

Content list available at [JournalPark](http://JournalPark)

# Turkish Journal of Forecasting

Journal Homepage: [tjforecast.xyz](http://tjforecast.xyz)

## Fuzzy Clustering Based Regression Model Forecasting: Weighted by Gustafson-Kessel Algorithm

N. A. Erilli<sup>1,\*</sup><sup>1</sup>Sivas Cumhuriyet University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Sivas, Turkey

### ARTICLE INFO

#### Article history:

Received	10	August	2020
Revised	22	November	2020
Accepted	16	December	2020
Available online	23	December	2020

#### Keywords:

Fuzzy clustering analysis  
Gustafson-Kessel method  
Weighted regression  
Forecasting

### RESEARCH ARTICLE

### ABSTRACT

Regression analysis is one of the well-known methods of multivariate analysis and it is efficiently used in many research fields, especially forecasting problems. In order for the results of regression analysis to be effective, some assumptions must be valid. One of these assumptions is the heterogeneity problem. One of the methods used to solve this problem is the weighted regression method. Weighted regression is a useful method when one of the least-squares assumptions of constant variance in the residuals is violated (heteroscedasticity). This procedure can minimize the sum of weighted squared residuals to produce residuals with a uniform variance if the appropriate weight will be used. (homoscedasticity). In this study, the Gustafson-Kessel method, one of the fuzzy clustering analysis method, is used to determine weights for weighted regression analysis. Gustafson-Kessel's method is based on the minimization of the sum of weighted squared distances which is used Mahalanobis distance, between the data points and the cluster centres. With the fuzzy clustering method, each observation value is bound to the specified clusters in a specific order of membership. These membership degrees will be calculated as weights in the weighted regression analysis and estimation work will be done. In application, 5 simulation and 1 real-time data were estimated by the proposed method. The results were interpreted by comparing with Robust Methods (M and S estimator) and weighted with FCM Regression analysis.

 2020 Turkish Journal of Forecasting by Giresun University, Forecast Research Laboratory is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).

## 1. Introduction

Statistics is the science of extracting meaning from data. To determine the meaning of the data and to extract the information of interest from the data; Various statistical techniques have been developed for different types of data. No statistical method can remove or explain all the uncertainties in the world. However, we can use statistics to make all these uncertainties numerical or useful and to explain partially in this way. Perhaps the most well-known of these statistical methods is regression analysis. The purpose of the Regression Analysis is to reveal the relationship between a dependent variable and one or more independent variables. This relationship does not necessarily depend on a

\* Corresponding author.

E-mail addresses: [aerilli@cumhuriyet.edu.tr](mailto:aerilli@cumhuriyet.edu.tr) (Necati Alp Erilli)

function, but it may not have a cause-effect relationship. Traditional regression analysis methods require a linear structure for parameters. In other words, if the given observations are not linear, then this method cannot capture the natural movement of the data, so they are therefore not enough in most cases [10]. Regression analysis is based on some assumptions. The most important of these assumptions is that the form of the relationship between dependent and independent variables is known. In cases where the assumptions are not provided, the estimates made will not be good enough. In this case, regression methods that allow stretching the linearity assumption in parametric regression are needed to make better estimates. Nonparametric and semi-parametric regression methods can be given as examples of these issues.

Homoscedasticity is the other important assumption for regression analysis. Homoscedasticity describes a situation in which the variance in the error term is the same across all values of the independent variables. If this assumption fails, the model exhibits another problem named heteroscedasticity. Whenever the ordinary least-squares method is applied directly to the model which has heteroscedasticity, the  $t$  and  $F$  tests will give incorrect results and parameters will lose the Best Linear Unbiased Estimator assumptions [29]. We can summarize the problems that will occur as follows: Since the  $E(u) = 0$  and  $E(X'u) = 0$  assumptions remain valid, the no-bias property is preserved. Another problem is that the ordinary least squares (OLS) estimator loses its effectiveness. In this case, the problem of heteroscedasticity causes larger variance, and there will always be estimators with smaller variances. The calculated variance and standard errors of the OLS estimator will be based on an incorrect statement. Thus, the test statistics will give distorted and unreliable results. As a result, OLS calculates the  $t$ -values and  $F$ -values using an underestimated amount of variance. So, this problem can lead you to interpret that a model term is statistically significant when it is not significant [8]. The problem of heteroscedasticity can also arise in estimates made with time series. Especially in data such as inflation, stock prices, and exchange rates, volatility changes over time. Error term variances are related to both the error terms of the past periods and show fluctuations. Therefore, in the case of the heteroscedasticity problem, it is the best solution to use heteroscedasticity elimination methods instead of using OLS formulas.

When the homoscedasticity assumption is not valid, the most common way to balance the variances is transformation. When a transformation is not sufficient, variances may need to be modelled. If the variances of error terms are known in regression models with varying variances, parameter estimates can be made using weighted least squares. However, in practice, the variances of the error terms are not known. There are many studies in the statistical literature that use weight estimates and give results that are asymptotically equivalent to weighted least squares estimates [26]. Heteroscedasticity problem has been tried to be solved with methods that have achieved successful results in different data structures such as weighted OLS with quadratic residues, methods based on absolute residuals, methods based on sequential residuals, modified sequential residuals methods and re-weighted iterative OLS methods. In the case of constant variance errors in regression models, if the estimation is made using re-weighted least squares or some robust methods, standard methods for evaluating the estimation variances may not work. In such cases, applying resampling techniques (Bootstrap, Jackknife, etc.) to the data is an alternative method for different variance [27]. Besides, statistical transformations are seen as another alternative. The most used transformations are Box-Cox and Box-Tidwell transformations. Many regression methods that introduce nonlinear structure to the regression model have been proposed to solve such problems [30]. The weighted regression method is one of the typical examples of such methods. Weighted regression analysis is a method based on a calculation by assigning a weight to each data point. The basic idea is to give smaller weights to observations associated with higher variances to make the squared residuals smaller. Weighted regression aims to minimize the sum of errors in the weighted squares. When it is used the right weights, heteroscedasticity is replaced by homoscedasticity [8, 10].

When we take a brief look at the literature on weighted regression, we see studies in different model structures. The general purpose of all these studies is to suggest different weights to eliminate the heteroscedasticity problem. [21] used Hard Clustering and Fuzzy Clustering Based Regression Techniques to estimating the concrete compressive strength. They found that regression results with fuzzy clustering had fewer errors and heteroscedasticity. [17] studied on L-estimators for linear heteroscedastic models. [2] studied for weighted sample means and linear regression in a single factor. They found that each estimated weight should be based on at least ten degrees of freedom. [19] proposed a re-weighting estimation method for heteroscedastic mixed geographically weighted regression models, in which the variance function of the model errors is estimated by the kernel method with an adaptive bandwidth and the coefficients are re-estimated based on the weighted observations. [16] suggest some robust tests for heteroscedasticity based on regression quantiles. [13] proposed a test for heteroscedasticity based on least-squares residuals. [20] proposed an alternative regression-adaptive approach is to apply the difference-based estimator. [18] studied on a regression model based on SVR and ELM method which is proposed for both

noise variance prediction and smoothing procedure. [6] introduced a test for heteroscedasticity for the non-parametric regression model with multiple covariates. It is based on a suitable residual-based empirical distribution function. Thus, the residuals are constructed by using local polynomial smoothing. [12] proposed two robust weighting methods based on HLP detection measures (based on robust Mahalanobis distance) on robust heteroscedasticity consistent covariance matrix estimators. [32] develop a robust estimation method for the parameters in the arithmetic mean function by using the weighted least squares estimation. They propose a new estimator named delete-one-component for the single-index in the variance function based on absolute residuals.

In this study, a new weighted regression method based on Gustafson-Kessel's fuzzy clustering algorithm is proposed. The purpose of using fuzzy clustering weights is to eliminate the uncertainties in cluster membership in weighting made with classical clustering methods. In this way, it is planned to prevent the weighting matrix used in the calculation of model coefficients from being a singular matrix. [25] is seen as the most important study on this subject. In the study of [25], weights were determined using the Fuzzy C-Means method. The main purpose of using Gustafson-Kessel in this study is to investigate the effect of Mahalanobis distance on weights instead of the Euclidean distance used in Fuzzy C-Means. In the analysis, it was seen that the effects of the weights obtained with Mahalanobis distances on the model estimates gave more successful results, especially in data structures with high outliers.

## 2. Fuzzy Clustering Analysis

The use of fuzzy clusters in clustering analysis was first proposed by [1]. In fuzzy clustering, fuzzy techniques are used to clustering the data and with these methods, an object can be classified into more than one cluster, unlike the classical methods. Since such algorithms deal with the uncertainty of real numbers, it can help to reveal clustering patterns that are consistent with the experiences of daily life.

In this approach, if the clusters are not separated from one another, or if some units in their membership are uncertain in cluster membership, it appears to be an appropriate method. Fuzzy clusters are the functions modifying each unit between 0 and 1 named as the cluster membership of the units. The units which are very closed to each other hold their places in the same cluster according to their closest membership degrees. The advantage of fuzzy clustering over classical clustering methods is that it provides more detailed information on the data. If an object is a member of more than one set, it can also help assign the object to sets. But there will be too much output in the fuzzy analysis if there are too many individuals or variables, so that can be difficult to summarize, classify, or make a decision about the data [2]. Membership degrees indicating which cluster a given individual will enter is found by calculating the distance between the data point and the cluster centre. The cluster centre the data point is closer to, the greater the degree of membership for that cluster. In general, the aim is to ensure that the distance between data and cluster centres is minimum and membership degrees are maximum.

The most widely used method in fuzzy clustering analysis is the fuzzy c-means method, which was introduced by Bezdek [3, 4]. Gustafson-Kessel and Gath-Geva methods are other commonly used methods in fuzzy clustering analysis [9, 11]. In the literature, the applications of fuzzy clustering analysis to regression analysis are low and varied. [25] used fuzzy c-means method as the weights for the weighted regression analysis. [14] used fuzzy clustering as switching regression models and [15] used a fuzzy c-means method to support vector regression models.

### 2.1. Gustafson-Kessel fuzzy clustering method

The Gustafson-Kessel (GK) fuzzy clustering algorithm was first proposed in [11]. Gustafson-Kessel algorithm is a fuzzy clustering algorithm developed to identify ellipse-like clusters instead of global clusters. Other methods like Fuzzy C-Means or Gath-Geva methods cannot give satisfactory results in these kinds of clusters. Gustafson and Kessel extended the standard fuzzy clustering algorithm by employing an adaptive and effective distance norm to determine clusters of different geometrical shapes in one data set. They used Mahalanobis distance instead of Euclidean distance, unlike other fuzzy clustering methods.

Let  $\Sigma_i$  be the covariance matrix of the cluster,  $c_i$  be the center of the  $i$ -th cluster,  $u_{ij}$  be the membership degree and  $\beta$  be fuzziness index. For the  $i$ -th cluster, its associated Mahalanobis distance is defined in equation 1:

$$d^2(x_j, c_i, \Sigma_i) = (x_j - c_i)^T \Sigma_i^{-1} (x_j - c_i) \quad (1)$$

Then, the covariance matrices are computed as follow:

$$\Sigma_i = \frac{\Sigma_i^*}{\sqrt[\beta]{\det(\Sigma_i^*)}}; \quad \Sigma_i^* = \frac{\sum_{j=1}^n u_{ij} (x_j - c_i)(x_j - c_i)^T}{\sum_{j=1}^n u_{ij}} \quad (2)$$

The objective function is also defined as;

$$J(X, C, \Sigma, U) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^\beta d^2(x_j, c_i, \Sigma_i) \quad (3)$$

The objective function  $J(X, C, \Sigma, U)$  is then minimized under the following constraints:

$$\begin{aligned} 0 &\leq u_{ij} \leq 1, \forall i, j \\ 0 &< \sum_{j=1}^n u_{ij} \leq n, \forall i \\ \sum_{i=1}^c u_{ij} &, \forall j \end{aligned} \quad (4)$$

In this minimization problem, the centre  $c_i$  and the membership degrees  $u_{ij}$  are updated according to the expressions given in equation 5:

$$c_i = \frac{\sum_{j=1}^n u_{ij}^\beta x_j}{\sum_{j=1}^n u_{ij}^\beta}, \quad u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d(x_j, c_j)}{d(x_j, c_k)} \right)^{2/(\beta-1)}} \quad (5)$$

## 2.2. Fuzzy clustering validity index

The purpose of clustering analysis is to put similar objects into the same groups. In many clustering algorithms, it is hard to know the actual number of clusters or groups before the application if there is no prior knowledge. In studies based on real data, if the researchers do not have preliminary information about data or the optimal number of clusters, trial and error may be required to find the true number of clusters. This can lead to both time and cost loss.

Determination for the optimal number of clusters in clustering studies is generally called Cluster Validity. After clustering processes are carried out, the validity of the optimal number of clusters calculated can be determined. Many validity indices are used in Fuzzy Clustering literature like Partition Coefficient, Classification Entropy, or Xie-Beni [3, 31]. In this study, an artificial neural network-based validity index which is introduced by [7] is used.

## 3. Weighted Regression Analysis

Weighted regression analysis is a special method that you can use when the least-squares assumption of constant variance in the residuals is violated. Weighted least squares regression does not need a defined type of function to describe the relationship between dependent or independent variables. Weighted least squares picture the behaviour of the random errors defined with the econometric model. Therefore, it can be used with specific functions. These functions can have linear or nonlinear structures defined in parameters. The method works by combining nonnegative weights, associated with each data point, into the fitting criterion. The size of the weight criterion is also a summary of the information in the observation values. Predisposing the weighted fitting criterion used in parameter estimates allows the weights to strictly determine the assistance of each observation to the final and last parameter estimates. It should also be noted that the weights assigned to each observation are adjusted to the weights of other observations. Therefore, it should be kept in mind that groups with different absolute weight may have the same effects [5]. The biggest criticism of weighted least squares is that the theory behind this method requires the weights to be known exactly. Since this is not possible in real-life applications, efforts are made to solve this problem by using estimated weights.

It is difficult to assess the impact of using estimated weights. But on the other hand, experience shows that small changes in weights due to estimation generally do not affect a regression analysis or its interpretation. In this method, if the weights are estimated from a small number of repeated observations, the results of the analysis may be affected in a worse and unpredictable way than expected. This is particularly the case where the weights for outliers of the predictor or explanatory variables are estimated using only a small number of observations. Aware of this possible problem, the researcher should use weighted least squares when weights can be predicted with precision [4, 12]. We can write the model as  $Y = X\beta + \varepsilon$ , where now  $\varepsilon$  is assumed to be (multivariate) normally distributed with mean vector 0 and non-constant variance-covariance matrix:

$$\sigma_i^2 = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix} \tag{6}$$

If we define the reciprocal of each variance ( $\sigma_i^2$ ), as the weight with  $W_i = \frac{1}{\sigma_i^2}$ , then weighted matrix  $W$  be a diagonal matrix containing these weights as seen below:

$$W = \begin{pmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_n \end{pmatrix} \tag{7}$$

Finally, the weighted least squares estimate then can be given in equation 8:

$$\hat{\beta}_{WLS} = (X^T W X)^{-1} X^T W Y \tag{8}$$

The following three conclusions can be drawn here: First, each weight is inversely proportional to the error variance. This reflects the information in the observation values. Thus, an observation with a small error variance contains relatively more information than observation with a large error variance and therefore has a large weight. Second, the weights used in the analysis must be known or estimated down to a proportionality constant. Third, weighted least squares regression is not associated with a function used to describe the relationship between process variables. Instead, the weighted least squares show the behaviour of random errors in the model. Thus, different functional structures (linear or non-linear) can be used in parameters [28].

### 3.1. Weighted regression analysis using Gustafson-Kessel loadings

[13] proposed a new model for weighted regression using fuzzy clustering result obtained as classification of the data consisting of explanatory variables defined as below:

$$Y = U_k^{-1} X \beta_k + \hat{e}_k \tag{9}$$

where  $U_k^{-1} = \begin{pmatrix} u_{1k} & 0 & \dots & 0 \\ 0 & u_{2k} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nk} \end{pmatrix}$ ,  $\hat{e}_k = \begin{pmatrix} \hat{e}_{1k} \\ \vdots \\ \hat{e}_{nk} \end{pmatrix}$ .

The estimate of least squares of  $\beta_k$  in equation obtained can be given in equation 10.

$$\tilde{\beta}_k = \left( X^t (U_k^{-1})^2 X \right)^{-1} X^t U_k^{-1} Y \tag{10}$$

In this formula, the feature of Gustafson-Kessel fuzzy clustering is used and  $u_k$  shows the weights of objects to a cluster  $k$  so we can use  $u_k$  directly for the weights of weighted regression as given in the formula.

In this study, the Gustafson-Kessel method is used to determine weights for weighted regression analysis. Gustafson-Kessel's method is based on the minimization of the sum of weighted squared distances between the data points and the cluster centres. With the fuzzy clustering method, each observation value is bound to the specified clusters in a specific order of membership. These membership degrees will be calculated as weights in the weighted regression analysis and estimation work will be done. This method works by incorporating extra nonnegative constants, or weights, associated with each data point, into the fitting criterion. The size of the weights indicates the accuracy of the information between the observations. To obtain the parameter estimates, the contribution of the weights to the parameter estimates for each observation is determined by optimizing the fitting criterion of the weights.

## 4. Application

In the application part, the proposed Gustafson-Kessel algorithm is used for 5 different types of simulation and real-time data for testing the power of the method. All types of data have heteroscedasticity problems where all tested with Breusch-Pagan-Godfrey test ( $p > 0.05$ ).

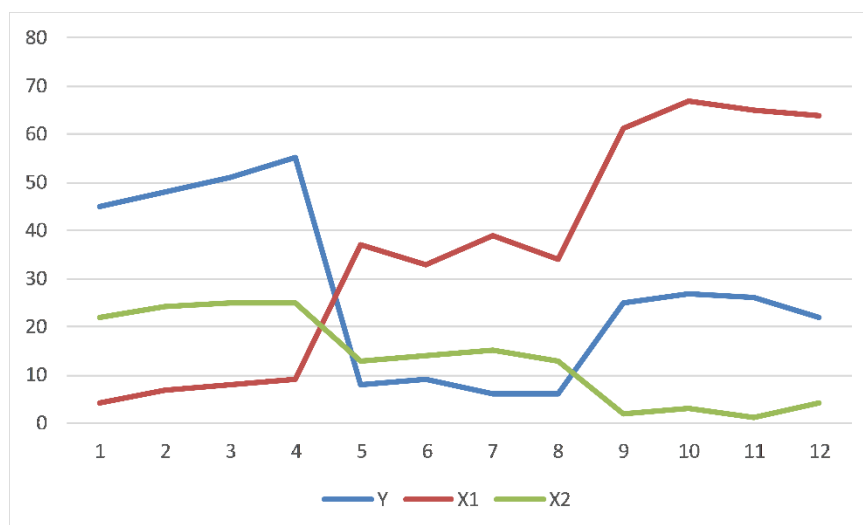
To test the power of the Gustafson-Kessel method, the results obtained are compared with some estimation methods. These methods are Weighted Regression (weighted by the inverse of std. Deviation and Fuzzy C-Means (FCM), Robust M-estimator, and S-estimator. For a comparison of all methods, the RMSE value is used. MATLAB 2009b and Eviews 9 package programs are used for calculations.

### 4.1. Simulation data

In the first part of the application, a data set of 3 variables and 12 observations was used. By changing the distribution of the variables in this data set, more data sets were obtained in 4 different structures and the effect of the method was investigated by applying the 6 methods mentioned. For each data structure, fuzzy clustering analysis was applied separately and weights for weight regression were obtained. Before fuzzy clustering was applied on each data, the appropriate number of clusters was determined with the artificial neural network cluster validity index. After the fuzzy clustering method was applied, 3 clusters were found for Simulation-1, Simulation-2, Simulation-5 data, and 2 clusters for Simulation-3, Simulation-4 data. Weights for weighted regression were calculated as given in equation 7 with the help of cluster membership degrees applied separately for each data with the Gustafson-Kessel method. By using the weights matrix obtained in this way, the coefficient estimations given in equation 8 were calculated. Simulation 1 data values are given in Table 1 and the line chart is given in Figure 1.

**Table 1.** Simulation-1 data

<b>Y</b>	45	48	51	55	8	9	6	6	25	27	26	22
<b>X<sub>1</sub></b>	4	7	8	9	37	33	39	34	61	67	65	64
<b>X<sub>2</sub></b>	22	24	25	25	13	14	15	13	2	3	1	4



**Figure 1.** Simulation-1 data line chart

The prediction models of the above 5 models for the data are given in Table 1 are given below.

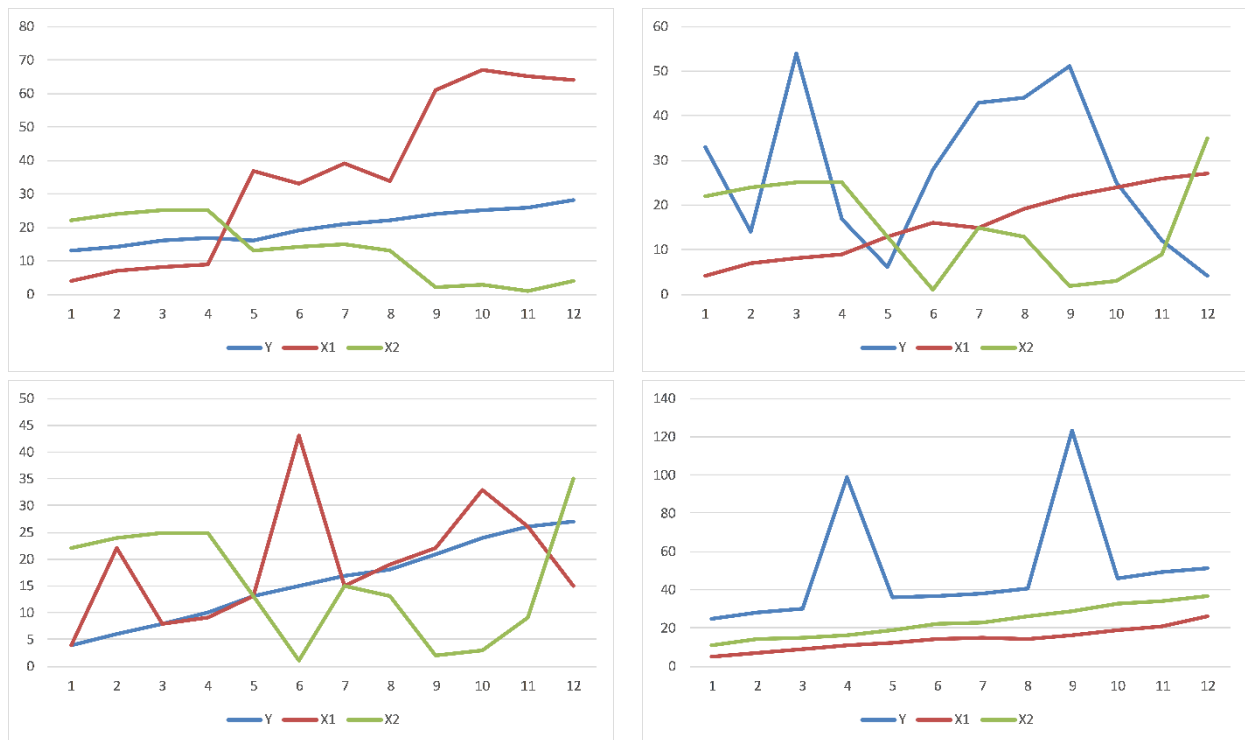
- Weighted Regression (Weighted with FCM):  $\hat{Y} = 50,5993 - 0,5374X_1 - 0,2850X_2$
- Weighted Regression (Weighted with GK):  $\hat{Y} = 49,5215 - 0,5215X_1 - 0,2551X_2$
- Weighted Regression (Weighted with  $\sigma^{-1}$ ):  $\hat{Y} = 1,110 - 0,284X_1 - 1,936X_2$
- M-Estimator:  $\hat{Y} = 14,7226 - 0,1066X_1 - 1,4247X_2$
- S-Estimator:  $\hat{Y} = 14,0995 - 0,1173X_1 - 1,4461X_2$

The RMSE values obtained from the model results given above are given in Table 2.

**Table 2.** RMSE results for simulation-1 data

Method	RMSE
Weighted with GK	14.4877
Weighted with FCM	14.4901
Weighted with $\sigma^{-1}$	17.9078
M-Estimator	17.9955
S-Estimator	18.0278

As for the results, weighted with GK value has the minimum RMSE. Similar to the above, the other 4 simulation data structures and the RMSE values obtained from the 5 methods applied are given in Figure 2 and Table 3.



**Figure 2.** Simulation 2-5 data structures

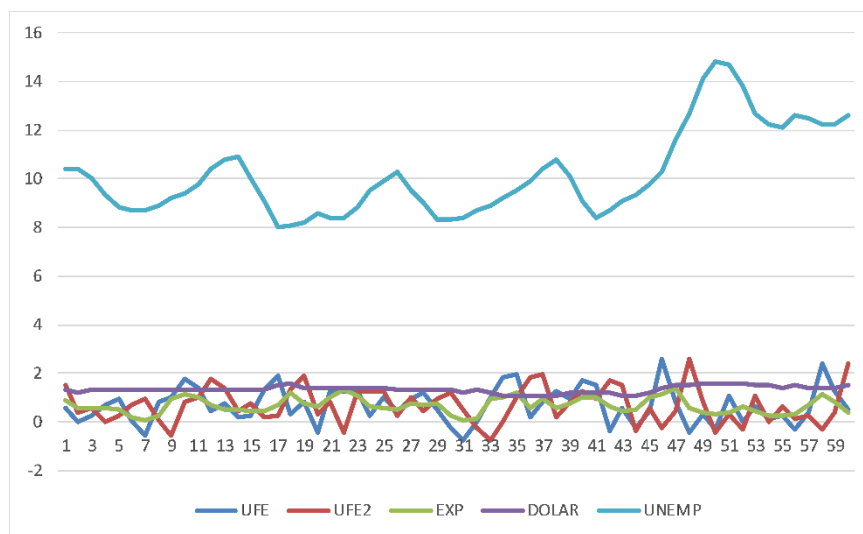
When we look at the RMSE results given in Table 3, we can say that the results of weighted regression with Gustafson-Kessel are the best when data has outliers. In other data structures, it is possible to say that if the corruptions increase, it is one step ahead of FCM or other robust methods.

## 4.2. Real-time data

Real-time data is Turkish inflation data and has 60 observations with 5 variables (2005:1-2009:12). In the data, Turkish inflation is the dependent variable and 2-term lagged inflation, expectation survey for inflation, dollar exchange rate, and the unemployment rate are the independent variables.

**Table 3.** RMSE results for simulation 2-5 data

Simulation-2 Results		Simulation-3 Results	
Method	RMSE	Method	RMSE
Weighted with $\sigma^{-1}$	1.7510	M-Estimator	15.2705
S-Estimator	1.7579	Weighted with GK	15.6779
M-Estimator	1.7594	S-Estimator	16.0595
Weighted with FCM	2.2359	Weighted with FCM	17.2468
Weighted with GK	2.2663	Weighted with $\sigma^{-1}$	23.1853
Simulation-4 Results		Simulation-5 Results	
Method	RMSE	Method	RMSE
Weighted with GK	6.6092	Weighted with GK	27.9788
Weighted with FCM	6.6131	Weighted with FCM	28.6355
M-Estimator	8.6324	Weighted with $\sigma^{-1}$	29.6729
Weighted with $\sigma^{-1}$	8.6777	M-Estimator	30.2578
S-Estimator	10.2063	S-Estimator	30.3111



**Figure 3.** Turkish inflation data with 5 variables

The RMSE results obtained from the above-mentioned 5 methods are given in Table 4. The result of Weighted with Gustafson-Kessel has the best RMSE value when compared with the other methods.

**Table 4.** RMSE results for Turkish inflation data

Method	RMSE
Weighted with GK	0.5527
M-Estimator	0.5529
Weighted with FCM	0.5627
S-Estimator	0.6162
Weighted with $\sigma^{-1}$	0.8773



## 5. Conclusion

In regression analysis, the heteroscedasticity problem reduces the reliability of forecast values. Many methods are proposed to solve such problems. The weighted regression analysis, which is made by adding a weight matrix to the calculation of the regression coefficients, is one of these and it aims to solve the problem of heteroscedasticity. In literature, the weight matrix is usually calculated by multiplying with  $\sigma^{-1}$  or  $\frac{1}{\sigma^2}$ .

Although there are different types of weighted regression methods to solve the problem of heteroscedasticity in the literature, there are very few weighting studies using fuzzy clustering algorithms. In this study, a new approach is proposed by taking [25] as a reference. With the proposed approach, a different fuzzy clustering method was used, and a new weight matrix was created in the weighted regression analysis. Thus, the effect of individual values on the weight matrix is minimized. In the study of [25], using the weights obtained from the Fuzzy Clustering loadings which are calculated from the FCM method, the often-used method in the fuzzy clustering analysis, is suggested. In this study, a weighted regression model is proposed which is based on weights obtained from the Gustafson-Kessel method in Fuzzy Clustering Analysis. The Gustafson-Kessel method was used in this study in contrast to the work done by the Fuzzy C-Means method in the literature.

The most important difference of this study from other similar studies is that when performing clustering analysis, Mahalanobis distance is used instead of Euclidean distance. Mahalanobis distance is used to detect outliers, in the development of linear regression models and classification studies. A point that has a greater Mahalanobis distance from the rest of the sample population of points is said to have higher leverage since it has a greater influence on the slope or coefficients of the regression equation. Regression methods can be used to determine if a particular case within a sample population is an outlier through the combination of two or more variable scores. Even for normal distributions, a datum can be a multivariate outlier even if it is not a univariate outlier for any variable, making Mahalanobis distance a more sensitive measure than checking dimensions individually.

The proposed method was applied to 5 simulations and a real dataset. It can be said that the proposed method gives very good results compared to other classical methods in nonlinear structures and models with high outliers. The results are more successful than the classical weighted regression and FCM weights to solve the problem of heteroscedasticity. In a summary, Gustafson-Kessel's weighted regression model can be a strong alternative to robust methods and FCM weighted method which is used to solve the heteroscedasticity problem.

## Acknowledgments

This study was presented at the 1st International Web Conference on Forecasting - IWCF'2017.

## References

- [1] R.E. Bellman, R. Kalaba, L.A. Zadeh, Abstraction and pattern classification, *J. Math. Anal. Appl.* (1966) 1-7.
- [2] T.R. Bement, J.S Williams, Variance of Weighted Regression Estimators When Sampling Errors are Independent and Heteroscedastic, *Journal of the American Statistical Association.* 64 (1969) 1369-82.
- [3] J.C. Bezdek, Cluster validity with fuzzy sets, *Journal of Cybernetics*, 3 (1947a) 58–73.
- [4] J.C. Bezdek, Numerical taxonomy with fuzzy sets, *Journal of Mathematical Biology.* 1 (1947b) 57–71.
- [5] R.J. Carroll, D. Ruppert, *Transformation and Weighting in Regression*, Chapman and Hall, New York, USA (1988)
- [6] J. Chown, U.U. Müller, Detecting heteroscedasticity in non-parametric regression using weighted empirical processes, *J. R. Statistical Methodology Series B*, 80 (2018) 951-974.
- [7] N.A. Erilli, U. Yolcu, E. Egrioglu, C.H. Aladag, Y. Oner, Determining the Most Proper Number of Cluster in Fuzzy Clustering by Artificial Neural Networks, *Expert Systems with Applications*, 38 (2011) 2248-2252.
- [8] J. Frost, *Regression Analysis: An Intuitive Guide*, Paperback (2019)
- [9] I. Gath, A.B. Geva, Unsupervised Optimal Fuzzy Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7 (1989) 773-780.
- [10] D.N. Gujarati, *Basic Econometrics*, 4th ed. McGraw Hill pub., USA (2002)
- [11] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with fuzzy covariance matrix, In *Proceedings of the IEEE CDC, San Diego*, (1979) 761–766.

- [12] M. Habshah, M. Sani, J. Arasan, Robust Heteroscedasticity Consistent Covariance Matrix Estimator based on Robust Mahalanobis Distance and Diagnostic Robust Generalized Potential Weighting Methods in Linear Regression, *Journal of Modern Applied Statistical Methods*, 17 (2018)
- [13] M.J. Harrison, B.P.M. McCabe, A test for heteroscedasticity based on least squares residuals, *J. Am. Stat. Assoc.*, 74 (1979) 494-500.
- [14] R.J. Hathaway, J.C. Bezdek, Switching Regression Models and Fuzzy Clustering, *IEEE Transactions fuzzy Systems*, 3 (1993) 195-204.
- [15] X. Hu, F. Sun, Fuzzy Clustering Based Multi-Model Support Vector Regression State of Charge Estimator for Lithium-ion Battery of Electric Vehicle, *IEEE Xplore*. (2009)
- [16] R. Koenker, G. Bassett, Robust tests for heteroscedasticity based on regression quantiles, *Econometrica*, 50 (1981) 43-61.
- [17] R. Koenker, Q. Zhao, L-Estimation for Linear Heteroscedastic Models, *Journal of Nonparametric Statistics*, 3 (1994) 223-235.
- [18] Z. Li, X. Hong, K. Hao, L. Chen, B. Huang, Gaussian process regression with heteroscedastic noises: A machine-learning predictive variance approach, *Chemical Engineering Research and Design*, 157 (2020) 162-173.
- [19] C.L. Mei, F. Chen, W.T. Wang, P.C. Yang, S.L. Shen, Efficient estimation of heteroscedastic mixed geographically weighted regression models, *The Annals of Regional Science*. (2020)
- [20] H.C. Müller, U. Stadtmüller, Estimation of heteroscedasticity in regression analysis, *Ann. Statist.*, 15 (1987) 610-25.
- [21] N.K. Nagwani, S.V. Deo, Estimating the Concrete Compressive Strength Using Hard Clustering and Fuzzy Clustering Based Regression Techniques, *The Scientific World Journal*. (2014)
- [22] J.V. Oliveira, W. Pedrycz, *Advances in Fuzzy Clustering and Its Applications*, John Wiley & Sons Inc. Pub., West Sussex, England (2007)
- [23] T.P. Ryan, *Modern Regression Methods*, Wiley, New York (1997)
- [24] M. Sato-Ilic, On Fuzzy Clustering based Regression Models, *NAI-ITS 2004 International Conference*, (2004) 216-221.
- [25] M. Sato-Ilic, T. Matsuoka, On an Application of Fuzzy Clustering for Weighted Regression Analysis, *Proceedings of the 4th ARS Conference of the IASC*. (2002) 55-58.
- [26] A. Shick, Improving weighted least squares estimates in heteroscedastic linear regression when the variance is a function of the mean response, *Journal of Statistical Planning and Inference*, 76 (1999) 127-144.
- [27] A.P. Verbyla, Modelling Variance Heterogeneity Residual Maximum Likelihood and Diagnostics, *Journal of the Royal Statistical Society*, 55 (1993) 493-508.
- [28] S. Weisberg, *Applied Linear Regression*, 3rd ed., Willy. (2005)
- [29] J.M. Wooldridge, *Introductory Econometrics: A Modern Approach*. 5th ed., South-Western Cengage Learning, USA. (2012)
- [30] J. Wu, Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis, *The Annals of Statistics*, Vol. 14, No 4. (2012)
- [31] L. Xie, G. Beni, A Validity Measure for Fuzzy Clustering, *IEEE Trans. On Pattern Analysis and Machine Int.* 13 (1991) 841-846.
- [32] J. Zhang, Y. Gai, B. Lin, X. Zhu, Nonlinear regression models with single-index heteroscedasticity, *Statistica Neerlandica*, 73 (2019) 292-316.