Jurnal Matematika UNAND
Vol. **14** No. **4** Hal. 320 – 332
Edisi Oktober 2025
ISSN : 2303–291X
e-ISSN : 2721–9410

https://doi.org/10.25077/jmua.14.4.333-332.2025

PERFORMANCE ANALYSIS OF PREDICTION INTERVAL ESTIMATORS IN RANDOM FOREST MODELS WITH CORRELATED PREDICTORS

MEISYATUL ILMA*, BAGUS SARTONO, HARI WIJAYANTO
IPB University
email : meisyatulilma2@gmail.com, bagusco@apps.ipb.ac.id, hari@apps.ipb.ac.id

Received June 19, 2025 Received in revised form July 15, 2025 Accepted August 8, 2025
Available online October 31, 2025

Abstract. *Uncertainty in prediction results is a critical aspect that must be addressed in regression modeling, particularly when there is a high correlation among explanatory variables. This study evaluates the performance of three prediction interval formation approaches within the Random Forest framework: Out-of-Bag Prediction Interval (OOB-PI), Quantile Regression Forest (QRF), and Split Conformal Prediction (SC). The evaluation was conducted through simulation studies involving various data structures, including different levels of correlation between variables, mean function shapes, and error distribution types. Further validation was carried out using data from the National Socio-Economic Survey (SUSENAS) for West Java Province in 2023. The results indicate that increasing the correlation among explanatory variables can enhance the efficiency and accuracy of prediction interval estimation. Among the methods assessed, OOB-PI exhibited the most balanced performance, achieving a prediction coverage rate close to 90% and yielding narrower interval widths compared to QRF and SC. These findings suggest that OOB-PI is an adaptive and efficient approach for modeling diverse data structures, including socioeconomic datasets with highly correlated predictors.*

Keywords: Multicollinearity, Prediction interval, Random Forest.

1. Introduction

Understanding the relationship between explanatory and response variables is a fundamental aspect of developing accurate predictive models. One commonly used statistical approach to examine the relationship between variables is regression modeling [1]. A major challenge in regression modeling lies in the presence of correlations among explanatory variables. High correlations among explanatory variables can lead to multicollinearity, a condition in which two or more explanatory variables are strongly linearly related, thereby complicating the identification of each

*Corresponding Author

variable's effect on the response variable [2]. One of the machine learning techniques effective in addressing multicollinearity issues is the Random Forest algorithm.

Random Forest is an ensemble learning algorithm that combines multiple decision tree models to produce more accurate predictions [3]. The primary advantage of Random Forest is its ability to handle multicollinearity more effectively than linear regression, as highly correlated variables do not always appear simultaneously in every decision tree [4]. Random Forest generates accurate predictions by averaging the results from all decision trees; however, the algorithm only provides a single-point prediction or point estimate. Point predictions offer only one specific value and often fail to capture the uncertainty inherent in the estimation process. Interval estimation is employed to address this limitation by providing a more comprehensive estimate through the quantification of uncertainty in Random Forest predictions.

The prediction interval formation methods employed in this study include Quantile Regression Forest (QRF). QRF generates a range of values that reflect prediction uncertainty, thereby providing deeper insights into data variability [5]. The Split Conformal Prediction (SC) method is also used to construct prediction intervals. This method offers probabilistic coverage guarantees without requiring specific distributional assumptions, making it suitable for various types of data, including those that do not conform to classical assumptions [6]. Another method used to construct prediction intervals within the Random Forest framework is the Out-of-Bag Prediction Interval (OOB-PI). The OOB-PI method utilizes the residuals from out-of-bag predictions without the need for additional validation data. OOB-PI leverages the empirical distribution of out-of-bag prediction errors as the basis for interval construction and has been shown to produce narrower prediction intervals than alternative methods, while still maintaining appropriate coverage in accordance with the desired confidence level [7].

The advantages and limitations of each prediction interval formation method indicate the need for further investigation to evaluate the performance of the three methods under various controlled data structure conditions. This study aims to assess the performance of OOB-PI, QRF, and SC methods in constructing prediction intervals within Random Forest modeling. The evaluation was carried out through a simulation study involving variations in correlation levels among variables, the shape of the mean function, and error distribution. Additionally, the methods were applied to data from the 2023 National Socio-Economic Survey (SUSENAS) for West Java Province. The analysis focused on validating the prediction interval coverage rate and the efficiency of interval width. The findings of this research are expected to contribute to the development of predictive methods that are adaptive to socio-economic data with correlated explanatory variable structures.

2. Methodology

This study adopts a quantitative approach consisting of two main stages: a simulation study and an application to empirical data. The simulation study was conducted to evaluate the performance of prediction interval formation methods under controlled data conditions, while the empirical application aimed to assess the reliability of the methods in the context of socioeconomic data that represent structures with high correlation among explanatory variables.

2.1. *Simulation Study*

The simulation was conducted by generating data from an additive error model as follows:

$$Y = m(\mathbf{X}) + \varepsilon. \tag{2.1}$$

This equation represents the relationship between a predictor vector $\mathbf{X} \in \mathbb{R}^p$ and a continuous response variable Y, with a deterministic component $m(\cdot)$ as the mean function, and ε as the random error term. The mean function used in this study is linear in form:

$$m(\mathbf{X}) = X_1 + X_2. \tag{2.2}$$

The linear form reflects a simple relationship commonly assumed in parametric models. The explanatory variables X were generated from a multivariate normal distribution with an exponential covariance structure, defined as:

$$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \text{with } \Sigma_{ij} = \rho^{|k-j|}; \quad k, j = 1, 2, \cdots, p. \tag{2.3}$$

The correlation parameter ρ was varied across three levels: low ($\rho = 0.1$), medium ($\rho = 0.5$), and high ($\rho = 0.9$). These variations reflect different degrees of multicollinearity typically encountered in empirical data. Furthermore, the error distribution was varied to assess the robustness of each method in handling prediction uncertainty. The types of error distributions used in the simulation are as follows:

(1) Homoskedastic Error: $\varepsilon \sim \mathcal{N}(0, 1)$ with constant variance across all observations.
(2) Heavy tailed Error: $\varepsilon \sim t_3/\sqrt{3}$, which has heavier tails compared to the normal distribution.
(3) Heteroskedastic Error: $\varepsilon \sim \mathcal{N}(0, \sigma_i^2)$ with variance depending on the value of the mean function:

$$\sigma_i^2 = 0.5 + 0.5 \cdot \frac{|m(X_i)|}{\mathbb{E}[|m(\mathbf{X})|]}. \tag{2.4}$$

The $\mathbb{E}[|m(\mathbf{X})|]$ function is computed based on the average absolute value of $m(X)$ from the simulated sample. The training sample size was varied across four levels: $n = 500$, 1000, 2000, and 5000. The test sample size was fixed at $n_0 = 1000$ to ensure consistent performance evaluation. Each simulation scenario was repeated 500 times to obtain performance estimates that are statistically stable and reliable.

The Random Forest model was calibrated through a grid search over the entry and node size parameters before being applied, in order to obtain the configuration with the best predictive performance. This tuning process was carried out using validation data through a 5-fold cross-validation scheme, repeated five times to enhance estimation stability. Each parameter combination was evaluated based on the mean absolute error (MAE) on the validation set. The optimal configuration obtained was then consistently applied across all replication scenarios to ensure that performance comparisons among methods were conducted under equivalent modeling conditions. The number of trees in the Random Forest algorithm ntree was set to 2,000, ensuring that the prediction results were more stable and less affected by random fluctuations across replications.

The performance of the three prediction interval estimation methods was evaluated based on two key metrics after the model was built, the coverage rate, which represents the proportion of actual values Y_i that fall within the marginal prediction interval at the 90% confidence level; and the prediction interval width, which reflects the efficiency of the interval estimation. All simulation analyses were conducted using the R Studio software, utilizing the randomForest, quantregForest, and conformalInference packages.

2.2. *Empirical Data Study*

The data used in this study were obtained from the 2023 National Socio-Economic Survey (SUSENAS) for West Java Province, provided by Statistics Indonesia (BPS). Six regencies/cities in West Java Province with relatively high correlations among explanatory variables were selected as the focus of the empirical study. The selection was made based on an initial analysis of the maximum correlation between explanatory variables, using a threshold criterion of greater than 0.7. The list of selected districts/cities included in the analysis is presented in Table 1.

Table 1. List of districts/cities included in the analysis

City/District	n	**Maximum Correlation**
Cirebon City	462	0.85
Bogor District	964	0.81
Tasikmalaya District	823	0.81
Cimahi City	438	0.79
Indramayu District	685	0.75
Banjar City	452	0.74

The response variable in this study is the household's monthly expenditure on food consumption, measured in Indonesian Rupiah. A total of 24 socioeconomic explanatory variables were used to explain the variation in food expenditure. These explanatory variables are classified into six substantive categories of socioeconomic indicators.

This structure is supported by empirical findings indicating that socioeconomic factors such as education, occupation, and housing conditions significantly influence household food consumption patterns [8]. The set of explanatory variables consists of 19 numeric variables and 5 categorical variables. Detailed information about the explanatory variables is presented in Table 2.

Table 2. Description of Study Variables

Factor	Variable	Type	Code
Non-Food Expenditure	Monthly Housing Expenditure	Numeric	X1
	Monthly Electricity Expenditure	Numeric	X2
	Annual Education Expenditure	Numeric	X3
	Annual Transport and Accommodation Expenditure	Numeric	X4
	Monthly Vehicle Fuel Expenditure	Numeric	X5
	Monthly Household Fuel Expenditure	Numeric	X6
	Monthly Telecommunication Expenditure	Numeric	X7
	Monthly Personal Care and Beauty Expenditure	Numeric	X8
	Annual Clothing Expenditure	Numeric	X9
	Annual Medical Treatment Expenditure	Numeric	X10
	Annual Household Equipment Expenditure	Numeric	X11
	Annual Vehicle Tax Expenditure	Numeric	X12
	Annual Health Insurance Expenditure	Numeric	X13
Housing	Housing Ownership Status	Categorical	X14
	Housing Area Size	Numeric	X15
	Main Water Source	Categorical	X16
Demographics	Household Size	Numeric	X17
	Number of Children Aged 0–4	Numeric	X18
Household Head	Age	Numeric	X19
	Gender	Categorical	X20
	Education Level	Numeric	X21
	Employment Status	Categorical	X22
Geography	Village/Urban Classification	Categorical	X23
Income	Percentage of Working Household Members	Numeric	X24

2.3. *Data Analysis Procedure*

The data analysis procedure in this study consists of four main stages: data exploration and preprocessing, regression modeling using the Random Forest algorithm, construction of prediction intervals, and model performance evaluation. The design was implemented to ensure that the prediction and interval estimation processes are carried out systematically, replicably, and in accordance with proper statistical principles. The details of each stage are explained as follows:

(1) Exploration and Preprocessing
The initial exploration involved descriptive analysis of both the response and explanatory variables. The correlation between explanatory variables was computed using the Spearman coefficient to detect potential multicollinearity. All categorical variables were encoded into dummy variables using the one-hot encoding technique. The response variable was log-transformed to stabilize variance and reduce distributional skewness.

(2) Regression Modeling with Random Forest
The predictive model was built using the Random Forest regression algorithm. Let $T_m(x)$ denote the prediction from the m-th tree and M the total number of trees. The final regression estimator $\hat{f}_{RF}(x)$ is defined as follows:

$$\hat{f}_{RF}(x) = \frac{1}{M}\sum_{m=1}^{M} T_m(x). \tag{2.5}$$

Equation (2.5) show that, x represents a new predictor vector, $T_m(x)$ is the prediction from the m-th tree, and M denotes the total number of trees in the ensemble.

The mtry and nodesize parameters were tuned through a grid search using a 5-fold cross-validation scheme repeated five times. The best configuration was determined based on the minimum Mean Absolute Error (MAE). The final model was built using a fixed number of trees, ntree = 2000, to ensure prediction stability.

(3) Construction of Prediction Intervals

Prediction intervals were constructed using three different approaches: OOB-PI, QRF, and SC. These approaches define the lower and upper bounds of the prediction interval based on different uncertainty estimation strategies.

A consistent confidence level of $1 - \alpha = 90\%$ was applied across all approaches. Accordingly, the prediction interval bounds were constructed using the $\alpha/2$ and $1 - \alpha/2$ quantiles of the prediction distribution from each method. The value of α was chosen by balancing the trade-off between coverage level and interval length efficiency. Further explanation of each approach is provided in the subsequent sections.

The OOB-PI method utilizes the out-of-bag prediction error as the basis for interval construction. The error is computed as:

$$e_i^{OOB} = y_i - \hat{f}_{OOB}(x_i), \tag{2.6}$$

where $\hat{f}_{OOB}(x_i)$ is the prediction from the trees that did not include x_i during training, and y_i is the actual value. The distribution of these errors is then used to calculate the $\alpha/2$ and $1 - \alpha/2$ quantiles, which are added to the point prediction $\hat{f}(x)$ to construct the prediction interval:

$$I(x) = \left[\hat{f}(x) + \hat{q}_{\alpha/2},\ \hat{f}(x) + \hat{q}_{1-\alpha/2}\right]. \tag{2.7}$$

The QRF approach constructs prediction intervals by estimating the conditional cumulative distribution of Y given $X = x$, using weights $w_i(x)$ derived from the decision tree structure:

$$\hat{F}_{Y|X}(y|x) = \sum_{i=1}^{n} w_i(x) \cdot \mathbb{I}(y_i \leq y). \tag{2.8}$$

The indicator function $\mathbb{I}(y_i \leq y)$ takes the value 1 if $y_i \leq y$ and 0 otherwise. Based on this distribution, the τ-th quantile is defined as:

$$\hat{q}_\tau(x) = \inf\{y : \hat{F}_{Y|X}(y|x) \geq \tau\}. \tag{2.9}$$

Thus, the prediction interval is given by:

$$I(x) = \left[\hat{q}_{\alpha/2}(x),\ \hat{q}_{1-\alpha/2}(x)\right]. \tag{2.10}$$

The SC method (Split Conformal Prediction) uses a data-splitting approach, where the training set is divided into two subsets: one for model training and the other for uncertainty calibration. After the model is trained, absolute residuals are computed on the calibration subset as:

$$R_i = \left| y_i - \hat{f}(x_i) \right| . \tag{2.11}$$

The k-th quantile, where $k = \lfloor (n/2+1)(1-\alpha) \rfloor$, from the distribution of R_i, is used as the uncertainty margin d. The symmetric prediction interval around $\hat{f}(x)$ is then constructed as:

$$I(x) = \left[\hat{f}(x) - d,\ \hat{f}(x) + d \right] . \tag{2.12}$$

(4) Performance Evaluation
Performance evaluation was conducted based on two primary metrics: coverage rate and interval width. The coverage rate measures the proportion of actual values y_i that lie within the prediction interval $I(x_i)$, reflecting the validity of the interval with respect to the specified confidence level.

The interval width was computed as the average of $U_i - L_i$, the difference between the upper and lower bounds of the prediction interval, over all test observations. To ensure statistical reliability and mitigate the influence of random variation during model fitting, the entire evaluation procedure was repeated 50 times.

3. Discussion

3.1. *Simulation Study*

The simulation study aimed to evaluate the effect of correlation among explanatory variables on the performance of prediction interval construction. The simulated data in this study were designed to represent a variety of data structure conditions, ranging from ideal settings to those reflecting the complexity of real-world scenarios. These settings take into account the form of the linear mean function, variations in error distribution, and different sample sizes.

The evaluation results in Figure 1 show that OOB-PI and SC were able to maintain an average coverage rate close to the 90% target. Both methods also produced consistent results across simulation replications, including when the correlation among explanatory variables increased. QRF exhibited less stable performance and was more sensitive to changes in correlation, especially when the error distribution was non-homogeneous. As observed in cases of low to moderate correlation, QRF tends to produce overly wide prediction intervals. Coverage rate dropped significantly when the correlation reached high levels. This pattern highlights the limitations of QRF in handling multicollinearity and non-standard error distributions. The stable performance of OOB-PI and SC indicates that both methods are more reliable when applied to data with high correlation and non-normal error distributions. The coverage rate distributions for the three interval estimation methods are presented in Figure 1.

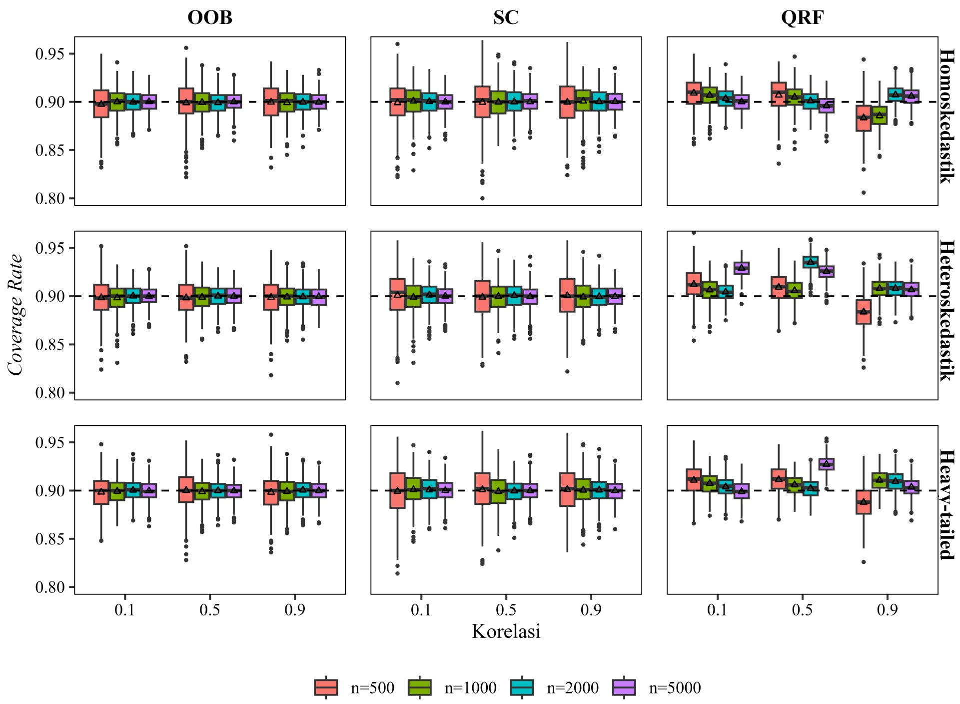

Figure. 1. Distribution of Coverage Rate for the linear mean function across variations in error types, correlation levels, and sample sizes.

The analysis considers not only the aspect of validity but also accounts for the efficiency of the prediction interval width. The interval width reflects the method's ability to convey predictive uncertainty, the narrower the resulting interval, the higher the efficiency of the method.

The analysis in Figure 2 shows that OOB-PI consistently produced the narrowest prediction intervals across all condition combinations. This indicates high efficiency as well as robustness to changes in correlation. SC ranked second, with slightly wider intervals than OOB-PI, but maintained stability across varying correlation levels and error types. In contrast, QRF consistently yielded the widest prediction intervals among the three methods. Although in some replications QRF appeared to produce narrow intervals, it showed a declining trend as correlation increased especially under the heavy tailed error distribution. This pattern aligns with previous findings that QRF experienced undercoverage at high correlation levels. The distribution of interval widths for all three methods under various conditions is shown in Figure 2.

Figure 2 illustrates that the improved efficiency of the QRF method is achieved at the expense of prediction interval validity. OOB-PI and SC maintain a more balanced trade-off between efficiency and reliability, making them more suitable for data with high correlation and non-normal error distributions. From the efficiency

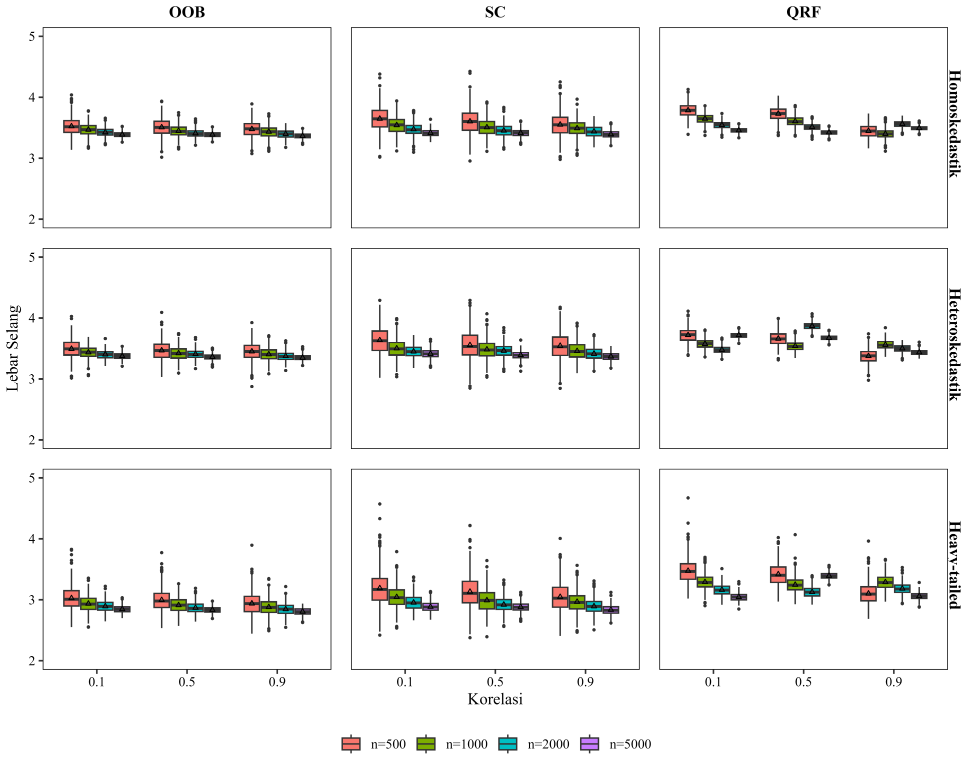

Figure. 2. Distribution of Interval Width for the linear mean function across variations in error types, correlation levels, and sample sizes.

perspective, OOB-PI emerges as the most favorable method due to its ability to construct narrow and stable prediction intervals across all conditions. Its consistent performance demonstrates superiority in maintaining interval compactness and robustness across different correlation levels and error distributions.

The overall evaluation of the two main metrics, namely the coverage rate and the interval width, indicates that the OOB-PI method delivers the best performance among the three methods evaluated. OOB-PI not only maintains prediction validity with a coverage rate close to the target level, but also excels in efficiency by producing the narrowest and most stable prediction intervals. The SC method also demonstrates good and relatively balanced performance, although it is less efficient compared to OOB-PI. The QRF method shows weaknesses in both aspects, with a less stable coverage rate and inefficiently wide intervals.

3.2. *Empirical Data Analysis*

Modeling was conducted using the Random Forest algorithm, configured based on optimal parameter settings obtained through repeated cross-validation. Once the regression model was established, prediction intervals were constructed using three approaches OOB-PI, SC, and QRF with a confidence level of 90%. The performance of these three methods was evaluated quantitatively using two main metrics: the coverage rate, which represents the proportion of actual values captured within the prediction intervals, and interval length, which indicates estimation efficiency. The evaluation results are presented visually in Figure 3 to facilitate interpretation and comparison of performance across methods.

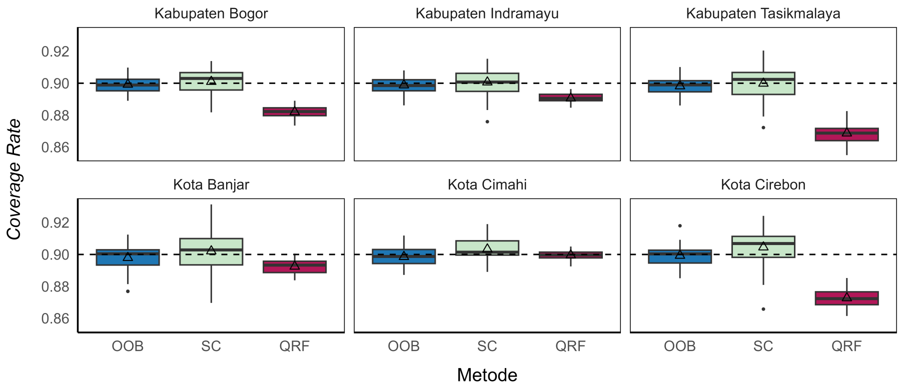

Figure. 3. Distribution of Coverage Rate for three prediction interval estimation methods across six districts/cities in West Java Province based on SUSENAS 2023 data.

Figure 3 shows that the OOB-PI method was able to maintain a coverage level close to the nominal 90% target across all analyzed districts and cities. The coverage stability of OOB-PI was generally good, with relatively low variability, although slight fluctuations were observed in Kota Banjar. SC demonstrated stable and consistent coverage, with a tendency to be conservative, as indicated by median coverage rate values slightly above 90% across all regions. The spread of SC's coverage values tended to be wider in certain areas, such as Kota Banjar and Kota Cirebon, indicating variability in maintaining coverage validity. The QRF method showed a consistent tendency toward undercoverage, with median coverage rate values falling below 90% across all districts and cities. The lowest performance of QRF was observed in Kabupaten Tasikmalaya and Kota Cirebon, indicating that this method is less stable in handling complex correlation structures among explanatory variables in empirical data.

The performance assessment of the method considers not only the coverage validity but also the efficiency of the prediction interval width as a critical aspect. This efficiency reflects the method's ability to convey predictive uncertainty accurately without producing overly wide intervals. A visualization of the interval width distributions for each method across the six districts and cities is presented in Figure 4.

Figure 4 shows that the OOB-PI method consistently produced the narrowest prediction intervals across all analyzed districts and cities, with low variation across replications. This indicates high predictive efficiency and estimation stability. The

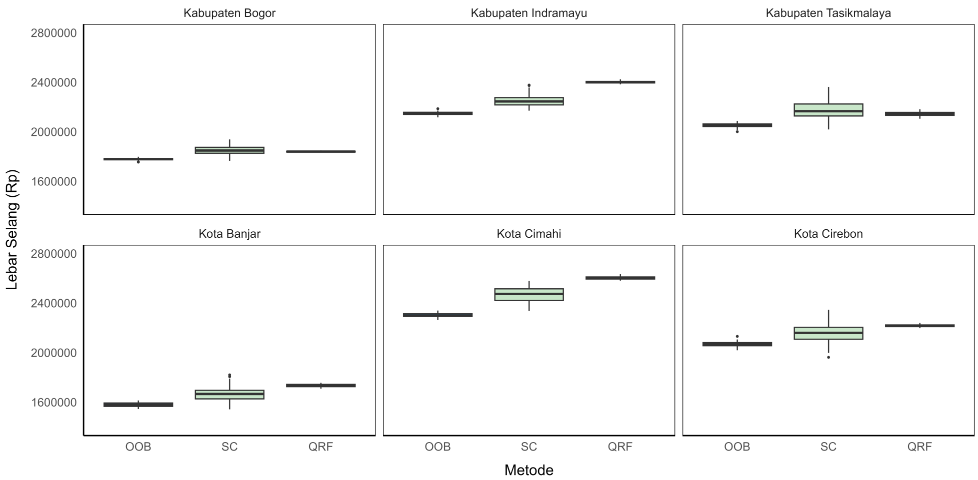

Figure. 4. Distribution of interval widths for three prediction interval estimation methods across six districts/cities in West Java Province, based on SUSENAS 2023 data.

SC method resulted in slightly wider intervals compared to OOB-PI, with higher variability—particularly in Kabupaten Indramayu and Kota Cimahi. In contrast, the QRF method tended to produce wider and less consistent intervals across regions. Although these wider intervals may appear to reflect higher uncertainty, they were not accompanied by improved coverage validity, as shown in previous results. These findings reinforce the indication that the wider intervals produced by QRF do not represent true predictive efficiency.

The overall results of the empirical analysis reinforce the findings from the simulation study. The OOB-PI method proved to be the most balanced and effective approach, as it successfully maintained a coverage rate close to the target while achieving high efficiency in interval width. The SC method can also be considered a reliable alternative due to its strong validity performance, although it was less efficient compared to OOB-PI. In contrast, QRF was found to be suboptimal, as it failed to achieve a proper balance between validity and efficiency.

The performance evaluation of the method in each district/municipality needs to be complemented by a further analysis that directly examines prediction accuracy at the household level. This approach aims to ensure that the prediction intervals produced by each method are not only valid and efficient in aggregate, but also reliable at the level of individual observations. The following visualization is presented in support of the analytical objective, namely to illustrate the characteristics of the prediction intervals based on empirical data from Banjar City. Predictions were performed for all households in the area; however, Figure 5 displays only 20 randomly selected observations as a representative sample to facilitate visual interpretation.

The visualization in Figure 5 compares the performance of the three methods in constructing prediction intervals and capturing actual values at the individual household level. The consistency of the OOB-PI method in producing relatively

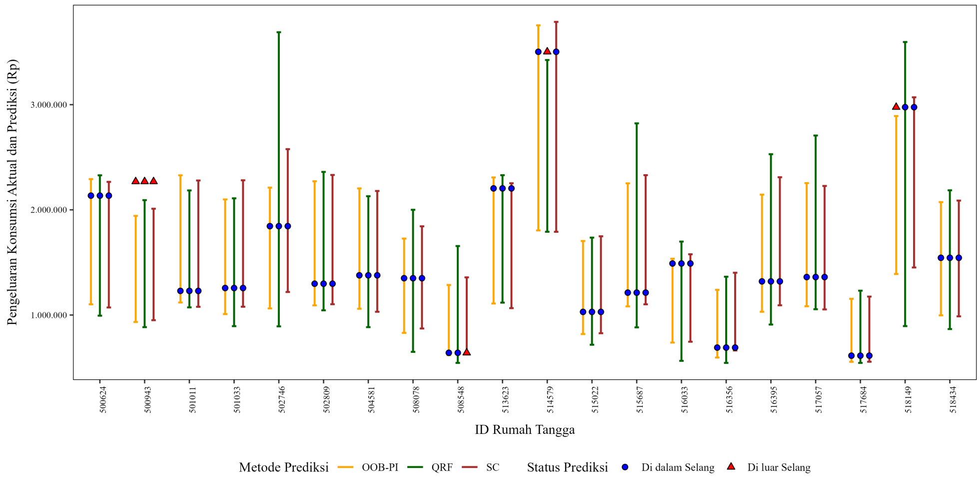

Figure. 5. Visualization of Prediction Intervals for 20 Randomly Selected Households in Kota Banjar Based on SUSENAS 2023 Data.

narrow intervals that cover most of the actual values demonstrates a balance between prediction efficiency and validity. The prediction intervals generated by the SC method also covered almost all actual values. However, the intervals were wider and showed higher variability across households, leading to lower efficiency compared to OOB-PI. The unstable performance of the QRF method, indicated by a number of actual values falling outside the prediction intervals despite the intervals not being particularly narrow, suggests that the weakness of QRF lies not only in efficiency but also in the accuracy of interval placement in capturing the true values.

Overall, this visualization further confirms that OOB-PI is the most reliable method for constructing accurate and efficient prediction intervals, both in simulated and empirical data contexts. The superiority of OOB-PI is reflected in its ability to maintain prediction coverage close to the nominal target while producing narrow and stable intervals. Therefore, OOB-PI can be recommended as a superior approach compared to SC and QRF in generating valid and efficient predictions, particularly in data contexts involving correlated explanatory variables and varying error structures.

4. Conclusion

The analysis results show that the Out-of-Bag Interval Prediction (OOB-PI) method achieved the best performance compared to Split Conformal Prediction (SC) and Quantile Regression Forest (QRF). The OOB-PI method consistently attained a coverage rate close to the 90% target while producing the narrowest prediction intervals, both in simulation and empirical data.

The performance of both OOB-PI and SC remained stable even as the correlation among explanatory variables increased, indicating that these methods are not significantly affected by multicollinearity in maintaining prediction validity and efficiency. In contrast, the QRF method showed a noticeable performance drop under high correlation, characterized by undercoverage and wider intervals, suggesting that it is less suitable for data with strong multicollinearity.

The application to the 2023 West Java SUSENAS data further confirmed findings from the simulation. This dataset reflects a real-world structure with high correlations among socioeconomic variables, as previously described. Across six regions in the study, OOB-PI consistently demonstrated superior performance in capturing actual values with efficient intervals. SC showed acceptable validity but was less efficient in some regions, while QRF failed to maintain reliable coverage and efficiency.

In conclusion, OOB-PI can be considered the most recommended method for constructing prediction intervals in Random Forest models, especially for data with complex structures and strong correlations among explanatory variables.

Bibliography

[1] James, G., Witten, D., Hastie, T., Tibshirani, R., 2021, *An Introduction to Statistical Learning*, Springer Texts in Statistics, Springer, New York.

[2] Gregorich, M., Strohmaier, S., Dunkler, D., Heinze, G., 2021, Regression with Highly Correlated Predictors: Variable Omission is Not the Solution, *International Journal of Environmental Research and Public Health*, Vol. **18**(8): 1 – 15.

[3] Bickel, P., Diggle, P., Fienberg, S., Gather, U., Olkin, I., Zeger, S., 2017, *Springer Series in Statistics*, Springer, Berlin.

[4] Fife, D. A., D'Onofrio, J., 2023, Common, Uncommon, and Novel Applications of Random Forest in Psychological Research, *Behavior Research Methods*, Vol. **55**(5): 2447 – 2466.

[5] Johnson, R. A., 2024, Quantile-Forest: A Python Package for Quantile Regression Forests, *Journal of Open Source Software*, Vol. **9**(93): 5976.

[6] Oliveira, R. I., Orenstein, P., Ramos, T., Romano, J. V., 2024, Split Conformal Prediction and Non-Exchangeable Data, *Proceedings of Machine Learning Research*, Vol. **25**: 1 – 15.

[7] Zhang, H., Zimmerman, J., Nettleton, D., Nordman, D. J., 2020, Random Forest Prediction Intervals, *Statistical Modelling*, Vol. **20**(5): 481 – 500.

[8] Harum, N. S., Aini, M., Risxi, M. A., Kartiasih, F., 2023, Pengaruh Sosial Ekonomi dan Kesehatan terhadap Pengeluaran Konsumsi Pangan Rumah Tangga Provinsi Jawa Tengah Tahun 2020, *Seminar Nasional Official Statistics*, Vol. **1**: 899 – 908.