ESKİŞEHİR TECHNICAL UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
A- APPLIED SCIENCES AND ENGINEERING

2020, Vol: 21, pp. 7-16, DOI: 10.18038/estubtda.818004

NONPARAMETRIC REGRESSION ANALYSIS BASED ON RATIONAL (Padé) APPROXIMATION FOR CENSORED-DATA

Dursun AYDIN [1], , Ersin YILMAZ [2, *]

[1] Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Mugla, Turkey
[2] Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Mugla, Turkey

ABSTRACT

This paper considers the estimation of a nonparametric regression model with randomly right-censored data. To estimate the model, rational (Padé) approximation based on truncated total least squares (P-TTLS) is used as a smoothing method. Because of censored, data points cannot be used directly in modeling process, a data transformation is needed for overcoming this problem. As known, synthetic data transformation assigns censored points as zero and gives additional magnitudes to uncensored ones associated with Kaplan-Meier distribution of the censored dataset. Thus, the differences between censored and uncensored observations grow which causes a kind of spatial variation in the shape of data. In this paper, to bring a solution to this problematic situation, P-TTLS is used that works well on spatial variation. Also, to see the performance of the P-TTLS on censored data modeling, a simulation study is carried out and it is compared with the benchmarked kernel smoothing (B-KS) method to observe how P-TTLS behaves.

Keywords: Pade approximation, Right-censored data, Kernel smoothing, Truncated total least squares

1. INTRODUCTION

Right-censored data is a common kind of data irregularity that researchers across in many application fields such as bioinformatics, industrial researches, biology and particularly in clinical trials and medical studies. Modeling right-censored data has critical importance when it comes from clinical trials, surgery, cancer researches or organ transplantation. Due to sensitivity of analysis, an estimation technique is needed that assures the correct

In this paper, censored survival times are considered in medical studies and our aim is to estimate right-censored data via nonparametric regression model. As known, censored data cannot be analyzed directly because of effects of censorship. To overcome this issue, there are three major techniques that attract attention in the literature. The first is the proportional hazards model proposed by Cox [1] that earned very much popularity in censored data cases because of its easy usability and flexible nature. There are some inferential studies about Proportional hazards model such as Cox [2], Tsiatis [3], Andersen and Gill [4] and so on. The second is the Kaplan-Meier weights (KMW) based regression model which is proposed by [5]. Then, [6-8] give some improvement and asymptotic of this model in the parametric context. Also, [9] and [10] used KMW for estimating the censored regression models. The third one is Synthetic data transformation based on a distribution estimation of the censoring variable which is the topic of this paper. There is a number of data transformation techniques proposed by authors such as [11-13] and their derivatives. In this study, synthetic data transformation defined by [12] is used. In this technique, magnitude of uncensored observation is increased according to Kaplan-Meier estimation of the distribution of censoring variable and it given zero to each censored data points. Thus, it provides an equality between expected values of completely observed and incomplete response values (see [12]). Details are indicated in Section 2.

*Corresponding Author: yilmazersin13@hotmail.com
Received: 29.10.2020 Published: 27.11.2020

Nonparametric regression model is widely used in censored data context because censored cases do not ensure the assumptions of the classical parametric methods. Some of important studies in the literature can be ordered as follows. [14] used Kernel and nearest neighbour methods to estimate the right-censored model, [15] studied the estimation of the regression model with nonparametric M-estimator and it support its study with bias correction and bootstrap procedure. [16] estimates the censored survival data by local linear smoothers (LOESS) and they make a detailed applications on that. In addition, [17-18] used nonparametric estimators to model the right-censored data. The distinguishable point of this paper among all of the mentioned studies is a using completely different smoothing method Padé based on truncated total least squares (P-TTLS). There are some important studies that used Pade approximation in numerical modeling under uncensored data such as [19]. The major reason for using P-TTLS is it can achieve the modeling of data that has a high variation structure (see [20]). It is known from the synthetic data transformation technique, there will be a huge difference between uncensored and censored data points which caused a high level of variation. Our expectation is that the P-TTLS method successfully estimates the non-parametric regression model under randomly right-censored data. Then, P-TTLS is compared with benchmark Kernel smoothing method to observe its behaviors and to evaluate its performance.

Rest of the paper is designed as follows. In section 2, randomly right-censored data and corresponding nonparametric regression model is introduced. Then synthetic data transformation technique is expressed. Sections 2.1-2.2 include estimation procedures of P-TTLS and Kernel smoothing methods based on a synthetic response variable. In section 3, some measurement tools are defined to evaluate the performances of mentioned methods. A Monte Carlo simulation study and its comparative outcomes are presented in Section 4. Discussion is given in Section 5.

2. MATERIAL AND METHODS

Suppose that $\{x_i, y_i\}_{i=1}^{n}$ be the completely observed pair of data where x_i's are the values of covariate that and y_i's are the response values. Let y_i's are censored by values of censoring variable c_i's that are independent from y_i's. In this case, incomplete observations can be written as follows

$$z_i = \begin{cases} y_i & if\ y_i \le c_i \\ c_i & if\ y_i > c_i \end{cases}, \quad \delta_i = \begin{cases} 1 & if\ y_i \le c_i \\ 0 & if\ y_i > c_i \end{cases} \tag{1}$$

where z_i's are the new response values that updated accordingly censorship and δ_i's involve information about existence of the censorship. Now, nonparametric regression model under randomly right-censored data can be written as follows

$$z_i = g(x_j) + \varepsilon_j,\ 1 \le j \le n \tag{2}$$

where $g(.)$ is an unknown smooth function and ε_i's denote the random error terms that distribute normally with zero mean and constant variance as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. As mentioned in Section 1, modelling techniques cannot be applied to model (2) directly because of censorship. To overcome this issue, response values z_i's are transformed to synthetic data points that include the effect of censorship via Kaplan-Meier estimator of c_i's distribution function $G(.)$. One of the most important causes for using synthetic data is incomplete response variable z and true response variable y have different expected values. Theoretically, synthetic data transformation provides equal expected values for both variables (see [21] for details). In this case, acquisition of synthetic data is given by

$$z_{jG} = \frac{\delta_i z_j}{1 - G(z_j)} \tag{3}$$

where $G(l_i|x_i) = P(c_i \le l_i|x_i),\ (l_i \in R)$. Here, instead of unknown $G(.)$, its Kaplan-Meier estimator $\hat{G}(.)$ is used. Thus, model (2) is rewritten for the synthetic responses

$$z_{jG} = g(x_j) + \varepsilon_{jG},\ \varepsilon_{jG} = z_{jG} - g(x_j),\ 1 \le j \le n \tag{4}$$

Here, ε_{jG} value(s) has the same assumption that given right after model (2) but for given distribution G. There are two common assumptions of distribution G for the validation of equation (4):

i) $y_i's$ and c_i's have to be independent
ii) $P(y_i \le c_i | y_i, x_i) = P(y_i \le c_i | y_i)$

Note that assumption (i) is a standard assumption for the meaningful right-censored data analysis. Assumption (ii) means that covariate x_i does not give any more information for given failure time whether data point censored or not. Due to given assumptions, it can be assured that $E(z_{jG}|x_j) = E(y_j|x_j) = f(x_j)$.

Since the problem of censorship is handled by (3), model (4) can be estimated. To achieve this objective $P - TTLS$ smoothing method is used which is the main concern of this paper. Also, Kernel smoothing method is introduced in Section 2.2 as a benchmark method.

2.1. P-TTLS Smoothing

As mentioned above, P-TTLS is used for the first time for modeling right-censored data via nonparametric model when response variable is transformed to synthetic data. In order to estimate nonparametric function $g(.)$ In model (4), procedure of P-TTLS is expressed in this section.

Let $g(x_j)$ is written by the form $g(x_j)_{[p,q]} = R(x_j)/Q(x_j)$ as follows

$$g(x_j)_{[p,q]} = \frac{R(x_j)}{S(x_j)} = \frac{r_0 + r_1x_j + r_2x_j^2 + \cdots + r_px_j^p}{s_0 + s_1x_j + s_2x_j^2 + \cdots + s_qx_j^q},\ p \le q, j = 1, \dots, n \tag{5}$$

It can be clearly seen that equation (1) is a rational function to be estimated that has $(p + q + 1)$ terms. In this case, estimation of $(p + q + 1)$ coefficients are needed to obtain estimate continuous function $g(x_j)$. A major issue is find the coefficients of the $R(x_j)$ and $S(x_j)$ for determined degrees of nominator and denominator p and q. Thus, making minimum the absolute difference between $g(x_j)_{[p,q]}$ and $g(x)$ which can be shown as $\{|g(x) - g_{[p,q]}(x)| \le \omega,\ \omega > 0\}$. In this case, equation (5) can be rewritten as

$$z_{jG} = g(x_j) \cong \frac{R(x_j)}{S(x_j)} = \frac{r_0 + r_1x_j + r_2x_j^2 + \cdots + r_px_j^p}{s_0 + s_1x_j + s_2x_j^2 + \cdots + s_qx_j^q} = g(x_j)_{[p,q]}, p \le q, j = 1, \dots, n \tag{6}$$

From (6), model coefficients to be estimated can be expressed as $\{r_a\}_{a=1}^{p}$ and $\{s_b\}_{b=1}^{q}$. Note that $s_0 = 1$ is decided to avoid indefinability of $g(x_j)_{[p,q]}$. To reach our goal, equation (6) is turned into the form that given by

$$\left[r_0 + r_1x_j + r_2x_j^2 + \cdots + r_px_j^p - s_1z_{jG}x_j - s_2z_{iG}x_j^2 - \cdots - s_qz_{jG}x_j^q\right] \cong z_{jG}, 1 \le j \le n \tag{7}$$

In order to simplify the understanding equation (7), its matrix and vector form can be shown as follows

$$\{\mathbf{g}_{[p,q]} = \mathbf{X\alpha}\} = \begin{bmatrix} 1 & x_1 & \dots & x_1^p & -z_{1G}x_1 & -z_{1G}x_1^2 & \dots & -z_{1G}x_1^q \\ 1 & x_2 & \dots & x_2^p & -z_{2G}x_2 & -z_{2G}x_2^2 & \dots & -z_{2G}x_2^q \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & \dots & x_n^p & -z_{nG}x_n & -z_{nG}x_n^2 & \dots & -z_{nG}x_n^q \end{bmatrix}_{(nxm)} \begin{bmatrix} r_0 \\ \vdots \\ r_p \\ s_1 \\ \vdots \\ s_q \end{bmatrix}_{(m\times1)} = \begin{bmatrix} z_{1G} \\ z_{2G} \\ \vdots \\ z_{nG} \end{bmatrix}_{(nx1)} \cong \mathbf{Z_G} \tag{8}$$

where $\mathbf{X}$ is a $(n \times m)$ matrix that includes elements of both polynomials $R(x_j)$ and $Q(x_j)$ and $(m = p + q + 1)$ under condition $n > m$. Vector of coefficients is $\boldsymbol{\alpha} = (r_a, s_b)^T$. It can be clearly realize that columns of $\mathbf{X}$ seem almost linearly dependent which is caused singularity in estimation process. So, one can say that $\mathbf{X}$ has an ill-posed problem that makes impossible to modelling. Also note that, data matrix $\mathbf{X}$ includes the terms with response values. Therefore, error contaminates both side of nonparametric model. In order to solve this problem, truncated total least squares $(TTLS)$, which is a regularization method proposed by Golub and Van Loan (1980), is merged with Padé approximation and thus, model estimation can be made by $P - TTLS$ smoothing method. From information given so far, it is clear that $P - TTLS$ estimate of $\mathbf{g}_{[p,q]}$ is obtained via estimation of the coefficient vector $\boldsymbol{\alpha} = (r_a, s_b)^T$.

The main idea of the $TTLS$ is that it gets rid of the smaller singular values of the augmented matrix $[\mathbf{X}, \mathbf{z}_G]$ (see [22-23] for details). In summary, purpose of the $TTLS$ is to diminish the effect of errors by truncating the singular values in the singular value decomposition (SVD) of the $[\mathbf{X}, \mathbf{z}_G]$. In order to estimate the vector of Padé coefficients $\boldsymbol{\alpha}$ by using $P - TTLS$, an algorithm is given below.

Algorithm for P-TTLS smoothing

Step 1. Calculate the SVD of $[\mathbf{X}, \mathbf{z}_G]$, which can be expressed as follows

$$[\mathbf{X}, \mathbf{y}] = \mathbf{U\Sigma V}' \tag{9}$$

where elements of diagonal matrix $\boldsymbol{\Sigma}$ are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{m+1} \geq 0$

Step 2. Determine a suitable truncation parameter $t \leq \min(m, rank\{[\mathbf{X}, \mathbf{y}]\})$.

Step 3. Block-partition the $(m+1) \times (m+1)$ matrix $\mathbf{V}$ as given below

$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}$, where $\mathbf{V}_{11} \in R^{m \times t}$ and $\left\{\mathbf{V}_{22} \equiv \left[v_{m+1,t+1}, \dots, v_{m+1,m+1}\right] \neq 0\right\} \in R^{1\times(m+1-t)}$

Step 4. Compute the Padé coefficients vector by

$$\widehat{\boldsymbol{\alpha}} = -\mathbf{V}_{12}(\mathbf{V}_{22})^{+} = -\mathbf{V}_{12}\frac{\mathbf{V}_{22}'}{\|\mathbf{V}_{22}\|_2^2} \tag{10}$$

where $(\mathbf{V}_{22})^{+}$denotes the pseudoinverse of $\mathbf{V}_{22}$ and $\widehat{\boldsymbol{\beta}}_{P-TTLS}^{t}$ shows the estimates of Padé coefficients.

Step 5. Estimate the $\mathbf{g}_{[p,q]}$ and fitted values $\hat{\mathbf{z}}_{P-TTLS}$ using with estimated vector of coefficients

$$\hat{\mathbf{z}}_{P-TTLS} = \hat{\mathbf{g}}_{[p,q]} = \mathbf{X}\widehat{\boldsymbol{\alpha}} = \mathbf{X}(\mathbf{P}_t\mathbf{X})^{+}\mathbf{P}_t\mathbf{z}_G = \mathbf{H}_t\mathbf{z}_G \tag{11),}$$

where $\mathbf{H}_t = \mathbf{X}(\mathbf{P}_t\mathbf{X})^{+}\mathbf{P}_t$ is the hat matrix for $P - TTLS$. Because of $\mathbf{H}_t$ calculates the orthogonal projection, it is also denoted as a projection matrix.

2.2. Kernel Smoothing

Kernel smoothing is very common nonparametric method that works with weighted average of the data. As indicated above kernel smoothing method (KS) used as a benchmark method for estimation of right-censored data based on a synthetic data transformation. Note that KS is used before for the same purpose in study [9] which can be inspected for the more details, and it has seen that KS has reasonable results on right-censored modelling. In this study, it used for testing the performance of proposed method $P - TTLS$. Suppose that $\hat{\mathbf{g}}_{KS}$ be a kernel estimate of the i^{th} censored-response value. Thus, a kernel smoother can be estimated as follows

$$\hat{\mathbf{z}}_{KS} = \hat{\mathbf{g}}_{KS} = \mathbf{W}\mathbf{z}_G = \sum_{j=1}^{n} w_{ij} z_{jG} \tag{12}$$

where $\hat{\mathbf{z}}_{KS}$ is a vector of fitted values obtained by kernel smoothing, $\mathbf{W}$ is a positive, semi-definite and symmetric smoother matrix formed by kernel weights w_{ij} given by [24] and [25], M is a number of kernels for each data point z_{iG} which is determined by researcher accordingly shape of data. Elements of smoother matrix $\mathbf{W}$ can be given by

$$w_{ij} = \frac{K_b(\frac{x-x_j}{b})}{\sum_{j=1}^{n} K_b\left(\frac{x-x_j}{b}\right)} = \frac{K_b(u)}{\sum_j K_b(u)} \tag{13}$$

where b represents the bandwidth parameter, $K_b(.)$ is a kernel function which determines the shape of the estimated curve and bandwidth parameter b controls the amount of avereging and finally, $\sum_j w_{ij} = 1$. Note that there are some basic propoerties of kernel function $K_b(.)$that are given by

$$\int_{-\infty}^{+\infty} K_b(u)\, du = 1,\ K_b(u) \geq 0 \text{ for all } u, \text{ and } K_b(u) = K_b(u\ -) \tag{14}$$

If Epanechnikov kernel function is inspected which is used in the simulation study of this paper,

$$K_b(u) = \frac{3}{4}(1-u^2), \qquad if\ \ |u| \leq 1$$

It can be seen that, it ensures the properties of the kernel weight function given in (14). In addtion, the consistency of the kernel function can be provided by consideration given below

i. If $|u| \to \infty$ and $E(z_G^2) < \infty$ then $\int_{-\infty}^{+\infty} |K_b(u)| du < \infty, uK_b(u) \to 0$
ii. Under assumptions of $b \to 0, nb \to \infty$ it can be shown that

$$\frac{1}{n}\left(\sum_{i=1}^{n} w_{ij} z_{iG}\right) = \hat{\mathrm{g}}_{KS}(x_i) \overset{p}{\to} \mathrm{g}(x_i) \tag{15}$$

The expression $\overset{p}{\to}$ represents the “convergence in probability”. From the information given so far, KS estimation of $\mathbf{g}$ based on synthetic responses can be written as in matrix and vector form

$$\hat{\mathbf{z}}_{KS} = \hat{\mathbf{g}}_{KS} = \mathbf{W}\mathbf{z}_G \tag{16}$$

Thus, KS estimation of right-censored nonparametric regression model is obtained by (16).

3. EVALUATION CRITERIA

In this section, two measurement tools are introduced to evaluate the performances of $P-TTLS$ and benchmarked KS methods on modelling right-censored data in a nonparametric context. This paper considers the “Mean Square Errors (MSE)”, which is the most common performance measure in the literature, and the “Relative Efficiency based on MSE (RE)” as evaluation metrics. Calculation of them are given below

$$MSE(\hat{\mathbf{g}}_M) = n^{-1}[\sum_{i=1}^{n}(\mathbf{z}_{iG} - \hat{\mathrm{g}}_M(x_i))^2] = n^{-1}[(\mathbf{z}_G - \mathbf{g}_M)^2] \tag{17}$$

where $\hat{\mathbf{g}}_M$ is a estimated nonparametric function for one of the mentioned methods. It replaces $\hat{\mathbf{g}}_{P-TTLS}$ for $P-TTLS$ estimation and $\hat{\mathbf{g}}_{KS}$ for KS estimation. Using the same notation, RE can be written as follows

$$RE(\hat{\mathbf{g}}_{M1}, \hat{\mathbf{g}}_{M2}) = \frac{MSE(\hat{\mathbf{g}}_{M1})}{MSE(\hat{\mathbf{g}}_{M2})} \tag{18}$$

where "$M1$" and "$M2$" denote the methods that should be compared. RE metric provides a relative comparison between two methods. Here, if $RE(\hat{\mathbf{g}}_{M1}, \hat{\mathbf{g}}_{M2}) < 1$, then it can be said that $\hat{\mathbf{g}}_{M1}$ is better estimator than $\hat{\mathbf{g}}_{M2}$.

4. SIMULATION STUDY

A nonparametric regression model is generated as follows for the simulation experiment:

- Sample size $n = (35, 100,300)$, and censoring level $CL = (0.04, 0.15, 0.40)$
- $\varepsilon_i \sim N(0, \sigma^2 = 1), i = 1, \dots, n$
- $x_i = 4\,(i - 0.5)/n,$
- $g(x_i) = \sin(-4.8x_i)\sin(1.4x_i)$
- $y_i = g(x_i) + \varepsilon_i,$
- $\delta_i \sim Ber(1 - CL)$
- $c_i \sim N(\mu_y, \sigma_y^2)^{[\delta_i = 0]}$ until $c_i > y_i$ otherwise $c_i \sim N(\mu_y, \sigma_y^2)^{[\delta_i = 1]}$ until $c_i \leq y_i$

Then, incomplete observations $z_i's$ are obtained by equation (1) and they are transformed to synthetic data points $z_{iG}'s$ by equation (3). As mentioned above, there are three sample sizes and three censoring levels are determined which means 9 different configurations for both $P - TTLS$ and KS methods. Note that each configuration is repeated 1000 times. Outcomes are given in following tables and figures.

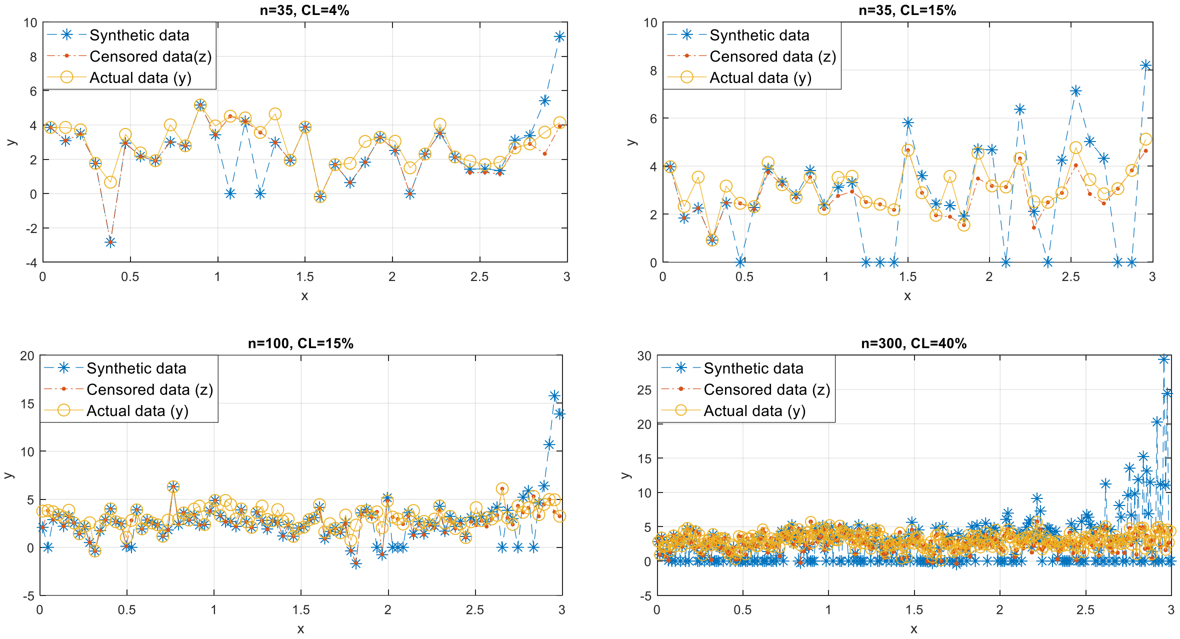

Figure 1. Completely observed (y_i), incompletely observed (z_i) & synthetic data points (z_{iG})

Figure 1 includes three kind of data used in this simulation experiment that are completely observed y_i's, right-censored z_i's and synthetic z_{iG}'s. These datasets are shown for four combinations because it is hard to show all of them. From Figure 1, the effect of synthetic data transformation can be clearly seen for three sample sizes and three censoring levels. One can realize that when CL is getting higher, because keeping expected values of synthetic data and actual data in balance as $E(z_G) \cong E(y)$ the uncensored ones are increased according to estimated Kaplan-Meier distribution of G.

Table 1 involves scores of MSE and RE for all simulation combinations and the best scores are indicated with the bold colors. It can be said that for this simulation design, $P - TTLS$ gives a superior MSE values and it dominates the KS smoothly. The reason for that can be claimed as $P-TTLS$ has a great performance between zeros (censored point in synthetic data context) and increased data points which are mentioned before. Note that, except one, all of RE values are smaller than one which means $P-TTLS$ relative efficiency is better than KS also.

Table 1. Performance scores for all configurations in terms of MSE and RE metrics

		MSE		
n	***CL(%)***	***P − TTLS***	***KS***	$RE(\hat{\mathbf{g}}_{P-TTLS}, \hat{\mathbf{g}}_{KS})$
35	4	**0.464**	0.8001	0.5799
	15	**2.982**	4.6184	0.6457
	40	**8.6743**	8.9112	0.9734
100	4	**0.3523**	0.5965	0.5906
	15	**1.1441**	1.4703	0.7782
	40	**5.5376**	7.6165	0.727
300	4	**0.1204**	0.2428	0.4959
	15	**0.363**	0.414	0.8767
	40	2.4667	**2.2659**	1.0886

Bold colored scores indicate the best performance.

However, $P-TTLS$ has some reasonable results, if outcomes of Table 1 and Figure 2 are inspected in detail, it can be seen that performance of $P-TTLS$ is getting worse for higher CLs which can be interpreted as P-TTLS has weaker than KS across the censorship. Of course, to test this hypothesis, a wider numerical experiments are needed that is a subject of future works associated with this paper.

In Figure 2, the boxplots of MSE values are presented with three panels. Each panel is formed for one sample size. In x-axis of panels, “P1” and “K1” denote boxplots of $P-TTLS$ and KS for CL=4% respectively. In a similar manner “P2” and “K2” represent the boxplots for CL=15%, “P3” and “K3” show boxplots for CL=40%. As can be seen in each panel, under heavy censorship boxplots are getting larger for both methods. Note also that, for each censoring level and sample sizes, $P-TTLS$ gives narrower boxplots than KS and it should be emphasize that $P-TTLS$ has less outlier MSE values. It can be interpreted as $P-TTLS$ provides more stable estimates in censoring case.

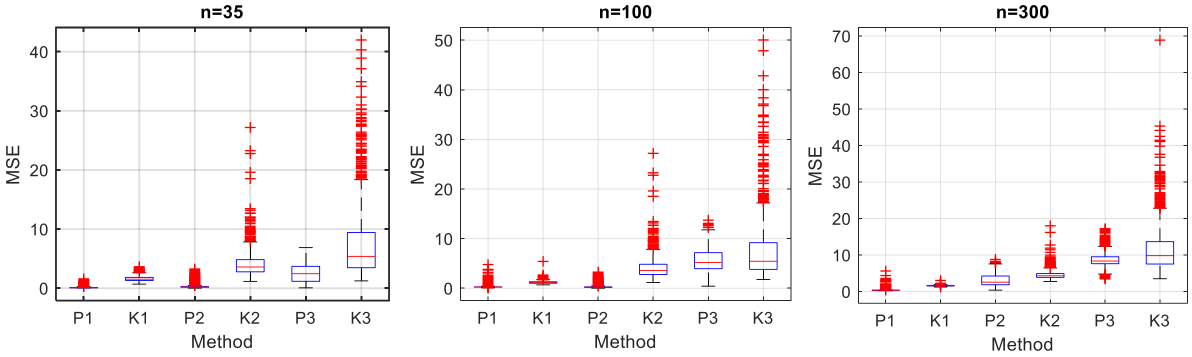

Figure 2. Boxplots for $RE(\hat{\mathbf{g}}_{P-TTLS}, \hat{\mathbf{g}}_{KS})$ values for different combinations

In Figure 3, estimated curves are seen with actual (completely observed) data points. At first, it should be indicated that estimates of P-TTLS and KS are closer to each other that is assured by MSEs in Table 1. Especially in low CL (4%) it is clear that both methods give their best and estimate the data very well. The important point is to see the performance of the methods under medium and heavy CLs. To achieve that three panels of Figure 3 are formed by CL=15% and CL=40%. Also, a plot for $n=300$ is not drawn because data points are too much and it cannot be available for visual analysis. Here, the reason for the wellness of the $P-TTLS$ can be explained as it tries to catch almost every data point. For the same reason, its performance decreases when censoring level increases. Note that the sensitivity of KS to censorship is less than $P - TTLS$. But in a general frame, $P - TTLS$ has better results.

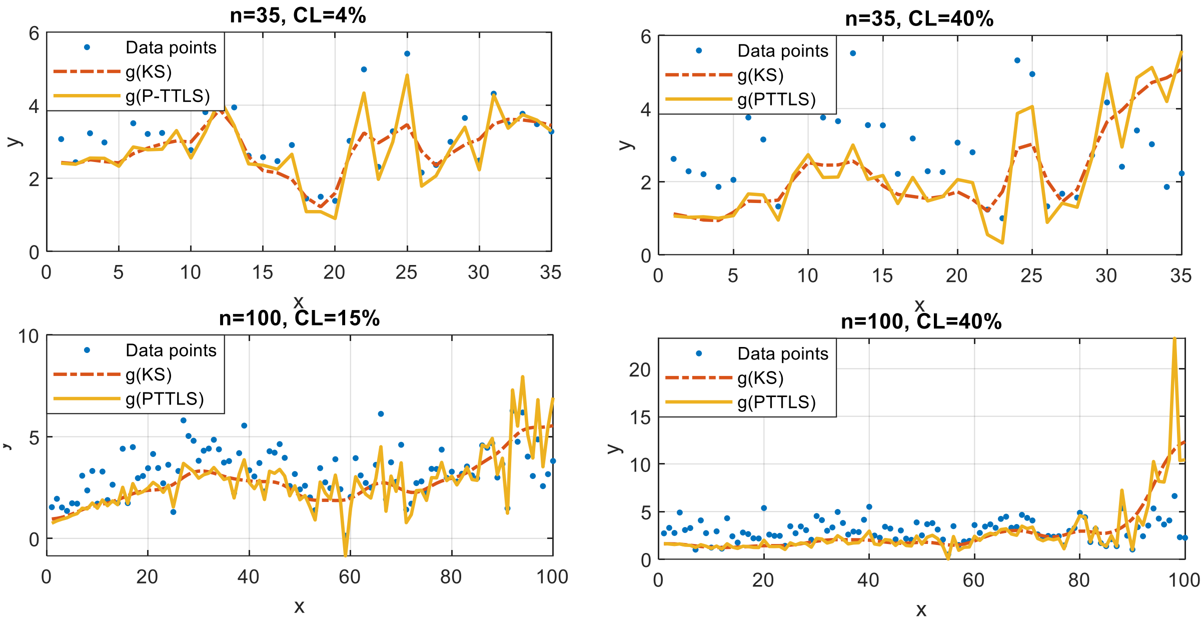

Figure 3. Estimated curves for $(n = 35, CL = 4\% - 40\%)$ and $(n = 100 - 300, CL = 15\%)$.

5. DISCUSSION

In this paper, we introduce a new modified smoothing method $P - TTLS$ using with synthetic data for estimation of the nonparametric regression model. Also, to evaluate its performance, KS method is used as a benchmark method. This main aim of this study is modelling randomly right-censored data optimally and see behaviors of P-TTLS in this context.

To realize the aim of this paper, a simulation study is made and results are presented in Section 4. In order to evaluate the performance of the methods, MSE and relative efficiency (RE) metrics are used. The outcomes of simulation study prove that $P - TTLS$ has ability for modelling randomly- right-censored data and it does better than KS. Also note that Figure 2-3 indicate that although P-TTLS has satisfying results, because of it tries to estimate every data point in the dataset, it is a sensitive method across the censorship. In this case, it can be said that KS is a more enduring method.

As a result, when all figures and tables are inspected, for prepared simulation study, $P - TTLS$ gives the best scores almost for all combinations. However, it should be said that making a real world data example provides a more reliable inferences on introduced method.

REFERENCES

[1] Cox DR. Regression models and life-tables, Journal of Royal Statistical Society, Series B, 1972; 34(2), 187-220

[2] Cox DR. Partial likelihood, Biometrika, 1975; 62(2), 269-276.

[3] Tsiatis AA. A large sample study of Cox's regression model, The Annals of Statistics, 1981; 9(1), 93-108.

[4] Andersen PK and Gill RD. Cox's regression model for counting processes: A large sample study, The Annals of Statistics, 1982; 10(4), 1100-1120.

[5] Miller RG. Least squares regression with censored data, Biometrika, 1976; 63, 449-64.

[6] Stute W and Wang J-L. The strong law under random censorship, The Annals of Statistics, 1993; 21(3), 1591-1607.

[7] Stute W. The central limit theorem under random censorship, The Annals of Statistics, 1995; 23(2), 422-439.

[8] Stute W. Consistent estimation under random censorship when covariables are present, Journal of Multivariate Analysis, 1993; 45(1), 89-103.

[9] Yılmaz E and Aydın D. Bandwidth selection problem for nonparametric regression model with right-censored data, Romanian Statistical Review, 2017; 2, 81-104.

[10] Orbe J and Virto J. Penalized spline smoothing using Kaplan-Meier weights with censored data, Biometrical Journal, 2018; 60(5), doi:10.1002/bimj.201700213.

[11] Buckley J and James I. Linear regression with censored data, Biometrika, 1979; 66(3), 429-436.

[12] Koul H, Susarla V and Van Ryzin J. Regression analysis with randomly right-censored data, The Annals of Statistics, 1981; 9, 1276-1288.

[13] Leurgans S. Linear models, random censoring and synthetic data, Biometrika, 1987; 74(2), 301-309.

[14] Dabrowska DM. Rank tests for independence for bivariate censored data, The Annals of Statistics, 1986; 14(1), 250, 264.

[15] Gross ST and Lai TL. Bootstrap methods for truncated and censored data, Statistica Sinica, 1996; 6, 509-530.

[16] Kim HT and Truong YK. Nonparametric regression estimates with censored data: Local linear smoothers and their applications, Biometrics, 1998; 54(4), 1434-1444.

[17] Ghouch AE, Van Keilegom I. Nonparametric regression with dependent censored data, Scandinavian Journal of Statistics, 2008; 35(2), 228-247.

[18] Osman M and Ghosh SK. Nonparametric regression models for right-censored data using Bernstein polynomials, Computational Statistics & Data Analysis, 2012; 56(3), 559-573.

[19] Zhang D and Cherkaev E. Pade approximations for identification of air bubble volume from temperature or frequency dependent permittivity of a two-component mixture, 2008; 16(4), 425-445.

[20] Aydın D and Yılmaz E. Truncation level selection in nonparametric regression usig Padé approximation, Communications in Statistics- Simulation and Computation, 2019; doi: 10.1080/03610918.2019.1565586.

[21] Aydın D and Yılmaz E. Modified spline regression based on randomly right-censored data: A comparative study, Communications in Statistics - Simulation and Computation, 2018; 47(9), 2587-2611.

[22] Fierro RD, Golub GH, Hansen PC and O'Leary DP. Regularization by truncated total least squares. SIAM Journal on Scientific Computing, 1997; 18(4), 1223–41.

[23] Sima DM and Huffel S. Level choice in truncated total least squares. Computational Statistics and Data Analysis, 2007; 52, 1208–22.

[24] Nadaray, EA. On nonparametric estimates of density functions and regression curves, Theory Appl. Probability, 1964; 10, 186–190,

[25] Watson GS. Smooth regression analysis, Sankhya, 1964; 26 (15), 175–184

.