



MODELLING NONLINEAR RELATION BY USING RUNNING INTERVAL SMOOTHER, CONSTRAINED B-SPLINE SMOOTHING AND DIFFERENT QUANTILE ESTIMATORS

*Burak DİLBER, Department of Statistics, Dokuz Eylül University, Turkey, burakdilber91@gmail.com

( <https://orcid.org/0000-0001-5055-8879>)

A. Firat ÖZDEMİR, Department of Statistics, Dokuz Eylül University, Turkey, firat.ozdemir@deu.edu.tr

( <https://orcid.org/0000-0003-4976-7168>)

Received: 21.07.2020, Accepted: 18.12.2020

*Corresponding author

Research Article

DOI: 10.22531/muglajsci.772523

Abstract

This paper compares the small-sample properties of two non-parametric regression methods, running interval smoother and constrained b-spline smoothing. The running interval smoother method deals with estimation of a conditional quantile (or a measure of location) using different estimators and here our focus is on Harrell-Davis and newly proposed NO quantile estimators. The constrained b-spline smoothing method uses the quantile regression estimator while obtaining conditional quantile estimates. Constrained b-spline smoothing and running interval smoother methods are compared with a simulation study by using theoretical distributions. Furthermore, the methods are examined graphically to understand how they can model the relationship between variables. Constrained b-spline smoothing and running interval smoother with NO estimator outperformed running interval smoother with Harrell-Davis estimator in terms of mean squared error.

Keywords: Non-parametric regression, Quantile estimators, RIS, COBS

HAREKETLİ ARALIK DÜZLEŞTİRİCİSİ, KISITLI B-SPLINE DÜZLEŞTİRME VE FARKLI KANTİL KESTİRİCİLERİ KULLANILARAK DOĞRUSAL OLMAYAN İLİŞKİNİN MODELLENMESİ

Özet

Bu makalede iki parametrik olmayan regresyon yöntemi, hareketli aralık düzleştiricisi ve kısıtlı b-spline düzleştirme yöntemlerinin küçük örneklem özelliklerinin karşılaştırması yapılmaktadır. Hareketli aralık düzleştiricisi yöntemi farklı kestiriciler kullanarak koşullu kantil (veya konum ölçüsü) değerinin tahmini ile uğraşır ve burada Harrell-Davis ile yeni önerilen NO kantil kestiricisine odaklanılmıştır. Kısıtlı b-spline düzleştirme yöntemi, koşullu kantil tahminleri elde ederken kantil regresyon tahmincisini kullanır. Kısıtlı b-spline düzleştirme ve hareketli aralık düzleştiricisi yöntemleri teorik dağılımlar kullanılarak elde edilen bir simülasyon çalışması ile karşılaştırılmıştır. Ayrıca, bu yöntemler, değişkenler arasındaki ilişkinin nasıl modellendiğini anlamak için grafiksel olarak incelenmiştir. Kısıtlı b-spline düzleştirme ve NO kestiricisi ile kullanılan hareketli aralık düzleştiricisi yöntemleri hata kareler ortalaması açısından Harrell-Davis kestiricisi ile kullanılan hareketli aralık düzleştiricisi yönteminden daha iyi performans göstermektedir.

Anahtar Kelimeler: Parametrik olmayan regresyon, Kantil kestiricileri, RIS, COBS

Cite

Dilber, B. and Özdemir, A. F., (2020). "Modelling Nonlinear Relation by Using Running Interval Smoother, Constrained B-Spline Smoothing and Different Quantile Estimators", *Mugla Journal of Science and Technology*, 6(2), 121-127.

1. Introduction

Whereas the assumption of linearity in regression analysis seems providing a good approximation of the population regression model in numerous situations, this is not always true in many cases. One way of arranging any possible curvature is to use some parametric model with quadratic or cubic terms but this might be unsatisfactory in terms of model adequacy criterions. Non-parametric regression is used to explain the nonlinear relationship among variables. In the

nonparametric regression analysis, the shape of the function is not pre-defined and there are no significant assumptions as in parametric regression analysis. In this way, flexibility is provided to reveal the relationship between variables. Despite to all these advantages, there are some difficulties in the application of nonparametric regression. First, a certain number is not given, but a large data set is needed and it requires intensive computer use.

The estimators used in the nonparametric regression are called smoothers. The basic idea of smoothing is to use

locally weighted mean. The estimated value of the dependent variable at a certain point of interest of x is determined by taking a weighted mean of points in the neighborhood of x . Two of these methods are running interval smoother (RIS) and constrained b-spline smoothing (COBS). The purpose of using these methods is to monitor the latent associations that might appear when analyzed in a way that is based on different quantile values. COBS and RIS methods enable this in a flexible and efficient manner.

The RIS method deals with some robust measure of location connected with the random variable y , given x . Many nonparametric regression estimators of this measure of location have been proposed. This method can also obtain the predicted value of the dependent variable using quantile estimators. This conditional quantile is estimated using different estimators and here the focus is on Harrell-Davis (HD) and NO quantile estimators [1,2].

The other non-parametric regression method is constrained b-spline smoothing (COBS). COBS is a very attractive method with some unique advantages. It facilitates robust function estimation via conditional median estimation of the dependent variable. It also provides computation of other conditional quantile functions which have gradually become an integral part of data analysis. The COBS method uses the quantile regression estimator proposed by Koenker and Bassett (1978) while obtaining conditional quantile estimates [3,4].

2. Description of the Methods

2.1. The Running Interval Smoother (RIS)

The values of independent variable (X) that are close to the point of interest of x is determined and denoted by X_i . Then, a conditional quantile of the Y_i values corresponding to these X_i is computed. For the running interval smoother, this conditional quantile might be estimated using different estimators.

MAD, median absolute deviation, is a robust measure of scale. It is the median of $|X_1 - M|, \dots, |X_n - M|$, where M is the usual sample median based on random sample X_1, \dots, X_n . When the parent distribution is normal, MAD estimates $Z_{0.75}\sigma$ where $Z_{0.75}=0.6745$ is the 0.75 quantile of the standard normal distribution. So it is rescaled and normalized version of MAD (MADN) estimates σ when sampling from a normal distribution. $MADN = MAD_x / 0.6745$.

The points X_i are close to x if

$$|x - X_i| \leq f * MADN \tag{1}$$

where $MADN$ is computed using X_1, \dots, X_n and f is a number between 0 and 1 and plays the role of a span. Let

$$N(x) = \{i = |x - X_i| \leq f * MADN\}. \tag{2}$$

That is, $N(x)$ indexes the set of all X_i values that are close to x . Let $\hat{\theta}_i$ be an estimate of some parameter of interest,

based on the Y_i values such that $i \in N(x)$. That is, use all of the Y_i values for which X_i is close to x . To get a graph of the regression line, calculate $\hat{\theta}_i$, the estimate of Y given that $x = X_i, i = 1, \dots, n$, and then plot the points $(X_1, \hat{\theta}_1), \dots, (X_n, \hat{\theta}_n)$ [5].

The span f controls the roughness of the line. As the f value increases a smooth will be a straight, horizontal line. However, if f is too close to zero, the result is a very ragged line. Often the choice $f = 1$ and $f = 0.8$ and give good results, but both larger and smaller values might be of interest particularly when n is small. A good method for an optimum f is to try out some values in an interactive graphics environment, the general strategy is to find the smallest f so that the plot of the points is reasonably smooth.

2.1.1. HD Quantile Estimator

A concern when estimating the q^{th} quantile with a single order statistics $\hat{x}_q = X_{(m)}, m = [qn + 0.5]$, is that its standard error can be relatively high. The problem is of particular concern when sampling from a light-tailed or normal distribution. A natural strategy for addressing this problem is to use all of the order statistics to estimate x_q , as opposed to a single order statistic, and several methods have been proposed. One such estimator was derived by Harrell and Davis (1982) [1]. To compute it, let Y be a random variable having a beta distribution with parameters $a = (n + 1)q$ and $b = (n + 1)(1 - q)$. That is, the probability density function of Y is

$$\frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} y^{a-1}(1 - y)^{b-1} \tag{3}$$

(Γ is the gamma function). Let

$$W_i = P\left(\frac{i-1}{n} \leq Y \leq \frac{i}{n}\right). \tag{4}$$

Then the Harrell-Davis estimate of the q^{th} quantile is

$$\hat{\theta}_q = \sum_{i=1}^n W_i X_{(i)}. \tag{5}$$

2.1.2. NO Quantile Estimator

With the aim of improving the performance in the lower and upper quantiles especially with small sample sizes, a new quantile estimator which is again a weighted average of all order statistics $(X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)})$ is introduced [2].

$$\begin{aligned}
 NO_q = & [B(0; n, q)2q + B(1; n, q)q]X_{(1)} \\
 & + B(0; n, q)(2 - 3q)X_{(2)} \\
 & - B(0; n, q)(1 - q)X_{(3)} \\
 & + \sum_{i=1}^{n-2} [B(i; n, q)(1 - q) \\
 & + B(i + 1; n, q)q]X_{(i+1)} \\
 & - B(n; n, q)qX_{(n-2)} \\
 & + B(n; n, q)(3q - 1)X_{(n-1)} \\
 & + [B(n - 1; n, q)(1 - q) \\
 & + B(n; n, q)(2 - 2q)]X_n
 \end{aligned} \tag{6}$$

where $B(i; n, q)$, $i = 0, 1, 2, \dots, n$ are the binomial probabilities with probability of success q and n is the sample size. This is a new quantile estimator and the performances of RIS with NO is totally unknown.

2.2. Constrained B-Spline Smoothing (COBS)

Regression splines are special functions defined by piecewise polynomials. The region defining the pieces is separated by a sequence of knots or breakpoints [6]. The aim is to force the piecewise polynomials to merge smoothly with the knots.

Splines can be named according to degrees. The simplest spline has degree 0 and it is also called a step function. The next simplest spline has degree 1 and it is also called a linear spline. The next spline is the quadratic spline of degree 2.

Constrained B-spline smoothing (COBS) method provides a way of dealing with quantiles [4,7]. This method is based on the estimated function of the constraints. In particular, COBS can include restrictions such as monotonicity, convexity, concavity and periodicity constraints. The COBS method uses the quantile regression estimator proposed by Koenker and Bassett (1978) while obtaining conditional quantile estimates [4].

Let $\rho_\tau(u) = u(\tau - I(u < 0))$, where the indicator function $I(u < 0) = 1$ if $u < 0$ otherwise $I(u < 0) = 0$. The goal is to estimate the τ^{th} quantile of y given x by finding a function $g(x)$ that minimizes

$$\sum \rho_\tau(y_i - g(x_i)) + \lambda \int |g''(x)| dx \tag{7}$$

based on the random sample $(x_1, y_1), \dots, (x_n, y_n)$ where λ is a scalar that controls smoothness and $\int |g''(x)| dx$ denotes the roughness penalty as L_1 norm of the second derivative of function. λ can be taken from zero to infinity.

3. Design of Simulation Study

RIS and COBS were compared via simulations based on $K = 10000$ replications. Sample sizes were $n=25$ and $n=50$. The model used for generating data was

$$Y = X + \lambda(X)\epsilon \tag{8}$$

where X is distributed $N(0,1)$. The distribution for the error term was taken to be one of four g-and-h distributions [8]. g-and-h distributions are well-known

distributions especially in robust statistical researches. The reason for using the g-and-h distribution is that it provides a simple method for generating observations from a wide variety of distributions, which include extreme departures from normality as measured by skewness and kurtosis. Let Z be a random variable which is generated from standard normal distribution. When $g \neq 0$, the transformation

$$X = \frac{\exp(gZ) - 1}{g} \exp(hZ^2/2) \tag{9}$$

and when $g = 0$, the transformation

$$X = Z \exp(hZ^2/2) \tag{10}$$

is used to generate data from g-and-h distribution.

The four error term distributions used here were the standard normal ($g = h = 0.0$), a symmetric heavy-tailed distribution ($h = 0.2, g = 0.0$), an asymmetric distribution with relatively light tails ($h = 0.0, g = 0.2$), and an asymmetric distribution with heavy tails ($g = h = 0.2$). Table 1 shows the estimated skewness (κ_1), kurtosis (κ_2) values and p-values of each distribution. D'Agostino test (P-DT) [9] for skewness and Bonett-Seier test of Geary's (P-BST) [10] for kurtosis are used to obtain p values of each distribution.

Table 1. Estimated skewness, kurtosis and p-values of tests of the g-and-h distributions

g	h	κ_1	κ_2	P-DT	P-BST
0	0	0	3	0.1907	0.6338
0	0.2	0	21.46	0.0000	0.0000
0.2	0	0.61	3.68	0.0000	0.0000
0.2	0.2	2.81	155.98	0.0000	0.0000

Three different variance patterns (VP) were used. For VP1, $\lambda(X) = 1$ for VP2, $\lambda(X) = |X| + 1$ and for VP3, $\lambda(X) = 1/(|X| + 1)$.

The criterion used to compare RIS and COBS was mean squared error that was estimated with

$$MSE = \frac{1}{nK} \sum_{k=1}^K \sum_{i=1}^n (\theta_{qik} - \hat{\theta}_{qik})^2 \tag{11}$$

where for the k^{th} replication, θ_{qik} is the true conditional q^{th} quantile of Y given $X = X_i$ and again $\hat{\theta}_{qik}$ is the estimate of θ_{qik} based on either RIS or COBS.

The methods were compared with three quantile values: $q=0.1$, $q=0.5$ and $q=0.9$, with sample sizes of $n=25$ and $n=50$. R programming language of version 3.5.2 was used for this simulation study. Here, COBS is applied via the **qsmcobs** function and RIS is applied via the **rungen** function. These functions are available in the **WRS2** package. For the COBS, $lambda = 0$ and for the RIS, $span = 0.8$. Typically, taking the span to be 0.8 suffices in terms of providing a relatively accurate estimate of conditional quantile, based on mean squared error. Here, $span = 0.8$ is assumed unless stated otherwise. On the other hand, RIS method is used with HD and NO estimators.

3.1. Simulation Results

Table 2. MSE values when ϵ was distributed as standard normal (Z)

		RIS			
		COBS	HD	NO	
n=25	q=0.1	VP1	2.249	3.113	2.601
		VP2	6.948	8.671	7.170
		VP3	0.984	1.684	1.400
	q=0.5	VP1	0.895	0.973	0.970
		VP2	2.862	3.100	3.083
		VP3	0.384	0.438	0.438
	q=0.9	VP1	2.253	3.105	2.598
		VP2	6.923	8.697	7.193
		VP3	0.979	1.669	1.390
n=50	q=0.1	VP1	2.528	3.248	2.923
		VP2	8.570	9.232	8.165
		VP3	1.052	1.723	1.557
	q=0.5	VP1	0.952	1.019	1.018
		VP2	3.238	3.384	3.375
		VP3	0.401	0.452	0.452
	q=0.9	VP1	2.536	3.240	2.919
		VP2	8.557	9.237	8.179
		VP3	1.052	1.726	1.561

Table 3. MSE values when ϵ was distributed as g=0.2 and h=0

		RIS			
		COBS	HD	NO	
n=25	q=0.1	VP1	2.170	2.993	2.603
		VP2	6.667	8.269	7.199
		VP3	0.941	1.623	1.398
	q=0.5	VP1	0.961	1.034	1.031
		VP2	3.081	3.310	3.290
		VP3	0.413	0.465	0.464
	q=0.9	VP1	2.651	3.579	2.844
		VP2	7.995	9.981	7.761
		VP3	1.157	1.873	1.490
n=50	q=0.1	VP1	2.433	3.137	2.884
		VP2	8.120	8.868	8.094
		VP3	1.008	1.682	1.546
	q=0.5	VP1	1.020	1.084	1.083
		VP2	3.474	3.620	3.610
		VP3	0.428	0.477	0.477
	q=0.9	VP1	2.904	3.628	3.183
		VP2	10.000	10.516	8.981
		VP3	1.198	1.883	1.667

Table 2 reports simulation results in which X and ϵ has standard normal distribution and Table 3 reports simulation results in which X has standard normal distribution and ϵ has asymmetric light tailed distribution. When q=0.5, COBS, RIS with NO and RIS with HD give close results. The RIS with NO estimator gives better results than COBS for VP2 and extreme quantiles as the sample size increases.

Table 4. MSE values when ϵ was distributed as g=0 and h=0.2

		RIS			
		COBS	HD	NO	
n=25	q=0.1	VP1	4.891	6.050	4.515
		VP2	15.091	17.833	13.309
		VP3	2.130	2.958	2.218
	q=0.5	VP1	1.948	2.036	2.029
		VP2	6.373	6.805	6.755
		VP3	0.828	0.879	0.878
	q=0.9	VP1	4.849	6.023	4.509
		VP2	14.978	17.716	13.233
		VP3	2.128	2.980	2.232
n=50	q=0.1	VP1	4.973	5.850	5.010
		VP2	17.614	17.894	14.914
		VP3	1.974	2.801	2.440
	q=0.5	VP1	2.070	2.141	2.139
		VP2	6.995	7.235	7.212
		VP3	0.868	0.918	0.917
	q=0.9	VP1	4.924	5.849	5.017
		VP2	17.523	17.730	14.806
		VP3	1.965	2.784	2.429

Table 4 reports simulation results in which X has standard normal distribution and ϵ has symmetric heavy tailed distribution. When q=0.5, COBS, RIS with NO and RIS with HD give close results. For extreme quantiles where VP2, RIS with NO estimator competes well with COBS as the sample size increases.

Table 5. MSE values when ϵ was distributed as g=0.2 and h=0.2

		RIS			
		COBS	HD	NO	
n=25	q=0.1	VP1	4.512	5.632	4.620
		VP2	14.274	16.867	14.037
		VP3	1.980	2.811	2.288
	q=0.5	VP1	2.176	2.267	2.258
		VP2	6.983	7.426	7.362
		VP3	0.920	0.971	0.970
	q=0.9	VP1	6.134	7.355	5.001
		VP2	19.738	22.292	14.768
		VP3	2.651	3.528	2.420
n=50	q=0.1	VP1	4.765	5.693	5.104
		VP2	16.667	17.443	15.478
		VP3	1.932	2.744	2.477
	q=0.5	VP1	2.340	2.407	2.404
		VP2	7.870	8.108	8.079
		VP3	0.988	1.035	1.035
	q=0.9	VP1	5.909	6.823	5.558
		VP2	22.187	21.263	16.620
		VP3	2.305	3.189	2.678

Table 5 reports simulation results in which X has standard normal distribution and ϵ has asymmetric heavy tailed distribution. When n=25 and extreme

quantiles, the performance of RIS with NO is improved in heavy-tailed distributions.

In terms of the distribution of the data generated in the simulation study, as the g and h values increase, the MSE values increase, especially in the extreme quantiles. On the other hand, one more simulation was performed when $n=500$ with 1000 replications and 36 experimental settings. In this simulation, RIS with NO gave the smallest MSE values 28 times out of 36 settings. But it can not be said that the MSE values were less than the ones computed when $n=25$ or 50. Each distributional setting lasted 10 hours on the average here.

4. Real Data Example

4.1. Example 1

Sochett et al. (1987) report data related to patterns of residual insulin secretion in children at the time they were diagnosed with diabetes [11]. A portion of the study was concerned with whether age can be used to predict the logarithm of C-peptide concentrations at diagnosis. This data set has 43 observations.

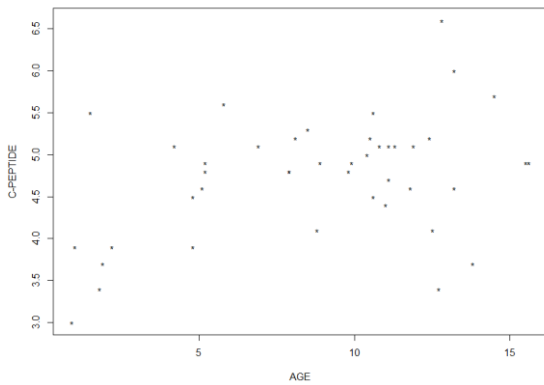


Figure 1. Association between age and c-peptide

Figure 1 shows the relationship between age and logarithm of C-peptide. There is an interesting association among these two variables. A weak but positive slope up to a certain age point 10 then the rest might be interpreted in different ways depending on the interest of the researcher. Using different quantile points simplifies this matter significantly.

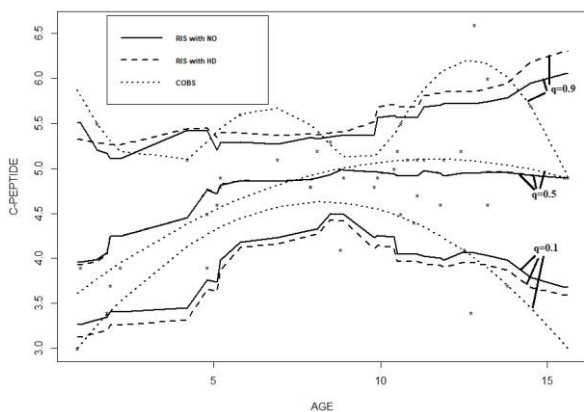


Figure 2. Comparisons of RIS with NO, RIS with HD and COBS (quantile values are 0.1, 0.5 and 0.9)

The data in Figure 2 is used to illustrate method RIS with NO, RIS with HD and COBS. In the COBS method, smoothing parameter (λ) equals 0. In the RIS with NO and RIS with HD, span (f) equals 0.8. For quantile regression line when $q=0.1$, notice that for age under 10 the plot suggests a positive association but over 10 methods suggest a negative association with C-peptide. For quantile regression line when $q=0.5$, notice that for age under 10 the plot suggests a positive association but over 10 methods suggest a little or no association with C-peptide. For quantile regression line when $q = 0.9$, notice that for age under 10 the plot suggests that there is little or no association but over 10 methods suggest a positive association with C-peptide. For COBS at $q = 0.9$, COBS indicated some unusually shaped regression lines. Here, there is concern about the shape of the regression line.

Table 6. MSE values of methods for example 1

	q=0.1	q=0.5	q=0.9
RIS with NO	1.063089	0.4055781	1.080137
RIS with HD	1.229652	0.4053521	1.229841
COBS	0.9410627	0.4118963	1.311917

MSE values of this data set for different quantiles are shown in Table 6. For $q = 0.9$, RIS with NO gives better results than the other methods. It is seen that the median gives close results. For $q = 0.1$, COBS gives more reliable results.

4.2. Example 2

Between March 12, 2020 and April 23, 2020 estimate of the number of daily cases occurred in Turkey COVID-19 can be modeled using these methods. The independent variable is days and the dependent variable is daily deaths. The data were obtained from the website of the World Health Organization [12].

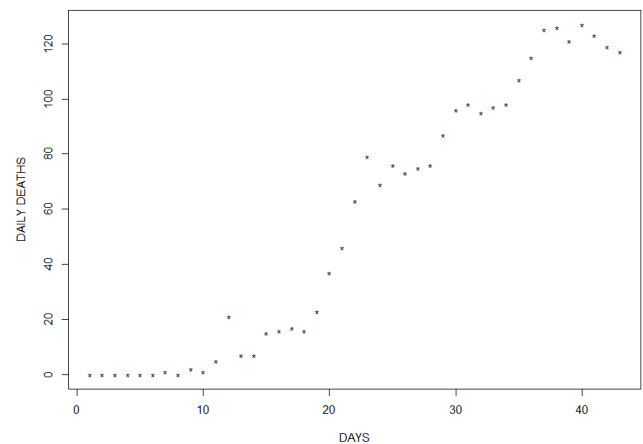


Figure 3. Association between days and daily deaths

Figure 3 shows the relationship between days and daily deaths. The relationship between variables is modeled using COBS, RIS with NO, RIS with HD and third order polynomial regression model (PR).

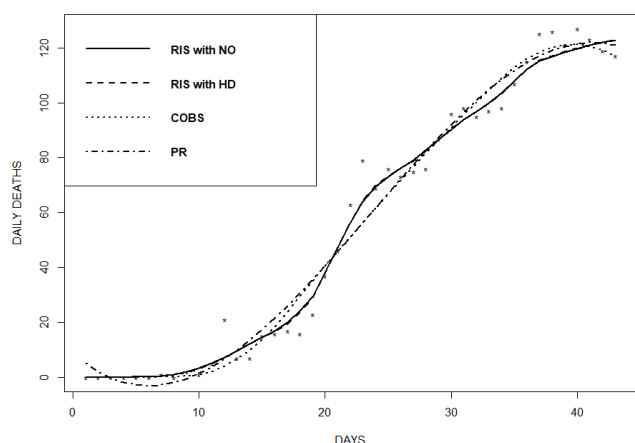


Figure 4. Comparisons of RIS with NO, RIS with HD, COBS and PR (quantile value is 0.5)

The data in Figure 4 is used to illustrate method RIS with NO, RIS with HD, COBS and PR. In the COBS method, smoothing parameter (λ) equals 0. In the RIS with NO and RIS with HD, span (f) equals 0.4. For quantile regression line when $q=0.5$, notice that the plot suggests a positive association with daily deaths.

Table 7. MSE values of methods for example 2

	RIS with NO	RIS with HD	COBS	PR
q=0.5	25.3764	23.8560	44.2608	48.5388

MSE values of this data set for median are shown in Table 7. It is seen that the RIS with HD and RIS with NO give more reliable results.

Table 8. Predicted value of daily deaths

DAYS	REAL VALUE	PREDICTED VALUE			
		COBS	RIS with NO	RIS with HD	PR
15/04/2020	107	113	108	108	112
16/04/2020	115	116	112	112	115
17/04/2020	125	118	115	115	117
18/04/2020	126	120	117	117	119
19/04/2020	121	121	118	119	121
20/04/2020	127	121	120	120	121
21/04/2020	123	121	121	121	122
22/04/2020	119	119	122	122	122
23/04/2020	117	117	123	123	121
24/04/2020	115	114	122	122	119
25/04/2020	109	110	121	121	117
26/04/2020	106	106	121	121	114
27/04/2020	99	101	119	120	111
28/04/2020	95	95	118	118	106
29/04/2020	92	88	117	117	101

Table 8 shows the estimated and real values of daily deaths between 15 April 2020 and 29 April 2020. It is

seen that COBS method are more successful especially when compared with the other methods.

5. Conclusion

In this paper, two nonparametric regression methods that can be used to model nonlinear relationship were compared. After introducing the fundamental structure of the conventional nonparametric regression methods, a simulation study was carried out by using g-h distributions with different skewness and kurtosis levels. The performance of the methods were also monitored with graphs using real data sets.

In the simulation study, the methods were compared in terms of mean squared error (MSE). In general, the COBS and the running interval smoother with NO quantile estimator gave better results than the running interval smoother with HD quantile estimator. That means they are more efficient than the other method. COBS and RIS with NO methods gave similar results for all g-h distributions but increasing the kurtosis and skewness levels (non-normality) resulted in better efficiency values for RIS with NO quantile estimator.

Under the normality, COBS and RIS with NO methods gave similar results so both can be preferred. When the distribution is asymmetric light tailed, in 15 of the 18 different cases examined, the COBS method had a lower MSE value, while in 3 of them the RIS with NO method had a lower value. However, both methods can be preferred because the MSE values are quite similar to each other. RIS with HD is not preferred. When the distribution is symmetric heavy tailed, in 12 of the 18 different situation examined, COBS method can be preferred, in other situations RIS with NO quantile estimator can be preferred. When the distribution is symmetric heavy tailed, in 12 of the 18 different situation examined, COBS method can be preferred, in other situations RIS with NO quantile estimator can be preferred. When the distribution is asymmetric heavy tailed, since COBS and RIS give similar results, both methods can be preferred. However, the RIS with NO method is recommended for VP2.

Performances of methods were also compared by sketching graphs. The type and degree of relationship among variables might change depending on the quantile points of interest. Using more than one quantile points let the researchers observe different patterns and make correct predictions accordingly. Preferring RIS or COBS enable this when analyzing nonlinear trends.

Moreover, predicted values of COVID-19 daily deaths for Turkey are obtained and it has been seen that COBS method makes very successful predictions.

Constrained B-Spline Smoothing (COBS) and Running Interval Smoother (RIS) with NO quantile estimator are suggested as efficient, flexible and robust tools of modelling nonlinear relationship among variables.

6. References

- [1] Harrell, F.E., Davis C.E., "A new distribution-free quantile estimator", *Biometrika*, 69, 635-640, 1982.
- [2] Navruz G., Özdemir A.F., "A new quantile estimator with weights based on a subsampling approach", *British J. of Mathematical and Statistical Psychology* (Early view) 2020.
- [3] Koenker R., Bassett G., "Regression quantiles", *Econometrica*, 46, 33-50, 1978.
- [4] He X., Ng P., "COBS: Qualitatively constrained smoothing via linear programming", *Computational Statistics*, 14, 315-337, 1999.
- [5] Wilcox R., *Introduction to Robust Estimation and Hypothesis Testing, 4th ed.* Academic Press Amsterdam, the Netherlands, 2017.
- [6] Hastie T., Tibshirani R., *Generalized Additive Models, 1st ed.* Chapman and Hall/CRC Press, London, 1990.
- [7] Koenker R., Ng P., "Inequality constrained quantile regression", *The Indian Journal of Statistics*, 67, 418-440, 2005.
- [8] Hoaglin D.C., "Summarizing shape numerically: The g-and-h distribution. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes*. New York, NY: Wiley-Interscience, 1985.
- [9] D'Agostino R.B., "Transformation to normality of the null distribution of G_1 ", *Biometrika*, 57, 3, 679-681, 1970.
- [10] Bonett D.G., Seier E., "A test of normality with high uniform power", *Computational Statistics and Data Analysis*, 40, 435-445, 2002.
- [11] Sochett E.B., Daneman D., Clarson C., Ehrich R.M., "Factors affecting and patterns of residual insulin secretion during the first year of type I (insulin dependent) diabetes mellitus in children", *Diabetes*, 30, 453-459, 1987.
- [12] World Health Organization, 2020 [Online]. Available: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.