ICIC Express Letters
Volume **18**, Number **2**, February **2024**

ICIC International ©2024 ISSN 1881-803X
pp. 105–113

SMALL AREA ESTIMATION WITH NORML1 PENALTY: A SIMULATION STUDY

Anang Kurnia[1], Novi Hidayat Pusponegoro[2,*]
Khairil Anwar Notodiputro[1], Agus Mohamad Soleh[1]
and Erni Tri Astuti[2]

[1]Department of Statistics
IPB University
Jalan Raya Dramaga, Kampus IPB Dramaga Bogor, West Java 16680, Indonesia
{ anangk; khairil; agusms }@apps.ipb.ac.id

[2]Department of Statistics
Politeknik Statistika STIS
Jalan Raya Otista No. 64C, Jatinegara East Jakarta, Jakarta 13330, Indonesia
erni@stis.ac.id
*Corresponding author: novie@stis.ac.id

Received March 2023; accepted June 2023

Abstract. *NormL1 regulation is the penalty used in the least absolute shrinkage and selection operator for linear models and linear mixed models. In today's big data era, a small area estimation (SAE) method requires the constraint of dimension of auxiliary information matrix. Prediction and area-specific effect variance are two major problems in the SAE. To ease the problems, an SAE model with the NormL1 (SAEL1) penalty is proposed. Simulations in the present paper illustrate general results about the performance of the proposed model and the accuracy of the proposed algorithms. Some results show that the SAEL1 is a viable alternative method for the shrinkage and selection of the fixed and area-specific effect coefficients. The SAEL1 provides better predictive value and ability of SAEL1 to shrink and select the fixed and area-specific effect coefficients that outperform.*

Keywords: Big data, Coordinate descent, False rate, Multi-start technique, Shrinkage

1. **Introduction.** Small area estimation (SAE) techniques utilize auxiliary variables to deliver a direct estimation. Two basic model types of the SAE are area-level and unit-level models based on the auxiliary information availability. In the current era of big data, database size and technology are experiencing rapid development. This offers high dimensional auxiliary variables in the SAE approach [1]. The problem in estimating parameters with a large number of variables is a strong correlation between two or more variables which can increase the mean squared error (MSE) [2]. Parsimony is an estimation problem since a model with fewer explanatory variables is easier to interpret [3]. The best model in high dimensional linear mixed effects models (LMMs) can be obtained not only by shrinking the estimation parameter, but also going to zero and selecting the fixed effect parameter at once. NormL1 regulation is the first penalty used in the least absolute shrinkage and selection operator (LASSO) introduced for linear models and some contributions have been extended to linear mixed models [4-8]. The SAE methods in use today require the constraint of dimension of auxiliary information matrix, i.e., $p < n$. The relaxation of dimensions of a dataset means a big challenge in the SAE as it simultaneously needs to evaluate the prediction accuracy and computational efficiency to settle issues on convergence of the estimator.

DOI: 10.24507/icicel.18.02.105

Small area refers to a group or geographical area with insufficient sample size to deliver a direct estimation of the parameters. The area-level model of SAE is more widely applied as it does not require unit- or individual-level information that may be confidential [9]. The area-level model of SAE according to the LMMs was proposed by Fay-Herriot (FH). With regard to the LMMs, variance in the FH model sub-populations can be explained through the fixed and area-specific effects which correspond to variance in the auxiliary and area-specific of sub-populations, respectively. Area-specific effects cannot be explained by the auxiliary variables and are assumed to be independent and identically normally distributed. The area-specific effects might increase the estimation variability; thus, a test for the presence of the area-specific effects was suggested [9,10]. The mixed distribution is also proposed to verify the area-specific effect inclusion in the model [11]. The speed of convergence in the SAE parameter estimate is also associated with the complexity of area effect vector [12]. In a situation of the large-scale data with observation that could be small area, sometimes, the number of small areas is relatively small compared to the total survey areas. The sparsity on area-specific effects can be imposed through setting zero for i-th large area while maintaining the nonzero value for i-th small area. It makes the normality assumption of area-specific effects can be violated. Therefore, reliable estimates can be obtained through an efficient selection of the true small areas (true nonzero random effects).

Area-specific effects play an important role in attaining reliable parameter estimates in the SAE; hence, their exclusion and inclusion in the model should be carefully examined to reach an optimum rate of convergence and accurate prediction. Shrinkage technique is indispensable in the SAE to obtain the SAE with reasonable inferential statistics [12]. An adaptive model for the SAE with automatic random effect selection (SARS) using a hard-ridge penalty for the SAE has been carried out [13]. However, the large-scale data challenges an estimation method that not only does select the area-specific effects, but also shrinks the value of area-specific coefficients at once to obtain the parsimony. To meet today's challenges, NormL1 is proposed as penalty [14]. The SAEL1 for area-specific effect selection empirically delivers the minimum MSE under each condition of auxiliary variable correlation, area effect variance components, and percentages of small area.

Prediction and area-specific effect variance are two major problems in the SAE. For such problems and big data challenges, a NormL1 penalty SAE model is proposed, which not only does shrink the parameter estimate, but also selects the fixed and the area-specific effects. The remaining parts of the present paper consist of the detailed methodology in Section 2, results and discussions of simulation in Section 3, and some concluding remarks and possible future study in Section 4.

2. **Methods.** To resolve the problems previously mentioned, we present the details of NormL1 penalty SAE methodology as follows.

2.1. **Small area estimation with random effects selection.** Fay-Herriot (FH) model based on linear mixed model with m number of small areas can be presented as

$$\boldsymbol{y} = \boldsymbol{X\alpha} + \boldsymbol{u} + \boldsymbol{e} \tag{1}$$

where $\boldsymbol{y}$ is $m \times 1$ vector of the parameters of inferential interest and the direct estimator $\hat{\boldsymbol{y}}$ assume available. The auxiliary information $\boldsymbol{X} = (X_1, X_2, \dots, X_m)^T$ is a known $m \times p$ matrix, $\boldsymbol{e}$ is a vector of independent sampling errors with mean vector 0 and variance matrix $\boldsymbol{R} = diag\big(\sigma_{ei}^2\big)$, σ_{ei}^2 representing the sampling variance of the direct estimators of the certain area, $i = 1, 2, \dots, m$. Therefore, $\boldsymbol{\alpha}$ is the $p \times 1$ vector of regression parameters, $\boldsymbol{u}$ is the $m \times 1$ vector of independent area-specific effects with zero mean and variance matrix $\sum_u = \sigma_u^2 I_m$. Empirical best linear unbiased prediction (EBLUP) is used in response variables predicting since σ_u^2 estimated with maximum likelihood/ML or restricted maximum likelihood/REML method. Thus, the parameter regression, area-specific estimate and the response prediction formula can be expressed as

$$\hat{\boldsymbol{\alpha}} = \left(\boldsymbol{X}_i^T \hat{\boldsymbol{V}}^{-1} \boldsymbol{X}_i\right)^{-2} \boldsymbol{X}_i^T \hat{\boldsymbol{V}}^{-1} \boldsymbol{y} \tag{2}$$

$$\hat{\boldsymbol{u}} = \hat{\boldsymbol{\sigma}}_u^2 \hat{\boldsymbol{V}}^{-1} \left(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\alpha}}\right) \text{ and } \hat{\gamma}_{iu} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \sigma_{ei}^2} \tag{3}$$

The EBLUP estimate values depend on the weighted factor γ_{iu} which corresponds to variance of the area-specific effects σ_u^2 and sampling error variance $\boldsymbol{R}$ [15]. The observed best predictor cannot be obtained if the shrinkage factor is not correctly identified. Regarding the sample size in survey areas, the area-specific effects term can be insignificant in the SAE model. However, eradicating the area-specific effects from the model may not be the best choice in terms of prediction accuracy; hence, it needs to obtain the small area estimates under such a high-dimensional dataset.

Set U of small areas with $u_i \sim N\left(0, \sigma_u^2\right)$ for $i \in U$ and $u_i = 0$ for $i \in U^C$. Given that $\boldsymbol{\sigma}_u^2$ and σ_{ei}^2 are unknown, then SARS model employs a penalized regression in optimizing the function of the least squares difference for estimating the FH parameter [13]. The SARS model assumes the sampling error variance $\boldsymbol{R} = diag\left\{\sigma_{ei}^2/n_i\right\}$ where n_i is the sample's size. The SARS model uses multiple penalties as combination of L0 and L2 referred to as hard-ridge penalty. In the SAE approach, assuming $p > n$ and the area-specific effects $\boldsymbol{u}$ are sparse, the hard-ridge penalty for selecting the fixed and area-specific effects is stated as

$$P_{02}\left(\boldsymbol{\beta}, \boldsymbol{u}; \lambda_{0\beta}, \lambda_{0u}, \eta_\beta, \eta_u\right) = \frac{\eta_\beta}{2}\|\boldsymbol{\beta}\|^2 + \frac{{\lambda_{0\beta}}^2}{2\left(1+\eta_\beta\right)}\|\boldsymbol{\beta}\|_0 + \frac{\eta_u}{2}\|\boldsymbol{u}\|^2 + \frac{{\lambda_{0u}}^2}{2\left(1+\eta_u\right)}\|\boldsymbol{u}\|_0 \tag{4}$$

in which $\lambda_{0\beta}$, λ_{0u} is tuning parameter for hard penalty to optimize the SARS prediction information criteria $\left(\lambda_{0u}, \lambda_{0\beta} \geq 0\right)$ and η_β, η_u tuning parameter for ridge penalty to select the fixed and area-specific effects. If σ_{ei}^2 are unknown, then $\boldsymbol{R} = diag\{1/n_i\}$, so the objective function of the SARS model with the area-specific effects selection is

$$\begin{aligned} \arg\min f\left(\boldsymbol{\beta}, \boldsymbol{u}; \lambda_{0\beta}, \lambda_{0u}, \eta_\beta, \eta_u\right) \triangleq{}& \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{u}\right)^T diag\left\{n_i\right\}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{u}\right) \\ &+ P_{02}\left(\boldsymbol{\beta}, \boldsymbol{u}; \lambda_{0\beta}, \lambda_{0u}, \eta_\beta, \eta_u\right) \end{aligned} \tag{5}$$

The SARS solves the issue of multiple tuning parameters through an iterative selection-estimation (SE) algorithm [13]. The selection step employs quantile thresholds to screen small area iteratively and the estimation step estimate $\boldsymbol{u}$ based on the true non-zero small area effects obtained from the previous step. The SARS optimization problem is challenging due to the non-convex and non-smooth feature of the hard-ridge penalty. Then, an iterative technique can be employed to solve the SARS problems.

2.2. **Small area estimation through area effects selection using LASSO method.** L1 or NormL1 is the penalty used in the least absolute shrinkage and selection operator (LASSO) for linear models. LASSO is a convex penalty that can be an alternative to shrink area-specific effects of the SAE; thus, the SAE model with the random area effect selection with LASSO penalty was proposed [14]. Based on the area-level model in Equation (1) and objective function of SARS model in (5), the proposed model's objective function can be formulated as

$$\arg\min f\left(\boldsymbol{u}; \lambda_u\right) \triangleq \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{u}\right)^T diag\left\{n_i\right\}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha} - \boldsymbol{u}\right) + \lambda_u\|\boldsymbol{u}\| \tag{6}$$

where λ_u is tuning parameter for LASSO penalty and $\lambda_u \geq 0$. Regarding computation, the SAE with LASSO penalty for area-specific effects selection utilizes coordinate gradient descent optimization approach so that area-specific effects estimator can be formulated as

$$\hat{\boldsymbol{u}} = \left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\right) - \left(\boldsymbol{W}^T\boldsymbol{W}\right)^{-1} sign(\boldsymbol{u})\lambda_u,\ \boldsymbol{W} = diag\left\{n_i\right\} \tag{7}$$

Based on Formula (7), we need the stopping criterion to obtain the optimum area-specific effects estimation. The SARS predictive information criteria (PIC) was proposed to overcome the less measurement criteria for other model prediction information related to the number of fixed and area-specific effects involved in the models [16,17]. Based on the penalty function and the predictive information criterion function above, the following is the SARS-PIC for area-specific effects of the area-level model with the LASSO to shrink parameter estimate and select the area-specific effects:

$$\boldsymbol{W}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{u}\|^2+P_1(\boldsymbol{u};\lambda_u);\ P_1(\boldsymbol{u};\lambda_u)=\sigma_e^2\left[J(\boldsymbol{u})+J(\boldsymbol{u})\log\left\{\frac{em}{J(\boldsymbol{u})}\right\}\right] \tag{8}$$

in which $J(\boldsymbol{u})$ is the number of area-specific effect rows which have coefficient values of parameters that are not equal to zero and it indicates the true small area or area with insufficient sample size.

2.3. **Penalized small area estimation using LASSO method.** Regardless of the number of the predictor p, model is already involved in a high-dimensional issue. An interesting question that has not been addressed in the SAE is how to obtain the small area estimates under such a high-dimensional dataset. Based on the FH model with a larger dimensionality of the predictors compared to the sample size and the area-specific effects $\boldsymbol{u}$ are sparse, the resulting objective function is then as follows:

$$\arg\min f(\boldsymbol{\beta},\boldsymbol{u};\lambda_\beta,\lambda_u)\triangleq(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{u})^T diag\{n_i\}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{u})+\lambda_\beta\|\boldsymbol{\beta}\|+\lambda_u\|\boldsymbol{u}\| \tag{9}$$

Using the coordinate gradient descent approach, the area-specific effect estimate is expressed in Formula (7) and the fixed effect estimate is expressed as

$$\hat{\boldsymbol{\beta}}=\left[(\boldsymbol{W}\boldsymbol{X})'(\boldsymbol{W}\boldsymbol{X})\right]^{-1}\left((\boldsymbol{W}\boldsymbol{X})'\boldsymbol{W}(\boldsymbol{y}-\boldsymbol{u})-sign(\boldsymbol{\beta})\lambda_\beta\right) \tag{10}$$

Formula (10) states that the penalty function in the proposed model is additive, so the predictive information criteria (PIC) for the LASSO-penalized SAE model are expressed as

$$\begin{aligned}&\boldsymbol{W}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}-\boldsymbol{u}\|^2+P_\beta(\boldsymbol{\beta};\lambda_\beta)+P_u(\boldsymbol{u};\lambda_u);\\&P_\beta(\boldsymbol{\beta};\lambda_\beta)=\sigma_e^2\left[J(\boldsymbol{\beta})+J(\boldsymbol{\beta})\log\left\{\frac{ep}{J(\boldsymbol{\beta})}\right\}\right];\\&P_u(\boldsymbol{u};\lambda_u)=\sigma_e^2\left[J(\boldsymbol{u})+J(\boldsymbol{u})\log\left\{\frac{en}{J(\boldsymbol{u})}\right\}\right]\end{aligned} \tag{11}$$

in which $J(\boldsymbol{\beta})$ and $J(\boldsymbol{u})$ are the number of fixed effect columns and the number of rows of area-specific effect that have with nonzero parameter coefficient, respectively.

3. **Simulation Study.** The study prepared to evaluate the performance of the proposed model and the characteristics of parameters estimate, we present the data simulation set and simulation result as follows.

3.1. **Data simulation setup.** The present paper performs simulations to illustrate general results about the performance of the proposed model estimator and the accuracy of the proposed algorithms. The simulation is structured with the modern data structures, which sets the number of observation or small area to 100 and the number of auxiliary variables to 200. The customized data are set with some different area-specific effect variance components, sparsity levels and correlation between auxiliary variables. Under the SAE model assumption, the auxiliary information X is independently drawn from a multivariate normal distribution with mean vector $\mathbf{0}$ and variance matrix $\boldsymbol{\rho}$ and the correlation structure of $\boldsymbol{\rho}=\left\{r^{|i-j|}\right\}$. The values of r are 0.2, 0.5 and 0.8, each of which represents the condition of the small correlated, moderately correlated and highly correlated data, respectively. The area-specific effects variance components $\left(\sigma_u^2\right)$ in the simulation are set to be 0.3 and 2 since those values represent small and big values of the variance with $\sigma_e^2 = 1$. Simulation is taken to emphasize that the model can be applied in large p set-up, even though vector $\boldsymbol{u}$ is sparse. The simulation data conduct the percentages of true nonzero fixed effects coefficients by 1% and the percentages of true small area number are 1%, 10%, 50% and 90% to all analysis areas. And, it sets the sample size of each area as $[1, 1000]$.

Under a sparsity assumption on $\boldsymbol{\beta}$, the initial points of $\boldsymbol{\beta}$ affect computational performance. The multi-start technique (MT) obtains initial points of $\boldsymbol{\beta}$ which has two phases: generation and improvement phase. The first phase is to generate random $\boldsymbol{\beta}$ initial vectors consisting of selecting 3% of columns $\boldsymbol{X}$ and predictors randomly than producing ordinary least square (OLS) $\boldsymbol{\beta}$ estimator. After H replication, there is random multiple initial $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{p\times H}$. The criterion to find the true nonzero elements for $\boldsymbol{\beta}$ in the improvement phase is revised. Instead of repeating the algorithm until there is no difference between old and new solutions, the probability of nonzero elements is used to update the random multiple initial $\hat{\boldsymbol{\beta}}$. $P\left(\hat{\beta}_j\right) = 1$ is obtained if every initial $\hat{\boldsymbol{\beta}}$ in j-th rows is nonzero, otherwise $P\left(\hat{\beta}_j\right) = 0$. Hence, the desired reduced information matrix $\boldsymbol{X}_{J\beta}$ is managed by selecting the column j of $\boldsymbol{X}$ if $P\left(\hat{\beta}_j\right) = 1$ so that the desired reduced information matrix $\boldsymbol{X}_{J\beta}$ and initial vector $\hat{\boldsymbol{\beta}}$ via OLS can be obtained.

Sparsity of area-specific effects is deeply associated with the sample size of each area. In fact, the rate of small areas is sometimes 1% or less. Furthermore, finding efficient initial points is crucial for enhancing the efficiency and effectiveness of computation; thus, an area-specific feature selection (AFS) method is revised to generate the efficient initial vector of $\boldsymbol{u}$. Given an error tolerance $\varepsilon > 0$, the maximum of sample weights M and sample weights of each areas n_i are taken. The initial points of u_i is obtained if $\tau_i \geq \varepsilon$ where $\tau_i = \frac{M}{n_i * T}$. Coordinate descent is considered to evaluate the performance of the proposed method and the algorithms accuracy, the missing rate (MR) which is the probability of undetected true nonzero elements and false alarm rates (FR) which is the probability of spuriously detected true zero elements are measured. MSE is employed to evaluate the performance of parameter estimate in the proposed model. The accurate prediction is achieved whenever the measurement above is smaller than the estimated model.

3.2. **Simulation results.** The modified simulation is run for 100 times, and then the SARS model and SAEL1 model (proposed) are compared. The MR and the FR of fixed effects and area-specific effects are computed. These results are reported in Figure 1 and Figure 2. The left-graph and the right-graph of Figure 1 are the MR plot of fixed effects and area-specific effects, respectively. Figure 1 shows that the MR values of SAEL1 model are close to zero. It indicates that the SAEL1 model performs almost none of nonzero fixed and area-specific effects are missing. Figure 1 also describes that the MR of the fixed effects in the SARS model are lower than 0.5. However, the MR of area-specific effects in the SARS model gets higher along with the increase in the proportion of true zero area-specific effects.

Figure 2 represents the FR of the fixed and area-specific effects. The plots show that SAEL1 model performs best in the shrinkage and selection of the fixed effect coefficients. However, the SAEL1 model does poorly shrink and select the area-specific effect coefficients since the values increase. The increasing pattern is in the direction of the correlation and the proportion of zero values with a slope of almost 45 degrees. While, the FR of the SARS model has higher values than the SAEL1's model.

In general, the comparison between the MR of the fixed and area-specific effects from the SARS model shows significant differences with the SAEL1 model. Otherwise, the FR of the fixed and area-specific effects from the SARS and the SAEL1 model has different

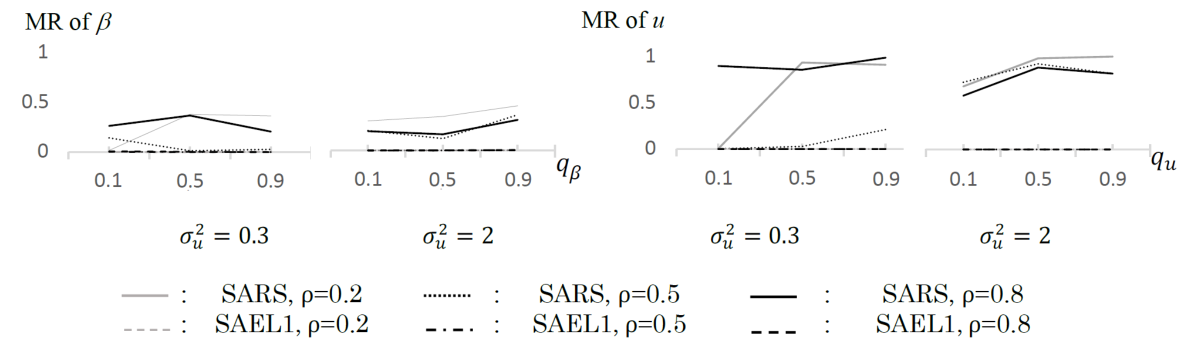

FIGURE 1. The MR of the SARS and SAEL1 model by correlation, true nonzero percentage of area-specific effects and its variance

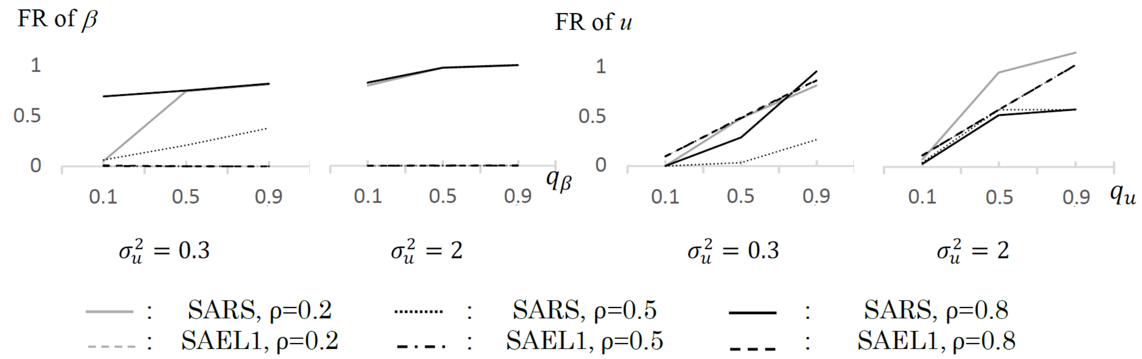

FIGURE 2. The FR of SARS and SAEL1 model by correlation, true nonzero percentage of area-specific effects and its variance

TABLE 1. Average MSE of the SARS and the SAEL1 models by auxiliary variables correlation, area-specific effect variance and true nonzero percentage

Models	ρ	$\sigma_u^2 = 0.3$			$\sigma_u^2 = 2$		
		q_u			q_u		
		0.1	0.5	0.9	0.1	0.5	0.9
SARS	0.2	0.1061	0.1191	0.1453	0.2064	0.3913	0.5032
SAEL1		0.0112	0.0124	0.0167	0.0128	0.0382	0.0935
SARS	0.5	0.0624	0.1072	0.1170	0.2135	0.1304	0.8963
SAEL1		0.0104	0.0323	0.0365	0.0140	0.0534	0.0972
SARS	0.8	0.1232	0.1304	0.1381	0.4868	2.0396	2.9328
SAEL1		0.0108	0.0455	0.0663	0.0144	0.0401	0.0762

conditions. The FR of the fixed effects from the SAEL1 model remains close to zero; while, for the FR of area-specific effects, the value is large and close to 1. In other words, the shrinkage and selection parameter using NormL1 outperform on the fixed effects, but perform poorly in the shrinkage and selection of the area-specific effect coefficients.

In this simulation, the performance of parameter estimate in the model evaluation is examined using average of the MSE and the results are presented in Table 1. Table 1 shows that the parameter estimate of SAEL1 outperforms the SARS model since it has smaller value of the average MSE in each condition of auxiliary variables correlation, area-specific effects variance and true nonzero percentage. As seen in Table 1, the average MSE of SARS model increases along with the increase of auxiliary variables correlation and true nonzero percentage either for homogeneous or heterogenous of area-specific effects coefficients. Particularly, the average MSE of SAEL1 model for small area-specific variances decreases along with an increase in the true nonzero percentage. Those values increase along with the increasing true nonzero percentage in the dataset which has higher area-specific variances. However, each true nonzero percentage the MSE values either of the SARS or SAEL1 model, is convergent.

The study aims to provide an introduction of the NormL1 method to solve the SAE. The application is illustrated using the simulated and estimate responses (Figure 3 and Figure 4).

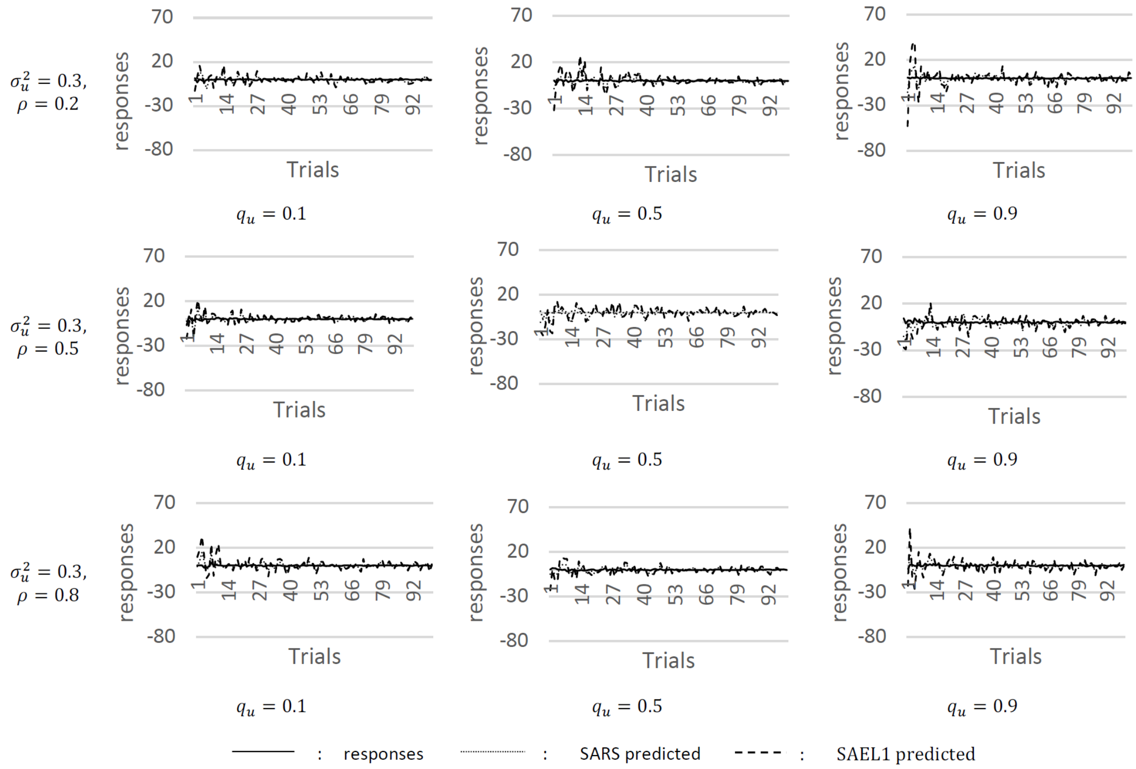

FIGURE 3. The predicted responses of SARS and SAEL1 model with $\sigma_u^2 = 0.3$ by auxiliary variable correlation and true nonzero percentage of area-specific effects

Figure 3 is the plot of the estimate responses of the SARS and SAEL1 models based on auxiliary variable correlation and true nonzero percentage with area-specific effect coefficients variance, $\sigma_u^2 = 0.3$. Figure 4 is the plot with area-specific effect coefficients variance, $\sigma_u^2 = 2$. Figure 3 confirms that the response predictions of the SARS model are more varied compared to the predicted values of the SAEL1 model. In general, the variation in the response prediction increases along with the true nonzero area-specific random effects' percentage. Thus, we conclude that the variation in the predicted value of the SAE model with penalty increases with the percentage of small areas of all the group or region in the study.

Figure 4 also describes more or less the same as Figure 3, but the variation is sharper at a high true nonzero percentage. This implies that the predicted responses of the SARS and SAEL1 models are insufficiently convergent in the dataset with higher area-specific effect variance.

4. **Conclusions.** This work proposes a method to shrink and select the fixed and area-specific effect for SAE with NormL1 (SAEL1). It also presents the evidences that the SAEL1 is a viable alternative method to shrink and select the fixed and area-specific effect

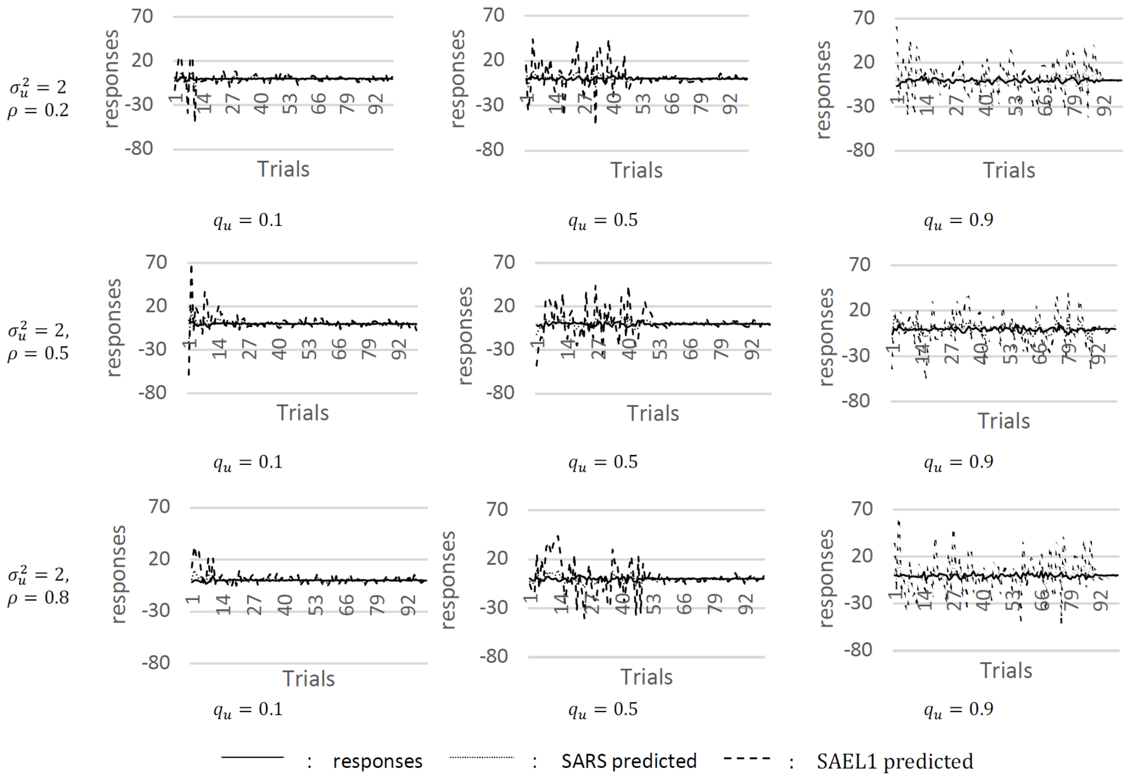

FIGURE 4. The predicted responses of SARS and SAEL1 model with $\sigma_u^2 = 2$ by auxiliary variable correlation and true nonzero percentage of area-specific effects

coefficients. The SAEL1 delivers better predictive value. The ability to shrink and select the fixed and area-specific effect coefficients of the SAEL1 model better than the SARS model, but the error in detecting the spurious zero of area-specific effects is high. Those encouraging results reported here suggest that a further study is needed to investigate some possibilities to utilize another penalty such as elastic net.

REFERENCES

[1] R. Kitchin, The opportunities, challenges and risks of big data for official statistics, *Statistical Journal of the IAOS*, vol.31, no.3, pp.471-481, 2015.

[2] R. Tibshirani, Regression shrinkage and selection via the LASSO, *Journal of the Royal Statistical Society: Series B (Methodological)*, vol.58, pp.267-288, 1996.

[3] T. Hastie, R. Tibshirani and M. Wainwright, *Statistical Learning with Sparsity: The LASSO and Generalizations*, Taylor & Francis Group, New York, 2016.

[4] H. D. Bondell, A. Krishna and S. K. Ghosh, Joint variable selection for fixed and random effects in linear mixed effects models, *Biometrics*, vol.66, pp.1069-1077, 2010.

[5] J. G. Ibrahim, H. Zhu, R. I. Garcia and R. Guo, Fixed and random effects selection in mixed effects models, *Biometrics*, vol.67, pp.495-503, 2011.

[6] J. Schelldorfer, P. Bühlmann and S. Van De Geer, Estimation for high-dimensional linear mixed-effects models using 1-penalization, *Scandinavian Journal of Statistics*, vol.38, pp.197-214, 2011.

[7] H. Peng and Y. Lu, Model selection in linear mixed effect models, *Journal of Multivariate Analysis*, vol.109, pp.109-129, 2012.

[8] N. Gauraha, Introduction to the LASSO: A convex optimization approach for high-dimensional problems, *Resonance*, vol.23, pp.439-464, 2018.

[9] J. N. K. Rao and I. Molina, *Small Area Estimation*, John Wiley & Sons. Inc., New Jersey, 2015.

[10] G. S. Datta, P. Hall and A. Mandal, Model selection by testing for the presence of small-area effects, and application to area-level data, *Journal of the American Statistical Association*, vol.106, pp.362-374, 2011.

[11] G. S. Datta and A. Mandal, Small area estimation with uncertain random effects, *Journal of the American Statistical Association*, vol.110, pp.1735-1744, 2015.

[12] G. Datta and M. Ghosh, Small area shrinkage estimation, *Statistical Science*, vol.27, no.1, pp.95-114, 2012.

[13] J. Lee, *Small Area Estimation with Random Effects Selection*, Ph.D. Thesis, Florida State University, Florida, 2017.

[14] N. H. Pusponegoro, A. Kurnia, K. A. Notodiputro, A. M. Soleh and E. T. Astuti, Small area estimation of sub-district's per capita expenditure through area effects selection using LASSO method, *Procedia Computer Science*, vol.179, pp.754-761, 2021.

[15] J. Jiang, T. Nguyen and J. S. Rao, Best predictive small area estimation, *Journal of the American Statistical Association*, vol.106, no.494, pp.732-745, 2011.

[16] T. Hastie, R. Tibshirani and J. Freidman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, California, 2009.

[17] Y. She, Selective factor extraction in high dimensions, *Biometrika*, vol.104, no.1, pp.97-110, 2017.