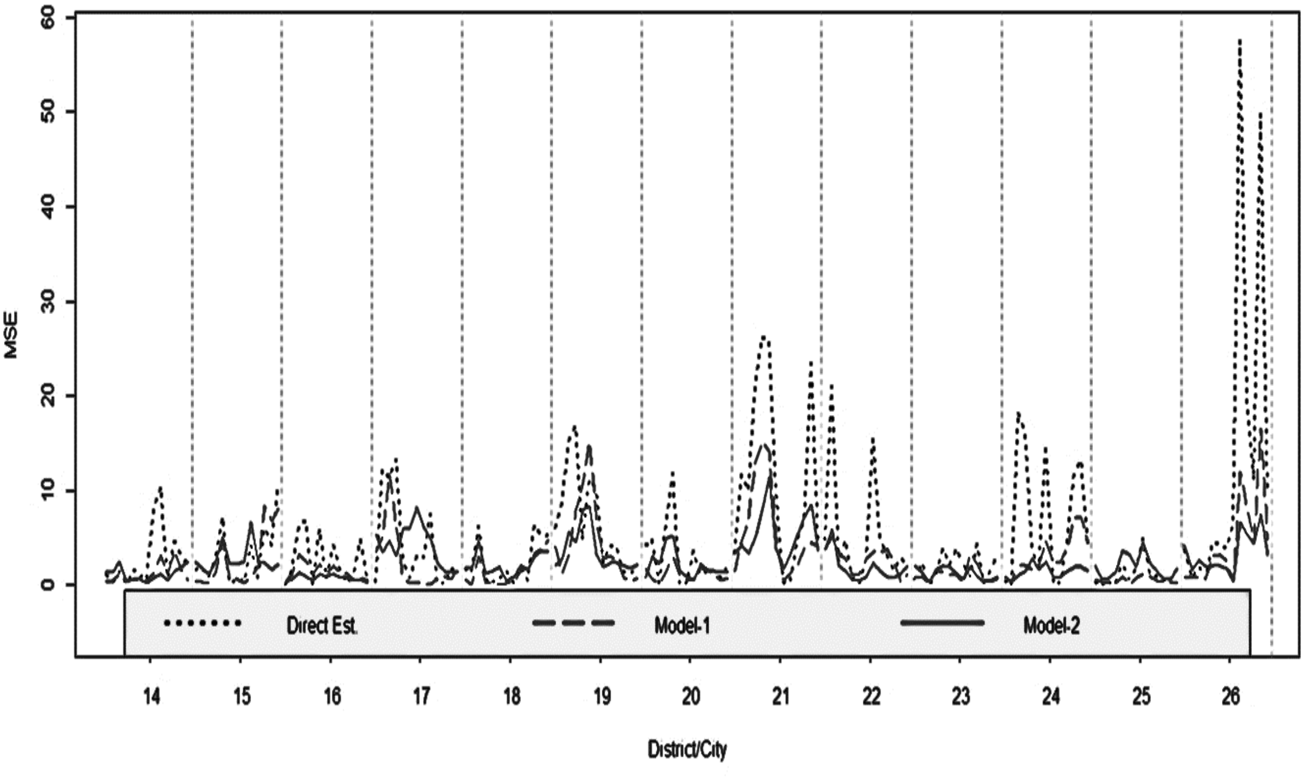

Fig. 2(B). The comparison of MSE estimation of the quarterly unemployment rate in West Java Province using direct estimation, Model-1, and Model-2 for district/city:14-26.