Fuzzy Rank-Based Clustering for Change Point Detection in Regression

Hazem Abd razaq and Mohammed Jasim Mohammed
Department of Statistics, College of Administration and Economics, University of Baghdad, 10071 Baghdad, Iraq
Hazem.Abd2201@coadec.uobaghdad.edu.iq, m.jasim@coadec.uobaghdad.edu.iq

Keywords: Fuzzy Change Points, Multistage Regression, Rank-Based Estimator, Fuzzy Rank-Based.

Abstract: Detecting change points in regression models is an essential statistical method for analysing phenomena that change abruptly or occur in multiple stages. Determining changes in data helps researchers to analyse variations and improve the accuracy of statistical and predictive models. This analysis becomes more critical when data contains outliers or errors that follow a heavy-tailed distribution, especially as traditional methods become less effective at detecting changes and estimating model coefficients. In this paper, we propose a new robust fuzzy method called FRBCP, which combines the fuzzy cluster algorithm with a rank-based estimator. This combination allows for dealing with fuzzy change points, reducing the influence of outliers, and achieving robustness of the estimate. The proposed method was compared with the fuzzy change point algorithm and Muggeo methods through numerical experiments and real data. The results showed that the proposed methods are satisfactory in detecting change points and estimating model coefficients, and are mainly effective when the data contains outliers or an error-heavy-tailed distribution.

1 INTRODUCTION

Regression analysis is among the most important statistical tools used to study the relationship between variables. Still, in many regression models, the relationship between the response variable and the explanatory variables changes at some location points, which are called Change Points (CPs), Knots, Break Points, or Thresholds [1] This represents a shift in the data pattern because the slope of the regression model changes, and these points become a tool for linking every two regimes in the model. Change point regression models are commonly applied in various fields, including medical research, econometrics, biology, and ecology.

Researcher Quandt [2], [3] was the first to study the switch regression problem in the years [1958, 1960]. The researcher explained the importance of estimating the location of the points where a change occurs from one model to another, many researchers have studied the detection of change points in regression models using different approaches; Sun et al [4] explained the procedures for the maximum likelihood to estimated change point, Teaming Xu[5] studied least squares method when we have two or more sub models, In [6], [7] the researchers studied robust Bayesian methodology in a segmented linear regression model, assuming that the variance is heterogeneous, Muggeo [8] proposed an alternative technique for estimating segmented regression model called linear reparameterization technique. Shi et al [9], Zhang and Li [10] used a rank-based approach with a reparameterization technique to estimate all model parameters, including the location of the change point, showing that the proposed methods are robust against outliers and heavy-tailed errors. Fuzzy clustering techniques are particularly suitable for identifying change points when boundaries between segmented regression are unclear [11]. Chang et al [12] introduced a fuzzy change point algorithm (FCP) for estimating the change Point using the fuzzy approach, and the researchers explained how fast it converges and the possibility of obtaining accurate estimates of the parameters without needing to know the distribution of the data. Although (FCP) is considered efficient in detecting change points, it is sensitive to outliers or errors in heavy-tailed distributions [13].

In this paper, we develop a robust procedure for switching regression models by connecting (FCP) with rank-based estimation, which is called FRB-CP. The proposed algorithm is more robust against outliers and heavy-tailed errors.

2 METHODOLOGIES

2.1 Change Points and Multistage Regression Model

In regression models, a change point is defined as the point at which the slope of the regression line changes and is divided into two or more regimes or parts. as explained in Figure 1, these points serve as connecting points between each adjacent pair of parts [14].

$$r = CPs + 1, r \geq 2,$$

where:

- r: number of regimes in the models.
- CPs: refers to the number of detected change points.

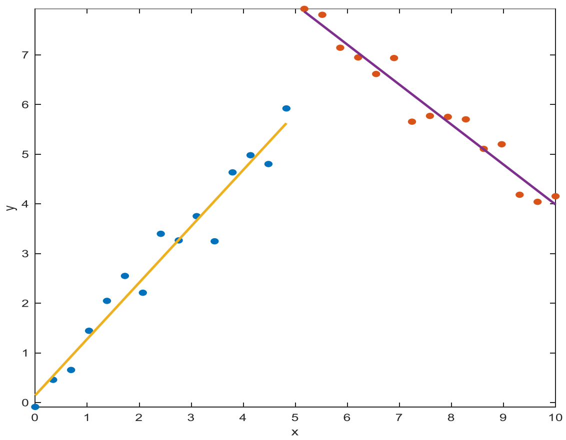

Figure 1: The division of the regression line into two parts. CP true 15, $y_1 = 1.2x, y_2 = 12 - 0.8x$.

Multistage regression is one of the most common problems faced by researchers when studying regression models. It occurs because of one or more change points that divide the nonlinear regression line into linear parts, where each part will have a different function. [15].

$$f(x) = \begin{cases} f_1(x:\beta_1) & x_0 \leq x \leq \delta_1 \\ f_2(x:\beta_2) & \delta_1 \leq x \leq \delta_2 \\ \vdots & \vdots \\ f_r(x:\beta_r) & \delta_{r-1} \leq x \leq x_n \end{cases}, \quad (1)$$

- δ: Represents the CP in the model, which must be estimated.
- r: The number of sub-models.
- β_r: The parameters of the model, which must be estimated.
- $f_j(x:\beta_j)$: a sub-linear regression model, $j = (1,2,\dots,r)$.

2.2 The Model

For any interval $(x_0,, x_n)$ on the real line, suppose a data set $D = \{x_1, x_2, \dots, (x_n, y_n)\}$. Arranged in increasing order according to the index of the observations, so we can determine the location change point by the order of data[13], and consider a regression model that includes one change point, ie, r=2, we can write switching regression as a simple two-line linear regression model as[15]:

$$fy_j = \begin{matrix} \beta_{0L} + \beta_{1L}x_j + e_{ijL} & x_0 \leq x_j \leq \delta \\ \beta_{0R} + \beta_{1R}x_j + e_{ijR} & \delta < x_j \leq x_n \end{matrix}, (2)$$

where $e_{ij} \sim N(0, \sigma^2), x_j \in \Re$ Are the independent variables, $(x_1, x_2, \dots, x_n) = (X_L', X_R')$.

$$X_L' = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_\delta \end{pmatrix},$$

$$X_R' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_{\delta+1}, & x_{\delta+2}, & \dots, & x_n \end{pmatrix},$$

and $y_j \in \Re$ are corresponding dependent variables,

$$Y_L' = (y_1, y_2, \dots, y_\delta),$$

$$Y_R{}' = (y_{\delta+1}, y_{\delta+2}, \dots, y_n).$$

If $\hat{\delta}_j$ are known:
and $x_j < \hat{\delta}_j < x_{j+1}, j = 1, \cdots, n$, then we can estimate the parameters for each half of the model by the unconstrained least squares method:

$$\hat{\beta}_L = (X_L{}'X_L)^{-1}X_L' \, Y_L,$$

$$\hat{\beta}_R = (X_R{}'X_R)^{-1}X_R' \, Y_R.$$

$$\text{SEE of all model} = SSE_L + SSE_R$$

However, in many cases, the change points are unknown, so we have three problems that researchers encounter when studying switching regression models.[12]:

- Problem 1: Determine the number of change points.
- Problem 2: Estimating the locations of change points and boundaries of all subsystems.
- Problem 3: Estimating all parameters of the regression model within each system.

If the number of change points and their locations are known, the third problem can be solved by using traditional regression methods for each regime. Hence, the first two problems are the main focus of the literature, which discusses the detection and estimation of change points; they are often studied separately because of their technical complexity. [13] For example, Bai [16] proposed that CPs are known, and Muggeo [8] assumed that CPs are continuous. In this study, we employ a fuzzy approach to detect a change point, which enables the change point to belong to each cluster with a different membership function.

2.3 Fuzzy Logic

In 1965, Zadeh introduced fuzzy set theory as a way to handle uncertainty. Sometimes, it is not possible to split elements into clearly separated groups; therefore, a fuzzy approach is used, as this method allows elements to belong to multiple groups simultaneously, each with different degrees of membership [17].

2.4 Membership Function

It is one of the fundamental concepts in fuzzy sets, used to represent the degree of membership of each element in the fuzzy set, which is a positive function within the interval [0, 1]. There are several types of membership functions, and depending on the type, different kinds of fuzzy sets are formed [17].

2.5 Fuzzy Change Points Algorithm FCP

Detecting a change point using a fuzzy approach was first introduced in 2015 by Chang et al. [12]. He assumes $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ are observations randomly selected from the switching regression system, and these points arranged according to the increasing order of x_j Instead of using probabilities to detect a change point, we employ a fuzzy approach, assuming that any one of the observations, except the last, could be a change point. $(X_1, y_1), (x_2, y_2), \dots, (x_{n-1}, y_{n-1})\}$ with membership $\alpha_1, \alpha_2, \dots, \alpha_{n-1}$ respectively where:

$$\sum_{i=1}^{n-1} \alpha_i = 1 , 0 \le \alpha_i \le 1,$$

Then we transform membership above into pseudo membership to each cluster, so if we have the j-th observation and we want to divide it into two clusters, the membership function will be as follows:

$$\mu_{rj} = \begin{cases} \sum_{i=j}^{n-1} \alpha_i, & \text{if } r = 1 \text{ and } 1 \le j \le n-1 \\ 0, & \text{if } r = 1, j = n \\ 1 - \mu_{1j}, & \text{if } r = 2 \end{cases} . \quad (3)$$

Then $\hat{\beta}_r$ are estimated by minimizing the objective function:

$$\hat{\beta}_r = \arg min \sum_{j=1}^{n} (\mu_{rj})^m (Y_j - X_j{}^T \beta_r)^2, \quad (4)$$

$$\hat{\beta}_r = (X^T U_r X)^{-1} X^T U_r Y , \qquad r = 1,2, \quad (5)$$

Where:

- $U_r \in R^{n \times n}$ A diagonal matrix having. $(\mu_{rj})^m$ as its j^{th} Diagonal elements.
- n: Number of observations of the model.
- μ_{ij}: Membership function.
- m: is a fuzziness index or the (weighting exponent) (m>1).

IF we have the i-th point $(x_i, y_i)$is a change point, then we can define $d_i{}^2$ As the total sum of squares of residuals in each regime:

$$d_i{}^2 = \sum_{r=1}^{2} (Y_{ir} - \hat{\beta}^T X_{ir})^T (Y_{ir} - \hat{\beta}^T X_{ir}). \quad (6)$$

Then update and find a new iteration of α as follows:

$$\alpha_i = \frac{\exp(-d_i{}^2/(m-1))}{\sum_{j=1}^{n-1} \exp(-d_j{}^2/(m-1))}. \quad (7)$$

After the convergence algorithm, we chose the largest membership as estimated to the CP.

3 RANK-BASED ESTIMATORS FOR REGRESSION

To achieve robustness, Rank based estimator is a commonly used in regression model, because This method used ranks of residual instead of their absolute values, which makes it less sensitive to noise from the classical assumptions of the normal distribution, in Change point detection recent studies have proved that the rank method gives more accurate and a robust estimation of all parameters of the model including the location of the change point. For example, in [9], [10], Researchers used a rank-based estimator with a reparameterization method, as explained by Muggeo. When the model contains one or multiple change points, it can be referred to as the iterative rank-based estimator. This estimator is based on the dispersion function given by Jaeckel and Jureckova [18], [19]. In linear models without change points, in switching regression, the rank-based estimator is arrived at by replacing the Euclidean norm in the objective function of the ordinary least squares estimator $\|e\|_2^2 \sum_{i=1}^{n} e_i^2$ With the following criterion [20]:

$$\|e\|_{\emptyset} = \sum_{i=1}^{n} \emptyset\left(\frac{R_i}{n+1}\right) e_i, \quad (8)$$

Where e_i:The residuals. R_i: Rank of residuals.$\emptyset(\cdot)$ it is a non-decreasing square objective function defined on the interval (0,1) and standardized such that

$$\int \emptyset(u)^2 du = 1.$$

$$\int \emptyset(u) du = 0.$$

The target function is usually chosen from a type error distribution. The Common usage is the Wilcoxon function:

$$\emptyset(t) = \sqrt{12}(t - 0.5),$$

or the signal target function:

$$\emptyset(t) = sgn(t - 0.5)\,.$$

The Wilcoxon function is more efficient in cases where the distribution of error is symmetrical or has particularly heavy tails, and it provides robust and effective estimation [10].

3.1 Fuzzy Rank-Based Clustering Algorithm for Change Point Regression Models (FRB-CP)

To make the parameter more robust we propose combine robustness in rank-based estimator with flexibility in (FCP) to estimate parameter and location change point in switching regression model, the main idea of the proposal replace sum of the squares of the residuals in each regimes in (FCP) algorithm with the rank dispersion function equation (8), as a result we get a new algorithm called FRB-CP, the proposed algorithm is more robust against outliers and heavy tailed errors. Hence, we choose the Wilcoxon function. The steps of the proposed method can be summarized as follows:

Determined a fuzzy index (m) and termination condition $\varepsilon > 0$.

Step 1: Initialize (α_i^0, $\hat{\beta}_i^{\,0}$) as iteration I=0.

Step 2: Using equation 3 to calculate μ_{rj}, $r = 1{,}2$ $j = 1{,}2, \ldots, n$.

Step 3: At the r-th step, update parameters $\hat{\beta}_r^{\,I+1}$ by fitting a rank-based regression:

$$\hat{\beta}_r^{\,I+1} = \arg\min \sum_{i=1}^{n} (\mu_{rj})^m \sqrt{12}\left(\frac{R_i^{\,I-1}}{n+1} - 0.5\right) d_i^{\,I-1}, \quad (9)$$

R_i: rank of the i-th residual d_i for each regime is defined as

$$d_i = (Y_{ir} - \hat{\beta}_r^{\,T} X_{ir}) \quad , r = 1{,}2,$$

$$Y_{i1} = \begin{bmatrix} Y_1 \\ Y_1 \\ \vdots \\ Y_i \end{bmatrix}, Y_{i2} = \begin{bmatrix} Y_{i+1} \\ \vdots \\ \vdots \\ Y_n \end{bmatrix},$$

$$X_{i1} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1P} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{i1} & \cdots & X_{ip} \end{bmatrix},$$

$$X_{i2} = \begin{bmatrix} 1 & X_{(i+1)1} & \cdots & X_{(i+1)P} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{np} \end{bmatrix}.$$

Step 4: find a new iteration of α_i as the same as (7).

Step 5: if $\left\|\hat{\beta}_r^{\,I+1} - \hat{\beta}_r^{\,I}\right\| < \varepsilon$ or, $\|\alpha_i^{\,I+1} - \alpha_i^{\,I}\| < \varepsilon$ Stop and give CP.

$\widehat{CP} = \arg\max\,(\alpha_i)$, else $I = I + 1$, go to step 2.

4 SIMULATION AND REAL DATA

4.1 Simulations

A series of numerical simulations using the MATLAB program was conducted to evaluate the performance of the algorithms employed in this study for detecting change points within linear regression models. We set the stopping condition. ε=0.000005 and the initial value,

$$\alpha_i^{\,0} = 1/C_{r-1}^{n-1}.$$

Simulation experiments are conducted according to the following stages:

1) Determined sample size n.
2) Generating explanatory variable (x) from the interval [0, 10]. The data is divided into non-overlapping subsections arranged in increasing order according to the index of the observations.
3) Generating random error according to the Box-Muller method in four different scenarios. In

case 1, the error is a normal distribution $e_i \sim N(0,1)$ in case 2, $95\% \, e_i \sim N(0, 0.3) + 5\% \, e_i \sim N(0, 3)$ in case 3, $90\% \, e_i \sim N(0, 0.3) + 10\% \, e_i \sim N(0, 3)$ in case 4, the errors $e_i \sim t_2$.

4) Generating the dependent variable Y_i based on the values of the explanatory variable X_i generated in 2, and the values of the random error e_i generated in 3, using switching regression model in formula 2, mentioned in the theoretical aspect in section II, three methods to estimating location change point and the coefficients of the model were compared FCP, FRB-CP, and segmented method(Seg) created by Muggo[8] each simulation replication with (R=1000) to analyze the performance of each method, The three methods are compared by MSE for each parameter, The simulation results are listed in the Tables 1, 2.

The results of the simulation experiments in Tables 1, 2, showed the following:

- Case 1: when $e_i \sim N(0,1)$The FCP method has a minimum value of MSE for estimating the location of the change point and parameters in each regime, compared with other methods.
- Cases 2,3: when the error contaminated ratio (5%, 10%) respectively, and case 4, when $e_i \sim (t_2)$The proposed method, FRB-CP, has obtained the lowest MSE for all parameters compared with other methods.

4.2 Real Data

The dataset provides information related to the Tigris River in the Baghdad governorate, specifically regarding bed load transport (kg/s) as a dependent variable Y, which is dependent on sediment discharge (m^3/s) as an independent variable x.

Table 1: Simulation result, cases error 1.2.3.4, n=30, cp=10.

n=30		$\widehat{\beta}, \widehat{CP}$	$\hat{\beta}_{0L}$	$\hat{\beta}_{1L}$	$\hat{\beta}_{0R}$	$\hat{\beta}_{1R}$	$\widehat{CP}$
Error		TRUE	1	0.5	-4.3	0.5	10
1	FCP	$\widehat{\beta}, \widehat{CP}$	1.02	0.485	-4.337	0.505304	10.03
		MSE	0.3213	0.1	0.642	0.012527	0.31
	FRB-CP	$\widehat{\beta}, \widehat{CP}$	1.1624	0.367	-4.246	0.494625	10.04
		MSE	0.604	0.203	1.147	0.021813	0.326
	Seg	$\widehat{\beta}, \widehat{CP}$	2.5385	-0.867	-4.77	0.555105	15.79
		MSE	2.6314	1.933	3.657	0.053139	36.79
2	FCP	$\widehat{\beta}, \widehat{CP}$	0.9996	0.505	-4.341	0.505593	10.02
		MSE	0.2509	0.064	0.401	0.008175	0.13
	FRB-CP	$\widehat{\beta}, \widehat{CP}$	1.0484	0.461	-4.286	0.498503	10.01
		MSE	0.1015	0.031	0.253	0.004972	0.13
	Seg	$\widehat{\beta}, \widehat{CP}$	2.5253	-0.854	-4.798	0.556525	15.72
		MSE	2.6657	2.14	6.879	0.085487	36.44
3	FCP	Mean	1.0074	0.495	-4.295	0.500353	10.14
		MSE	0.3614	0.104	3.007	0.03839	1.409
	FRB-CP	$\widehat{\beta}, \widehat{CP}$	1.0792	0.439	-4.261	0.494872	10.11
		MSE	0.1574	0.055	0.393	0.007602	1.154
	Seg	$\widehat{\beta}, \widehat{CP}$	2.5005	-0.814	-4.968	0.574928	15.82
		MSE	2.7822	2.178	10.87	0.137673	38.62
4	FCP	$\widehat{\beta}, \widehat{CP}$	1.0234	0.484	-4.191	0.488315	10.22
		MSE	0.3084	0.127	484.8	5.209698	2.715
	FRB-CP	$\widehat{\beta}, \widehat{CP}$	1.0739	0.435	-4.255	0.49208	10.19
		MSE	0.1626	0.071	0.916	0.032021	2.551
	Seg	$\widehat{\beta}, \widehat{CP}$	2.5613	-0.884	-4.775	0.554012	15.7
		MSE	2.6201	1.96	53.91	0.651655	34.95

Table 2: Simulation result, cases error 1.2.3.4, n=70, cp=10.

n=70		$\hat{\beta}, \widehat{CP}$	$\hat{\beta}_{0L}$	$\hat{\beta}_{1L}$	$\hat{\beta}_{0R}$	$\hat{\beta}_{1R}$	$\widehat{CP}$
Error	Method	true	1	0.5	-4.3	0.5	10
1	FCP	$\hat{\beta}, \widehat{CP}$	1.0203	0.504	-4.319	0.502486	10.01
		MSE	0.3598	0.578	0.09	0.002383	0.022
	FRB-CP	$\hat{\beta}, \widehat{CP}$	1.1848	0.177	-4.312	0.5015	10.01
		MSE	0.5633	1.007	0.142	0.003868	0.026
	Seg	$\hat{\beta}, \widehat{CP}$	2.5702	-2.777	-4.352	0.506944	15.66
		MSE	2.7225	10.9	0.143	0.003442	33.74
2	FCP	$\hat{\beta}, \widehat{CP}$	0.9981	0.514	-4.305	0.501494	10.01
		MSE	0.1864	0.294	0.059	0.001437	0.033
	FRB-CP	$\hat{\beta}, \widehat{CP}$	1.0437	0.436	-4.303	0.501012	10
		MSE	0.069	0.131	0.027	0.000645	0.036
	Seg	$\hat{\beta}, \widehat{CP}$	2.5461	-2.764	-4.315	0.502834	15.57
		MSE	2.5353	10.76	0.087	0.001971	32.05
3	FCP	$\hat{\beta}, \widehat{CP}$	1.0073	0.505	-4.34	0.507621	10.05
		MSE	0.3532	0.571	0.099	0.002417	0.121
	FRB-CP	$\hat{\beta}, \widehat{CP}$	1.0503	0.428	-4.313	0.502787	10.03
		MSE	0.0747	0.145	0.039	0.00098	0.072
	Seg	$\hat{\beta}, \widehat{CP}$	2.5555	-2.77	-4.361	0.510404	15.66
		MSE	2.6642	10.87	0.157	0.003479	33.7
4	FCP	$\hat{\beta}, \widehat{CP}$	0.9728	0.545	-5.258	0.597694	10.18
		MSE	0.3336	1.701	877.3	9.124844	7.126
	FRB-CP	$\hat{\beta}, \widehat{CP}$	1.0474	0.403	-4.289	0.498254	10.09
		MSE	0.1034	0.217	0.123	0.001842	4.492
	Seg	$\hat{\beta}, \widehat{CP}$	2.5261	-2.761	-4.631	0.534431	15.79
		MSE	2.5412	10.92	188.9	2.050768	43.88

Table 3: Real data for bed load at the site (CS10) using a formula (Van Rijn, 1984).

x	y
300	2.520
400	2.599
500	0.449
600	0.776
700	1.218
800	1.870
900	2.479
1000	3.083

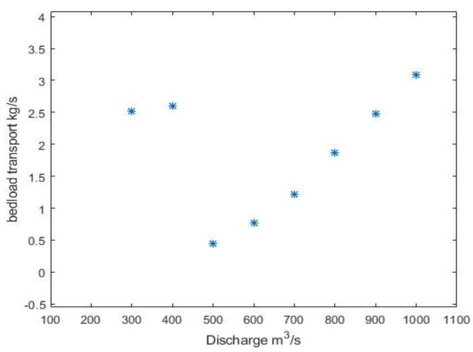

Figure 2: Spread of real data for the site (CS10) of the Tigris River.

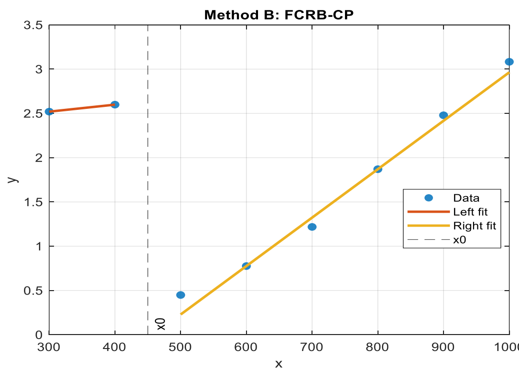

Figure 3: The Box Plot shows that the real data do not contain any outlier values.

The National Center for Water Resources Management in Iraq stated that it is possible to calculate and predict the riverbed load data using formulas from well-known and published equations in research. [21]. The researchers proposed and developed a wide series of formulas that predict the bed load, and they applied them to 16 sites along the northern part of the Tigris River in Baghdad governorate, and the researchers have shown that the site data (CS10) has shown a change in the slope of the discharge and this site includes eight observations, as shown in the Table 3, and spread of real data explained in Figure 2.

Table 4: Parameter and mean squared error for all models using real data.

B, cp	β_{0L}	β_{1L}	β_{0R}	β_{1R}	MSE	cp
FCP	2.283	0.001	2.411	0.005	0.016	2
FRB-CP	2.283	0.001	2.982	0.006	0.044	2
Seg	6.599	0.012	1.876	0.005	0.218	3

Where:

$$MSE = \frac{SSE_L + SSE_R}{(n - r(p+1))}, r = 2.$$

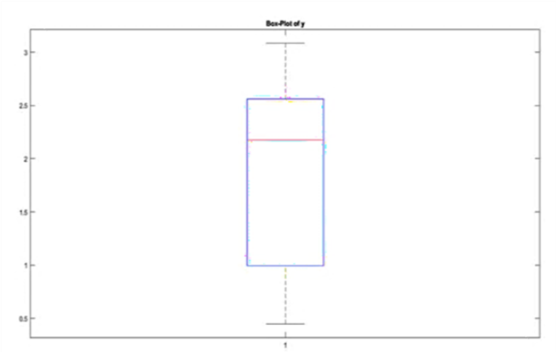

Figure 4: Fitting real data using FCP.

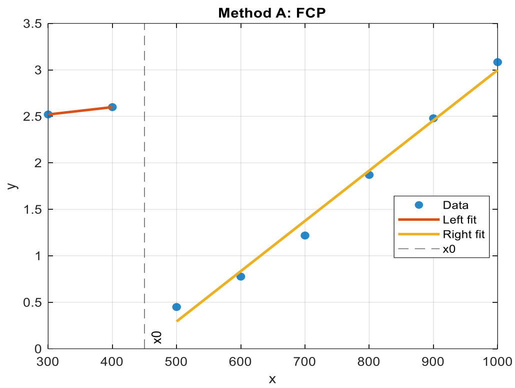

Figure 5: Fitting real data using FRB-CP.

4.3 Detection of Outliers in Real Data Using Box Plot

The box plot is a method for showing the number of outliers in the data and determining the spread of the data distribution around the median. The box plot of the dependent variable y has been created in applied data. Figure 3, shows the box diagram of the dependent variable for the applied data. It was found that it does not contain any outlier values. It also shows that the median of the values is approximately equal to 2.2, and that the distribution of data within the first and second Springs is skewed towards the left.

Also, we compared three methods FCP, FRB-CP, and Seg, for fitting real data in Table 4.

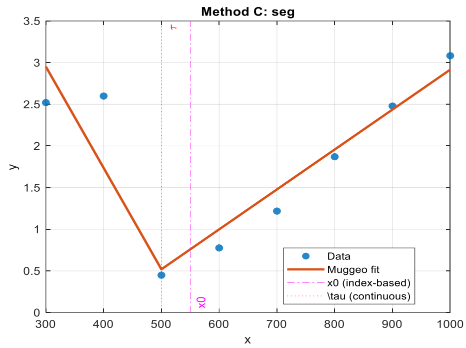

Figure 6: Fitting real data using Seg.

5 DISCUSSION

The goal of the study was to develop a robust method for detecting change points in regression models by combining a fuzzy approach with rank-based estimation. Simulation results demonstrated that the proposed method is more robust against outliers and heavy-tailed errors., Provides stable performance in small and medium samples. The application to real data from the Tigris River also demonstrated the method's ability to detect change points with high accuracy, confirming its validity for practical use in real-world data. It is also preferable in future research to compare the proposed method with other robust methods, especially at large sample sizes and under the influence of other heavy-tailed distributions, such as the Cauchy or Laplace distribution, to evaluate its performance.

6 CONCLUSIONS

In this paper, we introduce a new robust fuzzy method, called FRBCP, to detect change points in linear regression models. This method combines the fuzzy approach with robustness in a rank-based estimator. Evaluation was performed using a comprehensive set of simulated experiments and applied to real data, simulating various scenarios involving clean data as well as data contaminated with outliers or errors that have heavy-tailed distributions.

1) Performance in ideal conditions: Numerical experiments and applications to real data,

shown in Figures 4, 5 and 6, demonstrate that the performance of both FCP and FRB tends to converge when outliers are absent. This convergence confirms that integrating the rank-based estimator within the fuzzy framework does not impair the original effectiveness of the FCP algorithm in the absence of contaminants, but rather preserves its efficiency, especially in estimating change points, as shown in case 1.

2) Performance in non-ideal conditions contaminated data: The study also indicates that FRB-CP handles outliers more flexibly than other methods. In general, the proposed approach provides a superior evaluation of the model, especially when dealing with data contaminated by outliers or heavy tails.
3) The analysis real data bed load transport and discharge for the Tigris River's location (CS10), as shown in Table 4, confirms that both the first and second methods effectively detect the point of change in river discharge. This helps determine the flow systems essential for the maintenance of the channels

REFERENCES

[1] Y. Zhou, S. Gao, D. Guo, X. Wei, J. Rokne, and H. Wang, "A survey of change point detection in dynamic graphs," IEEE Trans. Knowl. Data Eng., vol. 36, no. 2, pp. 123-145, 2024.

[2] R. E. Quandt, "The estimation of the parameters of a linear regression system obeying two separate regimes," J. Am. Stat. Assoc., vol. 53, no. 284, pp. 873-880, 1958.

[3] R. E. Quandt, "Tests of the hypothesis that a linear regression system obeys two separate regimes," J. Am. Stat. Assoc., vol. 55, no. 290, pp. 324-330, 1960.

[4] L.-H. Sun, Z.-Y. Huang, C.-Y. Chiu, and N. Ning, "Detecting structural shifts and estimating change-points in interval-based time series," IEEE Trans. Knowl. Data Eng., to be published.

[5] T. Xu and Y. Wei, "A Least Squares Estimator for Gradual Change-Point in Time Series with m-Asymptotically Almost Negatively Associated Errors," Axioms, vol. 12, no. 9, p. 894, Sep. 2023.

[6] L. Cappello and O. H. M. Padilla, "Bayesian variance change point detection with credible sets," IEEE Trans. Pattern Anal. Mach. Intell., early access, 2024.

[7] S. G. Kang, W. D. Lee, and Y. Kim, "Bayesian multiple changing-points detection," AIMS Mathematics, vol. 10, no. 3, pp. 4662-4708, 2025.

[8] V. M. R. Muggeo, "Estimating regression models with unknown break-points," Stat. Med., vol. 22, no. 19, pp. 3055-3071, Oct. 2003,, doi: 10.1002/sim.1545.

[9] S. Shi, Y. Li, and C. Wan, "Robust continuous piecewise linear regression model with multiple change points," J. Supercomput., vol. 76, no. 5, pp. 3456-3480, 2020.

[10] F. Zhang and Q. Li, "Robust bent line regression," J. Stat. Plan. Inference, vol. 185, pp. 41-55, 2017.

[11] K.-P. Lu, S.-T. Chang, and M.-S. Yang, "Change-point detection for shifts in control charts using fuzzy shift change-point algorithms," Comput. Ind. Eng., vol. 93, pp. 12-27, 2016.

[12] S.-T. Chang, K.-P. Lu, and M.-S. Yang, "Fuzzy change-point algorithms for regression models," IEEE Trans. Fuzzy Syst., vol. 23, no. 6, pp. 2343-2357, 2015.

[13] K.-P. Lu and S.-T. Chang, "Robust fuzzy clustering algorithms for change-point regression models," Int. J. Uncertainty, Fuzziness Knowledge-Based Syst., vol. 28, no. 05, pp. 701-725, 2020.

[14] A. Roslyakova, "Modeling thermodynamical properties by segmented non-linear regression," Ph.D. dissertation, Ruhr University Bochum, Bochum, Germany, 2017.

[15] S. A. Julious, "Inference and estimation in a changepoint regression problem," J. R. Stat. Soc. Ser. D (The Statistician), vol. 50, no. 1, pp. 51-61, 2001.

[16] J. Bai, "Least squares estimation of a shift in linear processes," J. Time Ser. Anal., vol. 15, no. 5, pp. 453-472, 1994.

[17] R. Saatchi, "Fuzzy logic concepts, developments and implementation," Information, vol. 15, no. 10, p. 560, 2024.

[18] L. A. Jaeckel, "Estimating regression coefficients by minimizing the dispersion of the residuals," Ann. Math. Stat., pp. 1449-1458, 1972.

[19] J. Jurečková, "Nonparametric estimate of regression coefficients," Ann. Math. Stat., pp. 1328-1338, 1971.

[20] B. Brune, I. Ortner, and P. Filzmoser, "A rank-based estimation method for mixed effects models in the presence of outlying data," J. Data Sci., vol. 4, no. 7, pp. 123-145, 2024.

[21] A. A. Ali, N. Al-Ansari, Q. Al-Suhail, and S. Knutsson, "Spatial measurement of bed load transport in Tigris River," J. Earth Sci. Geotech. Eng., vol. 7, no. 4, pp. 55-75, 2017.