Science Journal of University of Zakho
DOI: https://doi.org/10.25271/sjuoz

Original Article

MODIFIED FUZZY-ROBUST RIDGE REGRESSION FOR MULTICOLLINEAR, OUTLIER-CONTAMINATED DATA

Vaman M. Salih[1], and Shelan S. Ismaeel[1]

[1] Department of Mathematics, College of Science, University of Zakho, Zakho, Kurdistan Region, Iraq.

**Corresponding author, E-mail*: shelan.ismaeel@uoz.edu.krd (Tel:+964-7504718846)

Received: 17, May, 2025

Accepted: 14, Jan, 2026

Published: 09, Apr, 2026

ABSTRACT

Multicollinearity is known to have a significant impact on the stability of linear regression parameter estimation, while the presence of outliers tends to compound this problem. Ridge regression helps to improve the multicollinearity problem, but it is highly sensitive to outliers. This paper proposes Modified Fuzzy Robust Ridge Regression (MFRRR), which modifies classical ridge regression by adapting the penalty parameter (k) through modified fuzzy robust estimators based on weighted residual membership functions. The method is evaluated under challenging data conditions involving simultaneous multicollinearity, outliers, and fuzzy uncertainty. Performance is assessed using both a real body fat dataset and Monte Carlo simulations with varying sample sizes $(N = 20, 50, 100, 150, 200)$, correlation levels $(\rho = 0.5, 0.8, 0.99)$, and contamination rates $(5\%, 10\%, 15\%, 20\%)$. MFRRR is compared to ordinary least squares (OLS), ridge regression, and robust ridge regression based on the mean absolute error (MAE) as an evaluation criterion. These findings indicate that MFRRR is always associated with smaller prediction errors and more reliable parameter estimates, especially when there is high multicollinearity and data contamination.

KEYWORDS: Multicollinearity, Outliers, Ridge regression, Robust regression, Fuzzy robust regression.

1. INTRODUCTION

The simultaneous presence of multicollinearity, outliers, and uncertainty in the data makes regression analysis very challenging and can often lead to unstable estimates of the parameters and unreliable inferential statements whenever the traditional methods like OLS are used. Fuzzy regression has been extensively used in modeling imprecise or uncertain data that cannot be properly addressed through standard numeric systems (Tsai & Wu, 2002). Within this framework, ridge-type regularization has been extended to fuzzy settings to alleviate multicollinearity and stabilize parameter estimates (Hong & Hwang, 2004; Hong *et al.*, 2004; Choi *et al.*, 2019; Rabiei *et al.*, 2019Kim & Jung, 2020; Karbasi *et al.*, 2021; Kareem & Mohammed, 2023). Nevertheless, the majority of fuzzy ridge regression models are mainly sensitive to multicollinearity and are sensitive to outliers. Simultaneously, robust regression and robust ridge-type algorithms have been designed so that uncertainty in data is mitigated independently (Akbari & Hesamian, 2019; Hesamian & Akbari, 2020; Farnoosh *et al.*, 2020; Bas & Egrioglu, 2024) although they do not offer an adaptive regularization mechanism based on fuzzy residual information. Thus, a major gap in the literature that exists to date with regard to a single methodology that concurrently considers fuzzy regression, resistance to outliers, and ridge regularization via an adaptive penalty framework is created. In an attempt to mitigate this deficiency, the current research comes up with MFRRR, a new regression structure using adaptive estimation of the shrinkage parameter k using fuzzy robust residuals obtained based on modified fuzzy membership functions.

The given methodology combines the benefits of fuzzy modeling, robust estimation, and ridge regularization into a single estimation process, and it aims at producing more reliable coefficient estimates in a scenario with data uncertainty. The key contributions of the current work are as follows: (i) the development of a novel fuzzy robust ridge regression framework with adaptive choice of penalty; (ii) high resistance to multicollinearity and aberrant observations; and (iii) comprehensive empirical

Access this article online

https://doi.org/10.25271/sjuoz.2026.14.2.1582

Science Journal of University of Zakho
Vol. 14, No. 02, pp. 209 –219 April-2026

Printed ISSN 2663-628X;
Electronic ISSN 2663-6298

validation by real-life data and Monte Carlo modeling, which altogether prove the high effectiveness of the proposed design as compared to the regular least squares, standard ridge regression and the current methods of robust ridge regression. The remainder of the paper is organized as follows: Section 2 presents the material and methods, including ridge regression, robust ridge regression, and the suggested methodology. Section 3 gives the practical implementation through numbers and the Monte Carlo simulation study. Finally, Section 4 provides the concluding remarks.

2. MATERIAL AND METHODS

Ridge Regression:

Ridge regression is a regularization technique applied to linear regression models to mitigate the effects of multicollinearity. It modifies the OLS estimator by introducing an ℓ_2 penalty that shrinks regression coefficients toward zero, thereby improving numerical stability. Prior to estimation, the response vector Y and the design matrix X are centered and standardized, and the intercept term is excluded from the penalty. The ridge regression estimator is obtained by minimizing the penalized loss function:

$$\min_{\beta} \|Y - X\beta\|_2^2 + k\|\beta\|_2^2 , \qquad (1)$$

where $X \in \mathbb{R}^{n\times p}$, $Y \in \mathbb{R}^{n}$, and $k > 0$ denotes the ridge penalty parameter. The closed-form solution is given by

$$\hat{\beta}_{Ridge} = \left(X^T X + kI_p\right)^{-1} X^T Y. \qquad (2)$$

To determine the penalty parameter k, the classical method of Hoerl and Kennard (1970) is employed:

$$k = \frac{p\,\hat{\sigma}^2_{\mathrm{OLS}}}{\hat{\beta}^T_{OLS}\,\hat{\beta}_{OLS}}, \qquad (3)$$

where p is the number of predictors and $\hat{\sigma}^2_{\mathrm{OLS}} = \frac{\sum_{i=1}^{n} {r_i}^2}{\mathrm{n-p}}$. Both $\hat{\sigma}^2_{\mathrm{OLS}}$ and $\hat{\beta}_{\mathrm{OLS}}$ are obtained from the OLS estimator.

Robust Ridge Regression:

Robust Ridge Regression (RRR) combines the idea of robust estimation with ridge regression in order to reduce the influence of anomalous data being used and still maintains the conventional ridge regression model. In this approach, the ridge penalty parameter is estimated to ensure a robust system is created instead of weighting the design matrix itself. This ensures that the numerical stability is maintained with the presence of multicollinearity, which at the same time reduces the sensitivity of the penalty term to extreme observations (Ismaeel *et al.*, 2024). The RRR estimator retains the standard ridge regression form

$$\hat{\beta}_{RRR} = \left(X^T X + k_{\mathrm{Robust}} I_p\right)^{-1} X^T Y, \qquad (4)$$

where k_{Robust} is a robust ridge parameter estimated using robust residual information. Analogous to Eq. (3), the robust version of the shrinkage parameter is computed as

$$k_{\mathrm{Robust}} = \frac{p\,\hat{\sigma}^2_{\mathrm{Robust}}}{\hat{\beta}^T_{\mathrm{Robust}}\,\hat{\beta}_{\mathrm{Robust}}}, \qquad (5)$$

where p is the number of predictors and $\hat{\sigma}^2_{\mathrm{Robust}} = \frac{\sum_{i=1}^{n} \omega(r_i)\, {r_i}^2}{\mathrm{n-p}}$.

Here, $\omega(r_i)$ denote robust weights derived from M-estimation. In this study, Tukey's bisquare loss function is employed, with corresponding influence function $\psi(\cdot)$, robust scale estimator s, and tuning constant $c = 4.685$. The robust residual weights $\omega(r_i) = \frac{\psi(r_i/(\mathrm{cs}))}{r_i/(\mathrm{cs})}$ are used only for estimating the penalty parameter k_{Robust}, while the coefficient estimation itself follows the classical ridge regression formulation.

The Proposed Method:

The MFRRR is a combination of ridge regression and modified fuzzy robust regression to handle the problems of multicollinearity and outliers simultaneously. In this method, the penalty parameter k is computed based on modified fuzzy robust regression estimators, while the final coefficient estimation preserves the classical ridge regression structure. This approach reduces the impact of both issues in linear regression models. The MFRRR estimator is given by

$$\hat{\beta}_{MFRRR} = \left(X^T X + k_{MFR} I_p\right)^{-1} X^T Y \qquad (6)$$

where k_{MFR} is an adaptive ridge penalty parameter determined using a modified fuzzy robust method. Analogous to Eq. (3), the modified fuzzy robust version of the shrinkage parameter is computed as

$$k_{MFR} = \frac{p\,\hat{\sigma}^2_{\mathrm{MFR}}}{\hat{\beta}^T_{MFR}\,\hat{\beta}_{MFR}}, \qquad (7)$$

where p is the number of independent variables and $\hat{\sigma}^2_{MFR} = \frac{\sum_{i=1}^{n} \mu(r_i)\, {r_i}^2}{\mathrm{n-p}}$, with $\mu(r_i)$ representing the degree of membership for each residual. Both $\hat{\sigma}^2_{MFR}$ and $\hat{\beta}_{MFR}$ are obtained through a modified fuzzy robust estimation procedure (as described in Salih & Ismaeel, 2025). These quantities are subsequently used to compute the ridge penalty k_{MFR}, while the final regression coefficients are estimated via the MFRRR estimator in Eq. (6), which retains the classical ridge structure with an adaptively estimated fuzzy–robust penalty parameter. The modified fuzzy robust component used to estimate k_{MFR} is summarized in the following algorithm:

Step 1: Express the fuzzy regression framework in this way:

$$Y = \beta_0 + \sum_{j=1}^{p} \beta_j X_j ,$$

where $X_i = (X_{li}, X_{mi}, X_{ri})$ and $Y_i = (Y_{li}, Y_{mi}, Y_{ri})$; $i = 1,2,3,4,\ldots, n$, are triangular fuzzy numbers. The crisp coefficients $\beta_0, \beta_1, \ldots, \beta_p$ are estimated by minimizing the following loss function:

$$\min r(b_0, b_1, b_2, b_3, \ldots, b_p) = \sum d\left(b_0 + b_1 X_{i1} + b_2 X_{i2} + b_3 X_{i3} + \ldots + b_p X_{ip}, Y_i\right)^2,$$

where *d(.,.)* denotes the distance between fuzzy numbers. The initial estimation of the coefficient vector is computed as:

$$\hat{\beta}_F = (X_L^T X_L + X_M^T X_M + X_R^T X_R)^{-1} (X_L^T Y_L + X_M^T Y_M + X_R^T Y_R).$$

Step 2: Determine the residual values r_i.

Step 3: Evaluate the Huber median from the absolute residuals and compute the distance:

$$D_i = \|u_i - v_i\|, \quad i = 1,2,3,4,\ldots, n,$$

where $\|\cdot\|$ is the Euclidean distance, $u_i = |r_i|$ and v_i is Huber median. Huber median is selected as a robust location estimator to reduce bias from extreme outliers, providing a more stable center than the arithmetic mean in contaminated datasets.

Step 4: Identify the membership function $\mu(r_i)$ according to the rule:

$$\mu(r_i) = \begin{cases} 1 & , |r_i| \le \text{median}(D_i) \\ \dfrac{\max(D_i) + \dfrac{\text{median}|r_i - \text{median}(r_i)|}{0.6745} - |r_i|}{\max(D_i) + \dfrac{\text{median}|r_i - \text{median}(r_i)|}{0.6745} - \text{median}(D_i)} & , \text{median}(D_i) < |r_i| < \max(D_i) + \dfrac{\text{median}|r_i - \text{median}(r_i)|}{0.6745} \\ 0 & , \text{otherwise} \end{cases}$$

The constant 0.6745 is derived from the 75th percentile of the standard normal distribution, ensuring a scale consistent with standard deviation for normally distributed data. The function smoothly down-weights influential residuals without discarding them.

Step 5: Construct the diagonal weight matrix:

$$\omega = diag(\mu(r_1), \mu(r_2), \mu(r_3), \ldots, \mu(r_n)).$$

Step 6: The $\hat{\beta}_{MFR}$ coefficient is estimated using the weighted fuzzy least squares method through the following formula:

$$\hat{\beta}_{MFR} = (X_L^T \omega X_L + X_M^T \omega X_M + X_R^T \omega X_R)^{-1} (X_L^T \omega Y_L + X_M^T \omega Y_M + X_R^T \omega Y_R).$$

Step7: Repeat Steps 2–6 until convergence. A tolerance of 0.000005 is used, which was found to provide stable convergence without excessive computational cost.

Remarks: Membership weights $\mu(r_i)$ directly influence both the residual scale $\hat{\sigma}^2_{MFR}$ and the ridge penalty k_{MFR}, integrating fuzzy robustness into the ridge framework through adaptive penalty selection.

3. APPLICATION

In this section, a numerical example and a Monte Carlo simulation study are presented to evaluate the performance of the proposed MFRRR method, comparing it with other approaches. In both the numerical example and the simulation study, the independent (X) and dependent (Y) variables were fuzzified using asymmetric triangular fuzzy numbers. Each fuzzy number was centered at its corresponding crisp value, with a left spread of 20% and a right spread of 10%. This fuzzification was applied only to the MFRRR method after generating the datasets, in order to evaluate its performance under fuzzy data conditions. The numerical analyses and Monte Carlo simulation studies were conducted using the R programming environment.

Numerical Example:

To evaluate the proposed model, the first 50 records of male body composition from the real-world dataset provided by Penrose *et al.* (1985) were analyzed. Age, height, weight, neck circumference, chest circumference, abdomen circumference, and hip circumference are the seven independent factors in the dataset, with the percentage of body fat serving as the response variable. Figure 1 depicts the scatter plot matrix for visualizing the relationship between each pair of variables in the body fat data set. The diagonal elements of this scatter plot matrix contain kernel density estimates of each variable. The upper triangle of this scatter plot matrix contains Pearson correlation coefficients and their significance levels, and the lower triangle contains scatter plots of each pair of variables. There is a strong positive relationship between weight and hip measurements ($r = 0.959$, $p < 0.001$), chest and abdomen measurements ($r = 0.942$, $p < 0.001$), and chest and hip measurements (r = 0.911, p < 0.001). The high correlation coefficients imply that there could be multicollinearity in these variables. There is also a positive relationship between weight and measurements of the abdomen ($r = 0.915$, $p < 0.001$) and chest ($r = 0.912$, $p < 0.001$). This could imply multicollinearity in the variables. Finally, height is highly unrelated to other variables. For example, abdomen measurements ($r = -0.052$) and hip measurements ($r = -0.045$). Clearly, height is independent of the data.

There are no obvious outliers and nonlinear trends in the scatter plots, and most observations seem to be relatively equally spaced. However, further diagnostic analyses should be performed to confirm that there are no influential observations or extreme values. To investigate this, a boxplot illustrating the distributions of the variables in the body fat dataset was generated. Figure 2 illustrates that the dependent variable and the independent variable of age do not exhibit any outliers; whereas, the remaining independent variables — weight, height, neck, chest, abdomen and hip — display outliers. Consequently, outliers are present in the body fat dataset. Furthermore, since the maximum variance inflation factor (VIF) value exceeds the commonly accepted threshold of 10, the dataset demonstrates a serious multicollinearity problem (Belsley, 1991). The calculated VIF values are 1.5628, 27.352, 1.5132, 5.3440, 10.892, 16.944, and 25.033. These results clearly confirm the presence of multicollinearity among independent variables.

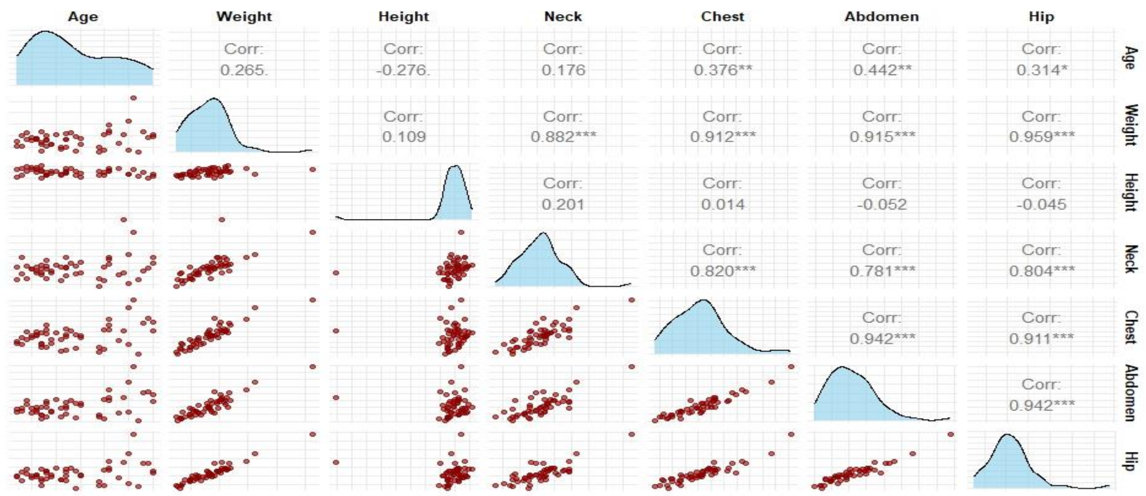

Figure 1: Pairwise scatter plot matrix with Pearson correlation coefficients among independent variables for the dataset on body fat.

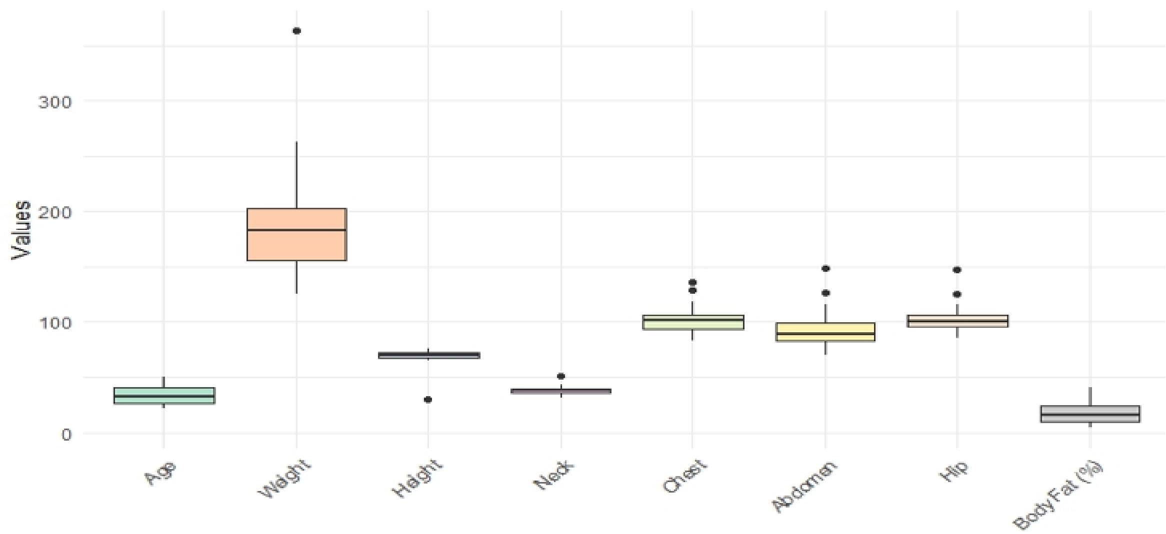

Figure 2: Boxplots of the variables in the body fat dataset.

Table 1 presents the parameter estimates, standard errors and MAE for the OLS, ridge regression, RRR and the proposed MFRRR methods applied to the body fat dataset.
The results clearly show that the OLS method performs poorly under such data conditions, as indicated by the high MAE value (3.510231) and unstable parameter estimates with large standard errors. This is expected due to the high sensitivity of OLS to multicollinearity and outliers, which heavily distort its estimation performance. Ridge regression and RRR substantially improve estimation stability, yielding much lower MAE values (0.051208 and 0.051173, respectively). In contrast, the proposed MFRRR method demonstrates the best overall performance. It achieves the highest predictive accuracy with a MAE of 0.049396. Moreover, its parameter estimates are less sensitive to multicollinearity and outliers, and its standard errors are smaller or comparable to those of Ridge and RRR, verifying the resilience of MFRRR to both issues. As shown in Fig. 3 and Fig. 4, OLS produces large deviations and high standard errors, particularly for β_0; whereas, MFRRR yields more stable estimates with lower standard errors, further demonstrating its robustness.

Table 1: Parameter estimates, standard errors and MAE of OLS, Ridge, RRR and MFRRR methods for the body fat dataset.

$\hat{\beta}_j$	OLS		Ridge Reg.		RRR		MFRRR	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
$\hat{\beta}_0$	-14.102	26.577	-0.0001	0.0093	-0.0001	0.0093	-0.0001	0.0090
$\hat{\beta}_1$	0.0258	0.0964	0.0123	0.0818	0.0134	0.0818	0.0159	0.0820
$\hat{\beta}_2$	-0.1327	0.0842	-0.5211	0.3438	-0.5237	0.3438	-0.5292	0.3439
$\hat{\beta}_3$	-0.1726	0.1258	-0.1233	0.0802	-0.1218	0.0803	-0.1187	0.0804
$\hat{\beta}_4$	-0.2013	0.4556	-0.0724	0.1522	-0.0718	0.1521	-0.0705	0.1521
$\hat{\beta}_5$	-0.0639	0.2014	-0.0693	0.2173	-0.0693	0.2173	-0.0692	0.2173
$\hat{\beta}_6$	1.0922	0.1838	1.6174	0.2710	1.6165	0.2710	1.6146	0.2709
$\hat{\beta}_7$	-0.1867	0.3097	-0.2264	0.3284	-0.2230	0.3285	-0.2160	0.3287
MAE	3.510231		0.051208		0.051173		0.049396	

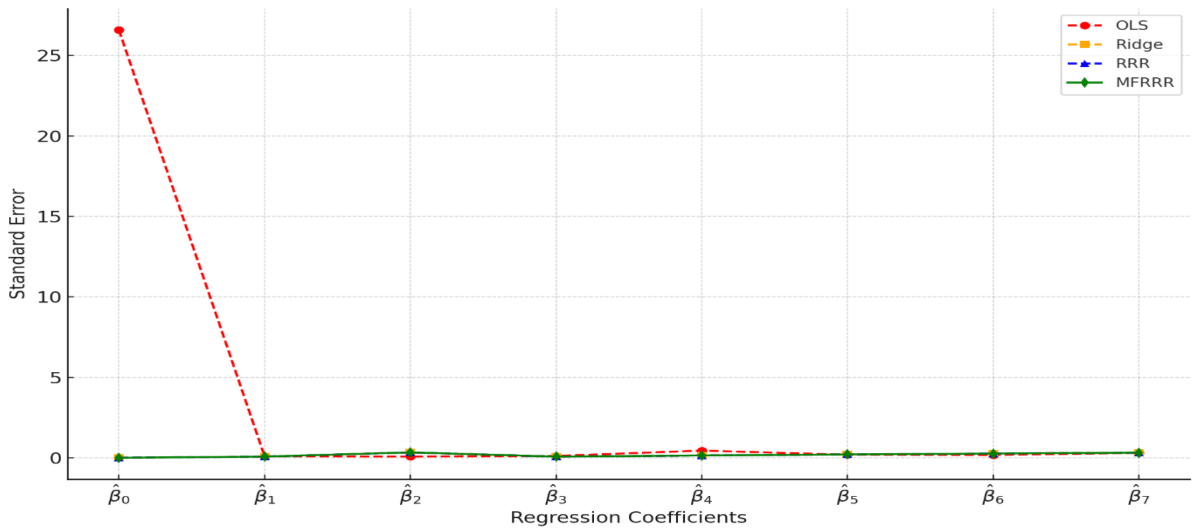

Figure 3: Parameter estimate standard errors for OLS, Ridge, RRR and MFRRR.

The OLS method exhibits extremely large standard errors, particularly for the intercept β_0, indicating severe instability caused by multicollinearity and the presence of outliers. Ridge regression and RRR substantially reduce estimation variability across coefficients. In contrast, the proposed MFRRR method yields consistently smaller and more stable standard errors for all parameters, demonstrating enhanced robustness and numerical stability.

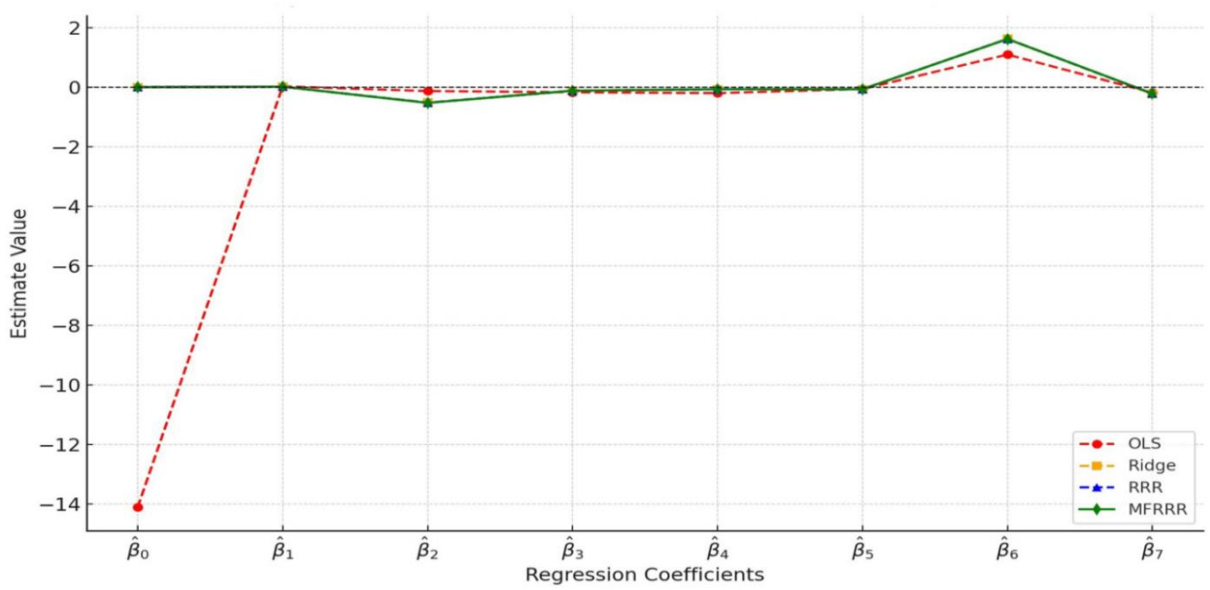

Figure 4: Parameter estimate stability for OLS, Ridge, RRR and MFRRR.

Stability here refers to the reduced variability and sensitivity of coefficient estimates across different regression methods. OLS shows pronounced coefficient fluctuations, especially for B_0 , reflecting its high sensitivity to multicollinearity and outliers. Ridge regression and RRR partially stabilize the estimates through regularization. The proposed MFRRR method provides the most stable coefficient estimates, with minimal deviation across parameters, confirming its effectiveness in simultaneously controlling multicollinearity and outlier influence.

Monte Carlo Simulation Study:

A Monte Carlo simulation study was carried out to assess the performance of the MFRRR method under challenging data conditions and compare it with other methods (OLS, Ridge Reg., RRR) by taking into consideration the simultaneous presence of multicollinearity and outliers. Simulated datasets were generated based on a linear regression model with three explanatory (independent) variables (p = 3) and varying sample sizes (N = 20, 50, 100, 150, 200). The response variable Y was generated according to the following model:

$$Y = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + r_i.$$

Each of the regression coefficients — the intercept and the three predictors — was assigned a value of 1. The error term r was independently drawn from a normal distribution with mean 0 and variance 1. To introduce high levels of multicollinearity, the predictor variables were generated through the transformation given below:

$$X_{ij} = \sqrt{1-\rho^2}\, V_{ij} + \rho\, V_{i(p+1)}, \quad i = 1,2,3,\ldots,n; \quad j = 1,2,3$$

where V_{ij} and $V_{i(p+1)}$ are independent normally distributed variables with mean zero and standard deviation 5 and ρ denotes the correlation level among explanatory variables. Three different levels of multicollinearity were examined by setting ρ = 0.5, 0.8, 0.99. To assess robustness under data contamination, outliers were introduced by replacing the first 5%, 10%, 15% and 20% of the observations in both X and Y with values drawn from a uniform distribution in the range [10, 20]. This contamination procedure was carried out across all combinations of sample sizes and levels of multicollinearity to ensure a full assessment of model performance.

Tables 2–4 present the MAE values obtained from the OLS, ridge regression, RRR and MFRRR methods under varying levels of multicollinearity (ρ = 0.5, 0.8, 0.99), different sample sizes (N = 20, 50, 100, 150, 200) and varying percentages of outliers (5%, 10%, 15%, 20%).

As shown in Table 2, under weak multicollinearity (ρ = 0.5), the OLS method consistently produced the highest MAE values. For example, with N = 20 and 15% outliers, OLS yielded a MAE of 4.1398. In contrast, even though the correlation was lower, the performance of ridge-based methods was very similar. Specifically, the MFRRR method had the lowest MAE, 0.0977, compared to Ridge (0.1024) and RRR (0.1010). This proves that, in the case of a weak multicollinearity, MFRRR is still superior in the sense of predictive power due to its inherent robust fuzzy structure.

Table 2: MAE values of the estimation methods when ρ = 0.5

N	% of outliers	OLS	Ridge Reg.	RRR	MFRRR
20	5	3.3579	0.1028	0.0886	0.0858
	10	5.4642	0.1067	0.0975	0.0959
	15	4.1398	0.1024	0.1010	0.0977
	20	4.7771	0.1015	0.1013	0.0972
50	5	4.0005	0.0588	0.0537	0.0524
	10	3.5692	0.0621	0.0601	0.0588
	15	4.8358	0.0707	0.0702	0.0676
	20	5.1109	0.0620	0.0620	0.0598
100	5	3.0866	0.0316	0.0307	0.0299
	10	4.4450	0.0400	0.0397	0.0384
	15	4.1767	0.0400	0.0400	0.0386
	20	4.4860	0.0409	0.0408	0.0393
150	5	3.2364	0.0208	0.0205	0.0198
	10	5.1165	0.0302	0.0299	0.0290
	15	4.9289	0.0305	0.0305	0.0294
	20	5.0182	0.0402	0.0402	0.0388
200	5	3.0564	0.0209	0.0205	0.0199
	10	4.5711	0.0311	0.0309	0.0299
	15	4.6900	0.0302	0.0300	0.0291
	20	5.0164	0.0308	0.0308	0.0297

As shown in Table 3, the trend continues with ρ= 0.8. Here as well, the OLS method performed poorly across all scenarios. For instance, at N = 200 with 15% outliers, OLS reached a MAE of 5.4978, while MFRRR resulted in a MAE of 0.0294—lower than Ridge (0.0305) and RRR (0.0303). This further confirms the consistency and efficiency of MFRRR in moderately collinear settings.

Table 3: MAE values of the estimation methods when ρ = 0.8

N	% of outliers	OLS	Ridge Reg.	RRR	MFRRR
20	5	3.1623	0.0744	0.0623	0.0606
	10	4.0188	0.0762	0.0680	0.0660
	15	4.6916	0.0718	0.0716	0.0689
	20	5.0119	0.0819	0.0819	0.0792
50	5	3.0931	0.0312	0.0291	0.0281
	10	4.6236	0.0505	0.0499	0.0483
	15	5.8931	0.0601	0.0600	0.0579
	20	6.7840	0.0602	0.0601	0.0581
100	5	4.1261	0.0315	0.0304	0.0296
	10	4.4916	0.0409	0.0406	0.0394
	15	5.8400	0.0412	0.0412	0.0398
	20	5.2908	0.0412	0.0411	0.0399
150	5	3.5594	0.0220	0.0215	0.0208
	10	5.4848	0.0307	0.0303	0.0294
	15	4.9912	0.0309	0.0308	0.0298
	20	5.3648	0.0362	0.0362	0.0349
200	5	3.5686	0.0186	0.0183	0.0177
	10	5.0869	0.0270	0.0267	0.0259
	15	5.4978	0.0305	0.0303	0.0294

	20	5.5502	0.0305	0.0305	0.0295

Table 4 shows the performance with high multicollinearity (ρ = 0.99), and the results indicate that the OLS method has the highest MAE values for all tested scenarios. In contrast, Ridge and RRR slightly reduce the errors, while MFRRR consistently achieves the lowest MAEs—for example, 0.0457 at N = 20 with 5% outliers. These findings confirm that MFRRR is the most robust and accurate among the compared methods.

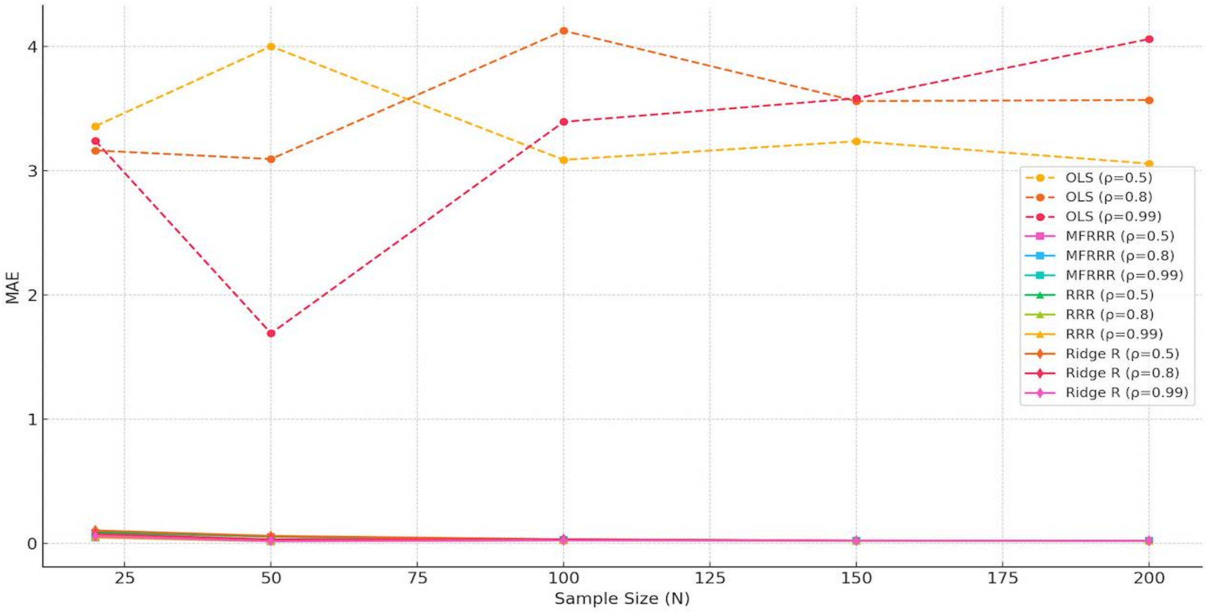

Table 4: MAE values of the estimation methods when ρ = 0.99

N	% of outliers	OLS	Ridge Reg.	RRR	MFRRR
20	5	3.2398	0.0562	0.0469	0.0457
	10	5.6050	0.0851	0.0804	0.0792
	15	8.1455	0.0922	0.0918	0.0892
	20	4.9047	0.0922	0.0931	0.0905
50	5	1.6899	0.0167	0.0163	0.0159
	10	4.7956	0.0506	0.0502	0.0488
	15	5.0330	0.0514	0.0505	0.0492
	20	6.4019	0.0612	0.0610	0.0589
100	5	3.3926	0.0231	0.0225	0.0219
	10	5.5629	0.0406	0.0401	0.0390
	15	6.1591	0.0406	0.0405	0.0392
	20	5.8936	0.0410	0.0410	0.0397
150	5	3.5816	0.0205	0.0201	0.0195
	10	5.0389	0.0308	0.0303	0.0294
	15	5.9174	0.0318	0.0317	0.0307
	20	6.1141	0.0380	0.0380	0.0367
200	5	4.0593	0.0203	0.0200	0.0194
	10	6.0186	0.0277	0.0275	0.0266
	15	6.0756	0.0302	0.0301	0.0291
	20	5.8045	0.0303	0.0302	0.0292

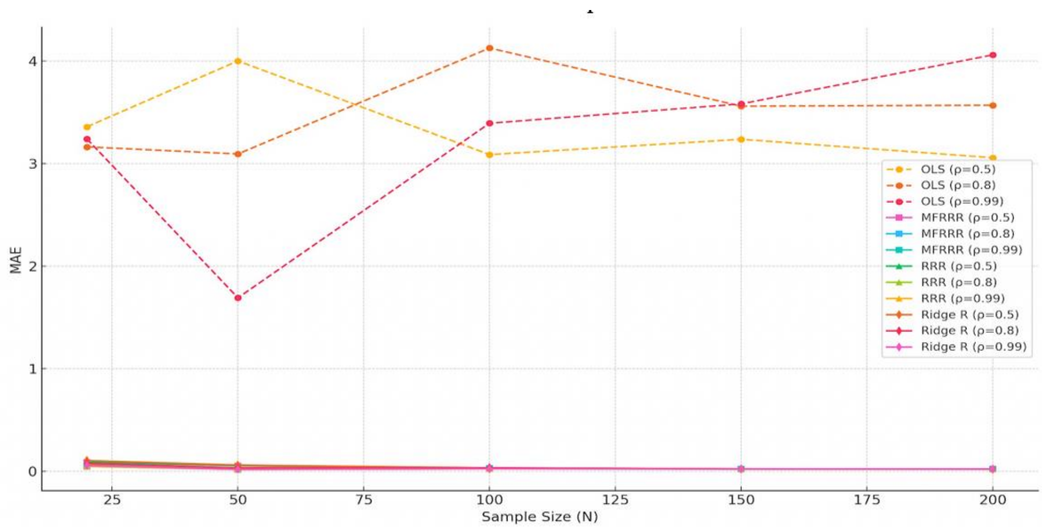

Figure 5: MAE values for 5% outliers across different ρ values for OLS, Ridge, RRR and MFRRR.

The OLS consistently produces the largest MAE values under increasing multicollinearity, indicating poor robustness to outliers. Ridge regression and RRR reduce prediction errors but remain sensitive at higher correlation levels. By contrast, a comparison with the MFRRR model indicates that it has the least MAE at all sample sizes and correlation levels and thus demonstrates better predictive stability and robustness in the presence of both multicollinearity and outliers.

CONCLUSION

In this paper, we introduced a new regression method coined MFRRR to exploit the limitations of classical regression methods in the presence of uncertainty and contaminated data. By incorporating modified fuzzy robust estimators in the ridge penalty framework, MFRRR is successful in improving both robustness and regularization. Its effectiveness in practice was supported through real-data experiments and thorough Monte Carlo simulations: MFRRR provided the smallest prediction errors with MAE = 0.049396, which were superior to ridge regression (MAE = 0.051208) and RRR (MAE = 0.051173) by approximately 3.5% and 3.4%, respectively, based on MAE. This convincing empirical evidence accentuates the potential of MFRRR for being a robust and flexible modeling paradigm in noisy data contexts. By being a hybrid system, MFRRR enhances robustness and offers flexibility for modeling real-world scenarios involving noisy data, imprecision, or ambiguity. Future studies would be helpful in focusing on the combination of MFRRR models with other machine learning paradigms or in applying the paradigm to high-dimensional data scenarios.

Acknowledgement:

The authors would like to thank the editor and the anonymous reviewers for their constructive comments and valuable recommendations, which have contributed to the improvement of the manuscript.

Ethical Statement:

This study used a publicly available dataset and simulated data. No human or animal subjects, or sensitive personal data, were involved; therefore, formal ethical approval was not required. The authors confirm that the study meets the necessary standards of academic integrity and transparency.

Author Contributions:

V. M. S. contributed to the conceptualization of the study, development of the mathematical framework, data collection, analysis, and investigation; S. S. I, contributed for overseeing the research direction, validating the mathematical results and proofs, providing critical intellectual guidance, both authors have read and agreed to the published version of the manuscript.

Conflict of Interest:

The authors declare no conflict of interest.

Funding:

This study did not receive specific funding from public, commercial, or non-profit organizations.

REFERENCES

Akbari, M. G., & Hesamian, G. (2019). A partial-robust-ridge-based regression model with fuzzy predictors-responses. *Journal of Computational and Applied Mathematics, 351*, 290–301. https://doi.org/10.1016/j.cam.2018.11.006

Bas, E., & Egrioglu, E. (2024). Robust picture fuzzy regression functions approach based on M-estimators for the forecasting problem. *Computational Economics, 65*(3), 2775–2810. https://doi.org/10.1007/s10614-024-10647-9

Belsley, D. A. (1991). *Conditioning diagnostics: Collinearity and weak data in regression*. Wiley-Interscience.

Choi, S., Jung, H.-Y., & Kim, H. (2019). Ridge fuzzy regression model. *International Journal of Fuzzy Systems, 21*. https://doi.org/10.1007/s40815-019-00692-0

Farnoosh, R., Ghasemian, J., & Solaymani Fard, O. (2020). Integrating ridge-type regularization in fuzzy nonlinear regression. *Computational and Applied Mathematics, 39*(2), 1–17. https://doi.org/10.1590/S1807-03022012000200006

Hesamian, G., & Akbari, M. G. (2020). A robust varying coefficient approach to fuzzy multiple regression model. *Journal of Computational and Applied Mathematics, 371*, 112704. https://doi.org/10.1016/j.cam.2019.112704

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hong, D. H., & Hwang, C. (2004). Ridge regression procedures for fuzzy models using triangular fuzzy numbers. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 12,* 145–159. https://doi.org/10.1142/S0218488504002746

Hong, D. H., Hwang, C., & Ahn, C. (2004). Ridge estimation for regression models with crisp inputs and Gaussian fuzzy output. *Fuzzy Sets and Systems, 142,* 307–319. https://doi.org/10.1016/S0165-0114(03)00002-2

Ismaeel, S. S., Midi, H., & Omar, K. M. T. (2024). A remedial measure of multicollinearity in multiple linear regression in the presence of high leverage points. *Sains Malaysiana, 53*(4), 907–920. https://doi.org/10.17576/jsm-2024-5304-14

Karbasi, D., Nazemi, A., & Rabiei, M. R. (2021). An optimization technique for solving a class of ridge fuzzy regression problems. *Neural Processing Letters, 53*, 3307–3338. https://doi.org/10.1007/s11063-021-10538-2

Kareem, R. E., & Mohammed, M. J. (2023). Fuzzy bridge regression model estimating via simulation. *Journal of Economics and Administrative Sciences, 29*(136), 60–69. https://doi.org/10.33095/jeas.v29i136.2607

Kim, H., & Jung, H.-Y. (2020). Ridge fuzzy regression modelling for solving multicollinearity. *Mathematics, 8*(9), 1572. https://doi.org/10.3390/math8091572

Penrose, K. W., Nelson, A., & Fisher, A. (1985). Generalized body composition prediction equation for men using simple measurement techniques. *Medicine & Science in Sports & Exercise, 17*(2), 189. https://journals.lww.com/00005768-198504000-00037

Rabiei, M., Arashi, M., & Farrokhi, M. (2019). Fuzzy ridge regression with fuzzy input and output. *Soft Computing, 23*, 4164. https://doi.org/10.1007/s00500-019-04164-3

Salih, V. M., & Ismaeel, S. (2025). Enhancing Parameter Estimation for Fuzzy Robust Regression in the Presence of Outliers. *Statistics, Optimization & Information Computing, 14*(4), 1795-1812. https://doi.org/10.19139/soic-2310-5070-2656

Tsai, T. R., & Wu, S. J. (2002). Fuzzy-weighted estimation in ridge regression analysis. *Journal of Information and Optimization Sciences, 23*(2), 259–271. https://doi.org/10.1080/02522667.2002.10698995

APPENDIX

Numerical Example – Body Fat Dataset

Table A1: Variables of the body fat dataset

No.	BodyFat	Age	Weight	Height	Neck	Chest	Abdomen	Hip
1	12.3	23	154.25	67.75	36.2	93.1	85.2	94.5
2	6.1	22	173.25	72.25	38.5	93.6	83	98.7
3	25.3	22	154	66.25	34	95.8	87.9	99.2
4	10.4	26	184.75	72.25	37.4	101.8	86.4	101.2
5	28.7	24	184.25	71.25	34.4	97.3	100	101.9
6	20.9	24	210.25	74.75	39	104.5	94.4	107.8
7	19.2	26	181	69.75	36.4	105.1	90.7	100.3
8	12.4	25	176	72.5	37.8	99.6	88.5	97.1
9	4.1	25	191	74	38.1	100.9	82.5	99.9
10	11.7	23	198.25	73.5	42.1	99.6	88.6	104.1
11	7.1	26	186.25	74.5	38.5	101.5	83.6	98.2
12	7.8	27	216	76	39.4	103.6	90.9	107.7
13	20.8	32	180.5	69.5	38.4	102	91.6	103.9
14	21.2	30	205.25	71.25	39.4	104.1	101.8	108.6
15	22.1	35	187.75	69.5	40.5	101.3	96.4	100.1
16	20.9	35	162.75	66	36.4	99.1	92.8	99.2
17	29	34	195.75	71	38.9	101.9	96.4	105.2
18	22.9	32	209.25	71	42.1	107.6	97.5	107
19	16	28	183.75	67.75	38	106.8	89.6	102.4
20	16.5	33	211.75	73.5	40	106.2	100.5	109
21	19.1	28	179	68	39.1	103.3	95.9	104.9
22	15.2	28	200.5	69.75	41.3	111.4	98.8	104.8

23	15.6	31	140.25	68.25	33.9	86	76.4	94.6
24	17.7	32	148.75	70	35.5	86.7	80	93.4
25	14	28	151.25	67.75	34.5	90.2	76.3	95.8
26	3.7	27	159.25	71.5	35.7	89.6	79.7	96.5
27	7.9	34	131.5	67.5	36.2	88.6	74.6	85.3
28	22.9	31	148	67.5	38.8	97.4	88.7	94.7
29	3.7	27	133.25	64.75	36.4	93.5	73.9	88.5
30	8.8	29	160.75	69	36.7	97.4	83.5	98.7
31	11.9	32	182	73.75	38.7	100.5	88.7	99.8
32	5.7	29	160.25	71.25	37.3	93.5	84.5	100.6
33	11.8	27	168	71.25	38.1	93	79.1	94.5
34	21.3	41	218.5	71	39.8	111.7	100.5	108.3
35	32.3	41	247.25	73.5	42.1	117	115.6	116.1
36	40.1	49	191.75	65	38.4	118.5	113.1	113.8
37	24.2	40	202.25	70	38.5	106.5	100.9	106.2
38	28.4	50	196.75	68.25	42.1	105.6	98.8	104.8
39	35.2	46	363.15	72.25	51.2	136.2	148.1	147.7
40	32.6	50	203	67	40.2	114.8	108.1	102.5
41	34.5	45	262.75	68.75	43.2	128.3	126.2	125.6
42	32.9	44	205	29.5	36.6	106	104.3	115.5
43	31.6	48	217	70	37.3	113.3	111.2	114.1
44	32	41	212	71.5	41.5	106.6	104.3	106
45	7.7	39	125.25	68	31.5	85.1	76	88.2
46	13.9	43	164.25	73.25	35.7	96.6	81.5	97.2
47	10.8	40	133.5	67.5	33.6	88.2	73.7	88.5
48	5.6	39	148.5	71.25	34.6	89.8	79.5	92.7
49	13.6	45	135.75	68.5	32.8	92.3	83.4	90.4
50	4	47	127.5	66.75	34	83.4	70.4	87.2

Note. This table presents the variables included in the body fat dataset used for analysis.