JOURNAL OF MECHANICS OF CONTINUA AND MATHEMATICAL SCIENCES
www.journalimcms.org

ISSN (Online) : 2454 -7190 Vol.-15, No.-2, February (2020) pp 212-223 ISSN (Print) 0973-8975

SUGGESTING MULTIPHASE REGRESSION MODEL ESTIMATION WITH SOME THRESHOLD POINT

Omar Abdulmohsin Ali[1]

[1]Department of Statistics, College of Administration and Economics,

University of Baghdad, Iraq

dromar72@coadec.uobaghdad.edu.iq

https://doi.org/10.26782/jmcms.2020.02.00019

Abstract

The estimation of the regular regression model requires several assumptions to be satisfied such as "linearity". One problem occurs by partitioning the regression curve into two (or more) parts and then joining them by threshold point(s). This situation is regarded as a linearity violation of regression. Therefore, the multiphase regression model is received increasing attention as an alternative approach which describes the changing of the behavior of the phenomenon through threshold point estimation. Maximum likelihood estimator "MLE" has been used in both model and threshold point estimations. However, MLE is not resistant against violations such as outliers' existence or in case of the heavy-tailed error distribution. The main goal of this paper is to suggest a new hybrid estimator obtained by an ad-hoc algorithm which relies on data driven strategy that overcomes outliers. While the minor goal is to introduce a new employment of an unweighted estimation method named "winsorization" which is a good method to get robustness in regression estimation via special technique to reduce the effect of the outliers. Another specific contribution in this paper is to suggest employing "Kernel" function as a new weight (in the scope of the researcher's knowledge).Moreover, two weighted estimations are based on robust weight functions named "Cauchy" and "Talworth". Simulations have been constructed with contamination levels (0%, 5%, and 10%) which associated with sample sizes (n=40,100). Real data application showed the superior performance of the suggested method compared with other methods using RMSE and R^2 criteria.

Keywords: Data-driven strategy, kernel, multiphase regression, robustness, threshold point, winsorization.

I. Introduction

The multiphase regression model or so-called "multi-stage" regression model is used in the form of a threshold regression model that allows the relationship between the dependent variable and the explanatory variable to be changed via breakpoint or threshold point in the explanatory variable value.These models detect

and investigate changes along the x-coordinate axis, and approximate the nonlinear relationship (if there is any) between the dependent variable and the explanatory variable.In this case, we are dealing with error distribution with heavy tail type, as well as dealing with the problem of "heterogeneity"[V] between subgroups (or strata) within the study sample and overcoming the problem resulting from it.

This topic becomes more interactive for researchers and scholars in various applied fields. For instance, a biological essay by (Julious, 2001) [VII] distinguished the patients according to their breathing quality at some threshold point which represents the change-over from aerobic (oxygen inhaled) to anaerobic (dioxide exhaled). In psychological studies, there is evidence that the risk of preterm delivery depends on the mother's stress only when it becomes above a specific thresholdpoint (Whitehead, 2002)[XIV].

Traditional estimations have been introduced to deal with this issue. Maximum likelihood estimator of this type of regression model has been presented by (Muggeo, 2003) [X].Trimming method proposed by (Liu, 2011) [IX]can be used as well, but this will be limited if the sample size is small because the neglected extreme (outliers) values will affect the regression model estimation under and above threshold point.

Therefore, wins orization presented by (Yale, 1976) [XV] seemed to be a preferable suggested methodology which can be employed with replacing certain proportions of extreme values by the maximum and/or minimum values at the above/ under thresholdboundaries instead of neglecting them without need to produce weights.

Robust approach (Fearnhead et al., 2017) [IV] was be used traditionally. Both Cauchy and Talworth weight functions (Dehnel, 2016)[III]were used with regular regression, but they can be employed as a suggested weighted approach to obtain the corresponding estimatorsof multiphase regression in this paper.

Klotsche, [VIII] introduced a nonparametric kernel representation instead of a multiphase regression parametric model. But, in this paper kernel function is employed as a weight function as a part of the new weighted estimator.

In this paper, the main goal of this paper was to suggest a new hybrid estimator obtained by aniterative heuristic algorithm which relies on data driven strategy that overcomes outliers. This idea was inspired by artificial intelligence algorithms where the estimations based onthe nearest neighbor criterion.

II. Methods and materials

II.i. Unweighted Methods

II.i.a. Maximum Likelihood Method

Maximum Likelihood method aims to obtain the parameter values which maximize the common likelihood function for a certain data. So, this will be the situation with multiphase regression model described by [X] as follows:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 U_i + \beta_3 V_i + e_i \quad ; \quad i = 1,2,3,\dots,n \qquad \dots (1)$$

where:

$$\beta_3 = \beta_2\left(c - c^{(0)}\right) \quad (2)$$

$$U_i = \left(x_i - c^{(0)}\right)_+ \quad , \quad V_i = -I\left(x_i > c^{(0)}\right) \quad (3)$$

$c^{(0)}$primary value of threshold point.

The Likelihood function of the model (1) associated with normal error term can be expressed by[VI]:

$$L(\beta_0, \beta_1, \beta_2, \beta_3, \sigma^2 / x_i, y_i) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{y_i - [\beta_0 + \beta_1 x_i + \beta_2 U_i + \beta_3 V_i]\}^2 . (4)$$

Logarithm of previous likelihood has to be derivedfor the corresponding parameters (β_0, β_1, β_2, and β_3), to get maximum likelihood estimator (MLE) of the parameters in the model (1) can be produced iteratively by the following steps:

i. Fixa seed estimate of the threshold point (say; $c^{(k)}$) falls into x range.

ii. Begin with k^{th}iteration, where is (k=1,2,3, …).

iii. Compute $U_i^{(k)}$and $V_i^{(k)}$ regarding equation (3).

iv. Fitting the model (1) to obtain parameter estimators of (β_0, β_1, β_2, and β_3).

v.Updating the threshold point through the formula below.

$$\hat{c} = c^{(k)} + \frac{\hat{\beta}_3}{\hat{\beta}_2} \quad (5)$$

vi. Resolve step (iii) iteratively to step (v). The stopping rule will be indicated by($\hat{\beta}_3 \approx 0$).

vii. The standard error of threshold point can be calculated according to Wald-type statistics[XI], i.e. SE($\hat{c}$), and then construct 95% confidence interval estimation of c.

It is worthy to mention that MLE coincides with ordinary least squares (OLS) estimation when data have normal distribution of the model (1) [VI].

II.i.b. Winsorization Method

Dixon, 1960 was the first researcher who introduced the winsoring operation tracing the work of Charles P. Winsor, 1940 [XII]. This method is originally based on reducing the effect of outliers on the regular (classical)linear regression model byreplacing the outliers themselves with other values that are closer to the original data rather than trimming them [XV].

The suggestion, here, is to employ this method on the multiphase regression model estimation for the first time (according to the scope of the researcher's knowledge) through the following steps:

i.Compute the estimation $\hat{y}_i$in the model (1) for the pairs (x_i ,y_i), i=1,2,…,n.

ii. Calculate the residuals

$$e_i = y_i - \hat{y}_i \qquad (6)$$

iii. The residuals arranged in ascending order such that $e_1 \le e_2 \le e_3 \le \cdots \le e_n$

iv. Determine the proportion (p_1%) (say; 20%) and its complement p_2=(1-p_1)% which is (80%) to determine the amount of winsorization for the outliers consequently.

v. Put the quantile values v_1 and v_2 that corresponding to the proportions p_1 and p_2 respectively.

vi. The winsoring (replacing) approach begins with substitution of the lower outliers by the value v_1 and the upper outliers by the value v_2.

vii. Calculate the new residuals e_i^* and then replace them in their original locations subjected to the following :

$$e^*_{i(winsor)} = \begin{cases} e_{g+1} , & i = 1,2,\dots,g \\ e_i , & i = g+1, g+2, \dots, n-g \\ e_i , & i = n-g+1, n-g+2, \dots, n \end{cases}$$

g: the number of points which have to be winsored at each corresponding outliers.

viii. Re-compute the new pair points (x_i ,y_i^*)where is

$$y^*_{i(winsor)} = \hat{y}_i + e^*_{i(winsor)} \qquad (7)$$

ix. Refine the estimation of model (1) again by considering $y^*_{i(winsor)}$values and obtain threshold point estimation as remarked in equation (5).

II.ii. Weighted Methods

In regular regression case, weighted estimation methods have robustnessfeature via lesseninglarge residuals due to existing of some violations such as outliers.But, this weightedmethodscan be suggested in multiphase regression model estimation with its threshold point estimation, too. The weighted method producesresistance against outliers with constraint of minimizing loss function $\rho(\cdot)$ instead of the traditional MLE. That can be done by getting the optimal solution for a specific data iteratively.

One can obtaina robust estimation of the parameters through minimizing $\sum \rho(e_i)$. Therefore, a weighted estimator can be gottenby the following procedure:

i. Starting with primary parameters $\hat{\theta}^{(k)} = (\hat{\beta}_0^{(k)}, \hat{\beta}_1^{(k)}, \hat{\beta}_2^{(k)}, \hat{\beta}_3^{(k)}, and\ \hat{c}^{(k)})$subjected to $\hat{\beta}_3^{(k)}$ starting with seed equal (0.01).

ii. Beginning with parameter estimators$\hat{\theta}^{(k+1)} = (\hat{\beta}_0^{(k+1)}, \hat{\beta}_1^{(k+1)}, \hat{\beta}_2^{(k+1)}, \hat{\beta}_3^{(k+1)})$through applying it on model (1) as follows:

a. Initializing with traditional method$\hat{\theta}^{(\ell)} = (\hat{\beta}_0^{(\ell)}, \hat{\beta}_1^{(\ell)}, \hat{\beta}_2^{(\ell)}, \hat{\beta}_3^{(\ell)})$, to be applied on the model (1) which can be rewritten as:

$$\hat{y}_i^{(\ell)} = \hat{\beta}_0^{(\ell)} + \hat{\beta}_1^{(\ell)} x_i + \hat{\beta}_2^{(\ell)} U_i^{(\ell)} + \hat{\beta}_3^{(\ell)} V_i^{(\ell)} + e_i \quad (8)$$

b. Calculate the residuals:

$$e_i^{(\ell)} = y_i - \hat{y}_i^{(\ell)} \quad (9)$$

c. Updating parameter estimators $\hat{\theta}^{(\ell+1)} = (\hat{\beta}_0^{(\ell+1)}, \hat{\beta}_1^{(\ell+1)}, \hat{\beta}_2^{(\ell+1)}, \hat{\beta}_3^{(\ell+1)})$ as a new robust weighted estimator subjected to minimize:

$$\sum_{i=1}^{n} \rho\left(e_i^{*(\ell)}\right) \quad (10)$$

Where: $e_i^{*(\ell)} = \frac{e_i^{(\ell)}}{\hat{\sigma}}$

So, the weighted estimator will be obtained $\theta^{\ell+1}$ as follows:

$$\hat{\theta}^{\ell+1} = (X'WX)^{-1}X'WY \quad (11)$$

Where: W is a diagonal square matrix with order n×n which its elements are $w\left(e_i^{*(\ell)}\right)$.

where is $w\left(e_i^{*(\ell)}\right) = \frac{\psi\left(e_i^{*(\ell)}\right)}{\left(e_i^{*(\ell)}\right)}$

Here, three distinct weight functions will be suggested to be employed which are:

Cauchy: $W(e^*) = \frac{1}{1+\left(\frac{|e^*|}{a}\right)^2}$, where (a=2.38.5)

Talworth: $W(e^*) = \begin{cases} 1 & if \quad |e^*| < a \\ 0 & if \quad 0 \end{cases}$, where (a=2.795)

Kernel (Epanechnikov): $W(e^*) = \frac{3}{4}(1 - e^{*2})$, $I_{\{|e^*| \leq \infty\}}$

Kernel has a special contribution as the first time to estimate the multiphase regression model with its threshold point (in the scope of the researcher's knowledge).

d. Loop iterations from step (a) to step (c) until justifying the convergencecriterion, i.e.,

$$\left|\hat{\theta}^{(\ell+1)} - \hat{\theta}^{(\ell)}\right| < 0.0001$$

Then, the new threshold point will be satisfying $\hat{\theta}^{(k+1)} = \hat{\theta}^{(\ell+1)}$

Omar Abdulmohsin Ali

iii. Repeat new run for threshold point cregarding equation (5) to be rewritten as:

$$\hat{c}^{(k+1)} = \hat{c}^{(k)} + \frac{\hat{\beta}_3^{(k+1)}}{\hat{\beta}_2^{(k+1)}} \qquad (12)$$

iv. Continue in resolving the above approach with a new threshold point cto until getting($\hat{\beta}_3 \approx 0$).

II. iii. Suggested (Hybrid) Method

The suggested a new hybrid estimator obtained by an ad-hoc algorithm which relies on data driven strategy**.**

i. Set the estimation of the five parameters (β_0, β_1, β_2, β_3, and c) which are obtained by the previous estimation unweighted methods (MLE and Winsorization), or by the weighted methods (Talworth, Cauchy, and Kernel).

ii. Compute RMSE criterion value of predicted $\hat{y}$values corresponding to the previous estimation methods.

iii. Compare RMSE values that computed in step (ii) above, to determine the three best methods from the five estimation methods mentioned above. Later, the best method will be determined among the last three methods to set the parameters ($\beta_{0\,Best}$, β_{1Best}, $\beta_{2\,Best}$, β_{3Best}, and c_{Best}) to rely on the upcoming comparisons.

iv. Calculate the standard error associated with each parameter of (β_{0i}, β_{1i}, β_{2i}, β_{3i}, and c_i), where is: i=1,2,3 which represent the best three candidates methods according to the following:

$$\mathrm{SE}(\hat{\theta}_{Best}) = \sqrt{\frac{\sum_{i=1}^{k}(\hat{\theta}_i - \bar{\theta})^2}{k-1}} \qquad (13)$$

k: represents the number of the candidate methods (which is here equal to 3).

$\hat{\theta}_i$: denotes any one parameter estimation of (β_{0i}, β_{1i}, β_{2i}, β_{3i}, and c_i) due to the candidate method (i).

v. Construct 95% confidence limits of the best method of the five parameters in the light of the three candidate estimation methods above subjected to the following formula:

$$\mathrm{C.I.}(\hat{\theta}_{Best}) = \hat{\theta}_{Best} \pm 1.96SE(\hat{\theta}_{Best}) \qquad (14)$$

vi. Generate (500) simulated numbers within under/ above the confidence limits for each parameter$\hat{\theta}_{Best(j)}$ such that: j=1,2,3, …, 500 the subscript represents the position of the generated value for the parameter$\hat{\theta}_{Best}$and getting different candidate models consequently.

vii. Linear combinations of the generated parameter values from the previous step have been selected, i.e., $(500)^5$distinct linear combinations have been obtained from

the different suggested models and finally get new one suggested primary model which is $(\beta_{0j1}, \beta_{1j2}, \beta_{2j3}, \beta_{3j4}$, and $c_{j5})$, where j1, j2, j3, j4, and j5=1,2,3, …, 500

Through the estimates of the primary model, an error estimate is extracted for this model and thus weights are obtained, based on each method: Cauchy, Talworth, and Kernel to be done for three weight matrices W_i , (i=1, 2, 3) according to the corresponding weight matrix computing methods described previously.

viii. The RMSE criterion is calculated again and for each estimate resulting from the above hybridization. In other words, extract the corresponding RMSE value for each weighted method from the weights mentioned above.

ix. The best estimate from step (iii) has been replaced by one of the three hybrid estimates resulting from the previous step (viii) according to the RMSE criterion when one of them is preferred, so, the new optimal estimate will be achieved and called$\hat{\theta}_{suggest(Hybrid)}$.

III. Simulation

Simulations were illustrated to reflect the performance of thesuggested(Hybrid) estimator compared with other weighted and unweighted estimators.Two different sample sizes were used which are (n=40, 100). In addition, regarding contamination proportions (0%, 5%, and 10%) of outliers with two distinct types of distribution which are normal distribution N(0,10) and t-distribution with (df=3)have been used.

Simulation experiments were constructed by using MATLAB program,(version13.a) with 500 replicates. Data generating were described by [II] and [I]as expressed below briefly.

$$y_i = 3.5 + 0.5x_i + I_{i2}(x_i - 5) + e_i$$

where:

$$I_{i2} = I(x_i \geq c)$$
$$I_{i1} = 1 - I_{i2}$$
$$i = 1, 2, 3, \ldots, n$$

Threshold point starting with seed (c = 5)

Table 1: RMSE according to Simulation experiment with (n=40)

Contamination / Method	No Contamination 0 %	Contamination with N(0,10)		Contamination with t_3	
		5 %	10 %	5 %	10 %
	RMSE	RMSE	RMSE	RMSE	RMSE
MLE	0.732346	0.975492	0.987352	1.586500	1.633336
Talworth	0.61848	0.933686	0.947751	1.057946	1.070116
Cauchy	0.555409	0.800723	0.814659	0.993533	1.017555
Kernel	0.072480	0.090321	0.093455	0.159381	0.159386
Winsor	0.406861	0.610063	0.607958	0.723076	0.729099
Suggest (Hybrid)	0.014278	0.034990	0.038739	0.042983	0.046199

From the numerical results in Table (1) one can notice that the suggested (Hybrid) method marks the best result with minimum RMSE valuefor all contamination levels and then followed by Krenel, Winsor, Cauchy, Talworth, and MLE respectively.

Table 2: Threshold point estimation with its uncertainty

According to simulation experiment initialed with (c=5) & (n=40)

Method	No Contamination 0 %	Contamination with N(0,10)		Contamination with t_3	
		5 %	10 %	5 %	10 %
MLE	5.142963 * (0.492892)** [4.176895, 6.109031] ***	5.148747 (0.624669) [3.924395, 6.373099]	5.034605 (0.633037) [3.793852, 6.275357]	5.249884 (1.00428) [3.281496, 7.218272]	5.233144 (1.026566) [3.221074, 7.245214]
Talworth	5.237626 (0.440624) [4.374004, 6.101249]	5.237638 (0.614853) [4.032526, 6.44275]	5.14254 (0.613948) [3.939201, 6.345879]	5.527807 (0.773678) [4.011398, 7.044217]	5.143222 (0.70721) [3.75709, 6.529354]
Cauchy	5.159864 (0.41203) [4.352286, 5.967442]	5.165285 (0.544177) [4.098698, 6.231872]	5.100855 (0.549116) [4.024587, 6.177123]	5.205368 (0.697836) [3.837609, 6.573126]	5.194741 (0.715873) [3.791629, 6.597852]
Kernel	5.180420724 (0.856246957) [3.502176689, 6.85866476]	5.264041025 (0.182245922) [4.906839017, 5.621243033]	5.128293309 (0.20300668) [4.73040021, 5.526186402]	5.354173 (0.403978) [4.562376, 6.14597]	5.253428 (0.257219) [4.749278, 5.757578]
Winsor	5.12125 (0.27167) [4.588777, 5.653723]	5.143117 (0.384765) [4.388977, 5.897257]	5.049218 (0.390592) [4.283659, 5.814778]	5.220224 (0.46453) [4.309744, 6.130703]	5.178688 (0.475462) [4.246782, 6.110594]
Suggest (Hybrid)	5.237658 (0.098103) [5.045375, 5.42994]	5.164776 (0.123257) [4.923192, 5.406359]	5.267419 (0.119869) [5.032475, 5.502362]	5.280329 (0.200681) [4.886995, 5.673663]	5.349435 (0.196493) [4.964309, 5.734562]

*: Threshold point estimator

**: Standard error of threshold point estimator

***: The lower limit (L.L.) and upper limit (U.L.) of the threshold point estimator

Considering table (2) above, the suggested method has best performance in threshold point estimating according to its corresponding standard error through all contamination levels. While other methods were as follows Krenel, Winsor, Cauchy, Talworth, and MLE respectively for (5% and 10%) contamination levels. The surprising result was with the priority of non-contamination (0%) level where Winsor method acts better than Kernel, i.e., the priority was Winsor, Krenel, Cauchy, Talworth, and MLE repectively.

Table 3: RMSE according to Simulation experiment with (n=100)

Contamination / Method	No Contamination 0 %	Contamination with N(0,10)		Contamination with t_3	
		5 %	10 %	5 %	10 %
	RMSE	RMSE	RMSE	RMSE	RMSE
MLE	0.724611	0.970704	0.983758	1.549166	1.558282
Talworth	0.614732	0.916347	0.922522	1.041511	1.062790
Cauchy	0.553868	0.786537	0.797801	0.990907	1.012590
Kernel	0.056678	0.072279	0.072298	0.135242	0.138733
Winsor	0.403021	0.596022	0.621543	0.705826	0.723446
Suggest (Hybrid)	0.009007	0.022309	0.022739	0.026818	0.029127

Through all the contamination levels (0%, 5%, and 10%) in table (3), the lowest value of RMSE associated with the suggested (Hybrid) method at first. While the second best value went toKernel method where it made a good data fitting. Then the rest of the methods come as follows: Winsor, Cauchy, Talworth, and MLE respectively.

Table 4: Threshold point estimation with its uncertainty

according to simulation experiment initialed with (c=5) & (n=100)

Contamination / Method	No Contamination 0 %	Contamination with N(0,10)		Contamination with t_3	
		5 %	10 %	5 %	10 %
MLE	5.011661 (0.265178) [4.491911, 5.531411]	5.012154 (0.361047) [4.304502, 5.719806]	5.015838 (0.363649) [4.303086, 5.728591]	5.01128 (0.586097) [3.862529, 6.160031]	5.072953 (0.610075) [3.877207, 6.26870]
Talworth	5.090419 (0.238242) [4.623464, 5.557374]	5.019137 (0.347334) [4.338363, 5.69991]	5.053097 (0.354144) [4.358974, 5.74722]	5.007508 (0.394131) [4.235011, 5.780005]	5.093475 (0.411925) [4.286103, 5.900848]
Cauchy	5.011712 (0.22341) [4.573828, 5.449597]	5.017738 (0.317465) [4.395506, 5.639971]	5.011116 (0.32064) [4.382661, 5.63957]	5.010419 (0.398527) [4.229306, 5.791532]	5.032835 (0.412225) [4.224874, 5.840797]
Kernel	5.057801 (0.049002) [4.961758, 5.153844]	5.108689 (0.060649) [4.989817, 5.22756]	5.024318 (0.060209) [4.90631, 5.142327]	5.128281 (0.098501) [4.935219, 5.321343]	5.148906 (0.100279) [4.952359, 5.345454]
Winsor	5.011795 (0.149062) [4.719634, 5.303955]	5.013168 (0.225938) [4.57033, 5.456007]	5.010947 (0.229021) [4.562067, 5.459828]	5.010826 (0.260473) [4.500299, 5.521353]	5.04989 (0.269552) [4.521569, 5.578212]
Suggest (Hybrid)	5.083278 (0.088378) [4.910057, 5.256499]	5.11909 (0.140052) [4.844588, 5.393591]	5.005005 (0.130048) [4.750111, 5.2599]	5.09332 (0.172802) [4.754629, 5.432012]	5.141562 (0.154977) [4.837807, 5.445318]

Note table (4) the values of the standard error of threshold point estimator the suggested method indicates the best result with all contamination levels. Furthermore,

the remaining methods were as follows Krenel, Winsor, Cauchy, Talworth, and MLE followed respectively for (0%, 5%, and 10%) contamination levels.

IV. Real Data

Real example was applied to bed-load data [XIII], [XVI]. Bed-load transport represents the response (y) variable, while the discharge sediments represent the explanatory variable (x) variable.

Table 5: RMSE and R^2 according to Real Data Application

Method	MLE	Talworth	Cauchy	Kernel	Winsor	Suggest (Hybrid)
RMSE	0.155623	0.007258	0.011441	0.006275	0.004475	0.0000028
R^2	0.637849	0.975055	0.962275	0.992967	0.999077	0.999958

Obviously, from table (5) it is shown that the suggested method has the smallest RMSE and the largest R^2values compared with other estimation methods. Moreover, Winsor method indicates the second lower value of RMSE compared with Kernel, and larger R^2 value as well. But, Talworth weighted indicates a lower result than Cauchy weighted methodaccording to RMSE result in addition to indicating a larger R^2 result than Cauchy.

Table 6: Threshold point estimation with its uncertainty

according to Real Data Application (c=5)

Characteristics	Method					
	MLE	Talworth	Cauchy	Kernel	Winsor	Suggest (Hybrid)
Threshold point $\hat{c}$	4.8341322	4.5942150	4.775470	4.6965780	4.5942150	4.8126050
Standard Error for $\hat{c}$	(0.4977456)	(0.0341250)	(0.0592800)	(0.0199870)	(0.0197960)	(0.0191090)
95% Confidence intervals of $\hat{c}$	[3.858551, 5.809713]	[4.773805, 4.851406]	[4.659282, 4.891659]	[4.629693, 4.763462]	[4.555042, 4.633389]	[4.556763, 4.631668]

According to the results of table (6) the values of the standard error of threshold point estimator the suggested method indicates the best result. Also, the remaining methods were as follows Winsor, Kernel, Talworth, Cauchy, and MLE came respectively.

V. Conclusions

From the above results listed through tables (1-6), conclusions can be indicated briefly as below.

V.i. Simulation

i.The main advantage of the suggested (Hybrid) method was its interactive achievements. It can be noticed that RMSE rhythm is stable in all simulation results for the priority according to all contamination levels (0%, 5%, and 10%) and both

sample sizes (n=40, 100) to yield always the lowest values compared with other methods from the tables (1) and (3).

ii. The other results of the suggested (Hybrid) coincided with its superiority in tables (2) and (4) which marked the minimum standard error of threshold point estimation.

iii. According to tables (1) and (3), Talworthweighted method was very close to MLE unweightedmethodfor RMSE values. While in tables (2) and (4), Talworth weighted method was close to Cauchy weighted method concerning standard error values.

V.ii. Real Data

i.The superiority of the suggested (Hybrid) method still steady according to the real data application according to minimum RMSE associated with maximum R^2 values from table (5). Also, (Hybrid) method yields the best results of closeness in estimating to the real data behavior around the threshold point represented by its standard error value from table (6).

ii. Winsor unweightedmethod showed the better results compared with Kernel weighted method in tables (5) and (6) which is the opposite result in simulation.

iii. Talworth weight showed the better performance than Cauchy in tables (5) and (6) which is the contrary in simulation.

References

I. Acitas, S. and Senoglu, B., (2020). "Robust change point estimation in two-phase linear regression models: An application to metabolic pathway data". Journal of Computational and Applied Mathematics, Vol. 363, pp 337–349.

II. Chen, C.W.S., Chan, J. S.K., Gerlach, R., and Hsieh, W. Y.L., (2011). "A comparison of estimators for regression models with change points". Stat Comput, Vol. 21, pp 395–414.

III. Dehnel, G., (2016). "M-Estimators in Business Statistics".Statistics in Transition new series, Vol. 17, No. 4, pp 1–14.

IV. Fearnhead, P. and Rigaill, G., (2017). "Changepoint Detection in the Presence of Outliers".Journal of the American Statistical Association, Vol. 114, No. 525, pp 169-183.

V. Ganocy, S. J. and Sun, J., (2015). "Heteroscedastic Change Point Analysis and Application to Footprint Data".Journal of Data Science, Vol. 13, pp 157-186.

VI. Hernandez, E.L., (2010). " Parameter Estimation in Linear-LinearSegmentedRegression. M.Sc. thesis, Department of Statistics, Brigham Young University,

Omar Abdulmohsin Ali

VII. Julious, S.A., (2001). "Inference and Estimation in a Change point Regression Problem". The Statistician, Vol. 50, Part 1, pp 51-61.

VIII. Klotsche, J. and Gloster, A. T., (2012). "Estimating a Meaningful Point of Change:A Comparison of Exploratory Techniques Based on Nonparametric Regression". Journal of Educational and Behavioral Statistics Vol. 37, pp 579-600.

IX. Liu, Z., (2011). "Empirical Likelihood Method for Segmented Linear Regression".Ph.D. Dissertation, Faculty of the Charles E. Schmidt, College of Science, Florida Atlantic University, USA.

X. Muggeo, V. M. R., (2003). "Estimating regression models with unknown break-points", Statist.Med., Vol. 22, pp 3055–3071.

XI. Muggeo, V. M. R., (2017)."Interval estimation for the breakpoint in segmented regression: a smoothed score-based approach". Aust. N. Z. J. Stat. Vol. 59, No.3, pp 311–322.

XII. Pusparum, M., (2017). "Winsor Approach in Regression Analysis with Outlier".Applied Mathematical Sciences, Vol. 11, No. 41, pp 2031-2046.

XIII. Ryan, S.E. and Porth, L. S., (2007). "A Tutorial on the Piecewise Regression Approach Applied to Bedload Transport Data". General Technical Report RMRS-GTR-189. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 41 p.

XIV. Whitehead, N., Hill, H.A., Brogan, D.J. and Blackmore-Prince, C., (2002). Exploration of threshold analysis in the relation between stressful life events and preterm delivery". American Journal of Epidemiology Vol. 155, pp 117–124.

XV. Yale, C. and Forsythe, A.B., (1976). "Winsorized Regression", Technometrics, Vol.18 No.3, pp 291-300.

XVI. Zhang, F., Li, Q.,(2017)."Robust bent line regression". J. Statist. Plann. Inference, Vol.185,pp41-55.