Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 2, 2193-2203
https://doi.org/10.28919/jmcs/5490
ISSN: 1927-5307

STACKING METHOD FOR DETERMINING WEIGHTS IN PARTIAL LEAST SQUARES MODEL AVERAGING

ANANG KURNIA[1,*], BAGUS SARTONO[1], MUHAMMAD ARNA RAMADHAN[1],
SEPTIAN RAHARDIANTORO[1,2], YUSMA YANTI[3]

[1]Department of Statistics, IPB University, Bogor 16680, Indonesia

[2]Department of Human Ecology, Okayama University, Okayama 700-8530, Japan

[3]Department of Computer Science, Pakuan University, Bogor 16144, Indonesia

Abstract: Model averaging has been developed to improve prediction accuracy in high dimensional regression. This approach is a weighted linear combination of some regression models. Two important procedures in model averaging are construction the candidate models and determination the weights. This paper evaluated performance of partial least squares model averaging (PLSMA) with some weights and proposed stacking as a method for determining the weights. Stacking determines the weights by minimizing least squares error over candidate models. Our proposed method (stacked-PLSMA) was evaluated in a simulation experiment and compared to equal weight, Akaike information criterion (AIC) weight, and Bayesian information criterion (BIC) weight. The result showed that stacked-PLSMA yielded smaller prediction error with high consistency than the other weights.

Keywords: high dimensional regression; model averaging; partial least squares; stacking

2010 AMS Subject Classification: 68T09, 62R07.

---

*Corresponding author
E-mail address: anangk@apps.ipb.ac.id
Received January 27, 2021

1. INTRODUCTION

Model averaging is an alternative approach used to handle high-dimensional regression which the number of predictors exceeds the sample size. The main concept of model averaging is construct several models (called candidate models) and combines them to estimate the final model. The goal in model averaging is to improve prediction accuracy. Unlike model selection which only choose the best model for prediction, model averaging approach combines a class of candidate models by giving weights to each candidate model. There are two important procedures in model averaging, construction of candidate models and determination the weights.

In constructing candidate models, there are some ways such as random partition of predictors using the Hierarchical Adaptive Random Partitioning (HARP) algorithm [1, 2], partition predictors based on marginal correlation with response variable [3] so each candidate model contains predictors that has similar correlation with response variable, and partial least squares model averaging [4]. For determining the weights for candidate models, there are some ways such as weights based on Akaike information criterion (AIC) [5, 6], weights based on Bayesian information criterion (BIC) [7, 8], weights based on Mallows' Cp [9, 10], weights based on cross-validation criterion [11], and weights based on unbiased estimator of risk [12].

In recent years, many researchers have discussed and applied model averaging. In econometrics, there are [13, 14]., the author [15] applied model averaging in design of experiment. In genomics, there are [16, 17]. The other studies about model averaging are discussed by [18, 19].

Our previous study [4], we have proposed partial least squares model averaging (PLSMA) as an approach for constructing the candidate models. In the present study, we discuss about determination the weights for PLSMA. We proposed stacking which developed by [20, 21] as a method to determine the weights. This method determines the weights by minimizing squared error between the actual value of response variable and the predicted value. By simulation study using R software, we assess the performance of stacked-PLSMA and compared to PLSMA with the other existing weights, including equal weight, AIC weight, and BIC weight. We use RMSEP (root mean squared error of prediction) and correlation between the actual value of response variable and the predicted value as the performance measures for each method.

The paper is organized as follows. Section preliminaries describes model averaging, including construction of candidate models, determination the weights, and stacking or stacked regression and also the procedure of simulation study. The next section is the main result that describes the result of this study through simulation study. Then the conclusion is presented in the last section.

2. PRELIMINARIES

2.1. Model Averaging

Suppose $f_k(X), k = 1, 2, ..., K$ be the set of candidate models and w_k be the weights corresponding to $f_k(X)$, where $0 \le w_k \le 1$ and $\sum_k w_k = 1$. The model averaging can be written as

$$f(X) = \sum_{k=1}^{K} w_k f_k(X) \tag{1}$$

Let y is $n \times 1$ vector of the response variable and X is $n \times p$ matrix of the predictors. For linear regression model, K candidate models contain m predictors where m is smaller than p and n. The model is defined as

$$f_k(X) = X_k \beta_k + \varepsilon \tag{2}$$

or

$$y_k = X_k \beta_k + \varepsilon \tag{3}$$

and β_k is estimated with ordinary least squares (OLS), so

$$\hat{\beta}_k = (X_k^{'} X_k)^{-1} X_k^{'} y \tag{4}$$

then the model averaging estimator can be expressed as

$$\hat{f}(X) = \sum_{k=1}^{K} w_k \hat{f}_k(X) = \sum_{k=1}^{K} w_k \hat{y}_k$$

There are two procedures in model averaging, construction of candidate models and determination the weights.

1) Construction of candidate models: Construction of candidate models is first procedure in model averaging. This paper applies partial least squares model averaging (PLSMA) for constructing candidate models developed by [4]. PLSMA transforms the origin predictor into new variables through partial least squares (PLS) and candidate models are constructed based on new variables. Then the algorithm of PLSMA can be described below:

   Step 0 Randomly permute the order of the observations.

   Step 1 Split data into two parts, $Z^{(1)} = (X_i, y_i), 1 \le i \le \frac{n}{3}$ and $Z^{(2)} = (X_i, y_i), \frac{n}{3} + 1 \le i \le n$

   Step 2 Resampling 75% observations in $Z^{(1)}$.

   Step 3 Do PLS to get new predictor variables.

   Step 4 Repeat Steps 2-3 for k times. Estimate $\hat{f}_k(X)$ based on $Z^{(1)}$ for $1 \le k \le K$.

   Step 5 For each candidate model, compute the weights.

   Step 6 Compute the predictions for each candidate model using the remaining half of data $Z^{(2)}$ by $\hat{y}_k = X_k \hat{\beta}_k$.

   Step 7 Compute the final prediction based on equation (5).

   In addition, for determining the weights will be presented in the next section.

2) Determination of weights: The second procedure in model averaging is determination the weights of candidate models. There are many choices of weights such as equal weight, AIC weight, and BIC weight. In this paper, we will compare these weights with the weight based on stacking method.

   a. Equal weight: The standard weight for model averaging is equal weight. Each candidate model is given the same weight, $w_k = \frac{1}{k}$ then the final prediction is $\hat{y} = \frac{1}{k} \sum_{k=1}^{K} \hat{y}_k$.

   b. AIC and BIC weight: Two types of weights are measured based on information criterion, Akaike information criterion (AIC) and Bayesian information criterion

(BIC). Both AIC and BIC weights are defined as $w_k = \frac{\exp\left(-\frac{1}{2}AIC_k\right)}{\sum_{k=1}^{K}\exp\left(-\frac{1}{2}AIC_k\right)}$ and

$w_k = \frac{\exp\left(-\frac{1}{2}BIC_k\right)}{\sum_{k=1}^{K}\exp\left(-\frac{1}{2}BIC_k\right)}$, where $AIC_k = -2\log L + 2p$; $AIC_k = -2\log L + p\log n$

with L is likelihood function. The higher weights are given to the better models.

2.2. Stacking

Wolpert [20] introduced stacked generalization as a technique to achieve a prediction accuracy which is a high as possible. Similar to model averaging, stacked regressions or stacking is defined a method for combining linear combination of predictors (called candidate models) to improve prediction accuracy [21]. Unlike model averaging which is giving the weights based on information criterion, stacking determines the weights by minimizing squared error between the actual value of response variable and the predicted value.

Given y is $n \times 1$ vector of the response variable and X is $n \times p$ matrix of the predictors. The model is defined by equation (1). In this term, the $\{w_k\}$ gotten by minimizing squared error:

$$\sum_n\left(y_n - \sum_k w_k f_k(X)\right)^2 \tag{5}$$

Then the result of estimator $\sum_k w_k f_k(X)$ will yield lower prediction error.

2.3. Simulation Study

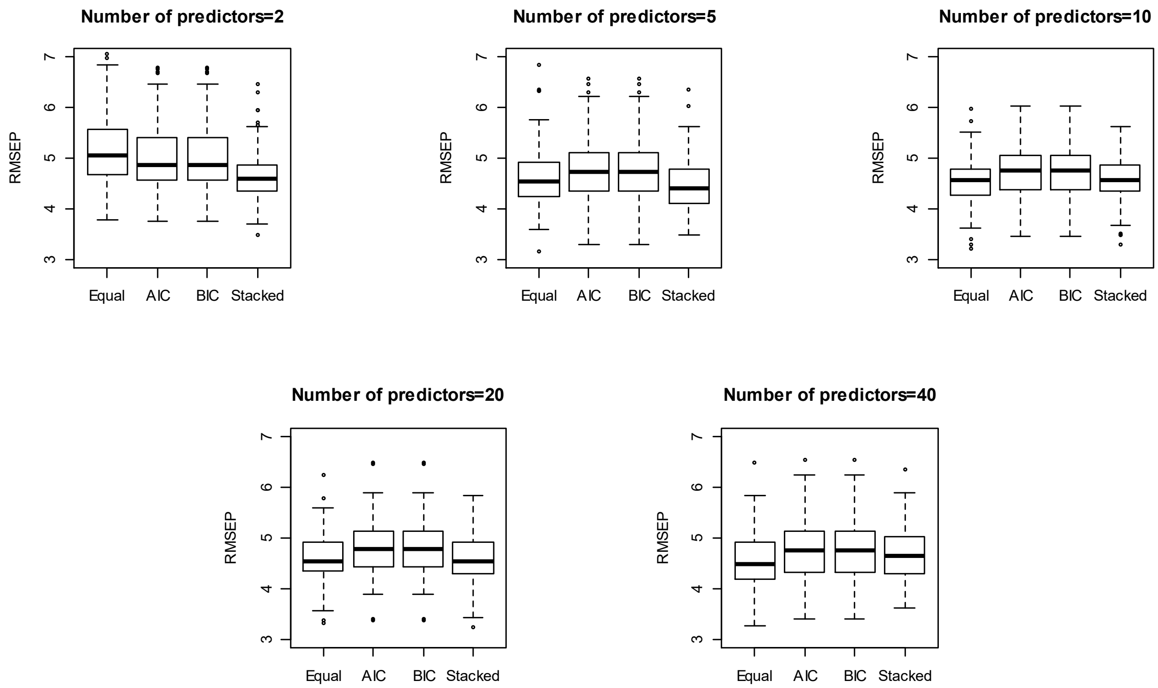
This paper proposes stacking as a method to determine the weights in partial least squares model averaging (PLSMA) that developed by Ramadhan et. al. [4]. We design a simulation study to evaluate the performance of our proposed method. In this simulation study, we adopted the settings in [3] with several modifications. By using R software, we generate the predictor variables $p = 2000$ from the multivariate normal distribution with mean 0 and covariance matrix $S = (s_{ij})$ with $s_{ij} = \rho^{|i-j|}$, $\rho = 0.95$, and set the sample size $n = 120$. The random effect is generated from normal distribution with mean 0 and standard deviation 2. We set the coefficient $\beta_0 = 60$ and assume that only 50 predictors have contributions in prediction the response variable. The true predictors $\beta_1 = 1$ be spaced evenly, $i = 20(j-1)+1, j = 1,2,...,50$.

To implement model averaging, we set two conditions number of candidate models $K = \{5, 20\}$ and the number of predictors in each candidate model $m = \{2, 5, 10, 20, 40\}$. The performance of our proposed method, stacking (called stacked-PLSMA) will be evaluated by measuring RMSEP (root mean squared error of prediction) and correlation between the actual value of response and the predicted value. Then, these performance measures will be compared to the other weights such as equal weight, AIC weight, and BIC weight.

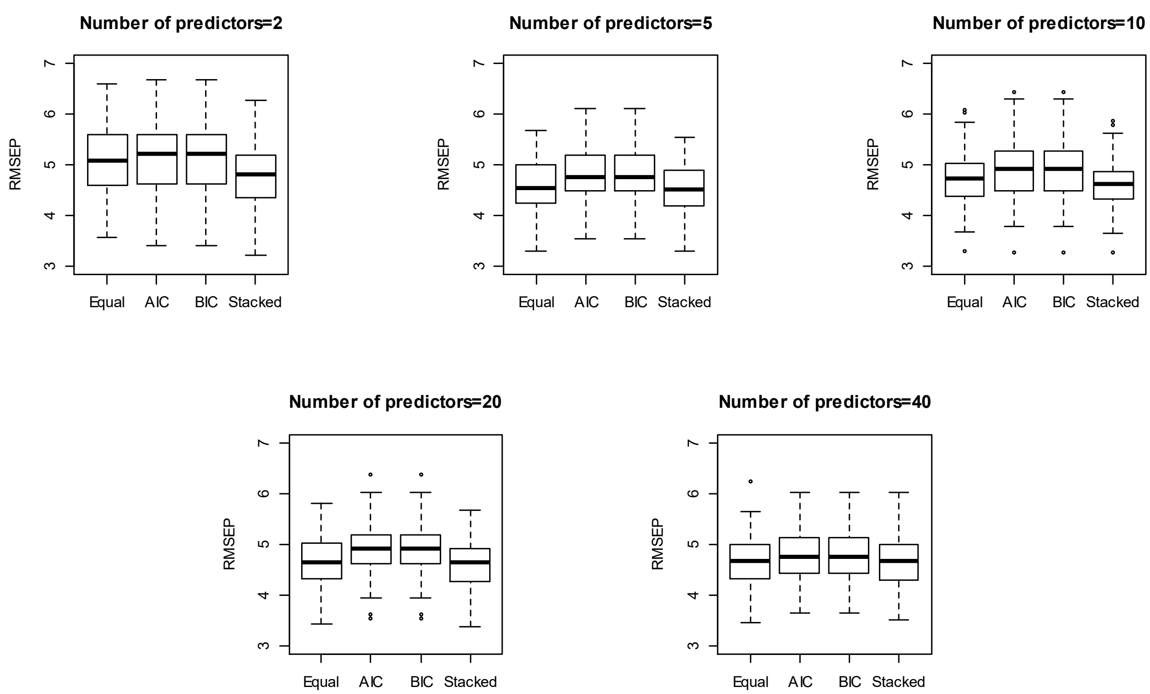
3. MAIN RESULTS

Figure 1 displayed the results of performance measure after 100 simulation runs. For number of candidate models $K = 5$ that shown in (a), when the number of predictors $m = 2$, three types of weights, equal weight, AIC weight, and BIC weight yielded same performance of RMSEP while stacking yielded the smaller RMSEP. The similar form of boxplots are produced in some conditions of $m = \{5, 10, 20\}$ and each weights produced smaller variances of RMSEP than $m = 2$. For $m = 40$, there were almost no difference performance measure RMSEP in all types of weights.

Figure 1 part (b) displayed the performance measure RMSEP for $K = 20$. The boxplots in (b) showed that the performances are similar to the result in (a), but when $K = 20$ the variance of RMSEP is smaller than $K = 5$. So, we can say that the variance of prediction error decreases when number of candidate models increases. Both AIC weight and BIC weight produced the same performance. The performance of equal weight is better than AIC and BIC weight in some conditions. Stacking produced the smallest RMSEP in all conditions. Although, the equal weight (in some conditions) has similar RMSEP with stacking, stacking yields the smaller variance than the equal weight, so stacking has higher accuracy.

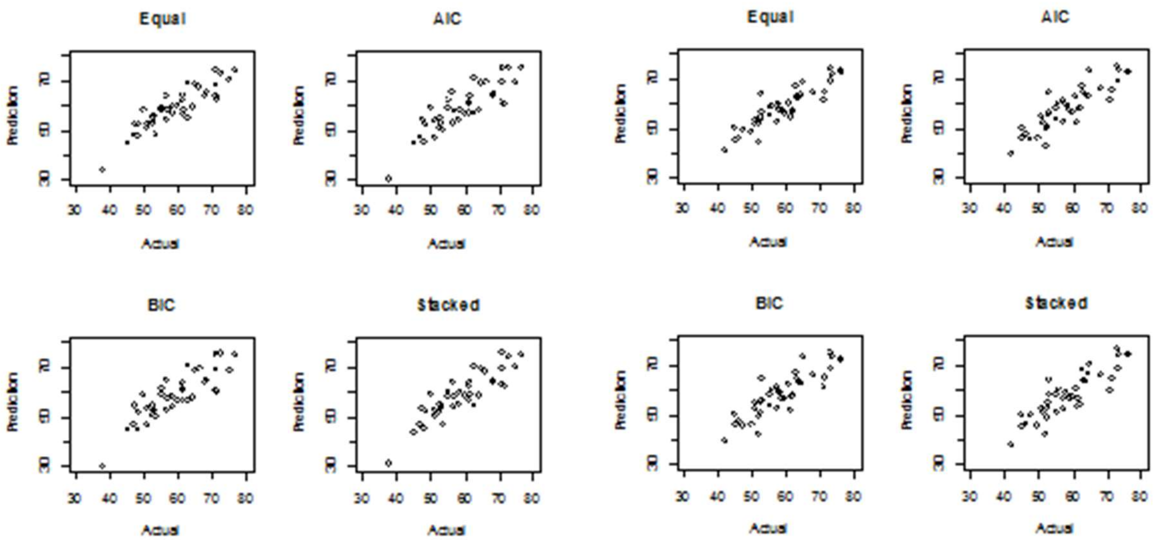


(a) Number of candidate models = 5



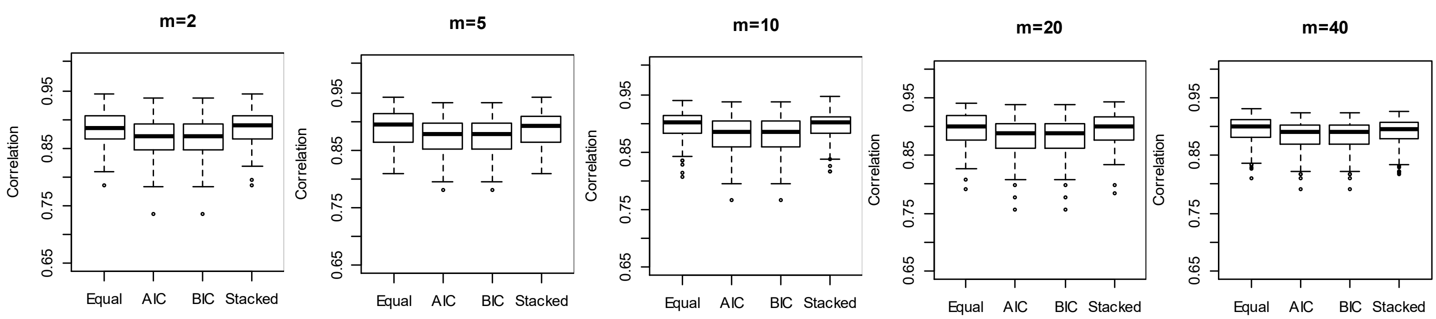
(b) Number of candidate models = 20

Figure.1 Boxplots of the performance measure RMSEP of number candidate models (a) and (b) after 100 simulation runs

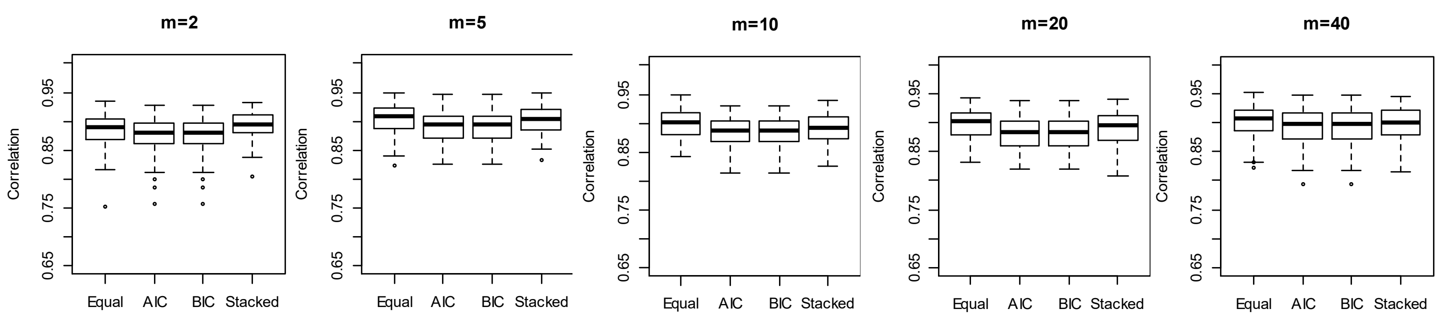


(a) Number of candidate models = 5 (b) Number of candidate models = 20

Figure.2 Scatterplots of actual and prediction for number candidate models (a) and (b)



(a) Number of candidate models = 5



(b) Number of candidate models = 20

Figure.3 Boxplots of the performance measure correlation of number candidate models (a) and (b) after 100 simulations

Besides RMSEP, the prediction accuracy can be measured by correlation between the actual value of response variable and the predicted value. The good prediction performance will have high correlation with the actual value while the bad prediction performance is shown by low correlation. The scatter plots of the actual value and the predicted values are displayed in Figure 2. Both (a) and (b) showed the linear pattern in each weight. The patterns indicated that the actual value and the predicted value have high correlation.

Table. I Comparison of correlation between actual value and prediction for $K = 5$

m	Weights			
	Equal	**AIC**	**BIC**	**Stacked**
2	0.884	0.868	0.868	0.885
5	0.888	0.873	0.873	0.888
10	0.895	0.879	0.879	0.895
20	0.896	0.880	0.880	0.894
40	0.893	0.884	0.884	0.891

Table. II Comparison of correlation between actual value and prediction for $K = 20$

m	Weights			
	Equal	**AIC**	**BIC**	**Stacked**
2	0.885	0.875	0.875	0.893
5	0.905	0.890	0.890	0.902
10	0.898	0.885	0.885	0.892
20	0.898	0.882	0.882	0.891
40	0.901	0.892	0.892	0.896

Table I and Table II showed the comparison of correlation between the actual value and the predicted value after 100 simulation runs. Each weight produced high correlation with the values are above 0.85. Both AIC and BIC weights yielded same correlation in all conditions but the values are smaller than correlation produced by equal weight and stacking. In some conditions, the equal weight has correlation greater than stacking, but the difference is not significant. We show the boxplots of correlations after 100 simulations that produced by partial least squares model averaging with equal weight, AIC weight, BIC weight, and stacking in Figure 3. All results of simulation study have been shown that our proposed method, stacking have produced the better prediction performance than the other weights.

4. CONCLUSIONS

In recent years, the analysis of high-dimensional regression has attracted a lot of attention. Model averaging has been developed as an alternative to model selection in the case where the number of predictors exceeds the number of observations. Our research proposed stacking as a method to determine the weights in partial least squares model averaging. As shown in simulation study, our proposed method (stacking) yielded the better prediction accuracy than the other existing weights. This research focused in determining the weights for partial least squares model averaging (PLSMA). As a future direction, more investigations are needed on how to determine the optimal number of candidate models to be average.

CONFLICT OF INTERESTS

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

REFERENCES

[1] M.R. Banan, K.D. Hjelmstad, Self-organization of architecture by simulated hierarchical adaptive random partitioning, in: [Proceedings 1992] IJCNN International Joint Conference on Neural Networks, IEEE, Baltimore, MD, USA, 1992: pp. 823–828.

[2] M.P. Perrone, Improving regression estimation: Averaging methods for variance reduction with extensions to general convex measure optimization, PhD Thesis, Brown University, 1993.

[3] T. Ando, K.-C. Li, A Model-Averaging Approach for High-Dimensional Regression, J. Amer. Stat. Assoc. 109 (2014), 254–265.

[4] M.A. Ramadhan, B. Sartono, A. Kurnia, Partial least squares in constructing candidates model averaging. Int. J. Sci. Res. Sci. Eng. Technol. 4(1) (2018), 1459-1463

[5] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, in: E. Parzen, K. Tanabe, G. Kitagawa (Eds.), Selected Papers of Hirotugu Akaike, Springer New York, New York, NY, 1998: pp. 199–213.

[6] G. Claeskens, N.L. Hjort, Model selection and model averaging, Cambridge Books, Cambridge University Press,

2008.

[7] A.E. Raftery, D. Madigan, J.A. Hoeting, Bayesian Model Averaging for Linear Regression Models, J. Amer. Stat. Assoc. 92 (1997), 179–191.

[8] P.J. Brown, M. Vannucci, T. Fearn, Bayes model averaging with selection of regressors, J. R. Stat. Soc., Ser. B, Stat. Methodol. 64 (2002), 519–536.

[9] C.L. Mallows, Some Comments on C_p, Technometrics. 42 (2000) 87–94.

[10] B.E. Hansen, Least Squares Model Averaging, Econometrica. 75 (2007), 1175–1189.

[11] Z. Yuan, Y. Yang, Combining Linear Regression Models: When and How? J. Amer. Stat. Assoc. 100 (2005), 1202–1214.

[12] H.Wang, X. Zhang, G. Zou, Frequentist model averaging estimation: a review. J. Syst. Sci. Complex. 22(4) (2009), 732–748.

[13] A. Ullah, H. Wang, Parametric and nonparametric frequentist model selection and model averaging. Econometrics. 1 (2013, 157-179.

[14] E. Moral-Benito, Model averaging in economics: An overview. J. Econ. Surv. 29 (2015), 46-75.

[15] D.T. Salaki, A. Kurnia, B. Sartono, Model Averaging Method for Supersaturated Experimental Design, IOP Conf. Ser.: Earth Environ. Sci. 31 (2016), 012016.

[16] S. Rahardiantoro, B. Sartono, A. Kurnia, Model Averaging for Predicting the Exposure to Aflatoxin B 1 Using DNA Methylation in White Blood Cells of Infants, IOP Conf. Ser.: Earth Environ. Sci. 58 (2017), 012019.

[17] D. Posada, T.R. Buckley, Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53(5) (2004), 793-808.

[18] S. Rahardiantoro, K.A. Notodiputro, A. Kurnia, Prediction Intervals of Model Averaging Methods for High-Dimensional Data. IOP Conf. Ser.: Earth Environ. Sci. 187(1) (2018), 012045

[19] D.T. Salaki, A. Kurnia, A. Gusnanto, I.W. Mangku, B. Sartono, Model averaging, an alternative approach to model selection in high dimensional data estimation. Forum Stat. Komputasi: Indonesian J. Stat. 20(2) (2015), 116-119.

[20] D.H. Wolpert, Stacked generalization. Neural Networks, 5(2) (1992), 241-259.

[21] L. Breiman, Stacked regressions. Mach. Learn. 24(1) (1996), 49-64.