

# **Published by**

Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 7 (2025) 2314

Journal of the **Nigerian Society** of Physical **Sciences** 

# Optimizing precision farming: enhancing machine learning efficiency with robust regression techniques in high-dimensional data

Nour Hamad Abu Afouna, Majid Khan Majahar Ali 🕩



School of Mathematical Sciences, Universiti Sains Malaysia 11800 USM, Penang, Malaysia

#### Abstract

Smart precision farming leverages IoT, cloud computing, and big data to optimize agricultural productivity, lower costs, and promote sustainability through digitalization and intelligent methodologies. However, it faces challenges such as managing complex variables, addressing multicollinearity, handling outliers, ensuring model robustness, and enhancing accuracy, particularly with small to medium-sized datasets. To overcome these obstacles, reducing retraining time and resolving the complexity issue is essential for improving the machine learning algorithm's performance, scalability, and efficiency, especially when dealing with large or high-dimensional datasets. In a recent study involving 435 drying parameters and 1,914 observations, two machine learning algorithms - Ridge and Lasso - were employed to analyze and compare the impact of two variable selection techniques, specifically the regularization methods Ridge and Lasso, before and after addressing heterogeneity in highly ranked variables (50, 100, 150, 200, 250, 300). Additionally, robust regression methods such as S, M, MM, M-Hampel, M-Huber, M-Tukey, MM-bisquare, MM-Hampel, and MM-Huber were applied. The results demonstrated that the robust methods, when applied to Ridge and Lasso, achieved the highest efficiency, with the smallest values for MAPE, MSE, SSE, and the highest R<sup>2</sup> values, both before and after accounting for heterogeneity. As a result of the study, the best models are the Ridge model with the MM bisquares before heterogeneity, the Ridge model with the MM method after heterogeneity, and the Lasso model with the MM method before heterogeneity and the Lasso model with MM Hampel after heterogeneity.

DOI:10.46481/jnsps.2025.2314

Keywords: Lasso, Ridge, M-estimation, MM-estimation, Robust Regression

Article History:

Received: 15 August 2024

Received in revised form: 18 October 2024 Accepted for publication: 31 October 2024 Available online: 27 December 2024

© 2025 The Author(s). Published by the Nigerian Society of Physical Sciences under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Communicated by: B. J. Falaye

#### 1. Introduction

Precision farming is a crucial development in agricultural operations, completely altering the method by which humans

\*Corresponding author: Tel.: +60 14-954 3405

Email address: majidkhanmajaharali@usm.my (Majid Khan Majahar

Ali 🔘)

approach harvests and resource efficiency. The current procedure employs advanced data analysis and technology to modify the techniques of agriculture to the specific requirements of certain fields and harvests. The application of mathematical models to simulate and predict agricultural results based on enormous amounts of data is critical to precision farming's efficiency. Figure 1 shows how IoT systems work. They collect data such as moisture content, temperature, humidity, and solar radiation, send it to the cloud, and process it. Farmers and users can then view the results on apps to optimize agricultural processes and increase production [1]. However, the accuracy and utility of these models are heavily dependent on the selection of significant variables and their ability to deal with data variances, such as outliers, which may influence results and restrict decision-making.

Machine learning (ML) has transformed variable selection in precision farming by providing robust instruments for analyzing large volumes of data and identifying complex patterns. Ridge regression and Lasso are two significant advances in machine learning for variable selection. Ridge Regression, commonly known as  $L_2$  regularization, stabilizes regression findings by penalizing coefficient size while focusing on multicollinearity and overfitting. Lasso, also known as  $L_1$  regularization, allows for both variable selection and coefficient reduction, which is especially effective for datasets with a large number of associated features.

The main components of precision farming are illustrated in Figure 2, which outlines the structured workflow, including data collection, preprocessing, analysis, testing and validation Despite these developments, the use of irrelevant or weakly described models can be harmful to precision agriculture. Models that fail to appropriately select significant factors and control outliers can cause a number of important issues according to Ref. [2].

- Reduced Predictive Accuracy: Insignificant models might ignore crucial correlations, leading to erroneous forecasts. This might result in a lack of agricultural ideas, affecting productivity and resource efficiency.
- Resource Misallocation: Ineffective models can result in inaccurate recommendations for nutrient, treatment, and water applications. This misallocation not only affects operational efficiency, but also raises expenses and may have a severe influence on the environmental sustainability of seaweed farming.
- Compromised decision-Making: Models that don't account for the intricacies of agricultural data might produce inaccurate results. This can weaken farmers' trust in data-driven suggestions, leading to reluctance to use precision farming methods.
- Risk of Overfitting: Insignificant models might be overfitting to noise or irrelevant characteristics in data, leading to large variance and insufficient generalization to new data. This can reduce the robustness of predictions and make the model not as accurate in various situations.
- Insufficient Data Processing: Models that cannot handle high-dimensional data or outliers might result in higher computing costs and processing delays. This inefficiency may restrict the scalability of precision agricultural technologies.

Beyond these technical challenges, the effective application of precision farming models has significant implications for broader community well-being. Accurate and robust models, as informed by ML frameworks like the one depicted in Figure 3 can lead to substantial improvements according to Ref. [3]:

- Precision farming may improve food security by improving crop yields and resource usage, leading to a more consistent and predictable supply, which is crucial for both local and global food security.
- Improved model precision can minimize agricultural input waste, reduce environmental effects, and improve sustainable farming practices.
- Efficient agricultural approaches based on accurate models can reduce costs and increase profitability for farmers, thereby benefiting the agricultural industry.
- Education and Knowledge Sharing using effective techniques and technology may boost local expertise and creativity in agriculture.

Investigate the association between precision farming and machine learning, particularly the impact of using irrelevant models on agricultural practices and community results. Discuss how complex methodologies like Ridge Regression and Lasso improve model reliability and variable selection, resulting in higher prediction accuracy and decision-making. This discussion aims to illustrate machine learning's important possibility of improving precision farming while additionally supporting sustainable agricultural growth and community well-being.

#### 2. Literature review

Several previous studies have employed robust regression analysis. For example, according to Mukhtar et al. [4, 5] used robust regression methods, including Tukey Bi-Square, Hampel, and Huber, to compare the impact of different regression algorithms (Ridge, Lasso, Elastic Net, Random Forest, Support Vector Machine, and Boosting) on forecasting an efficient model using 30 high-ranking variables. Similarly, according to Ibidoja et al. [6] applied robust regression techniques (M Bi-Square, M Hampel, and M Huber) to evaluate the impact of various regression algorithms (Random Forest, Support Vector Machine, Bagging, and Boosting) on forecasting models for 15, 25, 35, and 45 high-ranking variables. In a subsequent study, according to Ibidoja et al. [7] utilized robust regression methods (S, M, MM, M Bi-Square, M Hampel, and M Huber) to assess the impact of different regression algorithms (Ridge, Lasso, Elastic Net, Random Forest, Support Vector Machine, Bagging, and Boosting) on forecasting models for 45 high-ranking variables, both before and after addressing heterogeneity. The previous studies such as: according to Mukhtar et al. [4, 5] used robust regression methods such as Tukey Bi-Square and M-Hampel in precision farming; however, this research advances the field by using Ridge and Lasso regularization with robust regression approaches. This combination facilitates more effective dealing with high-dimensional data and multicollinearity, distinguishing our technique from previous studies. These studies are summarized in Table 1, which provides an overview of the literature review.

This paper primarily focuses on analyzing and comparing the impact of two variable selection techniques—the regression regularization algorithms Ridge and Lasso—both before and after addressing heterogeneity in highly ranked variables (50, 100, 150, 200, 250, 300). Subsequently, robust regression methods, including S, M, MM, M-Hampel, M-Huber, M-Tukey, MM-bisquare, MM-Hampel, and MM-Huber, will be applied. The study aims to evaluate and compare the performance of these regularization and robust regression algorithms in forecasting an efficient model using metrics such as Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Sum of Squares Error (SSE), and R-square  $R^2$ .

Robust regression is a statistical technique designed to handle outliers and leverage points in regression models, which can otherwise lead to biased estimates when using traditional methods like Ordinary Least Squares (OLS). Outliers can cause data to deviate from normality, making OLS estimators unreliable according to Ref. [8]. Additionally, robust regression techniques can be particularly advantageous in dealing with heteroscedasticity, where the variance of errors varies across observations. Various robust estimators, including robust versions of logistic regression, ridge estimators, Lasso, and elastic net techniques, have been developed to enhance efficiency and accuracy in such scenarios according to Ref. [9]. Robust regression provides a more reliable alternative to traditional regression methods, especially in datasets with outliers and heteroscedasticity, ensuring more accurate and efficient parameter estimation. This paper, applied robust regression techniques to address outliers, including S-estimation, M-estimation, MMestimation, M-bi square, M-Hampel, M-Huber, MM-Hampel, MM-Huber, and MM-Tukey methods.

The application of robust techniques is based on their indicated efficiency in addressing outliers and heterogeneity, especially in large and high-dimensional datasets. These techniques have been efficient at significantly reducing errors such as MAPE, MSE, and SSE while increasing  $R^2$ , particularly after reducing data heterogeneity. Research using these robust methodologies indicates improved model performance for accuracy and stability, finding them appropriate for situations where data variability and outliers may significantly impact predictions. This confirms their utilization in the research to ensure accurate predictions within the field of precision agriculture, where environmental variables often supply noise and variability according to Ref. [10].

Recent studies have increasingly concentrated on robust regression in high-dimensional contexts, specifically in addressing multicollinearity via combining Ridge and Lasso with robust methodologies. Mukhtar *et al.* [4] utilized hybrid models that combine Ridge and robust regression techniques to enhance predictive accuracy in agricultural datasets, whereas according to Rahayu *et al.* [11] employed similar methods for proficiency data, illustrating the effectiveness of MM and Sestimators for handling outliers and improving model stability. These studies demonstrate an increasing trend in using hy-

brid models for improving variable selection and prediction efficiency in complex datasets. Using hybrid techniques improves the theoretical framework of precision agriculture by solving both regional and dataset-specific challenges.

### 3. Methodology

#### 3.1. Flowchart of study

Figure 4 presents the flowchart of methodologies used to achieve the study's objectives. It shows the inclusion of all possible models up to the second order and the testing of various assumptions. Ridge and Lasso machine learning techniques are used to select 50, 100, 150, 200, 250, and 300 parameters because feature selection ranks important variables but does not indicate the number of significant factors. Insignificant parameters are excluded, and parameters showing heterogeneity are subsequently included in the modified model. Following this, validation metrics such as mean absolute percentage error (MAPE), mean squared error (MSE), sum of squared error (SSE), and R-squared (R<sup>2</sup>) are computed. Hybrid models are then developed for before, after, and modified heterogeneity using robust methods and machine learning models. The robust methods applied include the S-estimator, M-estimator, MMestimator, M-bi square, M-Hampel, M-Huber, MM-Hampel, MM-Huber, and MM-Tukey methods. Finally, validation metrics are computed using the 2-sigma and 3 sigma limits to determine the number of outliers.

The current investigation aims to improve on and build upon the research performed by Ibidoja, which used up to 45 variables, by initiating with 50 variables and next increasing the total an increase of 50 to evaluate the effect on the model's efficiency, finally selecting 100, 150, 200, 250, and 300 variables. The selection of these significant variables is motivated by their significant role in improving model efficiency, especially in high-dimensional data environments. Research indicates that including additional high-ranking variables significantly improves the predicted accuracy of robust regression models. This improvement is especially significant after solving the problem of heterogeneity when robust methodologies assist in handling the complexity caused by big variable sets. This work indicates methods for using Ridge and Lasso regularization methods to efficiently address multicollinearity and improve prediction accuracy, as shown by previous studies on precision farming datasets.

The validation measures used in this study mean absolute percentage error (MAPE), mean squared error (MSE), sum of squares error (SSE), and R<sup>2</sup> are crucial for evaluating the accuracy and reliability of the regression models. MAPE gives an accurate measure of prediction error concerning actual values, while MSE and SSE assist as indicators of the extent of inaccuracies in model predictions. R<sup>2</sup>, or the coefficient of determination, measures the amount of variation in the dependent variable that can be predicted from the independent variables. These metrics are commonly utilized in robust regression evaluations and are crucial for evaluating model efficacy, particularly in high-dimensional environments such as precision agri-

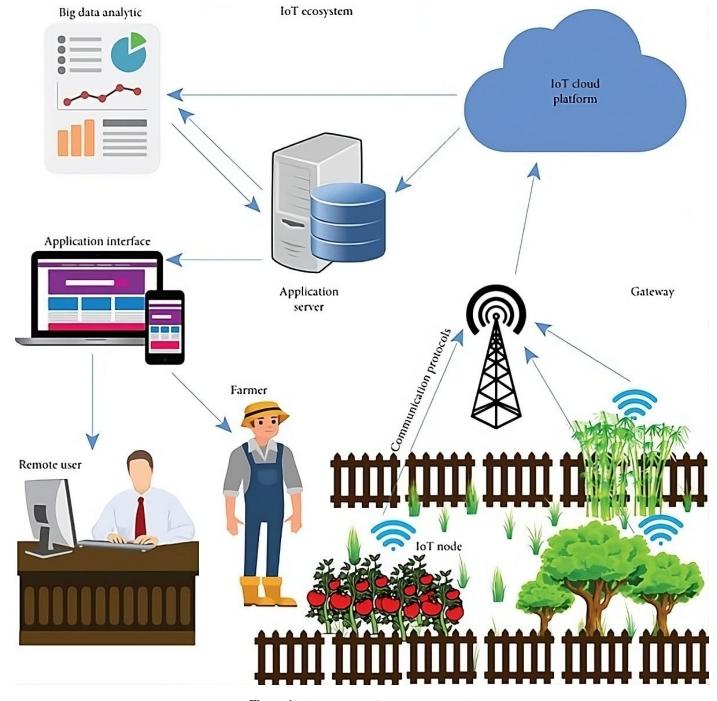


Figure 1: The structure of an IoT system [12].



Figure 2: The main components of precision farming [13].

culture, where reducing prediction error (MAPE, MSE, SSE) and maximizing model fit  $(R^2)$  are critical indicators of efficacy.

#### 3.2. Data description

The experimental drying process data for seaweed was collected using a v-Groove Hybrid Solar Drier (v-GHSD). The dataset comprises 1914 data points, featuring 29 independent variables and one dependent variable. Table 2 provides detailed information on the drying factors, which are critical due to the numerous sensors involved. This study examines the interaction

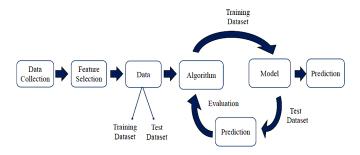


Figure 3: Machine learning blueprint [14].

effects among the variables, resulting in a total of 435 parameters when including second-order interactions. For instance, T2\*T4 denotes the interaction between T2 and T4, T5\*T10 indicates the interaction between T5 and T10, and T7\*T6 represents the interaction between T7 and T6. The dataset includes the main effects of 29 factors and the interaction effects of 406 variables, along with one dependent variable Y according to Ref. [15].

#### 3.3. Multiple Linear Regression (MLR)

Multiple linear regression is a statistical approach used for evaluating the impact of a predictor variable on a response variable. A Multiple Linear Regression (MLR) model is a regression model that includes multiple predictors  $x_1, x_2, x_3, \ldots, x_P$ . The formula for a Multiple Linear Regression (MLR) model is according to Ref. [16]:

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i + \epsilon_i,$$

or equivalently:

$$y_i = \beta_0 + \sum_{i=1}^p \beta_j x_{ij} + \epsilon_i, \tag{1}$$

where  $(y_i, x_i)$  are the values of the response and predictor variables in the *i*-th observation,  $\beta_0, \beta_1, \ldots, \beta_p$  are parameters, and  $\epsilon_i$  are error terms. The error  $\epsilon_i \sim N(0, \sigma^2)$  is a normally distributed random variable and is not mutually correlated according to Ref. [17].

### 4. Ordinary Least Squares (OLS) method

For the estimation of the parameters of the MLR modelin equation (1) using the Ordinary Least Squares (OLS) method, we minimize the sum of squared residuals (SSR). The SSR is given by:

SSR = 
$$\sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2$$
. (2)

From the SSR in equation (2), the OLS estimators for the coefficients can be computed using the formula:

$$\hat{\beta} = (X'X)^{-1}X'y. \tag{3}$$

#### 5. Heterogeneity

Heterogeneity refers to the variation of observations. The variability leads to incompatible forecasts and affects results according to Ref. [18]. Consider multiple linear regression (MLR):

$$Y_i = \beta_0 + \beta_1 T_{i,1} + \beta_2 T_{i,2} + \dots + a_i + \epsilon_i, \tag{4}$$

where  $Y_i$ , i = 1, 2, ..., n is the response value for the i<sup>th</sup> case (moisture content), estimates  $\beta$ 's are the regression coefficients for the predictor variables (drying parameters) T's, using equation (3)  $a_j$  denote heterogeneity, for j = 1, 2, ..., f. That is, the parameters that exhibit heterogeneity and  $\epsilon$  is the random error.

In equation 4 above, if the estimates of the regression equation are computed and a crucial variable is omitted, then the estimate  $\beta$  will be biased and inconsistent. It is also possible that some variables are correlated with the error term, which violates the assumption of regression. According to Ref. [19], the variance inflation factor in multiple regression is used to quantify the level of severity. The coefficient of determination can be written as:

$$R^2 = 1 - \frac{1}{\text{VIF}}.$$

If the  $R^2$  satisfies certain conditions, then the parameter is said to exhibit heterogeneity. According to Ref. [20] stated that the variance inflation factor in multiple regression is used to quantify the level of severity. It can be computed with  $R_i^2$ , where  $R_i^2$  for i = 1, 2, ..., p denote the quantity of determination between the ith variable  $x_i$  in the predictors matrix and the variables not related to it according to Ref. [21].

Let:

$$X^* = \begin{bmatrix} 1 & X_{11} & \dots & X_{1,p-1} \\ 1 & X_{21} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1} & \dots & X_{n,p-1} \end{bmatrix},$$

we can define:

$$X^{*'}X^* = \begin{bmatrix} n & 0' \\ 0 & r_{XX} \end{bmatrix},$$

so that  $r_{XX}$  is the correlation matrix representing the X variables. Since:

$$\sigma^{2}\{\hat{\beta}\} = \sigma^{2} \left( X^{*'} X^{*} \right)^{-1},$$
$$= \sigma^{2} \begin{bmatrix} 1/n & 0' \\ 0 & r_{YY}^{-1} \end{bmatrix},$$

the  $VIF_i$  for i=1,2,...,p-1 stands for the *i*-th diagonal element of  $r_{XX}^{-1}$ . If we show the proof for i=1, the rows and columns of  $r_{XX}$  can be permutated for the remaining *i*. Let:

$$X_{(-1)} = \begin{bmatrix} X_{12} & \dots & X_{1,p-1} \\ X_{22} & \dots & X_{2,p-1} \\ \vdots & \vdots & \vdots \\ X_{n2} & \dots & X_{n,p-1} \end{bmatrix}, \quad X_1 = \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \end{bmatrix},$$

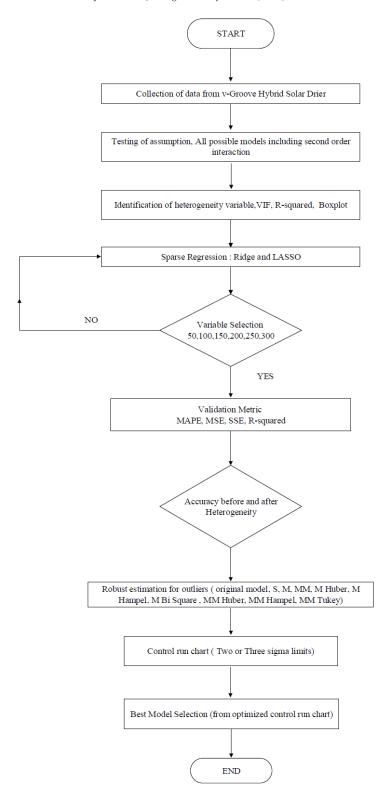


Figure 4: Methodology flowchart.

Using Schur's complement:

$$r_{XX}^{-1}(1,1) = \left(r_{11} - r_{1X_{(-1)}} r_{X_{(-1)}X_{(-1)}}^{-1} r_{X_{(-1)}1}\right)^{-1},$$

$$= \left(1 - \beta_{1X_{(-1)}}' X_{(-1)}' X_{(-1)} \beta_{1X_{(-1)}}\right)^{-1},$$

where  $\beta_{1X_{(-1)}}$  represents the regression coefficient of  $X_1$  on  $X_2, \ldots, X_{p-1}$ , excluding the intercept. For clarity,  $R_1^2$  and  $VIF_1$ 

are written as:

$$R_1^2 = \frac{\text{SSR}}{\text{SSTO}} = \frac{\beta'_{1X_{(-1)}} X'_{(-1)} X_{(-1)} \beta_{1X_{(-1)}}}{1} = \beta'_{1X_{(-1)}} X_{(-1)} \beta_{1X_{(-1)}},$$

and

$$VIF_1 = r_{XX}^{-1}(1,1) = \frac{1}{1 - R_1^2}.$$

#### 6. Regression learning

#### 6.1. Ridge Regression (RR)

Ridge regression is a valuable tool in agricultural research, particularly when dealing with high multicollinearity according to Ref. [22, 23]. The formula for ridge regression includes a penalty term added to the ordinary least squares method to address multicollinearity issues. This penalty term, controlled by a tuning parameter  $\lambda$ , shrinks the regression coefficients toward zero, reducing the impact of multicollinearity while maintaining the model's predictive power according to Ref. [24]. The coefficient of the ridge regression estimate  $\hat{\beta}^{RR}$  minimizes according to Ref. [25]:

$$L^{RR}(\beta) = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2,$$
 (5)

$$= SSR + \lambda \sum_{j=1}^{p} \beta_j^2,$$

where  $\lambda \geq 0$  is the regularization parameter controlling the shrinkage. Ridge regression estimates coefficients that make the SSR small and fit the data well. In equation (5) The term  $\lambda \sum_{j=1}^{p} \beta_{j}^{2}$  is the shrinkage penalty according to Ref. [21].

#### 6.2. Lasso Regression (LR)

Lasso regression, or Least Absolute Shrinkage and Selection Operator regression, is a type of linear regression that includes a regularization term for perform feature selection and prediction according to Ref. [26, 27]. Lasso regression eliminates irrelevant data, offering an excellent fit for prediction tasks without overfitting according to Ref. [27]. Lasso regularization also provides built-in feature selection by allowing coefficients to shrink towards zero according to Ref. [28]. The coefficient of the Lasso regression estimate  $\hat{\beta}^{Lasso}$  minimizes according to Ref. [27]:

$$L^{LR}(\beta) = \sum_{i=1}^{n} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{6}$$

$$= SSR + \lambda \sum_{j=1}^{p} |\beta_j|.$$

In equation (6) the Lasso utilizes an  $L_1$  penalty instead of an  $L_2$  penalty and Lasso will shrink the estimates of the coefficients towards zero according to Ref. [28].

#### 6.3. Robust regression

Robust regression is a technique used when the residuals do not follow a normal distribution or when outliers influence the model. It is a crucial tool for analyzing data affected by outliers, ensuring that the resulting models remain resilient against such outliers according to Ref. [29]. In this study, we applied robust regression techniques to address outliers, including S-estimation, M-estimation, MM-estimation, MM-bi square, MM-Hampel, MM-Huber, M-Hampel, M-Huber, and M-Tukey methods.

#### 7. Robust regression estimations

#### 7.1. S-Estimation

The robust regression model using S-estimation can eliminate up to 50% of outliers, resulting in a positive impact on other data according to Ref. [30]. The S-estimator is defined by:

$$\hat{\beta}_S = \min_{\beta} \hat{\sigma}_S(e_1, e_2, \dots, e_n),$$

where  $\hat{\sigma}_S$  is determined by the minimum scale of the robust estimation according to Ref. [31, 32]. The S-estimator minimizes the following:

$$\min \sum_{i=1}^{n} \rho \left( \frac{y_i - \sum_{j=0}^{p} \beta x_{ij}}{\hat{\sigma}_S} \right),$$

where  $\hat{\sigma}_S$  is computed as:

$$\hat{\sigma}_S = \begin{cases} \frac{\text{median}[e_i - \text{median}(e_i)]}{0.6745} & \text{if iteration} = 1\\ \sqrt{\frac{1}{nK} \sum_{i=1}^n w_i e_i^2}. & K = 0.199 & \text{if iteration} > 1 \end{cases}$$

The solution is found by differentiating with respect to  $\beta$ , resulting in:

$$\sum_{i=1}^{n} x_{ij} \cdot \rho' \left( \frac{y_i - \sum_{j=0}^{p} \beta x_{ij}}{\hat{\sigma}_S} \right) = 0, \quad j = 0, 1, 2, \dots, p$$

where p is a number of independent variables.  $\psi$  is a function that represents the derivative of  $\rho$ :

$$\psi(u_i) = \rho'(u_i) = \begin{cases} u_i \left[1 - \left(\frac{u_i}{c}\right)^2\right]^2, & \text{if } |u_i| \le c\\ 0 & \text{if } |u_i| > c \end{cases}$$

where c is a tunning constant.

#### 7.2. M-Estimation method

M-estimation is a robust regression method where the principle is to minimize the residual function. The M-estimator is defined according to Ref. [33]:

$$\hat{\beta}_{M} = \min_{\beta} \sum_{i=1}^{n} \rho \left( y_{i} - \sum_{j=0}^{p} \beta_{j} x'_{ij} \right).$$

## Standardized Residuals for Original Data

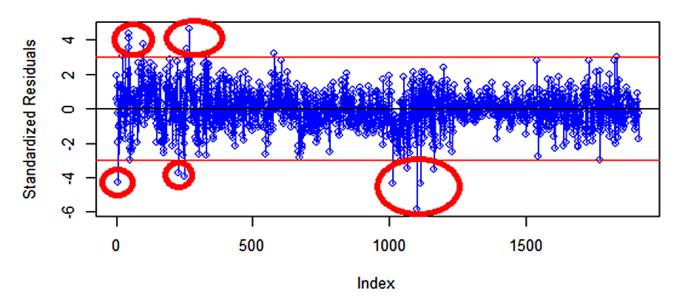


Figure 5: Scatter plot of standardized residuals for original data.

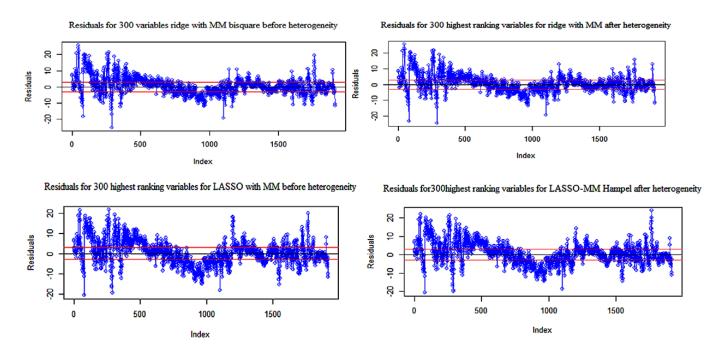


Figure 6: The residuals for the best model with a robust method for 300 high-ranking variables using a 3-sigma limit.

The objective is to solve:

$$\min_{\beta} \sum_{i=1}^{n} \rho(u_i) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}_{\text{MAD}}}\right) = \min_{\beta} \sum_{i=1}^{n} \rho\left(\frac{y_i - \sum_{j=0}^{p} \beta_j x'_{ij}}{\hat{\sigma}_{\text{MAD}}}\right),$$

where MAD is the median absolute deviation and  $\hat{\sigma}_{\mathrm{MAD}}$  is the

scaled median absolute deviation, computed as:

$$\hat{\sigma}_{\text{MAD}} = \frac{\text{median} |e_i - \text{median}(e_i)|}{0.6745} = \frac{\text{MAD}}{0.6745}$$

In this method:

- $\hat{\beta}_M$  is the estimated beta of the M-estimation.
- $\rho$  represents the weighted residuals.
- $e_i$  is the *i*-th residual.
- The function  $\rho$  determines the robustness of the estimator. Refer to Table 3 for detailed formulas.

#### 7.3. MM-Estimation

The MM-estimation procedure involves two steps. First, the regression parameters are estimated using S-estimation, which minimizes the scale of the residuals. Then, M-estimation is applied according to Ref. [34]. The MM-estimator is defined as:

$$\hat{\beta}_{MM} = \sum_{i=1}^{n} \rho_{1}' \left( \frac{y_{i} - \sum_{j=0}^{p} \beta_{j} x_{ij}}{\text{SD}_{MM}} \right) x_{ij} = 0,$$

where  $SD_{MM}$  is the standard deviation derived from the residuals of the S-estimation. The function  $\rho$  is based on the methods of Tukey, Hampel, and Huber. Detailed formulas for these robust regression methods are provided in Table 4.

#### 7.4. Metrics for model comparison

Metrics for model comparison are essential to assessing the suitability of a model. These metrics are crucial for determining whether a model is adequate. Common metrics include Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Sum Squares of Error (SSE), and R-squared (R²). These metrics measure the accuracy of the regression model in predicting the dependent variable within an acceptable range of accuracies. Model comparisons are typically made by considering the lowest MAPE, MSE, and SSE values, and the highest R² value [35]. The equations for these metrics are presented in Table 4, where:

- $Y_i$  is the actual value,
- $\bar{Y}$  is the mean value,
- $\hat{Y}_i$  is the predicted (estimated) value,
- n is the number of observations.

#### 8. Results and discussion

Based on Figure 5, the scatter plot of standardized residuals shows horizontal red lines at -3 and +3, which represent the threshold for residuals that are three standard deviations from the mean. Residuals outside this range (either below -3 or above 3) are flagged as potential outliers, suggesting that the model may not be fitting these data points well. The plot reveals several points that exceed the -3 to +3 range, indicating the presence of outliers. These outliers could have a significant impact on the model's predictions and may potentially distort the overall results.

Table 5 and Table 6 present metrics for model comparison for Ridge and Lasso regression using robust methods for 50,

100, 150, 200, 250, and 300 high-ranking variables, both before and after addressing heterogeneity. The evaluation metrics include Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Sum of Squares of Error (SSE), and R-squared (R²). The results are displayed for varying numbers of high-ranking variables: 50, 100, 150, 200, 250, and 300. To assess prediction accuracy, the predicted responses are compared to the actual responses for each regression model using validation methods. For all high-ranking variables, MAPE, MSE, and SSE decrease while R² increases as the number of high-ranking variables rises for both Ridge and Lasso across all robust methods, including M-estimation, S-estimation, MM-estimation, MM-bi square, MM-Hampel, MM-Huber, M-Hampel, M-Huber, and M-Tukey methods.

In Ridge regression, the MM Hampel method significantly outperformed other techniques for 50 high-ranking variables before addressing heterogeneity. The performance metrics for MM Hampel included a Mean Absolute Percentage Error (MAPE) of 8.801508, a Mean Squared Error (MSE) of 45.81388, a Sum of Squares of Error (SSE) of 87,687.77, and an R-squared (R<sup>2</sup>) of 0.8325343. However, after heterogeneity was accounted for, the M method emerged as the best performer, with a MAPE of 9.974874, an MSE of 48.1354, an SSE of 92,131.16, and an R-squared (R<sup>2</sup>) of 0.8240484. For 100 high-ranking variables, the MM method delivered significantly better results before addressing heterogeneity compared to other methods. The performance metrics for MM were: Mean Absolute Percentage Error (MAPE) of 7.889334, Mean Squared Error (MSE) of 34.66241, Sum of Squares of Error (SSE) of 66,343.86, and an R-squared (R<sup>2</sup>) of 0.8732968. Even after accounting for heterogeneity, the MM method continued to demonstrate superior performance, with a MAPE of 8.973277, an MSE of 39.5754, an SSE of 75,747.32, and an R-squared (R<sup>2</sup>) of 0.8553381. For 150 high-ranking variables, the MM method delivered significantly better results before addressing heterogeneity compared to other methods. The performance metrics for MM were: Mean Absolute Percentage Error (MAPE) of 7.562458, Mean Squared Error (MSE) of 34.19317, Sum of Squares of Error (SSE) of 65445.73, and an R-squared (R<sup>2</sup>) of 0.8750121. Even after accounting for heterogeneity, the MM method continued to demonstrate superior performance, with a MAPE of 8.273413, an MSE of 36.40819, an SSE of 69685.28, and an R-squared (R<sup>2</sup>) of 0.8669154. For 200 high-ranking variables, the M-Tukey method outperformed other methods before addressing heterogeneity. The performance metrics for M-Tukey included a Mean Absolute Percentage Error (MAPE) of 7.33001, a Mean Squared Error (MSE) of 33.59276, a Sum of Squares of Error (SSE) of 64,296.54, and an R-squared (R<sup>2</sup>) of 0.8772068. After accounting for heterogeneity, the MM Hampel method proved to be superior, achieving a MAPE of 8.010758, an MSE of 34.84692, an SSE of 66,697, and an R-squared (R<sup>2</sup>) of 0.8726224. For 250 high-ranking variables, the M-Tukey method significantly outperformed other approaches before addressing heterogeneity. The performance metrics for M-Tukey were a Mean Absolute Percentage Error (MAPE) of 7.310033, a Mean Squared Error (MSE) of 30.83707, a Sum of Squares of Error (SSE)

of 59,022.15, and an R-squared (R<sup>2</sup>) of 0.8872798. Even after accounting for heterogeneity, the M-Tukey method remained superior, achieving a MAPE of 8.098266, an MSE of 34.7557, an SSE of 66,522.41, and an R-squared (R<sup>2</sup>) of 0.8729558. For 300 high-ranking variables, the MM-bisquare method achieved notably better results than other methods before addressing heterogeneity. The performance metrics for MM-bisquare were a Mean Absolute Percentage Error (MAPE) of 6.826407, a Mean Squared Error (MSE) of 28.0242, a Sum of Squares of Error (SSE) of 53,638.32, and an R-squared (R<sup>2</sup>) of 0.8975618. After accounting for heterogeneity, the MM method still demonstrated strong performance, with a MAPE of 6.962468, an MSE of 29.09346, an SSE of 55,684.88, and an R-squared (R<sup>2</sup>) of 0.8936533.

In Lasso regression, for 50 high-ranking variables, the MM Huber method achieved notably better results before addressing heterogeneity compared to other methods. The performance metrics for MM Huber were: Mean Absolute Percentage Error (MAPE) of 8.968910, Mean Squared Error (MSE) of 44.51771, Sum of Squares of Error (SSE) of 85,206.9, and an R-squared  $(R^2)$  of 0.8372723. After accounting for heterogeneity, the MM bi-square method demonstrated superior performance with metrics of MAPE of 9.210072, MSE of 43.5384, SSE of 83,332.51, and R-squared (R<sup>2</sup>) of 0.840852. For 100 high-ranking variables, the MM Hampel method showed significantly better results before addressing heterogeneity compared to other methods. The performance metrics for MM Hampel were: MAPE of 8.800533, MSE of 45.13168, SSE of 86,382.03, and R-squared (R<sup>2</sup>) of 0.835028. After addressing heterogeneity, the MM bisquare method excelled with metrics of MAPE of 8.997942, MSE of 43.99379, SSE of 84,204.11, and R-squared (R<sup>2</sup>) of 0.8391874. For 150 high-ranking variables, the MM method achieved significantly better results before heterogeneity compared to other methods. The performance metrics for MM were: MAPE of 8.457516, MSE of 39.59102, SSE of 75,777.2, and R-squared (R<sup>2</sup>) of 0.8552811. After addressing heterogeneity, the MM Huber method demonstrated superior performance with metrics of MAPE of 8.482080, MSE of 38.69286, SSE of 74,058.14, and R-squared (R<sup>2</sup>) of 0.8585641. For 200 high-ranking variables, the MM Hampel method achieved notably better results before addressing heterogeneity compared to other methods. The performance metrics for MM Hampel were: MAPE of 8.333713, MSE of 37.74151, SSE of 72,237.25, and R-squared ( $R^2$ ) of 0.8620417. After accounting for heterogeneity, the MM method showed superior performance with metrics of MAPE of 8.468777, MSE of 39.51445, SSE of 75,630.66, and R-squared (R<sup>2</sup>) of 0.8555609. For 250 high-ranking variables, the M-Tukey method delivered significantly better results before addressing heterogeneity compared to other methods. The performance metrics for M-Tukey were: MAPE of 8.358520, MSE of 37.88064, SSE of 72,503.54, and R-squared ( $\mathbb{R}^2$ ) of 0.8615331. After addressing heterogeneity, the M method demonstrated superior performance with metrics of MAPE of 8.379303, MSE of 38.59054, SSE of 73,862.29, and R-squared (R<sup>2</sup>) of 0.8589381. For 300 high-ranking variables, the MM method achieved significantly better results before addressing heterogeneity compared to other methods.

The performance metrics for MM were: MAPE of 8.123120, MSE of 37.28122, SSE of 71,356.25, and R-squared ( $R^2$ ) of 0.8637242. After addressing heterogeneity, the MM Hampel method showed superior performance with metrics of MAPE of 8.197567, MSE of 37.64827, SSE of 72,058.79, and R-squared (R<sup>2</sup>) of 0.8623825. According to Ref. [36] suggests that model comparisons should be made based on the lowest values of RMSE, rRMSE, MAPE, MAD, and AIC, as well as the highest values of R<sup>2</sup> and adjusted R<sup>2</sup>. In this study, achieving accuracy and precision was determined by finding the models with the lowest MAPE, MSE, and SSE values, and the highest R<sup>2</sup> values. The R<sup>2</sup>, MAPE, MSE, and SSE are crucial metrics in regression analysis since they are specially formulated for evaluating model performance for continuous numerical data. R<sup>2</sup> measures the percentage of variation in the dependent variable explained by the independent variables, making it a vital statistic to evaluate the accuracy of a regression model's fit to the data. MAPE provides an accurate, obvious, percentage-based evaluation of model error, especially helpful in forecasting applications. MSE and SSE evaluate the extent of prediction errors by squaring the differences between actual and predicted values, giving them sensitivity to significant variations, which is crucial for ensuring that models avoid ignoring significant errors. These measures have the purpose of evaluating the accuracy of continuous predictions, in contrast to classification metrics like AUC (Area Under the Curve) or the F-score, which examine the efficacy of models predicting categorical outcomes. The AUC is irrelevant in regression assignments since it evaluates the relationship between true positive and false positive rates in binary classification, while the F-score heals the two factors, which are not relevant to continuous data. Consequently, regression measures focus on the minimization of the difference between observed and predicted continuous values, providing them more suitable than metrics produced for classification.

Table 7 presents metrics for model comparison between the original for Ridge and Lasso regression with the best model of robust methods for 50, 100, 150, 200, 250, and 300 highranking variables, both before and after addressing heterogeneity. The evaluation metrics include Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Sum of Squares of Error (SSE), and R-squared  $(R^2)$ . In the ridge regression for the 300 high-ranking variables before heterogeneity, the Original Model shows a MAPE of 7.063511 and an  $R^2$  of 0.9054084. In contrast, the MM Bisquare method significantly improves the MAPE to 6.826407, with an R<sup>2</sup> of 0.8957618. This suggests that the MM Bisquare method enhances prediction accuracy with only a minimal impact on the model fit, making it an excellent choice for this high-ranking variable set (50,100,150,200,250). After heterogeneity, the Original Model has a MAPE of 7.019137 and an R<sup>2</sup> of 0.9059521. The MM method improves the MAPE slightly to 6.962468, although with a marginally lower R<sup>2</sup> of 0.8936533. This indicates that while the MM method offers a modest improvement in accuracy, it comes with a slight reduction in the model fit, following a similar pattern observed before heterogeneity was addressed. The best model for the before heterogeneity of Ridge with MM bisquere and the after heterogeneity of Ridge with MM method

is shown in Figure 6 Previous studies have shown that MM estimation, which combines high breakdown point estimation (Sestimation) with M-estimation, outperforms S-estimation alone according to Ref. [33]. Additionally, research According to Ref. [34] introduced the Robust Ridge Regression estimator based on MM (RMM). This RMM, which incorporates a robust MM estimator, was found to outperform other methods across various disturbance distributions and levels of multicollinearity. This suggests that RMM is the most effective estimator for handling outliers and multicollinearity within the context of ridge regression. Similarly, according to Jeremia et al. [19] observed that addressing multicollinearity and outliers solely with Robust regression or Ridge regression is insufficient. Instead, Robust Ridge regression, which merges Robust regression with Ridge regression, effectively addresses both issues simultaneously. Their results demonstrated that integrating Robust regression with generalized Ridge regression results in a lower Mean Squared Error (MSE) compared to using Ridge regression alone. Since a lower MSE indicates a better estimator, it can be concluded that combining generalized Ridge regression with Robust regression is superior to using Ridge regression on its own.

Table 8. shows the comparison of the number and percentage of outliers exceeding 2-sigma and 3-sigma limits for Ridge and Lasso with robust regression, both before and after, for 50, 100, 150, 200, 250, and 300 high-ranking variables. For 2-sigma limits, the hybrid Ridge model with the Hampel estimator before heterogeneity showed the fewest outliers, totaling 74, which represents a 21% reduction compared to the original model. For 300 high-ranking variables, the hybrid Lasso model with the Hampel estimator after heterogeneity had the fewest outliers at 83, marking a 9% reduction compared to the original model. For 3-sigma limits, the hybrid Lasso model with the S estimator before heterogeneity had the smallest number of outliers at 17, reflecting a 26% reduction compared to the original model. After heterogeneity, the hybrid Lasso model with the S estimator had the fewest outliers at 16, also showing a 26% reduction compared to the original model. Figure 6. shows the residuals for the best model for Ridge and Lasso with the robust method for 300 high-ranking variables using a 3-sigma limit for before and after heterogeneity. The residual plots for Ridge and Lasso models, before and after accounting for heterogeneity, provide valuable insights into model performance. Before adjusting for heterogeneity, the residuals display noticeable patterns and varying spread, suggesting potential issues with model fit. After correcting for heterogeneity using MM and Hampel estimators, the residuals are more evenly distributed around zero, indicating improved model accuracy. However, some residual patterns and outliers persist, highlighting the need for further refinement, possibly by including additional variables or tuning the models. Overall, the adjustments for heterogeneity significantly enhance the model's reliability, though more work may be needed to fully address the remaining issues.

Table 9 presents a comparison between the results of this study and previous studies. Mukhtar *et al.* [4] highlighted challenges related to irrelevant variables and outliers across 30 high-

ranking variables, with the best hybrid model being Random Forest combined with Hampel, yielding a MAPE of 9.160917 and  $R^2$  of 0.838757. In another study, Mukhtar et al. [5] discussed the primary challenges of multicollinearity and outliers for the same set of variables, where the Lasso model with Hampel showed a MAPE of 9.17489 and  $R^2$  of 0.8230399. Ibidoja et al. [6] addressed outlier challenges for 15, 25, 35, and 45 high-ranking variables, with the Bagging model using M Bisquare for 45 variables achieving a MAPE of 8.151903 and  $R^2$  of 0.876975. According to Ibidoja et al. [7], challenges such as heterogeneity, multicollinearity, and outliers were addressed, and for 45 variables, the best hybrid model was Random Forest with Hampel (before heterogeneity), with a MAPE of 2.12589 and  $R^2$  of 0.9732063. After accounting for heterogeneity, Boosting with M Hampel gave a MAPE of 8.228835 and  $R^2$  of 0.5510545. Further, Ibidoja et al. [38] focused on heterogeneity and outliers, with Lasso using M Bi-square (single parameter added) for 45 variables achieving a MAPE of 8.149872 and  $R^2$  of 0.8845778. In this study, challenges involving heterogeneity and outliers were examined for 50, 100, 150, 200, 250, and 300 high-ranking variables. The Ridge model with MM Bi-square before heterogeneity for 300 variables showed the lowest MAPE (6.826407) and highest  $R^2$ (0.897561), followed by Ridge with MM after heterogeneity  $(MAPE = 6.962468, R^2 = 0.8936533), Lasso with MM before$ heterogeneity (MAPE = 8.123120,  $R^2 = 0.863724$ ), and Lasso with MM Hampel after heterogeneity (MAPE = 8.197567,  $R^2$ = 0.862382). Across 300 variables, this study demonstrated the best overall performance, with the lowest MAPE and highest  $R^2$  values.

Robust approaches are used in statistical modeling for dealing with challenges such as outliers. In research, outliers and variability in distributions are prevalent, and traditional regression models such as Ordinary Least Squares (OLS) can demonstrate significant sensitivity to these variables, resulting in incorrect or inefficient results. Robust methodologies, such Lasso and Ridge, supplemented with outlier-resistant approaches such S, M, MM, MM Bi-square, MM Hampel, MM Huber, M Hampel, M Huber and M Tukey, are specifically designed to solve these challenges by minimizing the impact of outliers and handling complex data structures more efficiently. These estimators effectively handle extreme values by changing them with more accurate estimates, so keeping the model's ability to generalize without bias from outliers. Methods such as Lasso and Ridge minimize multicollinearity by regularization, which penalizes significant coefficients and improves model stability among correlated variables. These effective techniques are crucial for improving prediction accuracy and providing more reliable ideas, particularly when the data is noisy or displays irregular patterns. Consequently, robust methodologies are crucial for constructing models capable of handling the complicated nature of real-world data without reducing performance.

Table 1: Summary of literature review.

Authors	Variables	Objectives	Evaluation Metrics	Results
According to Almetwally et.al [30]	The parameters consist of 3 and 6 variables, without any interactions.	Compare six estimation methods in robust regression, including M. Hampel, M. Bisquare, M. Huber, Sestimation, MM(S)-estimation, and MM estimation methods to determine the best estimation methods for regression models.	The estimation method uses bias and mean squared error (MSE) as its criteria.	The best three methods identified were M-estimation, MM(S)-estimation, and MM estimation methods.
According to Tirink et.al., [35]	l response variable 6 predictor variables Without Interaction	The study aims to compare the performance of robust estimators like the M (Huber and Tukey bi square) estimator, MM estimator, and LTS estimator in linear regression to estimate the optimum model in the presence of outliers in the dataset.	comparison criteria such as MSE, RMSE, rRMSE, MAPE, MAD, $R^2$ , $R^2_{adj}$ , and AIC	concluding that the M-Huber estimator showed more reliable
According to Singgih et al.,[20]	1 response variable 2 predictor variables Without Interaction	comparing M estimation, S estimation, and MM estimation to determine the best estimation method for robust regression.	Using residual standard error and adjusted r-square values	The robust regression model with S estimation was concluded to be the best model.
According to Mukhtar et al., [4]	1 response variable 29 predictor variables A total of 435 models with Interaction Using 30 variables	utilizes M-robust regression methods like M-bi square, M-Hampel, and M-Huber to handle outliers effectively, recommending random forest and M-Hampel models for efficient validation and analysis of big data.	validation metrics like sum square of error (SSE), mean absolute error (MAE), mean squared error (RMSE), mean absolute percentage error (MAPE), and R-Square	The study recommended that the best models for analyzing and comparing big data were random forest and M-Hampel due to their efficiency and minimal issues in validation.
According to Mukhtar et al., [5]	1 response variable 29 predictor variables A total of 435 models with Interaction Using 30 variables	compared the impact of three variable selection techniques in regularization regression algorithms, followed by robust regression using Tukey Bi-Square, Hampel, and Huber methods.	performance metrics such as MAE, RMSE, MAPE, SSE, R-square, and R-square Adjusted	The Lasso-Hampel method outperformed others

I	Authors	Variables	Objectives	Evaluation Metrics	Results
I	According	1 response variable 3 pre-	evaluate the performance of the	Using the criteria of es-	The proposed redescending M-
	to Khan	dictor variables Without	proposed redescending M-estimator	timation method mean	estimator in the paper provides
	et.al . [37]	Interaction	across different data generation scenarios, comparing it with existing redescending M-estimators like Huber, Tukey Biweight, Hampel, and Andrew-Sign function.	squared error (MSE)	highly robust and efficient estimates, performing almost as efficiently as ordinary least squares for normal data and highly resistant to outliers in contaminated datasets.
	According to Rahayu et al., [11]	1 response variable 3 predictor variables Without Interaction	comparing the M, MM, and S estimators in robust regression analysis on Indonesian literacy index data from 2018 to determine the most effective estimation method for estimating regression coefficients	Using residual standard error and adjusted resquare values	The S-estimator and MM-estimator were identified as the best methods due to having the smallest Residual Standard Error (RSE) values
13	According to Ibidojaa et al., [6]	I response variable 29 predictor variables A total of 435 models with Interaction Using 15,25,35,45 most significant variables	using machine learning algorithms like random forest, support vector machine, bagging, and boosting to select the significant parameters and then applying robust methods such as M Bi-Square, M Hampel, and M. Huber to develop the hybrid model for improved prediction accuracy and outlier reduction.	percentage of outliers outside the 2-sigma and 3-sigma	showed a significant reduction in outliers and better prediction accuracy for contaminated seaweed big data, with bagging M Bi-square performing the best.
	According to Ibidoja et al., [7]	1 response variable 29 predictor variables A total of 435 models with Interaction Using 15,25,35,45 most significant variables	The hybrid models are developed using robust methods such as M Bi-Square, M Hampel, M Huber, MM, and S, with validation metrics computed using 3-sigma limits to identify outliers.	percentage of outliers outside the 2-sigma and 3-sigma	The hybrid models, particularly random forest M Hampel and boosting M Hampel, were found to be the best for before and after heterogeneity, respectively.
I	According to Ibidoja et al., [38]	1 response variable 29 predictor variables A total of 435 models with Interaction Using 15,25,35,45 most significant variables	evaluates the proposed model's performance using ridge, LASSO, and Elastic net models, along with robust estimations like M Bi-Square, M Hampel, M Huber, MM, and S.	Evaluation metrics like MAPE, MSE, and R <sup>2</sup>	The hybrid model of sparse regression with 45 high-ranking variables and a 2-sigma limit effectively reduced outliers, outperforming other methods. LASSO BH shows the best performance with 45 high-ranking variables

Table 2: Representation of factors.

Symbols	Factors	Meanings
Y	Dependent	Moisture Content
H1	Independent	Relative Humidity (Ambient)
H5	Independent	Relative Humidity (Chamber)
PY	Independent	Solar Radiation
T1	Independent	Temperature (°C) Ambient
T2, T3, T4	Independent	Temperature (°C) Prior to Entering the Solar Col-
	•	lector
T5	Independent	Temperature (°C) Opposite the Down V-Groove
	•	(Solar Collector)
T6, T8	Independent	Temperature (°C) in Front of the Up V-Groove (So-
	•	lar Collector)
T7, T14, T15, T16, T21, T22	Independent	Temperature(°C) for the Solar Collector
T9, T10, T11, T12	Independent	Temperature (°C) Behind the Inside Chamber
T13, T17, T19	Independent	Temperature (°C) in Front of the Inside Chamber
T23, T25, T26, T27, T28, T29	Independent	Temperature (°C) from the Solar Collector to the
	•	Chamber

Table 3: Formulas for robust regression M, MM Method [5].

Methods	Objective Function	
Bisquare (Tukey's Bisquare)	$\rho(u_i) = \begin{cases} \frac{c^2}{6} \left[ 1 - \left( 1 - \left( \frac{u_i}{c} \right)^2 \right)^3 \right], & \text{if } i \\ \frac{c^2}{6}, & \text{if } i \end{cases}$	$ u_i  \le c$ $ u_i  > c$
	where $c = 4.685$ .	
	$\rho(u_i) = \begin{cases} \frac{u_i^2}{2}, \\ a u_i  - \frac{u_i^2}{2}, \\ \frac{-a}{2(c-b)}(c - u_i)^2 + \frac{a}{2}(b + c - u_i)^2 + \frac{a}{2}(b + c - u_i)^2 \end{cases}$ where $a = 2, b = 4, c = 8$	if $0 <  u_i  < a$
Hampel	$\rho(u_i) = \left\{ a u_i  - \frac{u_i^2}{2}, \right.$	if $a <  u_i  \le b$
	$\left(\frac{-a}{2(c-b)}(c-u_i)^2 + \frac{a}{2}(b+c-c)^2\right)$	$a$ ), if $b <  u_i  \le c$
Huber	$\rho(u_i) = \begin{cases} \frac{1}{2}u_i^2, & \text{if }  u_i  \le c\\ c u_i  - \frac{1}{2}c^2, & \text{if }  u_i  > c \end{cases}$	
Habel	$\int (u_i)^{-1} \left( c u_i  - \frac{1}{2}c^2, \text{ if }  u_i  > c \right)$	
	where $c = 1.345$	

Table 4: Metrics for model comparison [36].

Metrics	Equation
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left  \frac{Y_i - \hat{Y}_i}{Y_i} \right  \times 100$
Mean Squared Error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$
Sum of Squares of Error (SSE)	$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$
R-squared (R <sup>2</sup> )	$R^{2} = 1 - \frac{\sum_{i=1}^{n} (Y_{i} - \hat{Y}_{i})^{2}}{\sum_{i=1}^{n} (\hat{Y}_{i} - \bar{Y})^{2}}$

Table 5: Metrics for model comparison for ridge regression with robust method for 50, 100, 150, 200, 250, and 300 high ranking variables, before and after heterogeneity.

ML	Robust Method	High Ranking Variable		Before He	Before Heterogeneity			After Het	After Heterogeneity	
			MAPE	MSE	SSE	$R^2$	MAPE	MSE	SSE	$R^2$
Ridge	Original	50	9.459094	41.59782	79618.24	0.8479455	10.01975	45.19865	86510.21	0.8347832
			9.458448	41.59403	79610.98	0.8479593	10.28544	47.43193	90784.71	0.8266198
	M		9.088671	42.05232	80488.14	0.8462841	9.974874	48.1354	92131.16	0.8240484
	MM		9.030376	116.2844	222568.4	0.5749399	10.05090	51.9775	99484.94	0.8100041
	MM Bi-square		8.932346	45.83179	87722.04	0.8324689	10.05033	52.11318	99744.63	0.8095082
	MM Hampel		8.801508	45.81388	87687.77	0.8325343	10.05090	52.11169	99741.77	0.8095137
	MM Huber		8.918689	46.19288	88413.16	0.831149	10.05151	52.11069	99739.86	0.8095173
	M Hampel		9.292493	41.55148	79529.53	0.8481149	10.09944	47.70791	91312.93	0.8256110
	M Huber		9.094748	42.00605	80399.58	0.8464533	9.977855	48.12047	92102.58	0.8241029
	M Tukey		9.026400	43.94833	84117.1	0.8393536	10.06231	52.09507	96.60766	0.8095744
Ridge	Original	100	8.304651	33.36347	63857.68	0.8780449	8.998889	37.83964	72425.08	0.8616829
			8.304451	33.36423	63859.13	0.8780421	9.242874	39.23603	75097.75	0.8565787
	M		8.019361	33.6777	64459.11	0.8768963	9.060171	39.27781	75177.72	0.8564259
	MM		7.889334	34.66241	66343.86	0.8732968	8.973277	39.5754	75747.32	0.8553381
	MM Bi-square		8.522054	41.77504	79957.42	0.8472977	8.982879	39.97154	76505.52	0.8538901
	MM Hampel		7.948345	37.3687	71523.69	0.8634044	8.977533	39.74126	76064.77	0.8547319
	MM Huber		7.955227	37.58938	71946.08	0.8625977	9.046882	39.96533	76493.65	0.8539128
	M Hampel		8.078243	33.39565	63919.28	0.8779273	9.089914	38.99034	74627.51	0.8574767
	M Huber		8.023414	33.70079	64503.3	0.8768119	9.060665	39.24401	75113.04	0.8565495
	M Tukey		7.968238	36.11779	69129.45	0.8679769	8.985664	39.18361	74997.43	0.8567703
Ridge	Original	150	7.893903	30.60797	58583.65	0.8881172	8.511716	34.44371	65925.27	0.8740962
	S		7.89407	30.61025	58588.01	0.8881089	8.596673	35.02104	67030.27	0.8719859
	M		7.621724	31.12112	59565.83	0.8862415	8.363508	35.0202	67028.66	0.871989
	MM		7.5625	34.1932	65445.73	0.8750	8.2734	36.4082	69685.28	6998.0
	MM Bi-square		7.5667	34.2472	65549.08	0.8748	8.2919	36.8336	70499.55	0.8654
	MM Hampel		7.5975	34.3544	65754.37	0.8744	8.3018	36.7768	70390.8	0.8656
	MM Huber		7.5909	33.4998	64118.55	0.8775	8.2927	36.7897	70415.52	0.8655
	M Hampel		7.6362	30.9845	59304.24	0.8867	8.4111	35.2547	67477.53	0.8711
	M Huber		7.6210	31.0570	59443.11	0.8865	8.3861	35.3628	67684.42	0.8707
	M Tukey		7.5654	34.0423	65156.91	0.8756	8.2855	36.5149	69889.44	0.8665

Ridge Original 200  Ridge Original 200  MM Bi-square MM Hampel MM Huber M Huber M Huber M Huber M Hampel MM Hampel MM Huber M M Hampel MM Huber M M Hampel MM Huber M Huber M Huber M Hampel MM Huber M Huber	M	Pohiet Mathod	High	Panking		Refore He	Marogeneity			After Het	arogeneity.	
Ridge         Original         AMAPE         MSE         SSE         MAPE         MSE		POLICE INCOME	Variable				actogenety				COSCIICITY	
Ridge Original         200         7,6729         29,1637         55819,24         0.8934         8,1929         31,9778         61205.49           S         A         1,471         29,1618         55815.65         0.8934         8,1929         31,9778         61205.49           N         A         7,472         29,1618         55815.65         0.8934         8,2787         20,568.11           MM         Bisquare         7,372         29,1618         55805.09         0.8934         8,2790         32,558.7         6,668.7           MM Huber         7,3445         24,267         6,5837.42         0.8747         8,030         35,2398         67449.02           M Huber         7,340         2,233         53,538         6,296.54         0.8772         8,030         35,238         6,749.66           N         Humpel         7,340         3,4267         6,285.4         0.893         3,449         6,649           N         Humpel         7,340         3,548         6,296.54         0.873         8,168         6,649           N         M         Humpel         7,354         2,253         6,296.54         0.873         3,168         6,646           N         M					MAPE	MSE	SSE	$R^2$	MAPE	MSE	SSE	$R^2$
S         7,6723         29,1618         55815.65         0.8934         8,7792         32,5580         62315.98           MM         MM         7,3744         29,2168         55820,98         8,0867         32,5580         62315.88           MM         MM         7,374         3,57184         64530,94         8,0876         36,010         67011           MM         MBi-square         7,338         33,1241         64399,6         0,8789         8,0469         34,913         66823.86           MM         MBimpel         7,4248         3,51241         64390,0         0,8789         8,0469         34,913         66823.86           MM         Huber         7,4246         3,5714         64396,0         0,8789         8,0469         34,913         66823.86           MB Huber         7,4236         29,2551         55668,0         0,8930         8,0569         32,2596         62434,0           S         M Huber         7,625         28,7248         64296,4         0,872         8,0589         32,555         628,24           A         A         7,625         28,7248         64296,4         0,893         8,163         626,43           A         A         7,625	Ridge		200		7.6729	29.1637	55819.24	0.8934	8.1929	31.9778	61205.49	0.8831
MM         T,4141         29,2168         55920,98         0.8932         8,0887         32,7367         62658.11           MM Bisquare         MM Bisquare         7,3727         33,7525         6400.222         0.8746         8,0176         35,0119         65011           MM Hampel         7,3498         33,7184         64536.94         0.8767         8,0108         34,849         66697           MM Huber         7,3498         33,7184         64536.94         0.8767         8,0108         34,849         66697           MM Huber         7,3408         33,7184         64536.94         0.8767         8,0108         34,849         66697           M Huber         7,4236         29,347         50890         8,0560         32,786         6663,242           Kidge         Original         250         7,6253         28,7248         64296.34         0.8772         8,0390         34,835         6568.34           M         Tukey         7,3646         28,958         62499         8,0460         32,736         6566.34           M         M         Mampel         7,3646         28,958         62499         8,0460         32,736         6562.24           MM         Humpel		S			7.6723	29.1618	55815.65	0.8934	8.2792	32.5580	62315.98	0.8810
MMM         T3327         33.7325         6460222         0.8766         8.0176         35.0110         67011           MM Bisquare         T3380         33.1241         645309.6         0.8789         8.0469         34.913.6         66897           MM Hampel         T,4048         34.2472         65587.42         0.8779         8.0106         34.919.6         66897           MM Hampel         T,4236         29.2351         5595.60         0.8931         8.0569         32.2856         62884.51           M Huber         T,4226         29.2351         5595.60         0.8929         8.0660         32.7856         62884.51           Kidge         Original         7.3200         7.625         28.7248         54979.3         0.8950         31.6842         6663.46           S         ATUREN         7.3646         29.0292         55561.82         0.8949         8.0666         32.7856         6525.241           RMB         Incompanie         T,3349         31.0807         59488.5         8.123         33.0609         32.2881         6652.42           NMB         Incompanie         T,3349         31.0807         5243.28         8.0892         8.1666         32.448         655.741         688994		M			7.4141	29.2168	55920.98	0.8932	8.0587	32.7367	62658.11	0.8803
MM Bi-square         7,3380         33,1241         653996         0.8789         8,0469         34,9132         66823.86           MM Hampel         7,3408         33,7184         64356.94         0.8777         8,0108         34,8469         6697           MM Huber         7,4245         34,2672         28,747         8,0108         34,8469         66823.86           M Huber         7,4236         29,2351         55956.06         0.8921         8,0606         32,7857         62864.51           Sidge Original         250         7,6636         29,2347         56089.09         8,0829         8,0606         32,7857         65265.42           Ridge Original         250         7,6636         29,0292         5556.182         0.8959         8,1639         34,2857         6526.41           MM Bi-square         7,3546         29,0292         5556.182         0.8999         8,1233         31,0807         625.2241           MM Huber         7,3443         31,0807         5243.28         0.8942         8,148         655.1101           MM Huber         7,346         28,9568         58,432.28         0.8804         8,148         65.1416           MM Huber         7,3439         31,0807         52		MM			7.3727	33.7525	64602.22	0.8766	8.0176	35.0110	67011	0.8720
MM Hampel         7.4098         3.3.7184         64536.94         0.8767         8.0100         34.8469         66697           MM Hampel         7.2445         34.2672         65887.42         0.8747         8.0300         35.2398         67449.02           MM Huber         7.4236         29.2351         55956.06         0.8931         8.0309         32.8550         66943.46           M Huber         7.4236         29.2351         56089.09         0.8929         8.0666         32.7857         65696.33           Ridge Original         250         7.6656         29.0292         58.963         8.0299         34.2855         6552.42           MM Hampel         7.3646         28.9688         55423.28         0.8949         8.1469         6649.34           MM Hampel         7.3409         37.283         8.8843         8.1469         35.747         66898.94           M Huber         7.340         30.7253         8.8804         8.1469         35.747         6696.34           MM Huber         7.340         30.7253         8.8804         8.1469         35.468         45.466         7.1019         25.740         6649.451           M Huber         7.366         29.1857         58808.21		MM Bi-square			7.3380	33.1241	63399.6	0.8789	8.0469	34.9132	66823.86	0.8724
MM Huber         7.3445         34.2672         65587.42         0.8747         8.0300         35.2398         67449.02           M Huber         7.4236         29.2451         55956.06         0.8931         8.0509         32.8550         6288.1           M Huber         7.4236         29.2447         5596.06         0.8929         8.0606         32.8550         6286.33           Ridge         Original         250         7.6636         29.0248         54979.3         0.8929         8.0606         32.7567         6266.34           S         7.6636         29.0292         55561.82         0.8939         8.2995         32.6815         6552.42           MM         M         7.3646         28.0289         55581         8.1469         35.115         6525.41           MM         Hampel         7.3439         31.0807         59488.5         8.1469         35.115         6252.41           MM         Huber         7.3403         32.6927         6273.9         8.1466         35.808         65461.10           MM         Huber         7.3403         32.6927         62573.9         8.894         8.1466         35.220.41           M         Huber         7.360         29.		MM Hampel			7.4098	33.7184	64536.94	0.8767	8.0108	34.8469	<i>L</i> 6999	0.8726
M Hampel         7.4236         29.2351         5595.06         0.8931         8.0509         32.8550         62884.51           M Huber         7.4221         29.3047         56089.09         0.8929         8.0606         32.7567         65652.42           Ridge         Original         250         7.6536         29.3047         56089.09         8.0892         8.0606         32.7567         65652.42           Ridge         Original         250         7.6636         29.0292         55561.82         0.8939         8.2995         32.6815         60552.41           M         M         A         7.3446         28.9568         5542.32         0.8992         8.1233         33.0502         65522.41           MM         Bisquare         7.349         3.2.8087         6.8804         8.1469         35.1155         67211.01           MM         Hampel         7.3403         32.6927         6.2573.9         0.8805         8.1469         35.747         6808.94           MM         Huber         7.3403         32.627         6.2573.9         0.8805         8.1456         35.5115         6.2550.41           MM         Huber         7.3110         32.027         6.8805         8.1456		MM Huber			7.3445	34.2672	65587.42	0.8747	8.0300	35.2398	67449.02	0.8712
M Huber         7,4221         29,3047         56089.09         0.8929         8.0606         32.7567         62696.33           Ridge         Original         250         7,4221         29,3047         56089.09         0.8929         8.0606         32.7567         62696.32           Ridge         Original         250         7,6636         29,0292         55561.83         8.0239         31,6842         66252.41           MM         MM         Fight         7,3464         28,968         55423.28         0.8942         8.1630         31,6842         66252.41           MM         Mampel         7,3464         28,968         55423.28         0.8892         8.1649         35,1155         66252.41           MM         Mampel         7,3464         28,968         54243.28         0.8892         8.1450         35,115         66252.41           MM Huber         7,3403         31,0807         59488.5         0.8895         8.1456         35,246.84           M Huber         7,366         29,1324         5536.53         8.0942         8.1456         35,247         6646.171           M Tukey         7,3103         30,8371         59022.15         0.8873         8.1456         49245.04 </td <td></td> <td>M Hampel</td> <td></td> <td></td> <td>7.4236</td> <td>29.2351</td> <td>55956.06</td> <td>0.8931</td> <td>8.0509</td> <td>32.8550</td> <td>62884.51</td> <td>0.8799</td>		M Hampel			7.4236	29.2351	55956.06	0.8931	8.0509	32.8550	62884.51	0.8799
MTukey         7.3300         33.5928         64296.54         0.8772         8.0239         34.2855         65622.42           Ridge         Original         250         7.6255         28.7348         54979.3         0.8950         8.1630         31.6842         6663.46           NM         A         7.3646         29.0229         5556.82         0.8939         8.2995         32.6815         6652.41           MM         A         7.3646         29.0229         5556.82         8.1459         35.155         6721.10           MM         Bi-square         7.3439         31.0807         59488.5         0.8804         8.1469         35.155         6721.10           MM Hampel         7.3439         32.0827         62573.9         0.8804         8.1469         35.747         68089.94           M Huber         7.3406         29.022         5536.9         8.1469         35.747         68089.94           M Huber         7.3403         30.7253         58808.1         8.1469         35.747         68089.94           M Tukey         7.366         29.324         553.65.9         8.148         35.747         68089.1           M Tukey         7.306         29.324         553.65.9 <td></td> <td>M Huber</td> <td></td> <td></td> <td>7.4221</td> <td>29.3047</td> <td>56089.09</td> <td>0.8929</td> <td>8.0606</td> <td>32.7567</td> <td>62696.33</td> <td>0.8803</td>		M Huber			7.4221	29.3047	56089.09	0.8929	8.0606	32.7567	62696.33	0.8803
Original         250         7,6255         28,7248         54979.3         0.8950         8.1630         31.6842         60643.46           S         S         7,6636         29,0292         55561.82         0.8939         8.2995         32.6815         6255.41           M         A         7,3646         28,968         5542.328         0.8942         8.1295         32.6815         6255.41           MM         B         7,3646         28,968         5542.328         0.8942         8.1148         35.747         68089.94           MM         B         7,3403         32,6927         62753.9         0.8805         8.1466         35.8747         68089.94           MM         B         7,3403         32,6927         62573.9         0.8805         8.1486         35.747         68089.94           MM         B         30,000         7,361         29,1857         58805         8.1486         33,0443         6346.84           M         Huber         7,361         29,1857         58801.43         8.0983         34,7557         66522.41           M         Lukey         7,019         25,376.59         0.8975         0.9053         7,2311         26,804.51		M Tukey			7.3300	33.5928	64296.54	0.8772	8.0239	34.2855	65622.42	0.8747
S         7.6636         29.0292         55561.82         0.8939         8.2995         32.6815         6555.41           MM         MM         7.3646         28.9568         55423.28         0.8942         8.1223         33.0502         63258.02           MM         MM         7.3436         23.8887         62795.81         0.8801         8.1456         35.747         68089.94           MM Hampel         7.3403         32.6927         62573.9         0.8864         8.1148         35.5747         68089.94           MM Hampel         7.3403         32.6927         62573.9         0.8865         8.1456         35.8208         68561.71           M Huber         7.361         29.1857         58608.21         8.0984         34.7240         66461.71           Kidge         Original         30.0         7.3361         29.187         5808.21         8.0984         34.7240         66461.71           Ridge         Original         30.0         7.3636         28.9324         55376.59         8.9984         34.7557         66452.41           Ridge         Original         30.0         7.0635         25.8776         49529.72         0.9054         7.0191         25.7289         49245.04     <	Ridge		250		7.6255	28.7248	54979.3	0.8950	8.1630	31.6842	60643.46	0.8842
M         7.3646         28.9568         55423.28         0.8942         8.1223         33.0502         63258.02           MM         MM         7.3551         32.8087         6.795.81         0.8801         8.1469         35.1155         67211.01           MM Bi-square         7.3439         31.0807         59488.5         0.8864         8.148         35.5747         68089.94           MM Huber         7.3436         32.6927         6.25739         0.8805         8.1456         35.8208         68850.1           M Huber         7.3656         28.934         5580.143         0.8877         8.0984         3.47240         6646.171           Ridge         Original         300         7.3656         28.934         5536.143         0.8873         8.148         33.1737         63494.51           Ridge         Original         300         7.0635         25.9174         49605.94         0.9053         3.47557         6640.171           NM         Absquare         6.8821         26.282         50315.54         0.9053         7.021         25.093         55.044.7           MM Hampel         6.8874         26.282         50315.54         0.9039         7.0649         27.1179         555.04.1					7.6636	29.0292	55561.82	0.8939	8.2995	32.6815	62552.41	0.8805
MM         T.3551         32.8087         62795.81         0.8801         8.1469         35.1155         67211.01           MM Bi-square         7.3439         31.0807         59488.5         0.8864         8.1148         35.5747         6808.94           MM Hampel         7.3403         32.6927         62573.9         0.8805         8.1456         35.8208         68561           MM Huber         7.316         29.1857         58808.21         0.8877         8.0984         34.7240         66461.71           M Huber         7.366         29.1857         58808.21         0.8877         8.0984         34.7240         66461.71           M Huber         7.366         29.1857         58808.21         0.8877         8.0984         34.7557         66421.71           Kidge         Original         300         7.0635         25.9174         49605.94         7.0919         25.7289         49245.41           S         A         7.0635         25.9174         49605.94         7.0622         29.0935         5568.48           MM         Bi-square         6.8821         26.282         5692.72         6.9902         29.0107         5550.17           MM Huber         6.8846         27.8145		M			7.3646	28.9568	55423.28	0.8942	8.1223	33.0502	63258.02	0.8792
MM Bi-square         7.3439         31.0807         59488.5         0.8864         8.1148         35.5747         68089.94           MM Hampel         7.3403         32.6927         62573.9         0.8805         8.1456         35.8208         68561           MM Huber         7.3136         30.7253         58808.21         0.8877         8.0984         34.7240         66461.71           M Huber         7.3961         29.1857         55861.43         0.8873         8.1466         33.0443         6546.171           M Huber         7.3646         28.9324         55376.59         0.8873         8.1646         33.0443         6546.171           M Huber         7.3656         28.9324         55376.59         0.8873         8.1486         33.0443         6542.41           Ridge Original         300         7.0635         25.8176         49659.4         7.0191         25.7289         49245.04           S         A         7.0635         25.9174         49605.94         0.9039         7.0625         29.0935         5568.48           MM Bi-square         6.8824         27.7392         53638.32         0.8876         6.9942         27.193         557.193           MM Hubber         6.8887	16	MM			7.3551	32.8087	62795.81	0.8801	8.1469	35.1155	67211.01	0.8716
MM Hampel         7.3403         32.6927         62573.9         0.8805         8.1456         35.8208         68561           MM Huber         7.3136         30.7253         58808.21         0.8877         8.0984         34.7240         66461.71           M Hampel         7.3961         29.1857         55861.43         0.8933         8.1646         33.0443         66461.71           M Huber         7.3656         28.9324         55376.59         0.8942         8.1418         33.1737         63494.51           M Tukey         7.3100         30.8371         59022.15         0.8943         31.737         65322.41           Original         300         7.0635         25.8176         49529.72         0.9054         7.0191         25.7289         49245.04           S         5         25.9174         49605.94         0.9053         7.2311         26.8833         51397.15           M         6.8821         25.9174         49605.94         0.9053         7.0622         27.1953         5564.88           MM         6.8821         27.7392         53092.79         0.8986         6.9625         29.0935         5564.16           MM Huber         6.8847         27.8188         53245.12		MM Bi-square			7.3439	31.0807	59488.5	0.8864	8.1148	35.5747	68089.94	0.8700
MM Huber         7.3136         30.7253         58808.21         0.8877         8.0984         34.7240         66461.71           M Hampel         7.3961         29.1857         55861.43         0.8933         8.1646         33.0443         65246.84           M Huber         7.3636         28.9324         55361.43         0.8942         8.1418         33.1737         63494.51           M Tukey         7.3100         30.8371         59022.15         0.8873         8.0983         34.7557         6522.41           Original         300         7.0635         25.9174         49605.94         0.9054         7.0191         25.7289         49245.04           S         6.8821         25.9174         49605.94         0.9039         7.0622         27.1953         52051.75           MM         6.8821         26.2882         53092.79         0.8986         6.9625         29.0935         55684.88           MM Bi-square         6.8871         27.372         53022.12         0.8976         6.9902         29.4172         56304.47           MM Huber         6.8867         27.8148         53245.12         0.9039         7.0649         27.1179         51922.80           M Huber         6.8834		MM Hampel			7.3403	32.6927	62573.9	0.8805	8.1456	35.8208	68561	0.8691
M Hampel         7.3961         29.1857         55861.43         0.8933         8.1646         33.0443         63246.84           M Huber         7.3636         28.9324         55376.59         0.8942         8.1418         33.1737         63494.51           M Tukey         7.3100         30.8371         59022.15         0.8873         8.0983         34.7557         6522.41           Original         300         7.0635         25.8776         49529.72         0.9054         7.0191         25.7289         49245.04           S         7.0635         25.8776         49605.94         0.9063         7.2311         26.8533         51397.15           M         M         6.8821         26.2882         50315.54         0.9063         7.0622         27.1953         55684.88           MM         Bi-square         6.8301         27.7392         53092.79         0.8986         6.9025         29.0935         55684.88           MM Huber         6.8867         27.8745         53351.82         0.8991         7.0649         27.1279         54504.16           M Huber         6.8834         26.2932         50197.55         0.9041         7.0649         27.1978         52056.51           M Tukey </td <td></td> <td>MM Huber</td> <td></td> <td></td> <td>7.3136</td> <td>30.7253</td> <td>58808.21</td> <td>0.8877</td> <td>8.0984</td> <td>34.7240</td> <td>66461.71</td> <td>0.8731</td>		MM Huber			7.3136	30.7253	58808.21	0.8877	8.0984	34.7240	66461.71	0.8731
M Huber         7.3636         28.9324         55376.59         0.8942         8.1418         33.1737         63494.51           M Tukey         7.3100         30.8371         59022.15         0.8873         8.0983         34.7557         66522.41           Original         300         7.0635         25.8776         49529.72         0.9054         7.0191         25.7289         49245.04           S         7.0635         25.9174         49605.94         0.9053         7.2311         26.8533         51397.15           M         6.8821         26.2882         50315.54         0.9039         7.0622         27.1953         52051.75           MM         6.8821         26.2882         5315.54         0.8966         6.9625         29.0935         55684.88           MM         6.8264         28.0242         53638.32         0.8976         6.9902         29.4172         56304.47           MM         6.8887         27.8745         53351.82         0.8981         6.9914         28.4766         54504.16           M Huber         6.8467         27.8188         53245.12         0.8983         6.9748         29.0107         55256.51           M Huber         6.8834         26.2932		M Hampel			7.3961	29.1857	55861.43	0.8933	8.1646	33.0443	63246.84	0.8792
M Tukey         7.3100         30.8371         59022.15         0.8873         8.0983         34.7557         66522.41           Original         300         7.0635         25.8776         49529.72         0.9054         7.0191         25.7289         49245.04           S         7.0635         25.9174         49605.94         0.9053         7.2311         26.8533         51397.15           M         6.8821         26.2882         50315.54         0.9039         7.0622         27.1953         52051.75           MM         6.8301         27.7392         53092.79         0.8986         6.9625         29.0935         55684.88           MM Bi-square         6.8264         28.0242         53638.32         0.8976         6.9902         29.4172         56304.47           MM Huber         6.8887         27.8188         53245.12         0.8983         6.9748         29.0107         55526.51           M Humpel         6.8834         26.2932         50197.55         0.9041         7.0649         27.1279         51922.80           M Huber         6.8530         27.9441         53484.91         0.8979         7.0106         28.9429         55396.77		M Huber			7.3636	28.9324	55376.59	0.8942	8.1418	33.1737	63494.51	0.8787
Original         300         7.0635         25.8776         49529.72         0.9054         7.0191         25.7289         49245.04           S         7.0635         25.9174         49605.94         0.9053         7.2311         26.8533         51397.15           M         6.8821         26.2882         50315.54         0.9039         7.0622         27.1953         52051.75           MM         6.8301         27.7392         53092.79         0.8986         6.9625         29.0935         55684.88           MM Bi-square         6.8264         28.0242         53638.32         0.8976         6.9902         29.4172         56304.47           MM Hampel         6.8867         27.8188         53245.12         0.8983         6.9748         29.0107         55526.51           M Hampel         6.88716         26.2265         50197.55         0.9041         7.0649         27.1279         51922.80           M Huber         6.8834         26.2932         50325.09         7.0106         28.9429         53396.77		M Tukey			7.3100	30.8371	59022.15	0.8873	8.0983	34.7557	66522.41	0.8730
7.0635       25.9174       49605.94       0.9053       7.2311       26.8533       51397.15         6.8821       26.2882       50315.54       0.9039       7.0622       27.1953       52051.75         6.8301       27.7392       53092.79       0.8986       6.9625       29.0935       55684.88         square       6.8264       28.0242       53638.32       0.8976       6.9902       29.4172       56304.47         npel       6.8887       27.8745       53351.82       0.8981       6.9914       28.4766       54504.16         ser       6.8467       27.8188       53245.12       0.8983       6.9748       29.0107       55526.51         el       6.8716       26.2265       50197.55       0.9041       7.0649       27.1279       51922.80         el       6.8834       26.2932       50325.09       0.9039       7.0602       27.1978       52056.51         el       6.8530       27.9441       53484.91       0.8979       7.0106       28.9429       55396.77	Ridge		300		7.0635	25.8776	49529.72	0.9054	7.0191	25.7289	49245.04	0906.0
6.8821       26.2882       50315.54       0.9039       7.0622       27.1953       52051.75         quare       6.8301       27.7392       53092.79       0.8986       6.9625       29.0935       55684.88         quare       6.8264       28.0242       53638.32       0.8976       6.9902       29.4172       56304.47         npel       6.8887       27.8745       53351.82       0.8981       6.9914       28.4766       54504.16         er       6.8467       27.8188       53245.12       0.8983       6.9748       29.0107       55526.51         el       6.8716       26.2265       50197.55       0.9041       7.0649       27.1279       51922.80         6.8834       26.2932       50325.09       0.9039       7.0602       27.1978       52056.51         6.8530       27.9441       53484.91       0.8979       7.0106       28.9429       55396.77		S			7.0635	25.9174	49605.94	0.9053	7.2311	26.8533	51397.15	0.9018
quare       6.8301       27.7392       53092.79       0.8986       6.9625       29.0935       55684.88         quare       6.8264       28.0242       53638.32       0.8976       6.9902       29.4172       56304.47         npel       6.8887       27.8745       53351.82       0.8981       6.9914       28.4766       54504.16         er       6.8467       27.8188       53245.12       0.8983       6.9748       29.0107       55526.51         el       6.8716       26.2265       50197.55       0.9041       7.0649       27.1279       51922.80         6.8834       26.2932       50325.09       0.9039       7.0602       27.1978       52056.51         6.8530       27.9441       53484.91       0.8979       7.0106       28.9429       55396.77		M			6.8821	26.2882	50315.54	0.9039	7.0622	27.1953	52051.75	9006.0
quare         6.8264         28.0242         53638.32         0.8976         6.9902         29.4172         56304.47           npel         6.8887         27.8745         53351.82         0.8981         6.9914         28.4766         54504.16           ber         6.8467         27.8188         53245.12         0.8983         6.9748         29.0107         55526.51           el         6.8716         26.2265         50197.55         0.9041         7.0649         27.1279         51922.80           6.8834         26.2932         50325.09         0.9039         7.0602         27.1978         52056.51           6.8530         27.9441         53484.91         0.8979         7.0106         28.9429         55396.77		MM			6.8301	27.7392	53092.79	9868.0	6.9625	29.0935	55684.88	0.8937
npel 6.8887 27.8745 53351.82 0.8981 6.9914 28.4766 54504.16 e.e. 6.8467 27.8188 53245.12 0.8983 6.9748 29.0107 55526.51 e.e. 6.8716 26.2265 50197.55 0.9041 7.0649 27.1279 51922.80 e.e. 6.8834 26.2932 50325.09 0.9039 7.0602 27.1978 52056.51 e.e. 6.8530 27.9441 53484.91 0.8979 7.0106 28.9429 55396.77		MM Bi-square			6.8264	28.0242	53638.32	92680	6.9902	29.4172	56304.47	0.8925
el 6.8467 27.8188 53245.12 0.8983 6.9748 29.0107 55526.51 cl 8.8716 26.2265 50197.55 0.9041 7.0649 27.1279 51922.80 cl 8.834 26.2932 50325.09 0.9039 7.0602 27.1978 52056.51 cl 8.8530 27.9441 53484.91 0.8979 7.0106 28.9429 55396.77		MM Hampel			6.8887	27.8745	53351.82	0.8981	6.9914	28.4766	54504.16	0.8959
el 6.8716 26.2265 50197.55 0.9041 7.0649 27.1279 51922.80 6.8834 26.2932 50325.09 0.9039 7.0602 27.1978 52056.51 6.8530 27.9441 53484.91 0.8979 7.0106 28.9429 55396.77		MM Huber			6.8467	27.8188	53245.12	0.8983	6.9748	29.0107	55526.51	0.8940
6.8834       26.2932       50325.09       0.9039       7.0602       27.1978       52056.51         6.8530       27.9441       53484.91       0.8979       7.0106       28.9429       55396.77		M Hampel			6.8716	26.2265	50197.55	0.9041	7.0649	27.1279	51922.80	0.9008
6.8530 27.9441 53484.91 0.8979 7.0106 28.9429 55396.77		M Huber			6.8834	26.2932	50325.09	0.9039	7.0602	27.1978	52056.51	9006.0
		M Tukey			6.8530	27.9441	53484.91	0.8979	7.0106	28.9429	55396.77	0.8942

Table 6: Metrics for model comparison for LASSO regression with robust method for high ranking variables before and after heterogeneity.

		ing Variable		Before he	Before heterogeneity			After heterogeneity	rogeneity	
			MAPE	MSE	SSE	$R^2$	MAPE	MSE	SSE	$R^2$
LASSO	Original	50	8.958306	38.86046	74378.92	0.8579515	8.933586	38.57446	73831.51	0.8589969
	S		9.419857	43.19943	82683.71	0.8420911	9.626142	43.45914	83180.79	0.8411417
	M		9.156418	42.58668	81510.91	0.8443309	9.338847	43.09136	82476.86	0.8424861
	MM		8.969212	44.59114	85347.44	0.8370039	9.211808	43.51981	83296.92	0.8409199
	MM Bi-square		9.001022	43.63385	83515.19	0.8405031	9.210072	43.5384	83332.51	0.840852
	MM Hampel		8.970389	44.9011	85940.7	0.8358709	9.391685	49.62911	94990.12	0.8185883
	MM Huber		8.968910	44.51771	85206.9	0.8372723	9.215769	43.01301	82326.89	0.8427725
	M Hampel		9.268027	42.57249	81483.74	0.8443828	9.407373	42.95528	82216.41	0.8429835
	M Huber		9.328288	42.9958	82293.96	0.8428354	9.328288	42.9958	82293.96	0.8428354
	M Tukey		9.021315	43.65113	83548.27	0.8404399	9.215979	43.37112	83012.33	0.8414635
LASSO	Original	100	8.823839	37.7268	72209.09	0.8620954	8.811494	37.66045	72082.1	0.8623379
	S		9.184282	40.85047	78187.8	0.8506773	9.364906	41.35937	79161.84	0.8488171
	M		8.895828	40.50818	77532.66	0.8519285	9.102281	41.32916	79104.02	0.8489275
	MM		8.857082	42.598	81532.58	0.8442895	9.061962	42.2719	80908.42	0.8454815
	MM Bi-square		8.674510	42.20375	80777.97	0.8457306	8.997942	43.99379	84204.11	0.8391874
	MM Hampel		8.800533	45.13168	86382.03	0.835028	908620.6	42.79846	81916.25	0.8435567
	MM Huber		8.638455	42.09574	80571.25	0.8461254	9.114223	42.53842	81418.54	0.8445073
	M Hampel		8.976605	40.44503	77411.78	0.8521594	9.159449	41.08255	78632	0.849829
	M Huber		8.895465	40.50612	77528.71	0.851936	9.099129	41.29869	79045.69	0.8490389
	M Tukey		8.892657	42.42826	81207.69	0.84491	9.117924	42.21883	80806.85	0.8456755
LASSO	Original	150	8.347495	34.26593	65584.99	0.8747461	8.373180	34.29228	65635.42	0.8746498
	S		8.771521	37.51598	71805.59	0.862866	8.827855	37.50158	71778.02	0.8629187
	M		8.501082	37.17901	71160.62	0.8640978	8.597285	36.92832	70680.8	0.8650142
	MM		8.457516	39.59102	75777.2	0.8552811	8.505798	39.24539	75115.69	0.8565444
	MM Bi-square		8.463936	39.68487	75956.85	0.854938	8.575563	39.33908	75295	0.856202
	MM Hampel		8.460321	39.59161	75778.34	0.8552789	8.573668	39.41567	75441.59	0.855922
	MM Huber		8.503222	39.90662	76381.26	0.8541274	8.482080	38.69286	74058.14	0.8585641
	M Hampel		8.675947	38.02455	72778.98	0.8610071	8.612283	36.83739	70506.77	0.8653465
	M Huber		8.501100	37.17921	71161.01	0.864097	8.596594	36.90819	70642.27	0.8650877
	M Tukev		8.459070	39.60414	75802.31	0.8552331	8.584735	39.1486	74930.42	0.8568982

MAPE   MSE		М	Dobiet Mathod	Uich Donk		Dofore Uo	torogonoity			A flor Uata	mogonoiti,	
LASSO         Original         200         8.339226         33.78147           S         S         8.69938         36.44334           M         M         8.490187         36.59463           MM         Bisquare         8.428189         37.70382           MM         Hampel         8.333713         37.74151           MM         Huber         8.441823         37.74151           MM         Huber         8.441823         37.74151           M         Huber         8.441823         37.74151           M         Huber         8.441823         37.74151           M         Huber         8.48946         36.59841           M         Huber         8.378494         37.40204           A         S         8.673115         36.1386           M         Bissand         8.441201         39.14316           MM         Huber         8.383458         39.14316           M         Huber         8.455572         36.1026           M         Huber         8.455572         36.1436           M         Huber         8.45557         37.2468           M         Huber         8.451691         37.2468		ML		rugin Namk- ing Variable		Deloie He	ierogeneny			Aiter neterogenery	aogeneny	
LASSO         Original         200         8.339226         33.78147           S         8.69938         36.44334           M         8.490187         36.59463           MM         Bis-square         8.428189         37.67137           MM         Bis-square         8.441823         37.74151           MM         Huber         8.41823         37.74151           MM         Huber         8.441823         37.74151           M         Huber         8.44153         37.748393           M         Huber         8.554033         36.84411           M         Huber         8.378494         37.40204           LASSO         Original         250         8.308303         33.55028           S         8.673115         36.1386           M         Huber         8.441291         39.48344           MM         Huber         8.441291         39.48344           MM         Huber         8.358216         37.72468           M         Huber         8.358216         37.72468           M         Huber         8.441291         37.48064           LASSO         Original         30.0         8.278559         33.16					1	MSE	SSE	$R^2$	MAPE	MSE	SSE	$R^2$
S 8.69938 36.44334  M 8 8.490187 36.59463  MM Bi-square 8.441823 37.70382  MM Hampel 8.333713 37.74151  MM Huber 8.418473 37.68393  M Huber 8.489946 36.59841  M Tukey 8.378494 37.40204  LASSO Original 250 8.308303 33.55028  M M Bi-square 8.41291 39.48344  MM Huber 8.41291 39.48344  MM Huber 8.41291 39.48344  MM Huber 8.416911 37.77036  M Hampel 8.358216 37.72468  MM Huber 8.41691 37.77036  M Huber 8.451854 36.16061  S S S S S S S S S S S S S S S S S S S	Ι΄ _	LASSO	Original	200	8.339226	33.78147	64657.73	0.876517	8.328507	33.79357	64680.9	0.8764727
MM       8.490187       36.59463         MM       8.428189       37.67137         MM Hampel       8.333713       37.74151         MM Huber       8.415473       37.68393         MM Huber       8.454033       36.84411         M Huber       8.489946       36.59841         M Huber       8.489946       36.59841         M Tukey       8.378494       37.40204         AM       8.441291       39.48344         MM       8.455572       36.1402         MM       8.455572       36.1402         MM Huber       8.441291       39.48344         MM Huber       8.383458       39.14316         M Huber       8.383458       39.14316         M Huber       8.451854       36.06626         M Huber       8.451854       36.03626         M Huber       8.451854       36.13061         S       8.578550       37.8064         LASSO Original       300       8.278559       37.8064         LASSO Original       300       8.278559       37.8064         MM       8.43050       35.23839         MM       8.43050       35.23839         MM Huber       8			S		8.699938	36.44334	69752.55	0.8667869	8.771923	36.76979	70377.37	0.8655936
MM       8.428189       37.67137         MM Bi-square       8.441823       37.70382         MM Huber       8.333713       37.74151         MM Huber       8.415473       37.68393         M Huber       8.489946       36.59841         M Tukey       8.489946       36.59841         M Tukey       8.378494       37.40204         LASSO Original       250       8.308303       33.55028         M       8.471291       39.48344         MM Bi-square       8.358216       37.72468         MM Huber       8.4416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         LASSO Original       300       8.278559       33.16061         S       8.578759       35.28339         MM       8.233550       37.28122         MM       8.23300       35.32339         MM       8.233150       37.28122         MM Huber       8.48935       35.32339         M Huber       8.48935       35.32339         M Huber       8.48935       35.3005         M Hub			M		8.490187	36.59463	70042.12	0.8662339	8.566170	36.45279	69770.64	0.8667524
MM Bi-square       8.441823       37.70382         MM Hampel       8.333713       37.74151         MM Huber       8.415473       37.68393         M Huber       8.489946       36.59841         M Huber       8.378494       37.40204         LASSO       0riginal       250       8.308303       33.55028         S       8.673115       36.13586         M       8.455572       36.1402         MM Bi-square       8.385216       37.72468         MM Hampel       8.416911       37.77036         M Huber       8.451854       36.06626         M Tukey       8.57879       35.64102         M       8.57859       37.03852         MM       8.43050       35.32839         MM       8.232075       37.03852         MM       8.331507       36.30777         MM       8.48050       35.32839         M       8.48050       35.3005         M       8.48050       35.3005     <			MM		8.428189	37.67137	72103.01	0.862298	8.468777	39.51445	75630.66	0.8555609
MM Hampel       8.333713       37.74151         MM Huber       8.415473       37.68393         M Hampel       8.554033       36.84411         M Huber       8.489946       36.59841         M Tukey       8.308303       33.55028         S       8.673115       36.13586         M       8.45572       36.1402         MM       8.45572       36.1402         MM Bi-square       8.358216       37.72468         MM Hampel       8.41291       39.48344         M Huber       8.416911       37.77036         M Huber       8.511842       36.06626         M Tukey       8.57859       37.0804         LASSO       Original       300       8.57859       37.28122         MM       Bi-square       8.57859       37.03852         MM       Humpel       8.222075       37.03852         MM       Huber       8.43050       35.32839         MM       Humpel       8.486325       35.51796         M Huber       8.486325       35.51796         M Tukey       8.430501       35.32839         M Tukey       8.486325       35.20055         M Tukey       8.43050			MM Bi-square		8.441823	37.70382	72165.11	0.8621794	8.539521	37.55319	71876.8	0.86273
MM Huber       8.415473       37.68393         M Hampel       8.554033       36.84411         M Huber       8.489946       36.59841         M Huber       8.378494       37.40204         LASSO       0riginal       250       8.308303       33.55028         S       8.673115       36.13586         M       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Hampel       8.416911       37.77036         M Huber       8.416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.57879       35.64102         M       8.358520       37.88064         LASSO Original       300       8.278559       33.16061         S       8.57879       35.32839         MM       Bi-square       8.222075       37.03852         MM Hampel       8.331507       36.30777         MM Huber       8.489325       35.51796         M Huber       8.489325       35.5085         M Tuber       8.489325       35.5085         M Tuber       8.50014       35.5005			MM Hampel		8.333713	37.74151	72237.25	0.8620417	8.527913	37.7801	72311.11	0.8619006
M Hampel       8.554033       36.84411         M Huber       8.489946       36.59841         M Tukey       8.378494       37.40204         LASSO Original       250       8.308303       33.55028         S       8.673115       36.13586         M       8.455572       36.1402         MM       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Huber       8.416911       37.77036         M Huber       8.416911       37.77036         M Huber       8.451854       36.13317         M       4.48550       37.88064         LASSO Original       300       8.278559       37.60626         MM       8.43050       35.32839         MM       8.410ber       8.43050       35.32839         MM Huber       8.489325       35.51796         M Huber       8.489325       35.51796         M Huber       8.480501       35.32839         M Huber       8.480325       35.51796         M Huber       8.480501       35.50065			MM Huber		8.415473	37.68393	72127.04	0.8622521	8.526482	38.25761	73225.07	0.8601551
M Huber       8.489946       36.59841         M Tukey       8.378494       37.40204         LASSO       Original       250       8.308303       33.55028         S       8.673115       36.13586         M       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Hampel       8.383458       39.14316         M Huber       8.416911       37.77036         M Huber       8.451842       36.06626         M Huber       8.278520       37.88064         LASSO       0riginal       300       8.278559       33.16061         S       8.43050       35.32839         MM       Bi-square       8.222075       37.03852         MM Huber       8.186330       36.79287         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Huber       8.430501       35.32839         M Huber       8.430501       35.32839         M Huber       8.430501       35.2005         M Huber       8.430501       35.2006         M Huber       8.430501       35.2006         M Huber       8.430501			M Hampel		8.554033	36.84411	70519.62	0.865322	8.619524	36.61716	70085.25	0.8661515
LASSO       Original       250       8.308303       33.55028         S       8.673115       36.13586         M       8.455572       36.1402         MM       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Hampel       8.383458       39.14316         M Huber       8.416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.511842       36.06626         M Huber       8.511842       36.06626         M       8.57859       37.28122         MM       8.57879       35.32839         MM       8.13317       8.43050       35.32839         MM       8.331507       36.30777         MHuber       8.489325       35.51796         M Huber       8.480501       35.32839         M Tukey       8.480501       35.32839			M Huber		8.489946	36.59841	70049.36	0.8662201	8.566459	36.45606	692169	0.8667404
LASSO       Original       250       8.308303       33.55028         S       8.673115       36.13586         M       8.455572       36.1402         MM       Bi-square       8.455572       36.1402         MM       Bi-square       8.383458       39.14316         MM       Huber       8.416911       37.77036         M       Huber       8.511842       36.06626         M       Huber       8.451854       36.13317         M       Huber       8.278559       37.88064         LASSO       Original       300       8.278559       37.88064         LASSO       Original       300       8.278559       37.88064         M       8.43050       35.32839         MM       Bi-square       8.278799       35.2839         MM       Huber       8.43050       35.32839         M       Huber       8.43050       35.32839         M       Huber       8.43050       35.32839         M       Huber       8.43050       35.32839         M       4.43050       35.32839         M       4.43050       35.3005			M Tukey		8.378494	37.40204	71587.51	0.8632825	8.572767	37.45262	71684.32	0.8630976
S 8.673115 36.13586 M 8.455572 36.1402 MM Bi-square 8.358216 37.72468 MM Hampel 8.383458 39.14316 MM Huber 8.416911 37.77036 M Huber 8.451842 36.06626 M Huber 8.451854 36.13317 M Tukey 8.358520 37.88064  LASSO Original 300 8.278559 33.16061 S 8.578799 35.64102 M 8.123120 37.28122 MM Bi-square 8.222075 37.03852 MM Hampel 8.331507 36.30777 MM Hampel 8.489325 35.51796 M Huber 8.489325 35.51796 M Huber 8.489325 35.51796 M Huber 8.489325 35.51796	Ι΄ _	LASSO	Original	250	8.308303	33.55028	64215.23	0.8773621	8.309037	33.68248	64468.27	0.8768788
M       8.455572       36.1402         MM       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Hampel       8.383458       39.14316         MM Hampel       8.416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.451854       36.13317         M       8.358520       37.88064         LASSO Original       300       8.278559       35.64102         M       8.43050       35.32839         MM       Bi-square       8.43050       37.28122         MM Hampel       8.331507       36.79287         M Huber       8.489325       35.51796         M Huber       8.489325       35.32839         M Tukey       8.430501       35.32839			S		8.673115	36.13586	69164.05	0.8679108	8.711497	36.35499	69583.45	0.8671099
MM       8.441291       39.48344         MM Bi-square       8.358216       37.72468         MM Hampel       8.416911       37.77036         MM Huber       8.416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         LASSO Original       300       8.278559       33.16061         S       8.43050       35.32839         MM       8.123120       37.28122         MM Hampel       8.331507       36.30777         MM Huber       8.489325       35.51796         M Huber       8.480501       35.32839         M Tukey       8.430501       35.32839			M		8.455572	36.1402	69172.34	0.867895	8.379303	38.59054	73862.29	0.8589381
MM Bi-square       8.358216       37.72468         MM Hampel       8.383458       39.14316         MM Huber       8.416911       37.77036         M Huber       8.511842       36.06626         M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         LASSO Original       300       8.278559       33.16061         S       8.43050       35.32839         MM       8.123120       37.28122         MM Hampel       8.331507       36.30777         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Tukey       8.430501       35.32839			MM		8.441291	39.48344	75571.3	0.8556743	8.384646	38.35996	73420.96	0.859781
MM Hampel       8.383458       39.14316         MM Huber       8.416911       37.77036         M Hampel       8.511842       36.06626         M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         Original       300       8.278559       37.16061         S       8.43050       35.32839         MM       8.123120       37.28122         MM Hampel       8.222075       37.03852         MM Huber       8.331507       36.30777         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Tukey       8.26074       35.2005			MM Bi-square		8.358216	37.72468	72205.03	0.8621032	8.459702	37.67484	72109.64	0.8622854
MM Huber       8.416911       37.77036         M Hampel       8.511842       36.06626         M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         Original       300       8.278559       33.16061         S       8.578799       35.64102         M       8.43050       35.32839         MM Bi-square       8.222075       37.03852         MM Huber       8.331507       36.30777         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Tukey       8.56074       35.32839			MM Hampel		8.383458	39.14316	74920.01	0.8569181	8.390179	38.38325	73465.55	0.8596959
M Hampel       8.511842       36.06626       6         M Huber       8.451854       36.13317       6         M Tukey       8.358520       37.88064       7         Original       300       8.278559       33.16061       6         S       8.578799       35.64102       6         M       8.43050       35.32839       6         MM       Bi-square       8.123120       37.28122         MM       Hampel       8.331507       36.30777         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Tukey       8.26014       35.2005			MM Huber		8.416911	37.77036	72292.47	0.8619362	8.439615	38.56573	73814.8	0.8590289
M Huber       8.451854       36.13317         M Tukey       8.358520       37.88064         Original       300       8.278559       33.16061         S       8.578799       35.32839         M       8.43050       35.32839         MM       8.123120       37.28122         MM       4331507       36.30777         MM       8.331507       36.30777         MM       448052       35.32839         M       448032       35.32839         M       448032       35.32839         M       448030       35.32839         M       443050       35.32839         M       443050       35.32839			M Hampel		8.511842	36.06626	69030.82	0.8681653	8.589337	36.16109	69212.33	0.8678186
M Tukey       8.358520       37.88064         Original       300       8.27859       33.16061       0         S       8.578799       35.64102       0         M       8.43050       35.32839       0         MM       8.123120       37.28122         MM Hampel       8.222075       37.03852         MM Huber       8.331507       36.30777         M Huber       8.489325       35.51796         M Huber       8.430501       35.32839         M Tuber       8.26014       36.2005			M Huber		8.451854	36.13317	69158.9	0.8679207	8.491465	36.10925	69113.11	0.8680081
Original         300         8.278559         33.16061         6           S         8.578799         35.64102         6           M         8.43050         35.32839         6           MM         8.123120         37.28122         7           MM Hampel         8.331507         37.03852         7           MM Huber         8.186330         36.79287         7           M Huber         8.489325         35.51796         6           M Turkay         8.26014         35.32839         6			M Tukey		8.358520	37.88064	72503.54	0.8615331	8.480347	38.21429	73142.15	0.8603135
8.578799 35.64102 6 8.43050 35.32839 6 8.123120 37.28122 aquare 8.222075 37.03852 apel 8.331507 36.30777 er er 8.489325 35.51796 6 8.430501 35.32839 6		LASSO	Original	300	8.278559	33.16061	63469.42	0.8787864	8.235131	33.01727	63195.06	0.8793104
8.43050 35.32839 6 8.123120 37.28122 aquare 8.222075 37.03852 apel 8.331507 36.30777 er er 8.489325 35.51796 el 8.480325 35.51796 6			S		8.578799	35.64102	68216.92	0.8697197	8.636376	35.57731	68094.98	0.8699525
8.123120 37.28122 quare 8.222075 37.03852 apel 8.331507 36.30777 er 8.186330 36.79287 el 8.489325 35.51796 6 8.430501 35.32839 6			M		8.43050	35.32839	67618.53	0.8708625	8.285714	36.94616	70714.94	0.8649489
quare 8.222075 37.03852  apel 8.331507 36.30777  er 8.186330 36.79287  el 8.489325 35.51796  g 2.60714 35.32839			MM		8.123120	37.28122	71356.25	0.8637242	8.327026	37.18961	71180.91	0.864059
er 8.331507 36.30777 (er 8.186330 36.79287 2el 8.489325 35.51796 (er 8.430501 35.32839 (			MM Bi-square		8.222075	37.03852	70891.72	0.8646113	8.247804	36.88418	70596.33	0.8651755
er 8.186330 36.79287 3 el 8.489325 35.51796 0 8.430501 35.32839 0			MM Hampel		8.331507	36.30777	69493.06	0.8672825	8.197567	37.64827	72058.79	0.8623825
el 8.489325 35.51796 o 8.430501 35.32839 o 8.260714 36.20005			MM Huber		8.186330	36.79287	70421.56	0.8655093	8.285629	36.92877	70681.66	0.8650125
8.430501 35.32839 6			M Hampel		8.489325	35.51796	67981.37	0.8701695	8.555303	35.37433	67706.47	0.8706945
26,000 36 1170 36,00005			M Huber		8.430501	35.32839	67618.54	0.8708624	8.488345	35.25388	67475.93	0.8711348
0.2007.14 50.29005			M Tukey		8.269714	36.29005	69459.16	0.8673472	8.349048	36.27451	69429.42	0.867404

Table 7: Metrics for model comparison between original (Ridge and LASSO) regression models and the best robust model for 50, 100, 150, 200, 250, and 300 high ranking variables, before and after heterogeneity.

ing Vari- able  Ridge 50  100  150  200  200  200  1 ASSO 50	Original MM Hampel Original MM Original MM									
	Original MM Hampel Original MM Original MM Original									
	Original MM Hampel Original MM Original	MAPE	MSE	SSE	$\mathbb{R}^2$		MAPE	MSE	SSE	$\mathbb{R}^2$
	MM Hampel Original MM Original MM	9.459094	41.59782	79618.24	0.8479455	Original	10.01975	45.19865	86510.21	0.8347832
	Original MM Original MM	8.801508	45.81388	87687.77	0.8325343	M	9.974874	48.1354	92131.16	0.8240484
	MM Original MM	8.304651	33.36347	63857.68	0.8780449	Original	8.998889	37.83964	72425.08	0.8616829
	Original MM	7.889334	34.66241	66343.86	0.8732968	MM	8.973277	39.5754	75747.32	0.8553381
	MM	7.893903	30.60797	58583.65	0.8881172	Original	8.511716	34.44371	65925.27	0.8740962
		7.562458	34.19317	65445.73	0.8750121	MM	8.273413	36.40819	69685.28	0.8669154
	Original	7.672882	29.16366	55819.24	0.8933967	Original	8.192899	31.97779	61205.49	0.8831101
	M Tukey	7.330010	33.59276	64296.54	0.8772068	MM Hampel	8.010758	34.84692	<i>L</i> 6999	0.8726224
	Original	7.625524	28.72482	54979.3	0.8950008	Original	8.163046	31.68415	60643.46	0.8841834
	M Tukey	7.310033	30.83707	59022.15	0.8872798	M Tukey	8.098266	34.7557	66522.41	0.8729558
	Original	7.063511	25.8776	49529.72	0.9054084	Original	7.019137	25.72886	49245.04	0.9059521
	MM Bi-square	6.826407	28.0242	53638.32	0.8975618	MM	6.962468	29.09346	55684.88	0.8936533
	Original	8.958306	38.86046	74378.92	0.8579515	Original	8.933586	38.57446	73831.51	0.8589969
	M Huber	8.968910	44.51771	85206.9	0.8372723	MM Bi-square	9.210072	43.5384	83332.51	0.840852
100	Original	8.823839	37.7268	72209.09	0.8620954	Original	8.811494	37.66045	72082.1	0.8623379
	MM Hampel	8.800533	45.13168	86382.03	0.835028	MM Bi-square	8.997942	43.99379	84204.11	0.8391874
150	Original	8.347495	34.26593	65584.99	0.8747461	Original	8.373180	34.29228	65635.42	0.8746498
	MM	8.457516	39.59102	75777.2	0.8552811	MM Huber	8.482080	38.69286	74058.14	0.8585641
200	Original	8.339226	33.78147	64657.73	0.876517	Original	8.328507	33.79357	64680.9	0.8764727
	MM Hampel	8.333713	37.74151	72237.25	0.8620417	MM	8.468777	39.51445	75630.66	0.8555609
250	Original	8.308303	33.55028	64215.23	0.8773621	Original	8.309037	33.68248	64468.27	0.8768788
	M Tukey	8.358520	37.88064	72503.54	0.8615331	M	8.379303	38.59054	73862.29	0.8589381
300	Original	8.278559	33.16061	63469.42	0.8787864	Original	8.235131	33.01727	63195.06	0.8793104
	MM	8.123120	37.28122	71356.25	0.8637242	MM Hampel	8.197567	37.64827	72058.79	0.8623825

Table 8: Comparison of the number and percentage of outliers for 2-sigma and 3-sigma limits for Ridge and LASSO with robust regression, both before and after, for 50, 100, 150, 200, 250, and 300 high-ranking variables.

ML	Robust method	High	ranking	Before !	Before heterogeneity	After hete	After heterogeneity
		variable					
				$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)	$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)
Ridge	Original	50		94(4.9112)	24(1.2539)	96(5.0157)	19(0.9927)
		100		93(4.8589)	25(1.3062)	93(4.8589)	20(1.0449)
		150		90(4.7022)	27(1.4107)	92(4.8067)	23(1.2017)
		200		95(4.9634)	27(1.4107)	95(4.9634)	22(1.1494)
		250		92(4.8067)	27(1.4107)	90(4.7022)	22(1.1494)
		300		94(4.9112)	25(1.3062)	93(4.8589)	26(1.3584)
	S estimator	50		94(4.9112)	24(1.2539)	100(5.2247)	24(1.2539)
		100		93(4.8589)	25(1.3062)	96(5.0157)	21(1.0972)
		150		90(4.7022)	27(1.4107)	90(4.7022)	23(1.2017)
		200		95(4.9634)	27(1.4107)	92(4.8067)	23(1.2017)
		250		89(4.6499)	28(1.4629)	96(5.0157)	23(1.2017)
		300		95(4.9634)	25(1.3062)	98(5.1202)	26(1.3584)
	M estimator	50		101(5.2769)	32(1.6719)	96(5.0157)	25(1.3062)
		100		95(4.9634)	28(1.4629)	90(4.7022)	24(1.2539)
		150		98(5.1202)	33(1.7241)	92(4.8067)	29(1.5152)
		200		98(5.1202)	33(1.7241)	92(4.8067)	31(1.6196)
		250		87(4.5455)	33(1.7241)	88(4.5977)	30(1.5674)
		300		95(4.9634)	28(1.4629)	97(5.0679)	28(1.4629)
	MM estimator	50		102(5.3292)	39(2.0376)	70(3.6573)	19(0.9927)
		100		101(5.2769)	48(2.5078)	55(2.8736)	31(1.6196)
		150		106(5.5381)	42(2.1944)	65(3.3960)	29(1.5152)
		200		107(5.5904)	42(2.1944)	100(5.2247)	32(1.6719)
		250		98(5.1202)	39(2.0376)	92(4.8067)	36(1.8809)
		300		94(4.9112)	39(2.0376)	99(5.1724)	38(1.9854)
	M Bi-square	50		94(4.9112)	24(1.2539)	70(3.6573)	19(0.9927)
		100		117(6.1129)	51(2.6646)	91(4.7544)	26(1.3584)
		150		107(5.5904)	42(2.1944)	93(4.8589)	34(1.7764)
		200		108(5.6426)	43(2.2466)	99(5.1724)	34(1.7764)
		250		96(5.0157)	38(1.9854)	93(4.8589)	36(1.8809)
		300		100(5.2247)	37(1.9331)	96(5.0157)	36(1.8809)
	M Hampel	50		127(6.6353)	63(3.2915)	122(6.3741)	56(2.9258)
		100		98(5.1202)	33(1.7241)	90(4.7022)	26(1.3584)
		150		105(5.4859)	42(2.1944)	95(4.9634)	34(1.7764)
		200		111(5.7994)	44(2.2989)	98(5.1202)	36(1.8809)
		250		105(5.4859)	43(2.2466)	86(4.4932)	36(1.8809)
		300		74(3.8662)	42(2.1944)	95(4.9634)	36(1.8809)

He 2D (%)   He 3D (%)   He 3	ML	Robust method	High ranking variable	Before h	Before heterogeneity	After hete	After heterogeneity
M Huber 50 1127(6.6553) 64(3.3438) 170 1100 1100 1100 1100 1100 1100 1100				$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)	$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)
100   103(5.3814)   47(2.4556)   50   100   101(5.3814)   47(2.4556)   50   100   101(5.5904)   43(2.2466)   50   101(5.5904)   34(2.2466)   50   101(5.5904)   34(2.2466)   50   101(5.5904)   34(2.2466)   50   50   50   50   50   50   50   5		M Huber	50	127(6.6353)	64(3.3438)	70(3.6573)	19(0.9927)
150   105(54859)   43(2.2466)   250   104(54359)   43(2.2466)   250   104(54359)   43(2.2466)   250   104(54359)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8809)   36(1.8609)   36(1.			100	103(5.3814)	47(2.4556)	93(4.8589)	25(1.3062)
200 107(5.5904) 43(2.2466) 300 104(5.4366) 44(2.2466) 300 101(6.2769) 36(1.8309) 36(1.83			150	105(5.4859)	43(2.2466)	95(4.9634)	33(1.7241)
250 104(5436) 43(2.2466) 360 104(5436) 43(2.2466) 360 101(5.2769) 36(1.8809) 57(1.8107) 50 20(3.4807) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 200 20(4.8067) 20(1.5122) 50 20(2.48067) 20(1.5122) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(1.6719) 50 20(2.48067) 20(2.247) 20			200	107(5.5904)	43(2.2466)	102(5.3292)	34(1.7764)
Mampel estimator   300   101(5.2769)   36(1.8809)   56   1724   294(1.5152)   50   100   93(4.8589)   27(1.4107)   100   93(4.8589)   37(1.7241)   294(1.5152)   250   250   86(4.4932)   32(1.6719)   250   86(4.4932)   32(1.6719)   250   94(4.8677)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.4629)   26(4.9634)   28(1.9634)   26(4.9634)   28(1.9634)   26(4.9634)   28(1.9634)   26(4.9634)   28(1.9634)   26(1.2639)   26(1.2639)   26(1.2639)   26(1.2634)   26(1.2639)   26(1.2639)   26(1.2639)   26(1.2639)   26(1.2634)   26(1.2639)			250	104(5.4336)	43(2.2466)	92(4.8067)	35(1.8286)
Hampel estimator 50 99(5.1724) 29(1.5152) 100 93(4.889) 27(1.4107) 100 93(4.889) 27(1.4107) 100 93(4.889) 27(1.4107) 100 93(4.889) 27(1.4107) 100 93(4.889) 27(1.4107) 100 93(4.899) 27(1.4107) 100 93(4.932) 32(1.6719) 100 93(4.932) 32(1.6719) 100 93(4.932) 32(1.6719) 100 93(4.932) 32(1.6719) 100 93(4.934) 33(1.724) 100 93(5.1202) 33(1.724) 100 93(5.1202) 33(1.724) 100 93(5.1202) 33(1.724) 100 93(5.1202) 33(1.724) 100 93(5.1202) 100(5.2347) 31(1.6196) 1100 100(5.2347) 31(1.6196) 1100 100(5.2347) 31(1.6196) 1100 100(5.2347) 31(1.6196) 1100 100(5.2347) 33(1.2046) 1100 100(5.2347) 33(1.2046) 1100 100(5.2347) 33(1.2046) 1100 100(5.2347) 33(1.2046) 1100 100(5.2347) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.2047) 33(1.204822) 33(1.204829) 33(1.32484) 33(1.2047) 33(1.2047) 33(1.204829) 33(1.32489) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.32489) 33(1.3241) 33(1.			300	101(5.2769)	36(1.8809)	96(5.0157)	36(1.8809)
100 93(4.889) 27(1.4107) 150 92(4.8867) 33(1.7241) 200 91(4.7549) 29(1.5152) 250 86(4.4932) 23(1.6719) 300 92(4.8067) 28(1.4629) 301 101(5.2769) 32(1.6719) 302 92(4.8067) 28(1.4629) 103 98(5.1202) 33(1.7241) 203 98(5.1202) 33(1.7241) 204 98(5.1202) 33(1.7241) 205 87(4.5455) 33(1.7241) 206 96(5.0157) 31(1.6196) 207 100(5.2341) 44(2.2989) 208 100(5.2347) 34(1.6196) 209 100(5.2347) 34(1.6196) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 100(5.2347) 34(1.629) 200 99(5.1724) 24(1.2339) 200 99(5.1724) 24(1.2339) 200 99(5.1724) 27(1.4107) 200 92(4.9634) 27(1.4107) 200 92(4.9634) 27(1.4107) 200 92(4.8067) 26(1.384) 200 92(4.8067) 26(1.384) 250 93(4.899) 26(1.384) 250 93(4.899) 26(1.388) 250 93(4.899) 26(1.3884) 250 93(4.8899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 26(1.3884) 250 93(4.899) 2		Hampel estimator	50	99(5.1724)	29(1.5152)	93(4.8589)	23(1.2017)
150   92(4.8067)   33(1.7241)   50   8(4.492)   20(1.512)   8(4.492)   32(1.6719)   8(4.492)   32(1.6719)   8(4.492)   32(1.6719)   8(4.492)   32(1.6719)   8(4.492)   32(1.6719)   8(4.492)   32(1.6719)   8(4.962)   32(1.6719)   32(1.671384)   32(1.6719)   32(1.			100	93(4.8589)	27(1.4107)	92(4.8067)	21(1.0972)
250 91(4.7544) 29(1.5152) 86(4.4932) 32(1.6719) 870 86(4.4932) 32(1.6719) 870 92(4.8067) 28(1.4629) 87(4.9634) 28(1.4629) 87(4.9634) 28(1.4629) 87(4.9634) 28(1.4629) 87(4.9634) 28(1.4629) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241) 98(5.1202) 33(1.7241)			150	92(4.8067)	33(1.7241)	91(4.7544)	26(1.3584)
250 86(4.4932) 32(1.6719) 8  100 92(4.8067) 28(1.4629) 92(4.8067) 28(1.4629) 92(4.8067) 28(1.4629) 92(4.8067) 28(1.4629) 92(4.9634) 28(1.4629) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.5674) 92(5.0157) 30(1.6196) 92(5.0157) 30(1.6196) 92(5.0157) 30(1.6196) 92(5.0157) 92(5.0158) 92(5.0157)			200	91(4.7544)	29(1.5152)	89(4.6499)	32(1.6719)
300         92(4.8067)         28(1.4629)         95           100         95(4.9634)         32(1.6719)         95           1100         95(4.9634)         28(1.4629)         95           150         96(5.1037)         30(1.5674)         95           200         96(5.0157)         30(1.5674)         95           200         96(5.0157)         30(1.5674)         95           200         87(4.5455)         33(1.7241)         95           300         100         95(4.9634)         28(1.4629)         95           100         100         95(4.9634)         28(1.4629)         95           200         100         95(4.9634)         28(1.4629)         95           200         100         105(5.4859)         43(2.2466)         95           200         105(5.4859)         14(1.609)         96         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106         96         106			250	86(4.4932)	32(1.6719)	86(4.4932)	28(1.4629)
Huber 50 101(5.2769) 32(1.6719) 89(4.9634) 128(1.4629) 150 98(4.9634) 28(1.4629) 150 98(5.1202) 33(1.7241) 200 95(4.9634) 28(1.4629) 150 98(5.1202) 33(1.7241) 150 98(5.1202) 33(1.7241) 150 95(4.9634) 28(1.4629) 150 95(4.9634) 28(1.4629) 100 100(5.2247) 31(1.6196) 150 100(5.2247) 31(1.6196) 150 100(5.2381) 43(2.2466) 150 100(5.2381) 43(2.2466) 150 100(5.2381) 43(2.2466) 150 100(5.2381) 43(2.2466) 150 100(5.2381) 43(2.2466) 150 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 100(5.2381) 1100 100(5.2381) 100(5.2391) 1100 100(5.2381) 100(5.2391) 1100 100(5.2381) 100(5.2391) 1100(5.239			300	92(4.8067)	28(1.4629)	96(5.0157)	27(1.4107)
100   95(4.9634)   28(1.4629)   550   200   98(5.1202)   33(1.7241)   500   98(5.1202)   33(1.7241)   520   98(5.1202)   33(1.7241)   520   98(5.0157)   30(1.5674)   520   95(4.9634)   28(1.4629)   520   1005(5.2487)   31(1.6196)   520   105(5.24859)   43(2.2466)   520   105(5.24859)   43(2.2466)   520   107(5.5904)   44(2.2989)   520   107(5.5904)   44(2.2989)   520   90(5.1724)   38(1.9854)   520   90(5.1724)   38(1.9854)   520   90(5.1724)   38(1.9854)   520   90(5.1724)   38(1.9854)   520   90(5.1724)   28(1.4629)   520   90(5.1724)   28(1.4629)   520   90(5.1724)   28(1.4029)   520   90(5.1724)   28(1.4017)   520   90(5.1724)   28(1.4017)   520   86(4.4932)   29(1.3584)   520   92(4.8067)   26(1.3584)   520   92(4.8067)   26(1.3584)   520   93(4.8589)   26(1.3584)   26(1.		Huber	50	101(5.2769)	32(1.6719)	96(5.0157)	25(1.3062)
150   98(5.1202)   33(1.7241)   50   50   50   50   50   50   50   5			100	95(4.9634)	28(1.4629)	90(4.7022)	24(1.2539)
250 96(5.0157) 30(1.5674) 250 87(4.5455) 33(1.7241) 800 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(5.4859) 43(2.2466) 95(5.4859) 43(2.2466) 95(5.4859) 43(2.2466) 95(5.1724) 95(2.0376) 99(5.1724) 38(1.9854) 95(5.1724) 38(1.9854) 95(5.0157) 28(1.4629) 99(5.1724) 38(1.9854) 95(5.0157) 28(1.4629) 99(5.1724) 38(1.9854) 95(4.4932) 23(1.2017) 95(4.4932)			150	98(5.1202)	33(1.7241)	92(4.8067)	29(1.5152)
250 87(4.5455) 33(1.7241) 8 300 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4629) 95(4.9634) 28(1.4659) 95(4.9634) 28(1.4659) 95(4.9634) 28(1.4659) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.2466) 95(2.0376) 9			200	96(5.0157)	30(1.5674)	92(4.8067)	31(1.6196)
300         95(4.9634)         28(1.4629)         95           Tukey         50         100(5.2477)         31(1.6196)         95           100         105(5.4859)         43(2.2466)         84           150         106(5.5381)         43(2.2466)         84           200         107(5.5904)         44(2.289)         95           250         99(5.1724)         38(2.0376)         96           250         99(5.1724)         38(1.9854)         96           100         89(4.6499)         30(1.5674)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.2539)         96           200         91(4.7544)         24(1.107)         96           200         92(4.964)         20(1.0449)         96           200         92(4.8067)         17(0.8882)         96           200         92(4.80			250	87(4.5455)	33(1.7241)	88(4.5977)	30(1.5674)
Tukey 50 100(5.2247) 31(1.6196) 51 100 105(5.4859) 43(2.2466) 81 150 105(5.4859) 43(2.2466) 8200 105(5.5381) 43(2.2466) 8200 107(5.5904) 44(2.2989) 92(5.1724) 39(2.0376) 92(5.1724) 39(2.0376) 92(5.1724) 39(2.0376) 92(5.1174) 39(2.0376) 92(5.1174) 38(1.3854) 92(5.1174) 92(1.6729) 92(5.1177) 9200 92(4.4932) 92(4.12539) 92(4.4932) 92(4.12539) 92(4.1724) 92(4.13584) 92(4.1411) 92(4.10449) 92(4.1411) 92(4.10449) 92(4.1411) 92(4.10449) 92(4.1411) 92(4.10449) 92(4.1411) 92(4.10449) 92(4.1411) 92(4.10449) 92(4.13584) 92(4.1411) 92(4.14107) 92(4.8067) 92(4.3647) 92(4.3584) 92(4.3688) 92(4.3584) 92(4.3688) 92(4.3688) 92(4.3584) 92(4.3688) 9			300	95(4.9634)	28(1.4629)	97(5.0679)	28(1.4629)
100 105(5.4859) 43(2.2466) 150 106(5.5381) 43(2.2466) 200 107(5.5904) 44(2.2989) 250 99(5.1724) 39(2.0376) 300 100(5.2247) 38(1.9854) 100 89(4.6499) 30(1.5674) 150 89(4.6499) 21(1.0972) 250 99(5.1724) 24(1.2539) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 150 99(5.1724) 26(1.3584) 250 86(4.4932) 19(0.9927) 150 99(5.1724) 20(1.0449) 250 86(4.4932) 19(0.9927) 150 92(4.8067) 17(0.8882) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584)		Tukey	50	100(5.2247)	31(1.6196)	71(3.7095)	19(0.9927)
150 106(5.5381) 43(2.2466) 200 107(5.5904) 44(2.2989) 250 99(5.1724) 39(2.0376) 300 100(5.2247) 38(1.9854) 100 89(4.6499) 30(1.5674) 250 99(5.0157) 28(1.4629) 250 89(4.6499) 21(1.0972) 250 89(4.6499) 21(1.0972) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 150 99(5.1724) 26(1.3584) 250 86(4.4932) 19(0.9927) 250 85(4.441) 20(1.0449) 250 86(4.4932) 17(0.8882) 300 92(4.8067) 17(0.8882) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 29(1.5152) 150 94(4.9112) 27(1.4107) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584)			100	105(5.4859)	43(2.2466)	89(4.6499)	26(1.3584)
200 107(5.5904) 44(2.2989) 250 99(5.1724) 39(2.0376) 300 100(5.2247) 38(1.9854) 100 89(4.6499) 30(1.5674) 150 89(4.6499) 21(1.0972) 200 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 150 99(5.1724) 26(1.3584) 200 85(4.441) 20(1.0449) 200 85(4.441) 20(1.0449) 200 86(4.4932) 19(0.9927) 200 86(4.4932) 19(0.9927) 200 92(4.8067) 17(0.8882) 200 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584)			150	106(5.5381)	43(2.2466)	95(4.9634)	33(1.7241)
250 99(5.1724) 39(2.0376) 360 100(5.2247) 38(1.9854) 100 100(5.2247) 38(1.9854) 100 80(5.0157) 28(1.4629) 28(1.4629) 200 89(4.6499) 21(1.0972) 200 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 200 91(4.7544) 23(1.2017) 200 91(4.7544) 23(1.2017) 200 99(5.1724) 26(1.3584) 200 99(5.1724) 20(1.0449) 200 85(4.441) 20(1.0449) 85(4.441) 20(1.0449) 85(4.441) 20(1.0449) 86(4.4932) 19(0.9927) 86(4.4932) 19(0.9927) 86(4.4932) 27(1.4107) 200 92(4.8067) 27(1.4107) 200 92(4.8067) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 80(4.7022) 23(1.2017) 8			200	107(5.5904)	44(2.2989)	98(5.1202)	32(1.6719)
300         100(5.2247)         38(1.9854)           Original         50         96(5.0157)         28(1.4629)           100         89(4.6499)         30(1.5674)         9           150         89(4.6499)         21(1.0972)         9           200         91(4.7544)         24(1.2539)         9           250         86(4.4932)         23(1.2017)         9           300         91(4.7544)         24(1.2539)         9           100         99(5.1724)         24(1.2017)         9           150         99(5.1724)         26(1.3584)         9           150         95(4.9634)         20(1.0449)         8           250         86(4.4932)         19(0.9927)         9           300         92(4.8067)         17(0.8882)         10           150         92(4.8067)         27(1.4107)         10           150         94(4.9112)         27(1.4107)         10           250         93(4.8067)         26(1.3584)         9           250         93(4.8589)         26(1.3584)         9           250         93(4.8589)         26(1.3584)         9           250         93(4.8067)         23(1.2017)			250	99(5.1724)	39(2.0376)	94(4.9112)	36(1.8809)
Original 50 96(5.0157) 28(1.4629) 96(5.0157) 28(1.4629) 97(1.6674) 97(1.6072)			300	100(5.2247)	38(1.9854)	101(5.2769)	38(1.9854)
100 89(4.649) 30(1.5674) 150 89(4.6499) 21(1.0972) 200 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 23(1.2017) 300 99(5.1724) 26(1.3584) 150 99(5.1724) 27(1.4107) 150 99(5.1724) 20(1.0449) 200 85(4.491) 20(1.0449) 86(4.4932) 19(0.9927) 86(4.4932) 19(0.9927) 300 92(4.8067) 17(0.882) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017) 8	LASSO	Original	50	96(5.0157)	28(1.4629)	96(5.0157)	29(1.5152)
150 89(4.6499) 21(1.0972) 200 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 23(1.2017) 86(4.4932) 23(1.2017) 150 99(5.1724) 26(1.3584) 200 85(4.441) 20(1.0449) 250 86(4.4932) 19(0.9927) 800 92(4.8067) 17(0.8882) 0r 50 104(5.436) 27(1.4107) 150 92(4.8067) 27(1.4107) 150 93(4.8087) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017) 8			100	89(4.6499)	30(1.5674)	91(4.7544)	30(1.5674)
200 91(4.7544) 24(1.2539) 250 86(4.4932) 23(1.2017) 300 91(4.7544) 23(1.2017) 300 91(4.7544) 23(1.2017) 50 99(5.1724) 26(1.3584) 150 99(5.1724) 26(1.3584) 200 85(4.441) 20(1.0449) 250 86(4.4932) 19(0.9927) 800 92(4.8067) 17(0.8882) or 50 104(5.436) 27(1.4107) 150 92(4.8067) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017)			150	89(4.6499)	21(1.0972)	98(5.1202)	27(1.4107)
250 86(4.4932) 23(1.2017) 8 300 91(4.7544) 23(1.2017) 8 100 99(5.1724) 26(1.3584) 9 150 99(5.1724) 27(1.4107) 9 250 85(4.441) 20(1.0449) 8 250 86(4.4932) 19(0.9927) 8 250 86(4.4932) 17(0.8882) 9 27(1.4107) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 1100 101(5.2769) 29(1.5152) 150 93(4.8057) 26(1.3584) 250 93(4.8589) 26(1.3584) 8 250 93(4.8589) 26(1.3584) 8			200	91(4.7544)	24(1.2539)	90(4.7022)	26(1.3584)
300 91(4.7544) 23(1.2017) 50 99(5.1724) 26(1.3584) 100 99(5.1724) 26(1.3584) 150 99(5.1724) 27(1.4107) 150 95(4.9634) 20(1.0449) 200 85(4.441) 20(1.0449) 86(4.4932) 19(0.9927) 800 92(4.8067) 17(0.8882) 0r 50 104(5.436) 27(1.4107) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017) 8			250	86(4.4932)	23(1.2017)	87(4.5455)	24(1.2539)
or 50 99(5.1724) 26(1.3584) 99(5.1724) 100 99(5.1724) 27(1.4107) 150 95(4.9634) 20(1.0449) 200 85(4.441) 20(1.0449) 86(4.4932) 19(0.9927) 86(4.4932) 19(0.9927) 86(4.4932) 17(0.882) 92(4.8067) 17(0.882) 92(4.8067) 17(0.882) 92(4.8067) 17(0.882) 92(4.9112) 150 94(4.9112) 27(1.4107) 150 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 1300 90(4.7022) 23(1.2017) 8			300	91(4.7544)	23(1.2017)	91(4.7544)	24(1.2539)
100 99(5.1724) 27(1.4107) 150 95(4.9634) 20(1.0449) 200 85(4.441) 20(1.0449) 250 86(4.4932) 19(0.9927) 300 92(4.8067) 17(0.8882) or 50 104(5.4336) 27(1.4107) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017)		S estimator	50	99(5.1724)	26(1.3584)	97(5.0679)	25(1.3062)
150 95(4.9634) 20(1.0449) 200 85(4.441) 20(1.0449) 250 85(4.441) 20(1.0449) 86(4.4932) 19(0.9927) 86(4.4932) 17(0.8882) 250 92(4.8067) 17(0.8882) 27(1.4107) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 8300 90(4.7022) 23(1.2017) 8			100	99(5.1724)	27(1.4107)	100(5.2247)	26(1.3584)
200 85(4.441) 20(1.0449) 8 250 86(4.4932) 19(0.9927) 8 300 92(4.8067) 17(0.8882) 9 100 104(5.4336) 27(1.4107) 150 94(4.9112) 27(1.4107) 150 92(4.8067) 26(1.3584) 9 250 93(4.8589) 26(1.3584) 1300 90(4.7022) 23(1.2017) 8			150	95(4.9634)	20(1.0449)	95(4.9634)	18(0.9404)
250 86(4.4932) 19(0.9927) 8 300 92(4.8067) 17(0.8882)  or 50 104(5.4336) 27(1.4107) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017) 8			200	85(4.441)	20(1.0449)	85(4.441)	18(0.9404)
300     92(4.8067)     17(0.8882)     5       or     50     104(5.4336)     27(1.4107)       100     101(5.2769)     29(1.5152)       150     94(4.9112)     27(1.4107)       200     92(4.8067)     26(1.3584)       250     93(4.8589)     26(1.3584)       300     90(4.7022)     23(1.2017)			250	86(4.4932)	19(0.9927)	83(4.3365)	17(0.8882)
or 50 104(5.436) 27(1.4107) 100 101(5.2769) 29(1.5152) 150 94(4.9112) 27(1.4107) 200 92(4.8067) 26(1.3584) 250 93(4.8589) 26(1.3584) 300 90(4.7022) 23(1.2017)			300	92(4.8067)	17(0.8882)	91(4.7544)	16(0.8359)
101(5.2769) 29(1.5152) 94(4.9112) 27(1.4107) 92(4.8067) 26(1.3584) 9 93(4.8589) 26(1.3584) 90(4.7022) 23(1.2017) 8		M estimator	50	104(5.4336)	27(1.4107)	106(5.5381)	28(1.4629)
94(4.9112) 27(1.4107) 92(4.8067) 26(1.3584) 9 93(4.8589) 26(1.3584) 9 90(4.7022) 23(1.2017) 8			100	101(5.2769)	29(1.5152)	100(5.2247)	28(1.4629)
92(4.8067) 26(1.3584) 93(4.8589) 26(1.3584) 1 90(4.7022) 23(1.2017) 8			150	94(4.9112)	27(1.4107)	107(5.5904)	27(1.4107)
93(4.8589) 26(1.3584) 1 90(4.7022) 23(1.2017) 8			200	92(4.8067)	26(1.3584)	99(5.1724)	28(1.4629)
90(4.7022) 23(1.2017)			250	93(4.8589)	26(1.3584)	111(5.7994)	38(1.9854)
			300	90(4.7022)	23(1.2017)	88(4.5977)	33(1.7241)

5					6	
ML	Kobust method	Hign ranking variable	Belore	before neterogeneity	Aiter nete	Aiter neterogeneity
			$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)	$\mu \pm 2\sigma$ (%)	$\mu \pm 3\sigma$ (%)
	MM estimator	50	105(5.4859)	36(1.8809)	99(5.1724)	33(1.7241)
		100	113(5.9039)	34(1.7764)	107(5.5904)	25(1.3062)
		150	111(5.7994)	37(1.9331)	110(5.7471)	35(1.8286)
		200	109(5.6949)	35(1.8286)	108(5.6426)	32(1.6719)
		250	110(5.7471)	37(1.9331)	111(5.7994)	39(2.0376)
		300	102(5.3292)	34(1.7764)	96(5.0157)	36(1.8809)
	M Bi-square	50	107(5.5904)	34(1.7764)	99(5.1724)	33(1.7241)
		100	105(5.4859)	33(1.7241)	92(4.8067)	23(1.2017)
		150	109(5.6949)	38(1.9854)	107(5.5904)	33(1.7241)
		200	106(5.5381)	34(1.7764)	104(5.4336)	33(1.7241)
		250	109(5.6949)	34(1.7764)	112(5.8516)	35(1.8286)
		300	112(5.8516)	41(2.1421)	96(5.0157)	36(1.8809)
	M Hampel	50	113(5.9039)	36(1.8809)	109(5.6949)	43(2.2466)
		100	96(5.0157)	29(1.5152)	105(5.5381)	29(1.5152)
		150	109(5.6949)	37(1.9331)	109(5.6949)	31(1.6196)
		200	90(4.7022)	34(1.7764)	107(5.5904)	34(1.7764)
		250	123(6.4263)	37(1.9331)	112(5.8516)	37(1.9331)
		300	95(4.9634)	35(1.8286)	90(4.7022)	38(1.9854)
	M Huber	50	107(5.5904)	33(1.7241)	102(5.3292)	32(1.6719)
		100	107(5.5904)	34(1.7764)	107(5.5904)	26(1.3584)
		150	109(5.6949)	36(1.8809)	108(5.6426)	31(1.6196)
		200	105(5.4859)	33(1.7241)	108(5.6426)	36(1.8809)
		250	114(5.9561)	31(1.6196)	116(6.0606)	38(1.9854)
		300	97(5.0679)	34(1.7764)	98(5.1202)	32(1.6719)
	Hampel estimator	50	102(5.3292)	27(1.4107)	100(5.2247)	27(1.4107)
		100	98(5.1202)	29(1.5152)	103(5.3814)	27(1.4107)
		150	90(4.7022)	25(1.3062)	97(5.0679)	23(1.2017)
		200	90(4.7022)	24(1.2539)	90(4.7022)	23(1.2017)
		250	91(4.7544)	24(1.2539)	88(4.5977)	21(1.0972)
		300	87(4.5455)	21(1.0972)	83(4.3365)	19(0.9927)
	Huber	50	105(5.4859)	29(1.5152)	106(5.5381)	27(1.4107)
		100	101(5.2769)	29(1.5152)	100(5.2247)	28(1.4629)
		150	94(4.9112)	27(1.4107)	106(5.5381)	27(1.4107)
		200	91(4.7544)	26(1.3584)	99(5.1724)	28(1.4629)
		250	93(4.8589)	27(1.4107)	99(5.1724)	28(1.4629)
		300	90(4.7022)	23(1.2017)	86(4.4932)	21(1.0972)
	Tukey	50	109(5.6949)	35(1.8286)	102(5.3292)	32(1.6719)
		100	115(6.0084)	33(1.7241)	108(5.6426)	25(1.3062)
		150	110(5.7571)	39(2.0376)	110(5.7571)	30(1.5674)
		200	99(5.1724)	31(1.6196)	103(5.3814)	31(1.6196)
		250	109(5.6949)	33(1.7241)	110(5.7571)	36(1.8809)
		300	101(5.2769)	32(1.6719)	95(4.9634)	29(1.5152)

Table 9: Comparison of the results from this study with previous studies.

Authors	Size of Variables	Machine Learning	Robust Method	Hybrid Model	MAPE	$R^2$	Challenges
Mukhtar et al. [4]	30	Random Forest, Support Vector Machine, Boosting	Bi-square, Hampel, Hu- ber	Random forest with Hampel	9.160917	0.838757	Irrelevant variables and Outliers
Mukhtar et al. [5]	30	Ridge, Lasso, Elastic Net	Bi-square, Hampel, Huber	Lasso with Hampel	9.174890	0.823023	Multicollinearity and Outliers
Ibidoja <i>et</i> al. [6]	15, 25, 35, 45	Random Forest, Support Vector Machine, Bagging, Boosting	M Bi-square, M Hampel, M Huber	Bagging with M Bi-square	8.151903	0.876975	Outliers
Ibidoja et al., [7]	15, 25, 35, 45	Ridge, Random Forest, Support Vector Machine, Bagging, Boosting, Lasso, Elastic Net	M Bi-square, M Hampel, M Huber, MM	Random forest with Hampel (Before het- erogeneity), Boosting with M Hampel (After heterogeneity)	2.12589, 8.228835	0.9732063,	Multicollinearity and Outliers (Before and after heterogeneity)
Ibidoja et al., [38]	15, 25, 35, 45	Ridge, Lasso, Elastic Net	S, MM, M Bi-square, M Hampel, M Huber	Lasso with M Bi-square (Single parameter added)	8.149872	0.8845778	Outliers (Before, after heterogeneity and single parameter added)
This study	50, 100, 150, 200, 250, 300	Ridge, Lasso	S, M, MM, MM Bisquare, MM Hampel, MM Huber, M Hampel, M Huber, M Tukey	Ridge with MM bi squares (Before hetero- geneity), Ridge with MM (After heterogene- ity), Lasso with MM (Before heterogeneity), Lasso with MM Hampel (After heterogeneity)	6.826407, 6.962468, 8.123120, 8.197567	0.897561, 0.8936533, 0.863724, 0.862382	Outliers (Before and after heterogeneity)

#### 9. Conclusion

The results indicate that the top-performing hybrid models across various conditions were: the best model are Ridge model with the MM bi squares before heterogeneity, the Ridge model with the MM method after heterogeneity and the Lasso model with the MM method before heterogeneity, the Lasso model with MM Hampel after heterogeneity. These models showed better prediction accuracy (lower MAPE) arises from its ability to reduce the influence of outliers, leading to more reliable predictions for most data points. However, this robustness results in a model that captures slightly less overall variance, reflected in the lower R<sup>2</sup>. Conversely, the original model captures more variance by fitting to all data points, including outliers, but at the cost of prediction accuracy for the majority of the data. For 2 sigma, the best model before heterogeneity is the Ridge model with the Hampel estimator before heterogeneity, while after heterogeneity the Lasso model with the S estimator . additionally, for 3-sigma limits the best model is the Lasso model with the S estimator both before and after heterogeneity .These models showed significantly better performance. This study's novelty is the combination methodology utilizing Ridge, Lasso, and robust regression techniques, effectively solving important problems in precision farming, including outliers and multicollinearity. This method has shown higher efficiency comparing with standard regression methods by improving prediction accuracy and model stability, especially in high-dimensional datasets. Future study require be focused on improving these robust models to deal with larger and more complex data, in addition to investigating their applicability in different agricultural environments. Developing these hybrid methodologies will enable the improvement of forecasting models for various agricultural systems and improving decision-making processes in agriculture. It demonstrates that hybrid models, which combine Ridge and Lasso regression with robust techniques such as MM, Hampel, and S estimators, could significantly improve prediction accuracy in precision agriculture. These models improve by minimizing the impact of outliers and effectively addressing multicollinearity, resulting in more accurate predictions. By focusing on the most significant factors in high-dimensional datasets, farmers may more effectively identify which variables (such as soil conditions, weather, and crop features) have a significant effect on crop yields. This improved comprehension facilitates more efficient decision-making, allowing farmers to allocate resources with more accuracy while controlling variability between their agricultural land more efficiently. Improving the accuracy of prediction, these models immediately assist expense savings and profit addition. Optimized forecasts assist farmers to maximize resource allocation, including water, fertilizers, and labor, by selecting locations with the highest possibility of production improvement. This reduces unnecessary costs and reduces the wastage of resources. Moreover, minimizing the effect of outliers enables farmers to stay away from reacting to infrequent or severe occurrences, hence improving decision-making reliability. The end result includes higher crop yields, more efficient application of resources, less operating costs, and finally, improved profitability in precision agriculture.

#### Acknowledgment

The author(s) sincerely thank the Department of Mathematical Sciences at USM.

#### Data availability

The below link provides access to the dataset, which includes all relevant data used for the analysis presented in this paper. <a href="https://studentusm-my.sharepoint.com/:x:/g/personal/nourabuafouna\_student\_usm\_my/">https://studentusm-my.sharepoint.com/:x:/g/personal/nourabuafouna\_student\_usm\_my/</a> EUtn38i8wqRKlevsc10OknIBLKYqngop2GOH8OO7PCaZVg? e=bTOUOL.

#### References

- S. Ghosh & R. Dasgupta, "Machine learning and precision farming", *Machine Learning in Biological Sciences*, R. Dasgupta, Springer, Singapore, 2022, pp. 239–249. https://doi.org/10.1007/978-981-16-8881-2\_28.
- [2] U. M. Durdağ, "Minimum-variance-based outlier detection method using forward-search model error in geodetic networks", Geosci. Model Dev 17 (2024) 2187. https://doi.org/10.5194/gmd-17-2187-2024.
- [3] S. Mahanto, R. Chattopadhyay, S. Kundu & S. Kanthal, "Precision farming: innovations, techniques and sustainability", International Journal of Agriculture Extension and Social Development 7 (2024) 42. https://doi.org/10.33545/26180723.2024.v7.i4a.513.
- [4] M. Mukhtar, M. K. B. M. Ali, A. Javaid, M. T. Ismail & A. Fudholi, "Accurate and hybrid regularization-robust regression model in handling multicollinearity and outlier using 8sc for big data", Mathematical Modelling of Engineering Problems 8 (2021) 547. https://doi.org/10.18280/ mmep.080407.
- [5] M. Mukhtar, M. K. M. Ali, M. T. Ismail, F. M. Hamundu, Alimuddin, N. Akhtar & A. Fudholi, "Hybrid model in machine learning-robust regression applied for sustainability in agriculture and food security", International Journal of Electrical and Computer Engineering (IJECE) 12 (2022) 4457. https://doi.org/10.11591/ijece.v12i4.pp4457-4468.
- [6] O. J. Ibidoja, F. P. Shan, J. Sulaiman & M. K. M. Ali, "Detecting heterogeneity parameters and hybrid models for precision farming", Journal of Big Data 10 (2023) 130. https://doi.org/10.1186/s40537-023-00810-8.
- [7] O. J. Ibidoja, F. P. Shan, M. Mukhtar, J. Sulaiman & M. K. M. Ali, "Robust m-estimators and machine learning algorithms for improving the predictive accuracy of seaweed contaminated big data", Journal of the Nigerian Society of Physical Sciences 5 (2023) 1137. https://doi.org/10.46481/jnsps.2022.1137.
- [8] W. H. Nugroho, N. W. S. Wardhani, A. A. R. Fernandes & Solimun, "Robust regression analysis study for data with outliers at some significance levels", Mathematics and Statistics 8 (2020) 373. https://doi.org/10.13189/ms.2020.080401.
- [9] R. R. Wilcox, "Robust Regression", in *Introduction to Robust Estimation and Hypothesis Testing*, Eds. R. R. Wilcox, Elsevier, Los Angeles, California, 2022, pp. 577–651. https://doi.org/10.1016/b978-0-12-820098-8.00016-6.
- [10] Y. Sorek and K. Todros, "Robust regression analysis based on the k-divergence", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Seoul, , Republic of Korea, 2024, pp. 9511–9515. https://doi.org/10.1109/ICASSP48485.2024.10447931.
- [11] D. A. Rahayu, U. F. Nursholihah,G. Suryaputra & S. Surono, "Comparison of the M, MM and S estimator in robust regression analysis on indonesian literacy index data 2018", Journal of Sciences and Data Analysis 4 11 (2023) https://doi.org/10.20885/EKSAKTA.vol4.iss1.art2.
- [12] T. Qureshi, M. Saeed, K. Ahsan, A. A. Malik, E. S. Muhammad & N. Touheed, "Smart agriculture for sustainable food security using internet of things (IoT)", Wireless Communications and Mobile Computing 6 (2022) 608394. https://doi.org/10.1155/2022/9608394.

- [13] A. Ikram, W. Aslam, R. Aziz, F. Noor, G. Mallah, S. Ikram, M. A. Saeed, A. Abdullah & I. Ullah, "Crop yield maximization using an iot-based smart decision", Journal of Sensors 2022 (2022) 1. https://doi.org/10. 1155/2022/2022923.
- [14] L. Rabhi, N. Falih, L. Afraites & B. Bouikhalene, "A functional framework based on big data analytics for smart farming", Indonesian Journal of Electrical Engineering and Computer Science 24 (2023) 1772. https://doi.org/10.11591/ijeecs.v24.i3.pp1772-1779.
- [15] P. R. Kumar and M. K. M. Ali and O. J. Ibidoja, "Identifying heterogeneity for increasing the prediction accuracy of machine learning models", Journal of the Nigerian Society of Physical Sciences 6 (2024) 2058. https://doi.org/10.46481/jnsps.2024.2058.
- [16] S. Prasad, "Regression", in Advanced Statistical Methods, Eds. S. Prasad, Springer, Singapore, 2024, pp. 1–45. https://doi.org/10.1007/ 978-981-99-7257-9
- [17] D. C. Montgomery, E. A. Peck & G. G. Vining, "Introduction to linear regression analysis", John Wiley & Sons, Inc., New York, United States, 2021, pp. 71–78. https://content.e-bookshelf.de/media/reading/L-16125104-1a3a7c5bd1.pdf.
- [18] A. Zulkarnain, S. W. Rizki & H. Perdana, "Analisis regresi robust estimasi-MM dalam mengatasi pencilan pada regresi linear berganda", Bimaster: Buletin Ilmiah Matematika, Statistika Dan Terapannya 9 (2020) 123. http://doi.org/10.26418/bbimst.v9i1.38666.
- [19] N. E. Jeremia, S. Nurrohmah & I. Fithriani, "Robust Ridge regression to solve multicollinearity and outlier", Journal of Physics: Conference Series 1442 (2020) 012030. https://doi.org/10.1088/1742-6596/1442/1/ 012030.
- [20] M. N. A. Singgih & A. Fauzan, "Comparison of M estimation, S estimation, with MM estimation to get the best estimation of robust regression in criminal cases in Indonesia", Jurnal Matematika, Statistika Dan Komputasi 18 (2022) 251. https://doi.org/10.20956/j.v18i2.18630.
- [21] M. Mukhtar, M. K. M. Ali, M. T. Ismail, F. M. Hamundu, Alimuddin, N. Akhtar & A. Fudholi, "Hybrid model in machine learning–robust regression applied for sustainability in agriculture and food security", International Journal of Electrical and Computer Engineering 12 (2022) 4457. https://doi.org/10.11591/ijece.v12i4.pp4457-4468.
- [22] C. Lim, P. K. Sen & S. D. Peddada, "Robust nonlinear regression in applications", Journal of the Indian Society of Agricultural Statistics 67 (2013) 215. https://pubmed.ncbi.nlm.nih.gov/25580021/.
- [23] R. Finger & W. Hediger, "The application of robust regression to a production function comparison the example of swiss corn", IED Working Paper 2 (2009) 1. http://dx.doi.org/10.2139/ssrn.1430342.
- [24] P. Hasih, Y. Susanti & S. S. Handajani, "A robust regression by using huber estimator and tukey bisquare estimator for predicting availability of corn in karanganyar regency, indonesia", Indonesian Journal of Applied Statistics 1 (2018) 398. https://doi.org/10.13057/IJAS.V1II.24090.
- [25] F. Adewale, L. Olatunji & K. Ayinde, "Some robust ridge regression for handling multicollinearity and outlier", International Journal of Sciences: Basic and Applied Research (IJSBAR) 16

- (2014) 192. https://www.researchgate.net/publication/313724168\_Some\_Robust\_Ridge\_Regression\_for\_handling\_Multicollinearity\_and\_Outlier.
- [26] S. Peng, G. Tarr, S. Müller & S. Wang, "CR-Lasso: Robust cellwise regularized sparse regression", Computational Statistics & Data Analysis, 197 (2024) 107971 https://doi.org/10.1016/j.csda.2024.107971.
- [27] M. Xu, "Sales prediction based on lasso regression", Highlights in Science, Engineering and Technology 88 (2024) 343. https://doi.org/10.54097/p9hyrk70.
- [28] A. Khanna, F. Lu & E. Raff, "Sparse private lasso logistic regression", arXiv (2023) https://doi.org/10.48550/arXiv.2304.12429.
- [29] Y. Susanti, H. Pratiwi, S. Sulistijowati & T. Liana, "M Estimation, S estimation, and MM estimation in robust regression", International Journal of Pure and Applied Mathematics 91 (2014) 349. http://dx.doi.org/10.12732/ijpam.v91i3.7.
- [30] E. M. Almetwally & H. Mohamed and A. Almongy, "Comparison between M-estimation, S-estimation, and MM estimation methods of robust estimation with application and simulation", International Journal of Mathematical Archive 9 (2018) 55. https://www.researchgate.net/publication/328335899.
- [31] P. Rousseeuw & V. J. Yohai, "Robust regression by means of s estimators" in *Robust and Nonlinear Time Series Analysis*, Eds. J. Franke and W. Hardle and D. Martin, Springer, New York, 1984, pp. 256–274. https://doi.org/10.1007/978-1-4615-7821-5\_15
- [32] P. Exterkate, P. J. F. Groenen, C. Heij & D. van Dijk, "Nonlinear fore-casting with many predictors using kernel ridge regression", International Journal of Forecasting 32 (2016) 736. https://doi.org/10.1016/j.ijforecast. 2015.11.017.
- [33] J. Rougier, "Ensemble averaging and mean squared error", Journal of Climate 29 (2016) 8865. https://doi.org/10.1175/JCLI-D-16-0012.1.
- [34] J. Padrul, R. Dedi, D. Epha & S. Supandi, "Comparison of robust estimation on multiple regression model", Journal of Mathematics and Its Applications 17 (2013) 0979. https://doi.org/10.30598/ barekengvol17iss2pp0979-0988.
- [35] C. Tirink & H. Önder, "Comparison of M, MM and LTS estimators in linear regression in the presence of outlier", Turkish Journal of Veterinary & Animal Sciences 46 (2022) 420. https://doi.org/10.55730/1300-0128.
- [36] A. Tatlıyer, "The effects of raising type on performances of some data mining algorithms in lambs", Journal of Agriculture and Nature 23 (2020) 772. https://doi.org/10.18016/ksutarimdoga.vi.651232.
- [37] D. M. Khan, M. Ali, Z. Ahmad, S. Manzoor & S. Hussain, "A new efficient redescending m-estimator for robust fitting of linear regression models in the presence of outliers", Mathematical Problems in Engineering 2023 (2023) 1. https://doi.org/10.1155/2021/3090537.
- [38] O. J. Ibidoja, F. P. Shan & M. K. M. Ali, "Modified sparse regression to solve heterogeneity and hybrid models for increasing the prediction accuracy of seaweed big data with outliers", Scientific Reports 14 (2024) 17599. https://doi.org/10.1038/s41598-024-60612-7.