



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2021, 2021:90

<https://doi.org/10.28919/cmbn/6747>

ISSN: 2052-2541

AREA SPECIFIC EFFECTS SELECTION OF SMALL AREA ESTIMATION FOR CONSTRUCTION OF REGIONAL CONSUMER PRICE INDICES IN INDONESIA

NOVI HIDAYAT PUSPONEGORO^{1,2}, ANANG KURNIA^{1,*}, KHAIRIL ANWAR NOTODIPUTRO¹,

AGUS MOHAMAD SOLEH¹, ERNI TRI ASTUTI²

¹Department of Statistics, Faculty of Mathematics and Natural Science, IPB University, Bogor, 16680, Indonesia

²Department of Statistics, Politeknik Statistika STIS, Jakarta, 13330, Indonesia

Copyright © 2021 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract: Small area estimation (SAE) techniques are now widely employed to produce parameter estimates for smaller domains where sample sizes cannot be used to deliver direct estimation, such as regional consumer price indices (CPIs). Area-specific effects have important roles in attaining reliable parameter estimates in SAE. Thus their exclusion and inclusion in the model should be carefully examined to get accurate prediction. In practice, it is common to analyze large scale data in which the normality assumption of SAE models maybe violated due to sparsity of the area-specific effects. In this research, we consider the number of regional baskets of goods and services varies across districts. Hence, the calculation of regional CPIs needs different regional baskets of goods and services. To calculate such indices the normL1 penalty for area-specific effects selection is proposed. The response variables are the CPIs of January 2018 whereas the auxiliary variables are infrastructures and resources which are available in the village potential census in 2018 (PODES 2018). The proposed model fits to predict Regional CPIs by groups of expenditures

*Corresponding author

E-mail address: anangk@apps.ipb.ac.id

Received September 6, 2021

in the sampled districts based on the empirical relative RMSE and area-specific effect selection performance. Finally, based on the regional CPIs prediction values of all over districts/cities in Indonesia the general CPI of Indonesia during January 2018 can be calculated.

Keywords: coordinate descent; normL1; sparsity.

2010 AMS Subject Classification: 62J05, 62J07, 62P20.

1. INTRODUCTION

Consumer Price Indices (CPIs) are important indicators for monetary stability and inflation. In Indonesia, CPIs are officially calculated on the national level and 82 sampled districts in the urban area. However, the need of regional CPIs (RCPIs) all over districts is on demand as the district development grows. RCPIs show differences in the rate of regional inflation, different regional baskets of goods and services. Regional commodity baskets vary in amount and product of goods and services. In Indonesia, the commodity baskets are grouped by the 1999 Classification of Individual Consumption According to Purpose (COICOP). If the amount of the goods and services are considered as the sample size to the RCPIs prediction thus they may be poorly estimated because some districts have relatively small or zero amount of goods and services compared to the estimates at national level. Thus, one of the proposed solutions to this problem is to utilize the Small Area Estimation (SAE) method.

Small area estimation is estimation of population parameters for small areas or domains, where the sample size for a particular area or domain is too small to provide a valid estimate. The SAE area-level model to estimate CPIs has been demonstrated in various research papers such as to set up a regional basket with higher accuracy in the United Kingdom [1]. It utilized LASSO regression in auxiliary variables selection. SAE method is also applied to estimate consumer expenditure to construct the regional CPI [2], and to predict the district's CPI in West Java and Maluku Provinces, Indonesia [3].

Area-specific effects have important roles in attaining reliable parameter estimates in SAE

thus their exclusion and inclusion in the model should be carefully examined to get accurate prediction. Their values are insignificant if the observation have the similar and large sample size, therefore we need select the actual small area to get the reliable estimate. The selection process set up the area-specific effect as zero for a certain area with sufficient sample to deliver direct estimation, whereas it preserves the nonzero value for the small area. This sparsity of area-specific effects brings heavy tails and violation in its normality assumption [4, 5]. Thus, the existing SAE methods cannot handle the complexity of area-specific effects. Small area estimation with automatic random effects selection (SARS) model is a linear mixed model that developed to mitigate the random effects or area-specific effect selection issue [6].

SARS model utilized hard-ridge penalty to select the area-specific effects with an iterative selection-estimation algorithm. It employed normL0 as one of the components of the hard-ridge penalty that is discrete and non-convex. SAE with the normL1 penalty is developed to shrinkage the parameter estimate and select the area-specific effects. The research focused on developing a small area estimation model to obtain an accurate response prediction value. In the simulations, the proposed model brings out the smallest mean square error and good performance at shrinkage ability of the area-specific effects [7].

This study is conducted to investigate the parameter estimation accuracy of the SARS model and SAE with normL1 penalty for area-specific effects selection and select the relevant area-specific effects, to estimate Regional CPIs Group of Expenditures of sampled districts in Indonesia, and to predict the Regional CPIs Group of Expenditures of non-sampled districts in Indonesia. This paper is organized as follows: Section 2 discusses Small Area Estimation with Automatic Random Effects Selection (SARS) model, Section 3 presents the SAE with normL1 penalty as the proposed method, Section 4 describes the data set and preprocessing of constructing the regional CPIs in Indonesia and section 5 presents the results, and finally, Section 6 is about conclusions and future work of this study.

2. SMALL AREA ESTIMATION WITH AUTOMATIC RANDOM EFFECTS SELECTION (SARS) MODEL

The small area estimation area-level model is known as the Fay-Herriot (FH) model. It's a special form of the linear mixed model, which can be written as follow:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} + \mathbf{e} \quad (1)$$

where \mathbf{y} be the $m \times 1$ vector of the parameters of inferential interest and assume that the direct estimator $\hat{\mathbf{y}}$ is available, \mathbf{e} is a vector of independent sampling errors with mean vector $\mathbf{0}$ and known diagonal variance matrix $\mathbf{R} = \text{diag}(\sigma_{ei}^2)$, σ_{ei}^2 representing the sampling variance of the direct estimators of the i -th area. Moreover, $\boldsymbol{\beta}$ is the $p \times 1$ vector of fixed effects parameters of the auxiliary variables, \mathbf{u} is the $m \times 1$ vector of independent areas-specific effects with zero mean and $m \times m$ covariance matrix $\Sigma_u = \sigma_u^2 I_m$ where m is the number of small areas [8].

Based on the FH model that can be presented in the form of a linear model and expressed in formula (1), let U be a set of small areas. Then $u_i \sim N(0, \sigma_u^2)$ for $i \in U$ and $u_i = 0$ for $i \in U^c$ therefore we assumed that the area-specific effects u are sparse, an SAE method that was developed to mitigate this issue is SARS model. SARS is a small area modelling that considers the selection of area-specific effects employs the hard-ridge penalty as the penalized function is stated in formula (2).

$$\begin{aligned} P_{02}(\mathbf{u}; \lambda_{0u}, \eta_u) &= \frac{\eta_u}{2} \|\mathbf{u}\|^2 + \frac{\lambda_{0u}^2}{2(1 + \eta_u)} \|\mathbf{u}\|_0 \\ &= \sum_{i=1}^n \left(\frac{\eta_u}{2} u_i^2 + \frac{\lambda_{0u}^2}{2(1 + \eta_u)} 1_{u_i \neq 0} \right) \end{aligned} \quad (2)$$

where λ_{0u} is the tuning parameter for the hard penalty in order to optimized SARS prediction information criteria ($\lambda_{0u} \geq 0$) and η_u tuning parameter for ridge penalty in order to select the area-specific effects. Consider if σ_{ei}^2 are unknown then $\mathbf{R} = \text{diag}\{1/n_i\}$ so the objective functions of the SARS model with the area-specific effects selection is written in formula (3):

$$\begin{aligned} \arg \min f(\mathbf{u}; \lambda_{0u}, \eta_u) &\triangleq (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u})^T \text{diag}\{n_i\} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}) \\ &+ P_{02}(\mathbf{u}; \lambda_{0u}, \eta_u). \end{aligned} \quad (3)$$

Optimization problem in SARS is challenging due to the non-convex and non-smooth feature of the hard-ridge penalty. Thus, it is necessary to develop penalized SAE methods that not only select the area-specific effects but also shrinkage the value of their coefficients at once, such as utilize the normL1 penalty.

3. SMALL AREA ESTIMATION WITH NORML1 PENALTY FOR AREA SPECIFIC EFFECTS SELECTION MODEL

The dual penalty that is employed in the SARS model does not interact with each other. The normL2 penalty contributes to shrinkage estimation and the normL0 is solely engaged in sparsity. On the other hand, the normL1 penalty can fulfil the normL0 regularization and also shrinkage the parameter estimate to improve prediction performance. Based on the FH model in equation (1) and the objective function of the SARS model in equation (3), thus the objective functions of the small area estimation model with the area-specific effects selection with normL1 penalty can be written in formula (4) as follow:

$$\arg \min f(\mathbf{u}; \lambda_u) \triangleq (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u})^T \text{diag}\{n_i\} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{u}) + \lambda_u \|\mathbf{u}\|, \quad (4)$$

where λ_u is the tuning parameter for normL1 penalty and $\lambda_u \geq 0$. Regarding computation, it involves convex optimization thus area-specific effects estimate can be obtained by utilizing the coordinate gradient descent approach which is known as efficient algorithms [9].

Under assumption the sparsity of \mathbf{u} , based on the perspective of predictive learning [10] and predictive information criterion [11], SARS prediction information criterion (PIC) is used as criteria to achieve the optimal prediction tuning parameter. Since error model variance are unknown thus the application use SARS-PIC for free sampling error variance which can be expressed as in formula (5):

$$SARS_PIC_{free-em} = \frac{\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta} + \mathbf{u}\|_2^2}{(1 - \sigma_u^2 \Delta(\mathbf{u}))}, \quad (5)$$

where based on the simulation result and empirical experiment [6]. Thus, the variance of error model $\sigma_u^2 \in (2,3)$, and the value of $\Delta(\mathbf{u})$ is defined by:

$$\Delta(\mathbf{u}) = \frac{\sigma_e^2 [J(\mathbf{u}) + J(\mathbf{u}) \log\{e^m / J(\mathbf{u})\}]}{n\sigma_e^2}, \quad (6)$$

with $J(\mathbf{u})$ is the number of area-specific effects rows that have coefficient values of parameters that are not equal to zero and it indicate the true small area or area with insufficient samples.

4. DATA SETS

This paper focuses on constructing the RCPIs by the group of expenditures. Response variables of this paper were the group of expenditure RCPIs in January 2018, and the auxiliary variables information were regard to infrastructure & resources information from the Village Potential Census in 2018 (PODES 2018). Those are the ratio of villages with main income source of the majority of the population is agriculture, main income source of the majority of the population is industry/manufacture, market without permanent building, presence of a family of electric consumer & source of main street illumination is the state electricity company, sources of drinking water of majority of the families are branded bottled & refill water, sub-public health center, public health center, high school, university, fixed-route public transportation, and base transceiver station. The CPI in this paper uses the base year 2012 CPI = 100, based on the Cost of Living Survey (SBH) conducted by BPS during 2018. The survey in 2018 was implemented in 82 cities, consisting of 34 provincial capitals. The CPIs are classified by the group of expenditures based on the 1999 COICOP. The grouping of CPI (2012 = 100) 7 groups are as follows: Foodstuff (COICOP1); Prepared Foods, Beverages, Cigarette & Tobacco (COICOP2); Housing, Water, Electricity, Gas and Fuel (COICOP3); Clothing (COICOP4); Health (COICOP5); Education, Recreation & Sports (COICOP6); Transport, Communication and Financial Services (COICOP7).

Consumer price data collection in each city covers 225-462 types of goods and services from the results of the 2012 Cost of Living Survey (CLS) that are represented by 1-3 qualities/ brands for each commodity. The number of goods and services types for each city is different, it is selected based on certain criteria, following goods and services which are widely consumed by people in the respective city. The data collection of consumer prices needs some checklists with different survey periods (daily, weekly, biweekly, and monthly).

AREA SPECIFIC EFFECTS SELECTION OF SAE FOR CONSTRUCTION OF RCPIs

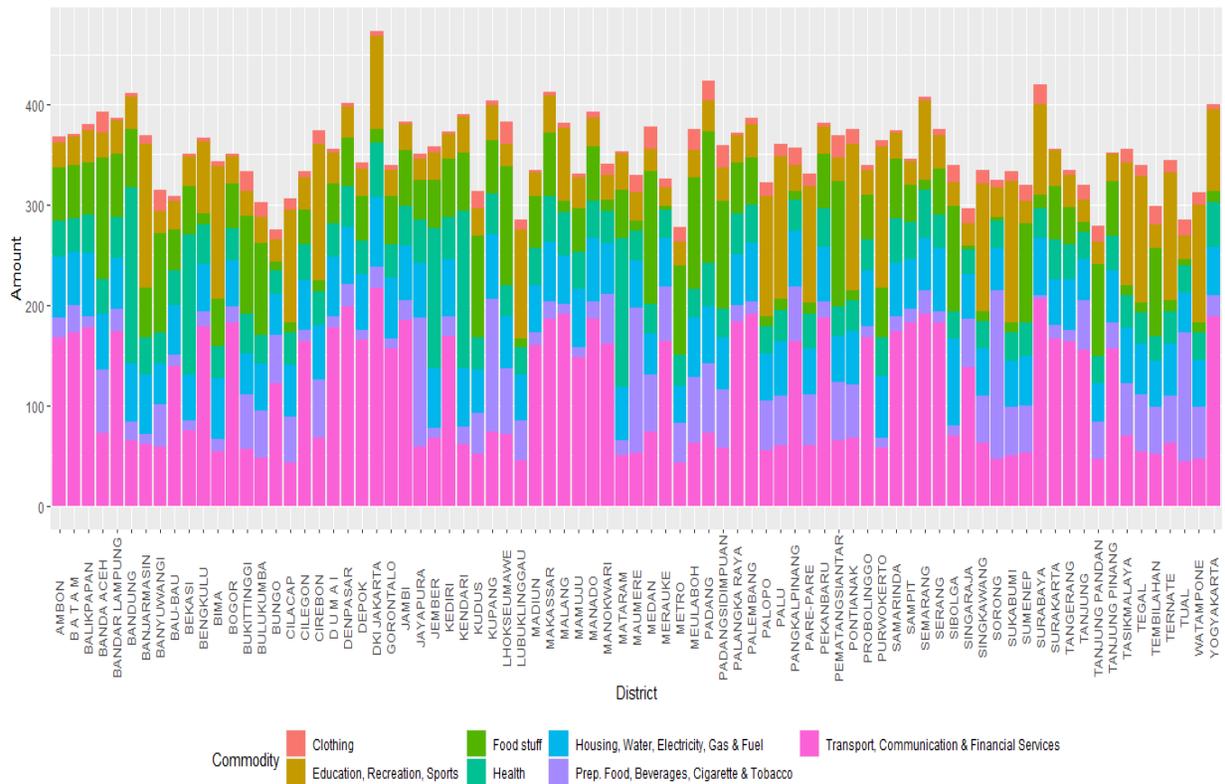


Figure 1. Goods and Services Distribution by Group of Expenditure and Cities

In the SAE model with area-specific effects selection, we take into account the amount of commodity in a regional basket of goods and services. It represents the sample size of each area, and their distribution by the group of expenditure and cities can be seen in Figure 1 above. Figure 1 shows the general result about the count of goods and services in sampled cities, the most commodity packages are in DKI Jakarta and mostly Transport, Communication and Financial Services is the maximum count of the commodity. It also describes a city with the largest count of the sampled commodity in each group, such as Medan has the largest count of the sampled commodity in Foodstuff. Sorong is the city with the largest count of the sampled commodity in prepare food, beverages, cigarette, and tobacco. Moreover, Sibolga, Padang Sidempuan, Bandung, Pontianak and DKI Jakarta are cities with the largest commodity packages in COICOP3 until COICOP7.

4.1. PREPROCESSING DATA

In Indonesia, CPI is calculated based on the result of consumer prices data processing in each city.

Direct estimators are needed to estimate the parameters for estimating the small area of CPI. The estimator was directly obtained from data collection of the consumer prices covers the goods and services whose quality/brands are generally consumed by the people in the respective city. The consumer price data are obtained from the selected respondents/retailers. This means that the probability of sampling districts/cities and commodities of goods or services is not known, even though this probability is needed to determine the weights for estimating a statistic.

The RCPIs of group expenditure statistics are estimated using a weighted average formula using the weighted results of the 2012 CLS. Furthermore, the direct estimator obtained by this method in this study is called the weighted proportional sampling (WPS) estimator and its mean square error can be written as follows:

$$\hat{y}_{gi}^{WPS} = \frac{1}{\sum_{j=1}^{n_{gi}} w_{gij}} \sum_{j=1}^{n_{gi}} w_{gij} \hat{y}_{gij}, \quad (7)$$

$$MSE(\hat{y}_{gi}^{WPS}) = \left[\frac{(n_{gi} - 1)^{-1} \sum_{j=1}^{n_{gi}} (\hat{y}_{gi} - \hat{y}_{gij})^2}{\left(\sum_{j=1}^{n_{gi}} w_{gij} \right)^2} \right] \sum_{j=1}^{n_{gi}} w_{gij}^2 \quad (8)$$

where w_{gij} and \hat{y}_{gij} are the weighted and direct estimate Regional CPI of the j -th sub group of expenditure, the i -th district and the g -th group of expenditure. Thus, n_{gi} the number of expenditure subgroups in the i -th district and the g -th group of expenditure with $j = 1, 2, \dots, n_{gi}$, $i = 1, 2, \dots, 82$ and $g = 1, 2, \dots, 7$.

Figure 2 below displays the distribution of Direct Estimation and WPS prediction of RCPIs by group of expenditure with WPS' average relative root MSE are shown in Table 1. It can be concluded that WPS Regional CPI of Education, Recreation, and Sports is the most accurate prediction statistics to publish Regional CPISs in Indonesia. On other hand, the WPS Regional CPI of Housing, Water, Electricity, Gas and Fuel has the most varied predictions to the published statistics.

Table 1. Average Relative RMSE of WPS Regional CPI of Group Expenditure

GROUP OF EXPENDITURE	RRMSE
FOOD STUFF	5.57283
PREPARED FOOD, BEVERAGES, CIGARETTE, AND TOBACCO	8.519799
HOUSING, WATER, ELECTRICITY, GAS AND FUEL	12.62479
CLOTHING	3.706992
HEALTH	9.851828
EDUCATION, RECREATION, SPORTS	4.955942
TRANSPORT, COMMUNICATION AND FINANCIAL SERVICES	8.98617

Based on Figure 2 and Table 1, it can be summarized that the WPS method obtains a precise prediction of regional CPIs of group expenditure. So that, WPS prediction values is employed as direct estimation Regional CPI of Group Expenditure in this study.

5. MAIN RESULTS

The first aim of this study is to investigate the prediction accuracy of SAE with the normL1 penalty and to select the relevant area-specific effects. SAE with normL1 penalty for area-specific effects selection (SAEL1) is the proposed model, thus we compare the proposed model with SARS and three EBLUPs for the small area. We consider the Fay-Herriot model on (1) with $\theta_i = X_i^T \boldsymbol{\beta} + u_i; i = 1, 2, \dots, n$. Assuming σ_u^2 and $\boldsymbol{\beta}$ are known, thus the best predictor (BP) of θ_i is given by:

$$\hat{\theta}_{i_{BP}} = \mathbf{X}_i^T \boldsymbol{\beta} + (1 - \gamma_i)(y_i - \mathbf{X}_i^T \boldsymbol{\beta}). \quad (9)$$

where $\gamma_i = \frac{\sigma_{ei}^2}{\sigma_u^2 + \sigma_{ei}^2}$. In the small area estimation, the sampling variances σ_{ei}^2 are assumed to be known, however both of σ_u^2 and $\boldsymbol{\beta}$ are unknown in practical. So that, $\boldsymbol{\beta}$ is estimated by replacing with its maximum likelihood estimator (MLE) for the EBLUP of $\boldsymbol{\theta}$ and the empirical best linear unbiased predictor (EBLUP) can be written as:

$$\hat{\theta}_{i_{EBLUP}} = \hat{\gamma}_i y_i + (1 - \hat{\gamma}_i) \mathbf{X}_i^T \hat{\boldsymbol{\beta}}. \quad (10)$$

with variance area-specific effects σ_u^2 can be estimated with maximum likelihood estimator (ML), restricted maximum likelihood estimator (REML) or Prasad and Rao estimator (PR). Thus

depending on estimation method of σ_u^2 , the EBLUPs are denoted as EBLUP-ML, EBLUP-REML and EBLUP-PR, respectively.

Table 2. Empirical Relative RMSE of Regional CPIs by Group of Expenditure of Sampled Cities

Group of expenditure	EBLUP-PR	EBLUP-ML	EBLUP-REML	SARS	SAEL1
Food Stuff	5.0236	14.8857	14.0906	16.5240	4.7463
Prepared Food, Beverages, Cigarette, and Tobacco	24.4385	25.6295	25.6102	17.4966	4.6325
Housing, Water, Electricity, Gas and Fuel	40.5030	51.8050	41.9377	8.2602	3.8801
Clothing	2.0833	2.4206	4.0035	12.2273	6.1060
Health	7.4877	10.8674	10.4267	11.2443	5.7130
Education, Recreation, Sports	14.4653	18.2210	15.9665	17.5133	7.4862
Transport, Communication and Financial Services	33.9654	43.2098	34.9490	11.3300	4.7015

The empirical relative RMSE are presented in Table 2 and we can investigate the prediction accuracy of the proposed model and previous SAE models. SAE with normL1 penalty for area-specific effects selection as the proposed model brings out the smaller empirical relative RMSE compared to other models. SAE with the normL1 penalty method obtains a selection of the area-specific effects to select the relevant area-specific effects. The proposed method brings out the rates of nonzero area-specific effects from the application data are about 10% by the group of expenditure RCPIs, respectively. Based on the empirical relative RMSE and area-specific effect selection performance, we can summarize that SAE with normL1 penalty fits to predict RCPIs by the group of expenditure in the sampled cities.

This study also derives the total RCPIs prediction of sampled cities in Indonesia. Figure 3 displays the distribution of the direct estimation RCPIs, and the prediction values by the SARS method and SAE with normL1 penalty method. Figure 3 carried out the comparison of prediction RCPIs, and the best accuracy results for the SAE method for area-specific effects selection methods used is SAE with normL1 penalty. Furthermore, we also construct the general CPI prediction values of Indonesia and the official general CPI of Indonesia in January 2018 that published by BPS-Statistics Indonesia is 132.1 [12]. The general CPI prediction values of the WPS

method, SARS method and SAE with normL1 penalty are 132.1653, 133.2927, and 131.7263, respectively. And, the standard deviation of general CPI prediction values of the WPS method is 4.76, SARS method is 16.55 and SAE with normL1 penalty method is 5.11. Based on application results, the SAE with normL1 model is quite good at prediction of regional CPIs by the group of expenditure, total regional CPIs moreover at prediction of Indonesia general CPI.

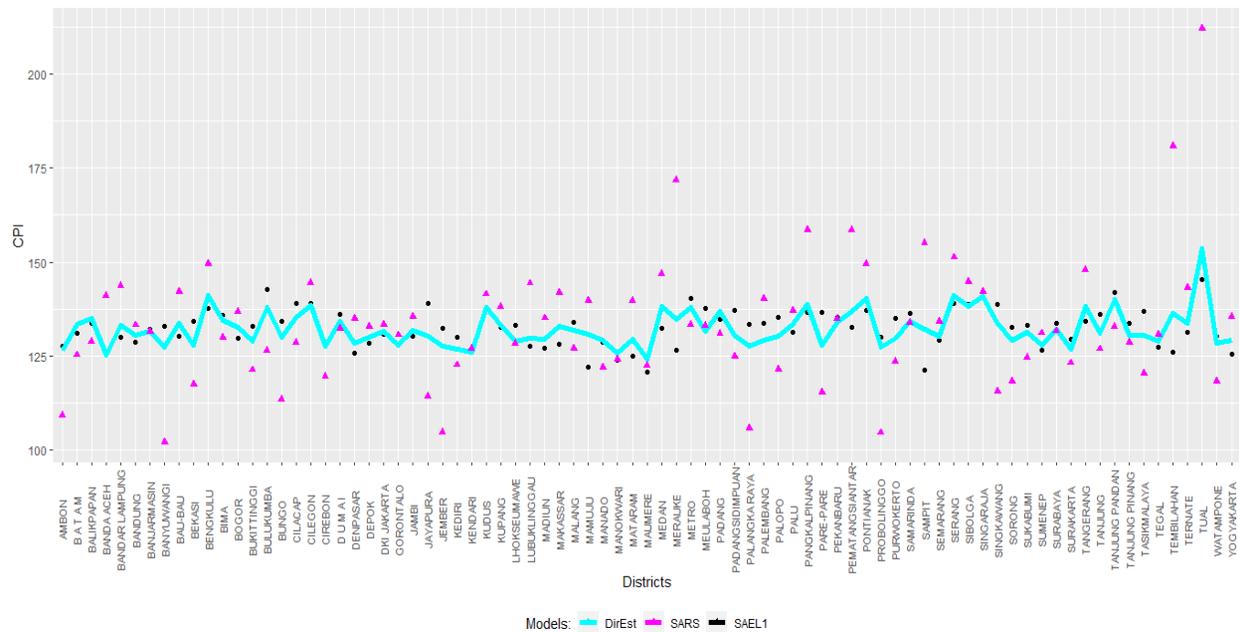
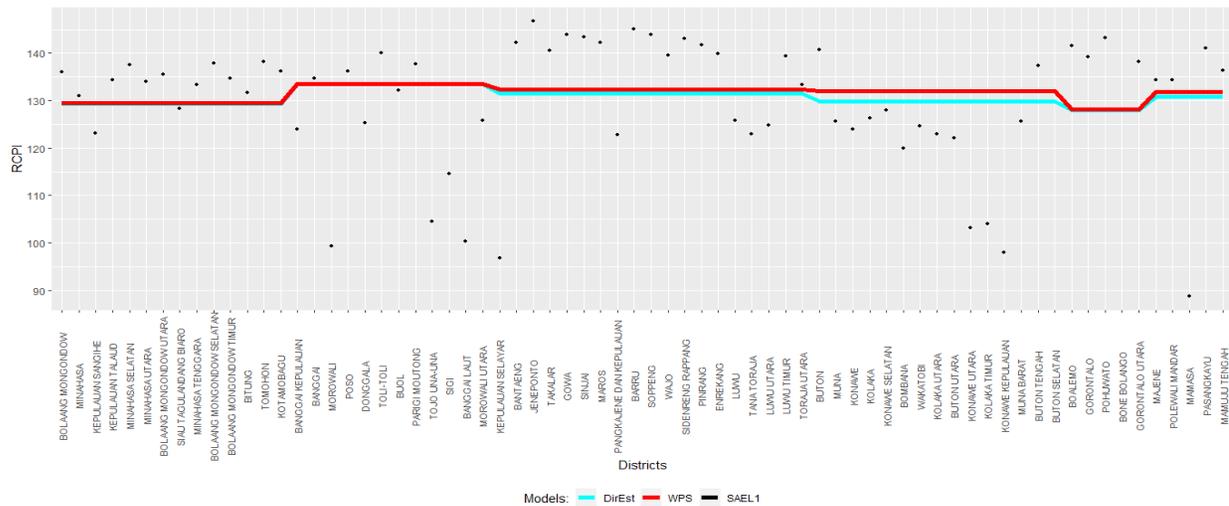
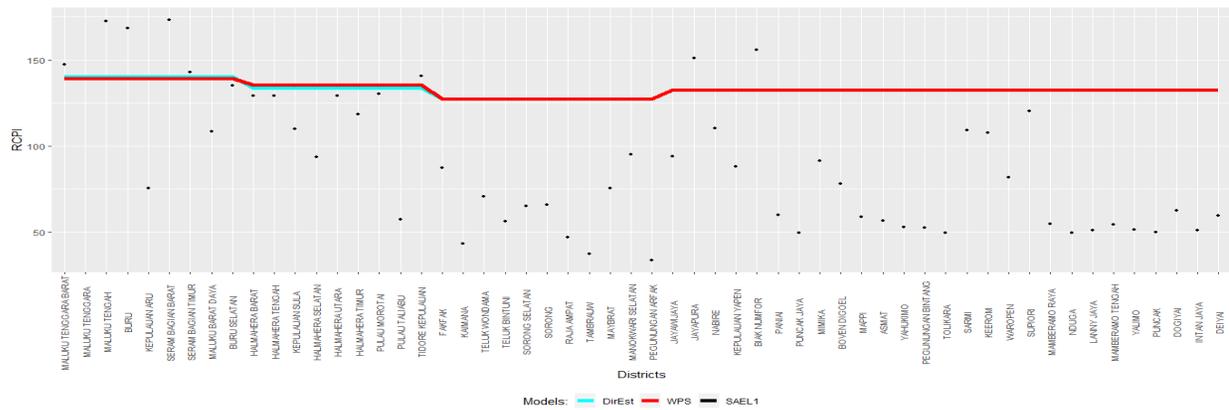


Figure 3. Regional CPIs Prediction of Direct Estimation, SARS and SAEL1 Method

In the last objective of the study, we predict Indonesia general CPI using RCPIs all over the districts/cities. To fulfil it, we derive regional CPIs by the group of expenditure and total regional CPIs of non sampled districts/cities in Indonesia. We employ the maximum value of area-specific effects of each province in formulation to predict the RCPIs by the group of expenditure of non sampled districts/cities. Furthermore, we also utilize the mean of weighted of sampled districts/cities by the group of expenditure to compute the total RCPIs prediction of non-sampled districts/cities. Figure 4 represents the prediction of total RCPIs prediction of districts/cities in Indonesia. Based on the direct estimate and WPS method, total RCPIs prediction of non-sampled districts/cities are the mean of those values for each province. Thus, the prediction values in the non-sampled districts/cities in each province are the same. Those are displayed as a straight line in Figure 4.



(d)



(e)

Figure 4. Distribution of Total Regional CPIs Prediction of district/cities in Indonesia by island: (a) Sumatera, (b) Java, Bali and Nusa Tenggara, (c) Kalimantan, (d) Sulawesi, and (e) Maluku and Papua

On the other hand, the RCPIs prediction values of SAE with normL1 penalty are vary. The results are categorized by island in Indonesia and displayed in Figure 4 above. Figure 4 displays price heterogeneity across provinces even more inter-regional in Indonesia. For Indonesia, the consumer price index measures the pure price change in a selected basket of goods and services (of constant quantity and quality) typically purchased by Indonesian households. Thus, the diversity suggest that some cities move aggregate prices more than others [13,14]. Figure 4 gives evidence that prices stability take place in districts/cities of the middle

region of Indonesia. And, heterogeneity of basket of goods and services prices occur in districts/cities of the frontier, the outermost and least developed regions often referred to as 3T (*terdepan, terluar dan tertinggal*) regions. Finally, based on the regional CPIs prediction values of all over districts/cities in Indonesia, we calculate the general CPI of Indonesia of January 2018 is 131.4059 with its standard deviation is 23.4684.

6. CONCLUSION

This paper constructs Indonesia general CPI using prediction values of RCPIs all over the districts/cities. The prediction RCPIs derive from SAE with the normL1 penalty method as the proposed model. The proposed model obtain good results on the prediction of RCPIs by the group of expenditure at sampled and non-sampled districts/cities. The proposed model also attain good performance at area-specific effects shrinkage and selection of application data, since it brings out the small value of nonzero area-specific effects rate. Furthermore, the usage maximum value of area-specific effects of each province in the prediction formula has good performance compared to the official CPIs that published by BPS-Statistics Indonesia. The prediction results display price heterogeneity over districts/cities. Yet, formal price tests are needed to investigate districts/cities in each province contains cities that contribute more to prices changes, since better price control in these leader cities will allow for faster convergence to price stability.

SAE with normL1 penalty method borrows auxiliary variables strength of 82 sampled cities to calculate the total RCPIs prediction of non-sampled district/cities. Heterogeneity of prediction RCPIs might be caused by the auxiliary variables selection of model, especially in this 'Big' data era. Database size and technology has developed rapidly, it offers high dimensional of auxiliary variables in the SAE approach. Future works will conduct to mitigate the issues.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

- [1] N. Würz, T. Schmid, P. Smith, N. Tzavidis, Construction of regional consumer price indices using small area estimation, in *Statistische Woche: 2017*, Rostock, Germany. 19 Sep 2017 - 22 Feb 2018.
- [2] J. Dawber, N. Würz, P. Smith, *The Development of regional CPIH : Assessment of small area estimation for regional expenditure weights*. 46 Newport (UK): Office for National Statistics, 2019.
- [3] A.O. Fengki, *Small area estimation methods for consumer price index statistics of regencies/municipalities in West Java and Maluku Provinces*. Master's Thesis, IPB University, Bogor, Indonesia, 2020.
- [4] D. Handayani, K.A. Notodiputro, A. Saefuddin, I. Wayan Mangku, A. Kurnia, Empirical best predictor for nested error regression small area models, *IOP Conf. Ser.: Earth Environ. Sci.* 187 (2018), 012036.
- [5] J. Jiang, T. Nguyen, J.S. Rao, Best Predictive small area estimation, *J. Amer. Stat. Assoc.* 106 (2011), 732–745.
- [6] J. Lee, *Small area estimation with random effects selection*. Dissertation, Florida State University, Florida, United States of America, 2017.
- [7] N.H. Pusponegoro, A. Kurnia, K.A. Notodiputro, A.M. Soleh, E.T. Astuti, Small area estimation of sub-district's per capita expenditure through area effects selection using LASSO method, *Procedia Computer Sci.* 179 (2021), 754–761.
- [8] J. N. K. Rao, and I. Molina, *Small area estimation 2nd edition*. New Jersey: John Wiley & Sons, Inc., 2015.
- [9] J. Schelldorfer, P. Bühlmann, S.V. De Geer, Estimation for high-dimensional linear mixed-effects models using ℓ_1 -penalization: high-dimensional mixed-effects models, *Scandinavian J. Stat.* 38 (2011), 197–214.
- [10] T. Hastie, R. Tibshirani, J. Freidman, *The elements of statistical learning: data mining, inference, and prediction*, Springer California, 2009.
- [11] Y. She, *Selective Factor Extraction in High Dimensions*, ArXiv:1403.6212 [Stat]. (2016).
- [12] BPS-Statistics Indonesia, *Consumer price indices in 82 cities in Indonesia (2012=100) 2018*. Jakarta, 2019.
- [13] B.P. Jangam, V. Akram, Do prices converge among Indonesian cities? an empirical analysis, *Bull. Monetary Econ. Bank.* 22 (2019), 239–262.
- [14] P.K. Narayan, Understanding Indonesia's city-level consumer price formation: implications for price stability, *Bull. Monetary Econ. Bank.* 22 (2019), 405–422.