

Article

A-Spline Regression for Fitting a Nonparametric Regression Function with Censored Data

Ersin Yılmaz ^{1,*} , Syed Ejaz Ahmed ² and Dursun Aydın ¹¹ Mugla Sıtkı Kocman University, Faculty of Science, Statistics, Muğla 48000, Turkey; duaydin@mu.edu.tr² Faculty of Science, Mathematics and Statistics, Brock University, Niagara Region, St. Catharines, ON L2S 3A1, Canada; sahmed5@brocku.ca

* Correspondence: ersinyilmaz@mu.edu.tr

Received: 1 May 2020; Accepted: 26 May 2020; Published: 29 May 2020



Abstract: This paper aims to solve the problem of fitting a nonparametric regression function with right-censored data. In general, issues of censorship in the response variable are solved by synthetic data transformation based on the Kaplan–Meier estimator in the literature. In the context of synthetic data, there have been different studies on the estimation of right-censored nonparametric regression models based on smoothing splines, regression splines, kernel smoothing, local polynomials, and so on. It should be emphasized that synthetic data transformation manipulates the observations because it assigns zero values to censored data points and increases the size of the observations. Thus, an irregularly distributed dataset is obtained. We claim that adaptive spline (A-spline) regression has the potential to deal with this irregular dataset more easily than the smoothing techniques mentioned here, due to the freedom to determine the degree of the spline, as well as the number and location of the knots. The theoretical properties of A-splines with synthetic data are detailed in this paper. Additionally, we support our claim with numerical studies, including a simulation study and a real-world data example.

Keywords: adaptive splines; nonparametric regression; right-censored data; synthetic data transformation

1. Introduction

Let (x_i, y_i) , $1 \leq i \leq n$ be a sample of observations where x_i 's are values of a one-dimensional covariate x and y_i 's denote the values of the completely observed response (lifetime) variable y . In medical studies such as clinical trials, y is often subject to random right-censoring and censored by a random variable c with c_i values representing the censorship times, i.e., patient withdrawal time. In this case, the observed response values at designed points x_1, x_2, \dots, x_n will be t_i 's, defined as

$$t_i = \min(y_i, c_i), \delta_i = \begin{cases} 1 & y_i \leq c_i \text{ (uncensored)} \\ 0 & y_i > c_i \text{ (censored)} \end{cases} \quad (1)$$

where δ_i 's are the values of the censoring indicator function that contains the censoring information. It should be noted that y_i , c_i , and t_i have distribution functions $F_x(s) = P(y_i \leq s)$, $G_x(s) = P(c \leq s)$, and $H_x(s) = P(t_i \leq s)$ for $s \in R$, respectively. Additionally, we assume that y_i 's and c_i 's are independent, which is a very common assumption of right-censored analysis (see [1,2]). Thus, the relationship between the distribution of t and (y, c) can be written as follows, in terms of corresponding survival functions:

$$1 - H_x(s) = [(1 - F_x(s)) \cdot (1 - G_x(s))] \quad (2)$$

This paper considers the problem of fitting a nonparametric regression function with right-censored data. Based on the condition in Equation (1) and assumption in Equation (2), the nonparametric regression model $y_i = f(x_i) + \varepsilon_i$ can be written as

$$t_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (3)$$

where $t_i = (t_1, t_2, \dots, t_n)$ are the right-censored response values that solve the censorship problem, $f(\cdot)$ is the smooth function to be estimated, and ε_i 's are the normally distributed random errors denoted as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$.

In the context of linear regression, the estimation of censored data is performed using the linear regression model proposed by [3]. Different estimators based on normal least squares for linear regression under right-censored data were introduced by [4–7]. In addition, some theoretical extensions are discussed by [8,9]. Note, also, that all the methods discussed by the above-mentioned authors are based on the assumption that there is a linear relationship between censored responses and independent variables. In real-world applications, it cannot be known whether the relationship between the responses and explanatory variables is linear. Although there are some processes to test linearity, these cannot be applied directly to censored data because they were designed based on uncensored data. In this scenario, a nonparametric regression model is widely preferred.

There are several various studies to estimate the model in Equation (3) in the literature. These existing approaches can be classified as spline-based methods, kernel smoothers, or local smoothing techniques. Spline-based techniques for right-censored data can be categorized as either smoothing splines ([10,11]) or regression splines [12]. Here, in terms of the estimation of the model in Equation (3), the difference between smoothing splines and regression splines can be expressed as being that smoothing splines have to use all unique data points as knots and, because of that, the variance of the model would be large as the fitted curve tries to pass both increased values and zeros. In regression splines, knot points can be freely determined. Regression splines perform better than smoothing splines for this reason already (see [12]). However, as is known, regression splines work based on truncated power basis polynomials, which force the method to work with a fixed degree. Studies about kernel smoothers include [13,14]. Research on local smoothing techniques can be found in [15,16]. In this study, an adaptive ridge estimator (or A-spline) is introduced based on a B-spline basis function to achieve the estimation of the model in Equation (3).

It is obvious that conventional regression estimators, whether nonparametric or not, cannot be used directly for modeling censored data. To solve this issue, there are three approaches being taken in the literature; these include using Kaplan–Meier weights [4], synthetic data transformation ([17,18]), and data imputation techniques. This paper focuses on synthetic data transformation, which is the most widely used technique in the literature. The main contribution of this technique is that it provides theoretically equal expected values of both synthetic data ($y_{i\hat{G}_x}$) and the completely observed response variable (y_i) based on a Kaplan–Meier estimator (\hat{G}_x) of the censoring variable (c_i) that can be expressed as $E(y_{i\hat{G}_x}) \cong E(y_i)$ by increasing magnitudes of uncensored observations and assigning zero values to the censored ones. Details about synthetic data transformation are given in Section 2.

The main motivation of this paper is to present a new nonparametric estimator to deal with synthetic response observations better than existing approaches. All of the methods given above have some restrictions when modeling synthetic data, which are indicated above. Because of these kinds of problems, we introduce a modified A-spline estimator, which has no boundary effects for the number of knots, location of knots, and degree of splines.

The A-spline proposed by [19] provides a sparse regression model that is easy to understand and interpret. A trademark of the A-spline that it can determine suitable knot points for B-splines by using adaptive ridge regression (see [20] for adaptive ridge regression), based on the approximation of the L_0 norm with an iterative procedure (see [19,21] for more details).

In Section 2, our methodology is presented with a synthetic data transformation, a B-spline regression, an adaptive ridge approach, and, finally, a modified A-spline estimator for the nonparametric regression model based on the synthetic responses. We also give an algorithm to obtain the introduced estimator. Section 3 involves the statistical and asymptotic properties of the obtained estimator. A simulation study and real-world data application are given in Sections 4 and 5, respectively. Finally, concluding remarks are presented in Section 6.

2. Materials and Methods

2.1. Synthetic Data Transformation

To account for the right-censored data in the estimation procedure, an adjustment must be applied to the censored dataset. Otherwise, the methods for estimating $f(x_i)$ cannot be applied directly. One of the most important reasons for this is that the right-censored response variable t_i and the actual response variable y_i have different expected values. As indicated in Section 1, to avoid this issue, synthetic data transformation is used. It can be calculated simply as follows:

$$t_{iG} = \frac{\delta_i t_i}{1 - G_x(t_i)} \tag{4}$$

where $G_x(\cdot)$ is the distribution of the censoring variable c_i , which is mentioned in the previous section. Note that because the distribution G_x is generally unknown, instead of G_x , its Kaplan–Meier estimate \hat{G}_x is used (see Koul et al. 1981), which can be formulated as

$$1 - \hat{G}_x(s) = \prod_{j=1}^n \left(\frac{n - j}{n - j + 1} \right)^{I[t_{(j)} \leq s, \delta_{(j)} = 0]}, \quad s \geq 0 \tag{5}$$

where $t_{(j)}$'s denote the ordered values of the response observations as $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ and $\delta_{(j)}$'s are the values ordered associated with $t_{(j)}$'s. Note that if the distribution G_x is taken arbitrarily, some values of t_i may be identical, which prevents the correct calculation of the Kaplan–Meier estimator. Therefore, the ordered values $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(n)}$ might not be unique. It should be emphasized that the Kaplan–Meier estimator gives an opportunity for ordering the t_j 's uniquely. In addition, it is a widely known property of the estimated distribution $\hat{G}_x(s)$ that its estimated distribution has jumps only at censored data points (see Paterson, 1977, and Kaplan and Meier, 1958).

After the acquisition of \hat{G}_x , the transformation in Equation (4) can be rewritten as

$$t_{i\hat{G}} = \frac{\delta_i t_i}{1 - \hat{G}_x(t_i)} \tag{6}$$

Thus, the model in Equation (3) is written by using the synthetic response variable $t_{i\hat{G}}$ as follows:

$$t_{i\hat{G}} = f(x_i) + \varepsilon_{i\hat{G}} \text{ where } \varepsilon_{i\hat{G}} = t_{i\hat{G}} - f(x_i), \quad i = 1, 2, \dots, n \tag{7}$$

It is important to mention that the error terms ($\varepsilon_{i\hat{G}}$), which depend on synthetic data, are random variables for given \hat{G}_x . Accordingly, it can be said that $n \rightarrow \infty, E(\varepsilon_{i\hat{G}}) \cong 0$. Consequently, at each design point x_i , the mean of the distribution function of $t_{i\hat{G}}$'s can be expressed as $E(t_{i\hat{G}}|x_i) = f(x_i)$. In addition, Lemma 1 assumes that synthetic and true response variables $t_{i\hat{G}}$ and y_i have identical expectations. It is known that the estimation of the smooth function $f(x_i)$ is a problem of estimating the expected value from the right-censored responses.

Lemma 1. *In a censorship context, incomplete observations with the associated censoring indicator variable $\{(t_i, \delta_i)\}_{i=1}^n$ are used to model the actual values of y_i using the regression function $f(x_i)$. In this manner, if the*

distribution of censoring variable G is known, then the conditional expectation of $f(x_i)$ can be expressed as $E[t_{iG}|x_i] = E[y_i|x_i] = f(x_i)$.

Proof of Lemma 1 is given in the Appendix A. In order to achieve the goal of this study, synthetic responses are modeled through a modified A-spline approach, which is formed by merging B-splines and the adaptive ridge penalty. Details are given in the next section.

2.2. B-Spline Approximation

Because our A-spline regression has been constructed based on B-splines, the necessary information and important basics are described in this section. Let $k = \{k_i | i \in \mathbb{Z}\}$ be a non-decreasing sequence given by

$$k_0 \leq k_1 \leq \dots \leq k_m \leq k_{m+1}$$

where m denotes the number of knots, k_i 's are the knot points, and (k_0, k_{m+1}) are the boundaries of the knots that cannot be counted as knot points. In this context, the B-splines of degree $q \geq 0$ are the piecewise polynomial function that has nonzero derivatives up to order $(q - 1)$ at each of the given knot points. From the properties of the B-splines, it can be said that $(m + 2q + 1)$ knots are needed for $(m + q)$ polynomial pieces. In this case, a B-spline can be described as a non-zero spline between interval $[k_i, k_{i+q+1}]$ where $(i > 0)$. Therefore, the i^{th} B-spline of degree q is notated as $B_{i,q}(x_i)$, and the calculation of it is given by

$$B_{i,q}(x_i) = \frac{x_i - k_i}{k_{i+q} - k_i} B_{i,q-1}(x_i) + \frac{k_{i+q+1} - x_i}{k_{i+q+1} - k_{i+1}} B_{i+1,q+1}(x_i), \quad q > 0 \quad (8)$$

To solve the recursive formula in Equation (8), see the algorithm described in [22]. Note that if $q = 0$, then $B_{i,0}(x_i) = I(k_i \leq x_i \leq k_{i+1})$. Some fundamental properties of B-splines are that:

- The B-spline consists of q -degreed and $(q + 1)$ polynomial pieces.
- Each spline function must be derivable up to $(q - 1)$ order.
- $(q + 1)$ B-splines are nonzero for given x_i .
- Each B-spline should be positive between intervals determined by $(q + 2)$ knot points.

From the information given above, a fitted smooth function $\hat{f}(x_i)$ for data synthetic data pairs $\{x_i, y_{i\hat{G}}\}_{i=1}^n$ can be written as a linear combination of B-splines for k knots by

$$\hat{f}(x_i) = \sum_{j=1}^k \hat{\alpha}_j B_{j,q}(x_i) \quad (9)$$

Equation (9) is useful only for a mathematical approximation. Note, also, that B-splines are a widely-used approximation for the estimation of a single-index (univariate) nonparametric regression model (see [22] for details). From this, a minimization problem emerges with a smoothness penalty written as follows:

$$PSS(\alpha; \lambda) = \sum_{i=1}^n \left\{ t_{i\hat{G}} - \sum_{j=1}^k \alpha_j B_{j,q}(x_i) \right\}^2 + \lambda \int_{x_{min}}^{x_{max}} \left\{ \sum_{j=1}^k \alpha_j B''_{j,q}(x) \right\}^2 dx \quad (10)$$

where $\lambda > 0$ is the smoothing parameter that controls the smoothness of the estimated curve. Checking the amount of the penalty term has a very crucial role in the accuracy of the model estimation. This is very similar to the smoothing parameter described by [23]. In B-spline regression, one important issue is the order of the derivative determined for the penalty term, because for its higher orders, some calculation problems may be exposed. Choosing the number and positions of the knots are very substantial decisions in the minimization of the problem in Equation (10), especially for right-censored datasets.

In this paper, setting the locations and numbers of the knots is a prior aim because it has a direct relationship with the accuracy of the estimated model, as mentioned above. To provide a suitable solution for this issue, an adaptive ridge penalty is used instead of the penalty term in Equation (10) proposed by [19]. Note, also, that the smoothing parameter is chosen by an improved AIC_c criterion, as proposed by [24]. In the next section, the adaptive ridge penalty is introduced.

2.3. Adaptive Ridge

The adaptive ridge method promises the best tradeoff between the goodness of fit (the left part of Equation (10)) and the number of knots, which provides a more powerful regression model. To achieve this purpose, it uses a large and equally spaced number of knots, then modifies the penalty term by using this number of knots.

Let a B-spline define the knot points k_1, k_2, \dots, k_m , and assume that for interval r^{th} knot, $\Delta^{q+1}(\alpha_r) = \mathbf{0}$. From that, the given knots are updated as $k_1, k_2, \dots, k_{r-1}, k_{r+1}, \dots, k_m$. Thus, the penalty term changes from the overall number of knots to the number of non-zero $(q + 1)$ order differences given by

$$\frac{\lambda}{2} \sum_{i=q+2}^m \|\Delta^{q+1}(\alpha_i)\|_0 \tag{11}$$

where $\|\Delta^{q+1}(\alpha_i)\|_0$ denotes the L_0 -norm of the difference term $\Delta^{q+1}(\alpha_i)$, which means that if $\Delta^{q+1}(\alpha_i) = \mathbf{0}$ then $\|\Delta^{q+1}(\alpha_i)\|_0 = 0$, and $\|\Delta^{q+1}(\alpha_i)\|_0 = 1$ otherwise. Here, $\lambda > 0$ is a smoothing parameter. The point of this penalty term is that it deletes the r^{th} knot and works by using the intervals $[k_{r-1}, k_r) \cap (k_r, k_{r+1}]$. Thus, the modeling process is completed using the remaining knot points.

Note that Equation (11) cannot be differentiable, which prevents the acquisition of the fitted model. The adaptive ridge method provides an approximation of the L_0 norm given in Equation (11) (see [21] for a more detailed discussion). The main idea of the adaptive ridge method is using weights to approximate the L_0 -norm. In this context, the penalized minimization criterion in Equation (10) is rewritten by using a weighted new penalty as follows:

$$WPSS(\alpha; \lambda) = \sum_{i=1}^n \left\{ t_{i\hat{G}} - \sum_{j=1}^k \alpha_j B_{j,q}(x_i) \right\}^2 + \frac{\lambda}{2} \sum_{j=q+2}^{q+m+1} w_j [\Delta^{q+1}(\alpha_j)]^2 \tag{12}$$

and the vector and matrix form of Equation (12) is

$$WPSS(\alpha; \lambda) = \|\mathbf{t}_{\hat{G}} - \mathbf{B}\alpha\|_2^2 + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D} \alpha \tag{13}$$

where $\mathbf{t}_{\hat{G}}$ is the vector of the synthetic response values, $\mathbf{W} = \text{diag}(w_j)$ and w_j 's represent positive weights, and \mathbf{D} involves the values of $\Delta^{q+1}(\alpha_j)$, which is the first difference operator and can be calculated as

$$\Delta(a_j) = a_j - a_{j-1} \text{ and } \Delta^q(a_j) = \Delta^{q-1}(\Delta(a_j)) \tag{14}$$

and $\alpha = \{\alpha_j\}_{j=1}^{q+k+1}$ is the vector of coefficients for the B-spline design matrix $\mathbf{B} = \left(\{B_{j,q}(x_i)\}_{i,j}, 1 \leq i \leq n, 1 \leq j \leq q + m + 1 \right)$, illustrated in Equation (8). In order to make more explicit the i^{th} row of the \mathbf{B} matrix, it is given as

$$\mathbf{B}_i = [B_{1,q}(x_i), B_{2,q}(x_i), \dots, B_{(q+m+1),q}(x_i)]$$

It should be noted that w_j 's provide the approximation of the penalty term in Equation (12) to the L_0 -norm and also in the adaptive-ridge procedure; weights are of crucial importance to the choice of a perfect location for the knots. To approximate to L_0 -norm, weights are determined by an iterative

process from the previous values of the coefficients α_j 's, which can be realized by the formula given in [19] as follows:

$$w_j = \left[(\Delta^{q+1}(\alpha_j))^2 + \gamma^2 \right]^{-1}, \gamma > 0 \tag{15}$$

where $\gamma > 0$ is constant, and it can be seen that the approximation $\|\Delta^{q+1}(\alpha_i)\|_0 \approx w_j(\Delta^{q+1}(\alpha_i))^2$ depends on γ .

Remark 1. As mentioned above, because the weights are determined by an iterative procedure using Equation (15), it is important to determine $\gamma > 0$ appropriately. If $(\Delta^{q+1}(\alpha_j)) < \gamma$ then the w_j 's obtained may be extremely large, causing $(\Delta^{q+1}(\alpha_j)) \approx 0$, and therefore, the resulting all penalty term is $w_j(\Delta^{q+1}(\alpha_i))^2 \approx 0$. However, if $(\Delta^{q+1}(\alpha_j)) \gg \gamma$, then the approximation of $\|\Delta^{q+1}(\alpha_i)\|_0 \approx w_j(\Delta^{q+1}(\alpha_i))^2$ is realized. In this matter, [20] obtain the value $\gamma = 10^{-5}$ after some numerical computations, which can be accepted as a suitable value of γ .

In the next section, the modified A-spline estimator is introduced based on the given adaptive-ridge penalty and synthetic response values.

2.4. Modified A-Spline Estimator

In this section, the estimation coefficient vector α is given, and, to provide a more precise and detailed explanation, an algorithm is presented. A modified A-spline estimator to estimate the right-censored nonparametric regression model in Equation (7) is obtained by minimizing Equation (13) after some algebraic operations that are given in Appendix A.1. In this case, the vector of the estimated coefficients of the A-spline regression ($\hat{\alpha}$) is computed by using the formula

$$\hat{\alpha} = (\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D}^T\mathbf{W}\mathbf{D})^{-1}\mathbf{B}^T\mathbf{t}_{i\hat{c}} \tag{16}$$

where $\lambda\mathbf{D}^T\mathbf{W}\mathbf{D}$ denotes the adaptive ridge-penalty, which involves both the difference-matrix \mathbf{D} and weight matrix \mathbf{W} . From here, fitted values for the model in Equation (7) can be obtained as

$$E[t_i|x_i] \cong E[t_{i\hat{c}}|x_i] = \hat{f}(x_i) = \mathbf{B}\hat{\alpha} = \mathbf{H}_A\mathbf{t}_{i\hat{c}} \tag{17}$$

where $\mathbf{H}_A = \mathbf{B}(\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D}^T\mathbf{W}\mathbf{D})^{-1}\mathbf{B}^T$ is a hat matrix. It should be emphasized that because of computational difficulties, instead of calculating values of matrix \mathbf{D} , the all penalty term $(\mathbf{D}^T\mathbf{W}\mathbf{D})$ is obtained by an iterative algorithm, which is the most efficient method. The algorithm for our modified A-spline estimator is shown in Algorithm 1.

Algorithm 1. Algorithm for the modified A-spline estimator $\hat{\alpha}$.

Input: covariate x_i , synthetic responses $t_{i\hat{c}}$, constant $\gamma = 10^{-5}$

output: $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{q+m+1})^T$

- 1: **Begin**
 - 2: Give initial values $\alpha^{(0)} = \mathbf{0}_{q+m+1}$ and $\mathbf{W}^{(0)} = \mathbf{I}$ to start iterative process
 - 3: **do** until converges weighted differences to L_0 - norm
 - 4: $\hat{\alpha}^{(s)} = (\mathbf{B}^T\mathbf{B} + \lambda\mathbf{D}^T\mathbf{W}\mathbf{D})^{-1}\mathbf{B}^T\mathbf{t}_{i\hat{c}}$
 - 5: $w_j^{(s)} = \left[(\Delta^{q+1}(\alpha_j^{(s)}))^2 + \gamma^2 \right]^{-1}$
 - 6: $\hat{\alpha} = \hat{\alpha}^{(s)}, \mathbf{W} = \text{diag}(w_j^{(s)})$
 - 7: **end**
 - 8: Obtain k_s by $(\Delta^{q+1}(\alpha))^{2\mathbf{W}^s}$
 - 9: Return $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_{q+m+1})^T$
 - 10: **End**
-

3. Statistical Properties of the Estimator

It is implicit that an A-spline estimator is a different kind of ridge-type estimator and is used for the estimation of the right-censored nonparametric regression model in this paper. It follows that the expressions given below can be written about using the random error terms of the model in Equation (3).

$$E[\varepsilon_i|x_i] = 0, \text{Var}[\varepsilon_i|x_i] = \sigma_\varepsilon^2 \mathbf{I} \tag{18}$$

However, in this paper, because of censoring, instead of employing the model in Equation (3), that in Equation (7), which involves synthetic responses, is used. In this case, the distribution properties in Equation (18) are changed depending on Lemma 1 and are rewritten as follows:

$$E[\varepsilon_{i\hat{G}}|x_i] \cong 0, \text{Var}[\varepsilon_{i\hat{G}}|x_i] = \sigma_{\varepsilon_G}^2 \mathbf{I} \tag{19}$$

where $\varepsilon_{i\hat{G}} = (t_i - \hat{t}_{i\hat{G}})$, $\sigma_{\varepsilon_G}^2$ is the variance of the right-censored nonparametric model based on the synthetic response variable, \mathbf{I} is the $n \times n$ identity matrix, and $\hat{t}_{i\hat{G}}$ denotes fitted values. It should be noted that the obtained estimator is a vector of coefficients $\hat{\alpha}$, and therefore, the quality of the model is measured partially based on the bias and variance of $\hat{\alpha}$. In this context, from the ordinary ridge regression method, the variance–covariance matrix of $\hat{\alpha}$ can be approximated by using $\sigma_{\varepsilon_G}^2$ as follows:

$$\text{Cov}(\hat{\alpha}) = \sigma_{\varepsilon_G}^2 \frac{1}{n} (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1} (\mathbf{B}^T \mathbf{B}) (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1} \tag{20}$$

If $M_A = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1}$, then the covariance matrix of the fitted values of the model can be given by

$$\text{Cov}(\hat{\mathbf{f}}) = \sigma_{\varepsilon_G}^2 \frac{1}{n} (\mathbf{B} M_A \mathbf{B}^T) (\mathbf{B} M_A \mathbf{B}^T)^T \tag{21}$$

Because of $\sigma_{\varepsilon_G}^2$ is generally unknown, it needs to be estimated as follows:

$$\hat{\sigma}_{\varepsilon_G}^2 = \frac{\|\mathbf{t} - \hat{\mathbf{f}}\|^2}{n - \text{tr}(\mathbf{H}_A)} \tag{22}$$

where $\text{tr}(\cdot)$ indicates the sum of the diagonal elements of a matrix. Additionally, the bias of $\hat{\alpha}$ is one of the quality measurements for the estimated model. In order to calculate the bias, the conditional expected values of the estimator $E[\hat{\alpha}|x_i]$ have to be obtained by

$$E[\hat{\alpha}|x_i] = (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1} \mathbf{B}^T \mathbf{B} \alpha \tag{23}$$

Following on from Equation (23), the bias can be written as

$$\begin{aligned} \text{Bias}(\hat{\alpha}) &= E[\hat{\alpha}|x_i] - \alpha = \left[(\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1} - (\mathbf{B}^T \mathbf{B})^{-1} \right] \mathbf{B}^T \mathbf{B} \alpha \\ &= (\mathbf{B}^T \mathbf{B} + \lambda \mathbf{D}^T \mathbf{W} \mathbf{D})^{-1} \mathbf{B}^T \mathbf{B} \end{aligned} \tag{24}$$

In this study, Equations (20)–(22) and (24) are used as quality measures to evaluate the performance of the estimated right-censored model. In addition, the mean squared errors (MSE) commonly employed in the literature is also used for measuring the quality of the fitted model. It is obtained as follows:

$$\text{MSE}(\hat{\mathbf{f}}) = \frac{1}{n} \left(\sum_{i=1}^n (t_i - \hat{f}_i)^2 \right) = \|\mathbf{t} - \hat{\mathbf{f}}\|^2 / n \tag{25}$$

3.1. Extended Properties of the Estimator

The modified A-spline estimator introduced in this paper is a smoothing technique that allows for the optimal selection of base functions, penalties, knot points, and the location of knots. It achieves that by using adaptive (weighted) ridge penalty via approximating the L_0 norm.

In this section, some large sample properties of the modified A-spline estimator are given under right-censoring. It is worth noting that the theoretical properties of the A-spline estimator have not been deeply inspected in the literature. There have been some important studies about adaptive ridge estimators, such as [20,21,25]. This section provides some initial inferences about the A-spline estimator in a nonparametric context and under censorship conditions.

Before we describe the asymptotic properties of the estimator, it should be emphasized that the flexibility of the A-spline estimator allows the choice of penalty and knot points, causing difficulties in the theoretical inferences. As is already known, the A-spline estimator is a specialized version of the P-splines proposed by [26]. Its major difference is that the A-spline changes the penalty terms using weights that are iteratively obtained and by approximating the L_0 -norm. Because of this, some assumptions and inferences are derived based on the known properties of P-splines.

The main function of the A-spline estimator is given in Equation (12), which can be rewritten as follows:

$$WPSS(\alpha; \lambda) = \sum_{i=1}^n \left\{ t_{i\hat{G}} - \sum_{j=1}^k \alpha_j B_{j,q}(x_i) \right\}^2 + \frac{\lambda}{2} \sum_{i=q+2}^m \|\Delta^{q+1}(\alpha_i)\|_{\tau} \tag{26}$$

where $\|\cdot\|_{\tau}$ denotes the τ -norm. To obtain substantial results, for this study, we assume that $\tau \rightarrow 0$ because solving L_0 -norm requires complex calculations. Accordingly, it can be said that minimizing Equation (26) has good potential for both estimating α_j 's and determining the optimal knot points, such as model selection for sufficiently large $\lambda > 0$. As is known from the literature, model selection with $\tau \rightarrow 0$ is realized by penalizing non-zero parameters, which is a limiting case of the bridge estimation introduced by [27] and given as

$$\lim_{\tau \rightarrow 0} \sum_{i=q+2}^m \|\Delta^{q+1}(\alpha_i)\|_{\tau} = \sum_{i=q+2}^m I[\Delta^{q+1}(\alpha_i) \neq 0] \tag{27}$$

For $\tau \geq 1$, the objective function in Equation (26) has a convex structure, and its global minimum can be obtained easily by using numerical algorithms. However, when $\tau \rightarrow 0$ and $\tau = 0$, the criterion in Equation (26) is no longer convex and its computation is non-trivial. In the L_0 -norm context, there is no guarantee of reaching a global minimum. Moreover, more than one local minimum could exist. Thus, there is no unique solution of this estimator, and it depends on the iterative process. In [21], it is shown that a minimum of 5 and maximum 40 iterations provide reasonable convergence of the estimator to real parameters.

When estimator $\hat{\alpha}$ is inspected asymptotically, although its objective function in Equation (26) is non-convex, calculations about asymptotic consistency can be guided. In this case, the following condition is assumed:

$$\mathbf{R}_n = \frac{1}{k} \sum_j^k \mathbf{B}_j \mathbf{B}_j^T \rightarrow \mathbf{R} \tag{28}$$

where \mathbf{R} is a non-negative definite matrix, and also assumed is that

$$\frac{1}{q+m+1} \max_{1 \leq j \leq k} \mathbf{B}_j^T \mathbf{B}_j \rightarrow 0 \tag{29}$$

In general, the obtained explanatory variables included by \mathbf{B} are scaled. Accordingly, all of the diagonals of \mathbf{R} are equal to 1. Note that it must be assumed that $\mathbf{B}_j^T \mathbf{B}_j$ and $(\mathbf{B}_j^T \mathbf{B}_j)^{-1}$ are nonsingular

matrices; consequently, \mathbf{R} are full rank matrices to the obtained identifiable properties. Using the conditions in Equations (28) and (29), the limiting behavior of estimator $\hat{\alpha}$ can be observed by inspecting the asymptotic state of affairs of the minimization problem in Equation (12). To see the consistency of $\hat{\alpha}_n$, the function is given as

$$U_n(\hat{\alpha}_n) = \sum_{i=1}^n \left\{ t_{i\hat{G}} - \sum_{j=1}^k \hat{\alpha}_{nj} B_{j,q}(x_i) \right\}^2 + \frac{\lambda_n}{2} \sum_{i=q+2}^m \|\Delta^{q+1}(\hat{\alpha}_i)\|_\tau \tag{30}$$

where $\hat{\alpha}_n$ is a consistent estimator for $\lambda_n = o(n)$. This result is confirmed by following theorem:

Theorem 1. *If \mathbf{R} is a full rank-matrix and $\frac{\lambda_n}{2} \rightarrow \lambda \geq 0$, then $\hat{\alpha}_n \xrightarrow{p} \operatorname{argmin}(U)$ where*

$$U(\hat{\alpha}_n) = (\hat{\alpha}_n - \alpha)^T \mathbf{R}(\hat{\alpha}_n - \alpha) + \lambda \sum_{i=q+2}^m \|\Delta^{q+1}(\alpha_i)\|_\tau \tag{31}$$

Thus, $\lambda_n = o(n)$ and $\hat{\alpha}_n$ is a consistent estimator of α . It could therefore be said that

$$|\hat{\alpha}_n| \rightarrow \alpha, \text{ as } n \rightarrow \infty \tag{32}$$

Proof of Theorem 1 is given in the Appendix A.

Because U_n is not convex due to the degree of norm $\tau \rightarrow 0$, and to ensure the accuracy of Equation (32), some additional notes are needed. Accordingly, it can be said that $\lambda_n = O(\sqrt{n})$ is essential for $\tau \rightarrow 0$. From that, if $\frac{\lambda}{\sqrt{n}} \rightarrow \lambda \geq 0$ and $\tau \rightarrow 0$, then it can be written that

$$\sqrt{n}(\hat{\alpha}_n - \alpha) \xrightarrow{d} \mathbf{R}^{-1}\mathbf{J} \sim N(0, \sigma^2\mathbf{R}^{-1}) \tag{33}$$

where \mathbf{J} has a distribution $N(0, \sigma_\varepsilon^2)$ and its elements consist of the random error terms ε_i 's.

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, and the experimental conclusions that can be drawn.

4. Simulation Study

In this section, a simulation study is carried out to see the behaviors of the modified A-spline estimator when estimating the right-censored nonparametric model. Before the results of simulation experiments, datasets for the different simulation combinations are generated using by the “simcensdata” function in the R software, which can be accessed via this link: <https://github.com/yilmazersin13/simcensdata-generating-randomly-right-censored-data>. Our data generation procedure, with accompanying descriptions, is given in Table 1.

Table 1. Data generation procedure with explanations.

Steps	Explanation
Step 1. Decide n_{obs} , n_{sim} , and \mathcal{L}	Sample size of simulated dataset and number of repetitions and censoring level, respectively
Step 2. Produce, t_i 's, $f(\cdot)$, and ε_i 's	Nonparametric covariate, real smooth function, and random error terms
Step 3. Obtain, y_i 's	Actual (complete-uncensored) data points
Step 4. Generate δ_i 's using Bernoulli distribution	Values of censoring indicator
Step 5. Obtain c_i 's i.i.d. with y_i 's	Censoring values that cut the actual lifetimes
Step 6. Find $t_i = \min(y_i, c_i)$	Obtain partly observed response values

For this simulation study, within the scope of Step 1, $n_{obs} = (35, 100, 350)$, $n_{sim} = 1000$, and the censoring levels $\mathcal{L} = (5\%, 20\%, 40\%)$. The nonparametric covariate and random errors in Step 2 are generated as $x_i = \theta(i - \frac{1}{2})/n$ and $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$, where θ is a constant that determines the shape of the curve. Note that, in this study, two different types of function are used to test the introduced method under various conditions. These functions are given below with their formulations as follows:

Panel (a) and Panel (b) represent two different datasets that were formed based on nonlinear functions f_1 and f_2 . The plots of Figure 1 are drawn for $n = 100$ and $\mathcal{L} = 20\%$. It should be noted that the optimal selection of numbers and the positions of knots are extremely important for the functions represented in these panels. In the context of synthetic data transformation, censored data points take zero values and completed points take higher values than they are. In this case, deciding the properties of knots will be crucial.

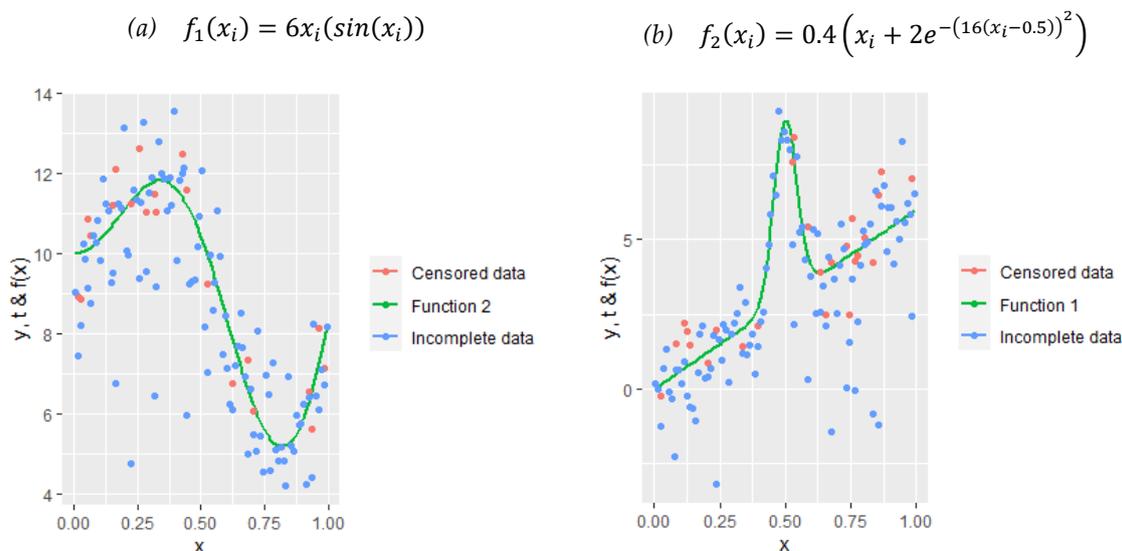


Figure 1. Scatter plots of both censored data and incomplete response data points over the smooth functions to be estimated by A-splines.

From the data generation procedure given above, the right-censored nonparametric model can be written as follows:

$$t_i = f_h(x_i) + \varepsilon_i, \quad i = 1, \dots, n_{obs}, \quad h = 1, 2 \tag{34}$$

Then, to use censorship information in the estimation process, a synthetic data transformation is done, as in Equation (6). Therefore the final model to be estimated, as given by the simulation experiments, is

$$t_{i\hat{G}} = f_h(x_i) + \varepsilon_{i\hat{G}}, \quad i = 1, \dots, n_{obs}, \quad h = 1, 2 \tag{35}$$

In this simulation study, for three sample sizes, three censoring levels, and two functions, 18 configurations are obtained. All the outcomes for the model in Equation (34) under these conditions are given in the following figures and tables.

Table 2 represents the scores of all the evaluation metrics for each of the simulation configurations. The results are inspected from three essential aspects in terms of the estimation performance of the A-spline estimator that are the effects of the sample size, censoring level, and shape of the data. For the first aspect, it can be seen from the table that $\hat{\sigma}_{\varepsilon\hat{G}}$, MSE and $Var(\hat{\alpha})$ decrease when the sample size increases. This can be interpreted as practical proof of the asymptotic convergence that is one of the main purposes of this simulation study. This interpretation is consistent for all censoring levels. The censoring level naturally affects the performance of the estimator contrary to sample size; however, there is a sensitive point, which depends on the reaction of the estimator to variation in the censoring level, which makes this paper significant. If the scores are inspected carefully, it can be clearly seen

that there are no huge differences between low and high censoring levels, which can also be seen in the figures given below. This case proves that the A-spline estimator achieves mitigation of the effect of the censoring level on selecting the optimal knot points, as expected. Finally, two different function types are used in this paper. f_1 has a shape that is similar to that of a sinus function and is not hard to catch for any smoothing technique. f_2 is an almost linear function but has one big peak; this is a challenge for the estimator, especially under censoring. The outcomes in Table 2 demonstrate this. Although the results for \hat{f}_1 are smaller than those for \hat{f}_2 , it can be said that the A-spline estimator shows a satisfactory performance for both datasets.

Table 2. Variances and biases of $\hat{\alpha}$, variance of the model ($\hat{\sigma}_{\varepsilon_{\hat{c}}}$), and MSE of \hat{f} for all simulation combinations.

Function Type		$\hat{f}_1(x)$				$\hat{f}_2(x)$			
\mathcal{L}	n	$\hat{\sigma}_{\varepsilon_{\hat{c}}}$	MSE	Bias($\hat{\alpha}$)	Var($\hat{\alpha}$)	$\hat{\sigma}_{\varepsilon_{\hat{c}}}$	MSE	Bias($\hat{\alpha}$)	Var($\hat{\alpha}$)
5%	35	1.2151	1.2064	0.9971	0.0911	1.2151	1.2064	0.9971	0.1347
	100	1.1929	1.1957	0.9979	0.0215	1.1733	1.1804	0.9979	0.0212
	350	1.2089	1.2186	0.9982	0.0052	1.1750	1.1861	0.9982	0.0051
20%	35	2.3410	2.5041	0.9971	0.1756	2.3410	2.5041	0.9971	0.2140
	100	2.3098	2.4858	0.9979	0.0417	2.3827	2.5881	0.9979	0.0430
	350	2.2610	2.4481	0.9982	0.0098	2.3724	2.5716	0.9982	0.1031
40%	35	3.7614	4.3230	0.9971	0.2822	3.7614	4.3230	0.9971	0.3465
	100	3.5616	4.1446	0.9979	0.0643	4.4640	5.1566	0.9979	0.0806
	350	3.4012	3.9752	0.9982	0.0147	4.5578	5.2489	0.9982	0.1980

Table 3 represents the comparative outcomes for the introduced estimator modified A-spline and commonly used SS and RS. The best scores are indicated with bold colored text. As can be seen, the results indicate that the modified A-spline estimator shows the best performance from a general perspective. Additionally, as mentioned in the introduction section, RS has smaller MSEs than SS. From here, it can be said that the introduced method gives more satisfying results, which can be explained by its adaptive nature. If Table 3 is inspected carefully, it can be realized that for the results obtained from $\hat{f}_2(x)$, the RS method has attractive outcomes when the censoring level is 40%. It is an understandable situation because of the shape of the function.

Table 3. MSE values for the A-spline, smoothing spline (SS), and regression spline (SS) methods to make comparisons.

Function Type		$\hat{f}_1(x)$			$\hat{f}_2(x)$		
\mathcal{L}	n	A-spline	RS	SS	A-spline	RS	SS
5%	35	1.2064	1.9622	1.9581	1.2064	1.6632	1.9527
	100	1.1957	1.2386	1.5765	1.1804	1.4327	1.7061
	350	1.2186	1.1975	1.2644	1.1861	1.2760	1.6197
20%	35	2.5041	3.2386	3.6733	2.5041	2.9084	3.1748
	100	2.4858	3.1953	3.2851	2.5881	2.6006	2.7318
	350	2.4481	2.5459	2.7270	2.5716	2.5591	2.4059
40%	35	4.3230	5.5459	5.5779	4.3230	4.5839	5.3468
	100	4.1446	4.6522	5.0570	5.1566	5.1013	5.1927
	350	3.9752	4.0440	4.0591	5.2489	4.9741	5.0198

Figure 2 aims to show how the modified A-spline estimator behaves when the sample size is exceedingly small, under various censoring levels. It is obvious that the estimation of f_1 is easier than that of f_2 , which is explained below. Figure 2 shows this more clearly. In addition, it can be said that the method is successful for even extremely small sample sizes ($n = 35$). This is an important

contribution of this method for right-censored data because in a medical dataset and especially in clinical observations, many data may frequently be unobtainable.

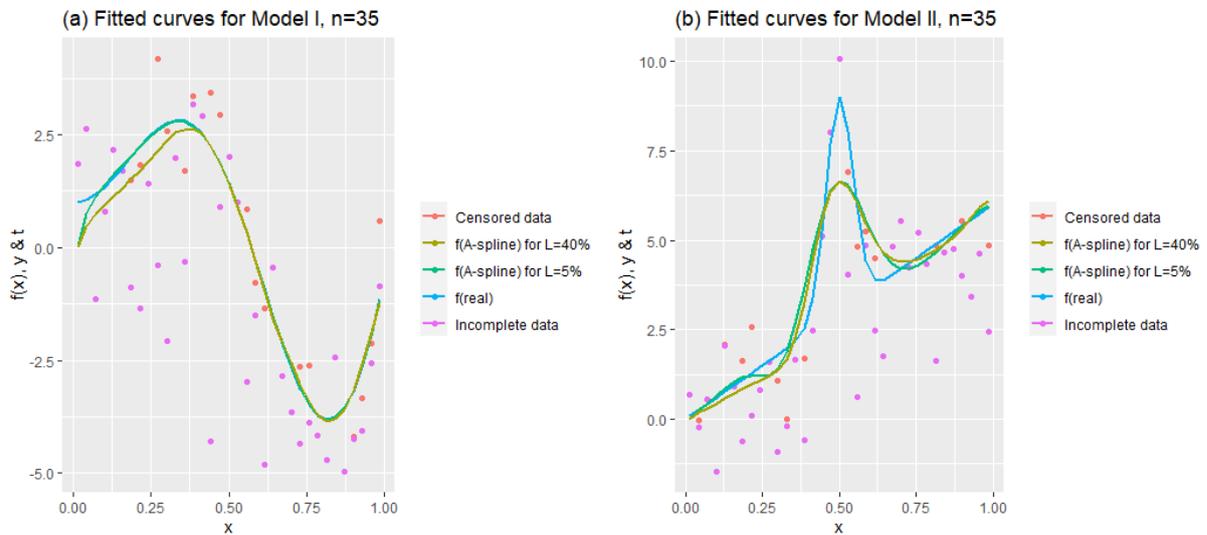


Figure 2. Fitted curves to see the performance of the estimator for $n = 35, \mathcal{L} = 5\%, 40\%$.

Figure 3 presents the effects of sample sizes by keeping the censoring level constant at 20%. Model I was obtained using f_1 ; similar fitted curves are obtained for $n = 100$ and $n = 350$, and these curves seem to be good representations of the data. This inference is also valid for Model II. Both plots show that the fitted curves successfully model right-censored data.

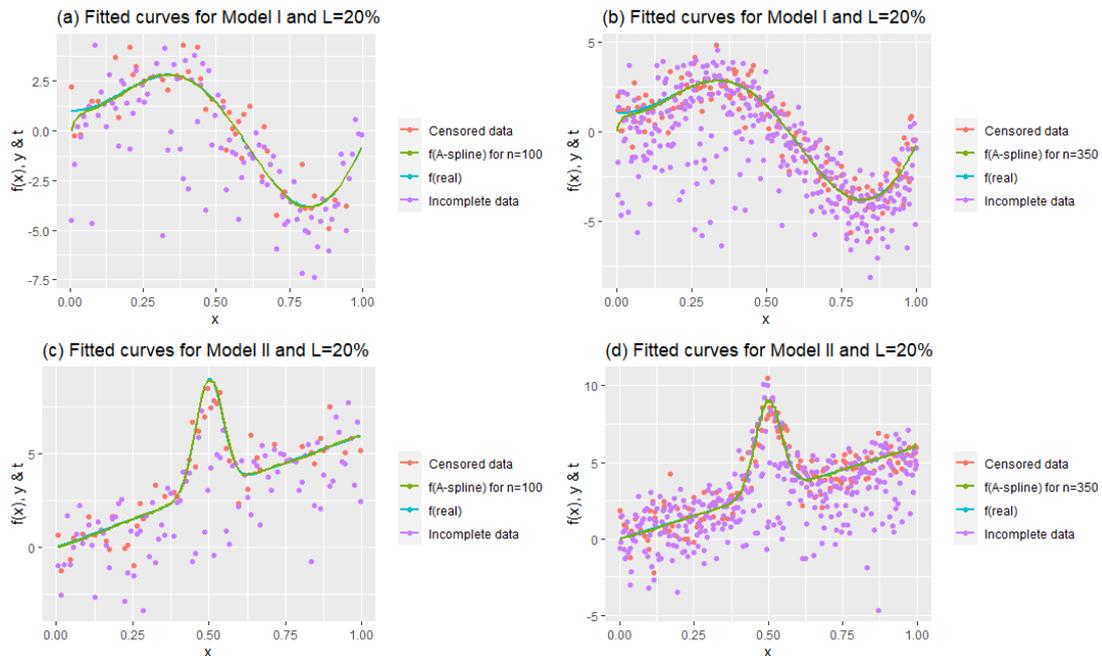


Figure 3. Fitted curves to see the performance of the estimator for $n = 100, 350, \mathcal{L} = 20\%$.

Figure 4 demonstrates how the method works under heavy censoring. To that end, fitted curves are shown for a moderate sample size together with the lowest and the highest censoring levels, 5% and 40%. As we expected, the A-spline estimator demonstrates its ability to handle data with zero values

obtained by synthetic data transformation, and it can be clearly seen that there is a difference between the two graphs. This inference is also supported by the results in Table 2.

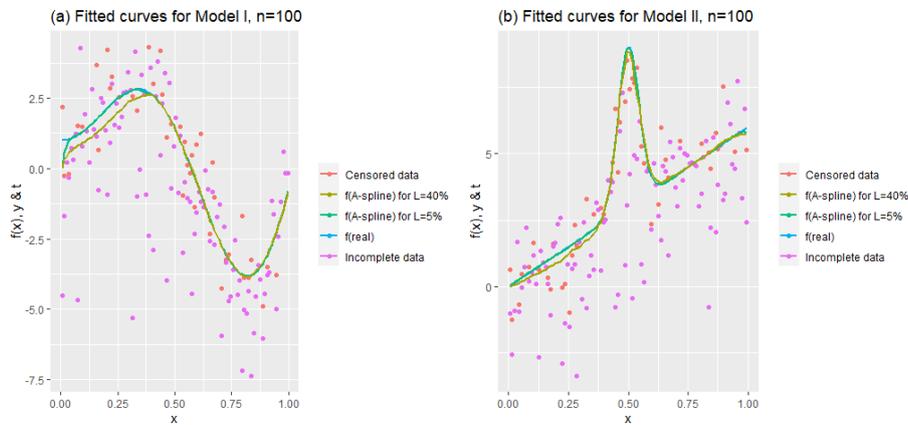


Figure 4. Fitted curves to observe the quality of the estimates for $n = 100$, $\mathcal{L} = 5\%$, 40% .

Figure 5 depicts bar plots of the measurement tools for both the estimated A-spline coefficients and the estimated model. In each panel, A1.5%, A1.20%, and A1.40% denote the obtained scores of the evaluation metric for $\mathcal{L} = 5\%$, 20% , and 40% , respectively, for $n = 35$. In a similar manner, A2.5%, A2.20%, and A2.40% represent the scores for $n = 100$ and all the censoring levels, and A3.5%, A3.20%, and A3.40% denote the results for $n = 350$ for all the censoring levels. The top panels of Figure 5 include bar plots for the bias values. As in Table 2, it can be seen here that the biases for the two models are very similar and, as expected, become smaller in larger samples. The panels in the middle show bar plots for the variances of the coefficients. The plots appear similar for the two models, but, as has been said before, because the estimation of Model II is more difficult than that of Model I, the y-axis is significantly wider in scope. The panel at the bottom is drawn for the MSE values of the estimated model, and it is similar to the variance plots. Essentially, these plots prove that the A-spline estimator can estimate the model by overcoming the effect of censorship in terms of various evaluation metrics.

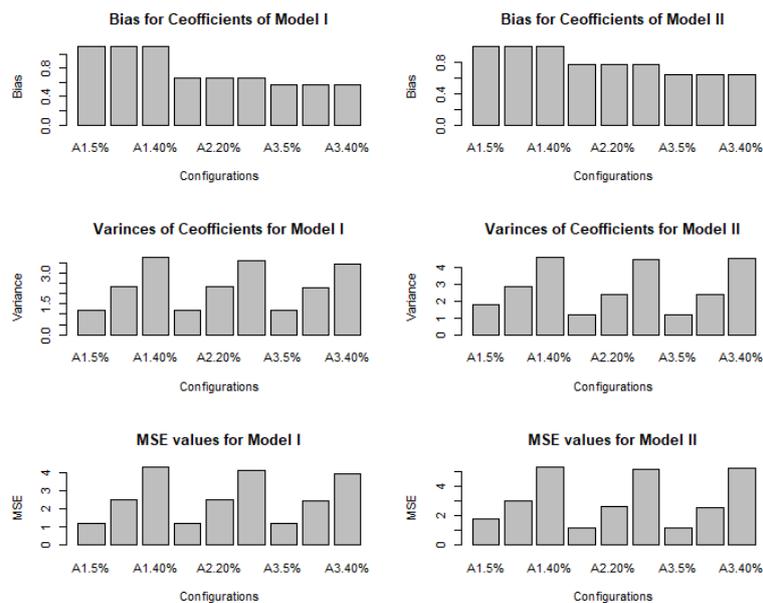


Figure 5. Bar plots of MSE values for each simulation repetition and their changes under different censoring levels.

5. Real Data Application

This section is prepared to show the performance of the modified A-spline estimator on real right-censored data. The dataset represents data from colon cancer patients in İzmir. The dataset involves the survival times, censoring indicator (δ_i), and albumin (i.e., the most common protein found in the blood) values of patients. To provide continuity, the logarithms of the survival times are considered as a response variable (*survival time*), and albumin is taken as a nonparametric covariate (*albumin*). The right-censored regression model is thus given by

$$t_i = \log(\text{survival time}_i) = f(\text{albumin}_i) + \varepsilon_i, \quad 1 \leq i \leq 97 \tag{36}$$

Note, also, that because t_i 's cannot be used directly in the estimation procedure, they have been replaced by the synthetic responses shown in Equation (6). The model in Equation (36) is thus rewritten as

$$t_{i\hat{C}} = \log(\text{survival time}_{i\hat{C}}) = f(\text{albumin}_i) + \varepsilon_{i\hat{C}} \tag{37}$$

The dataset contains information for 97 patients to be used for this analysis. However, the records of 32 of these patients are incomplete, containing right-censored observations; the data of the remaining 65 patients are uncensored (deceased). Consequently, in this dataset, the censoring level is $\mathcal{L} = 32.98$. The outcomes calculated for the model in Equation (37) are given the following table and figure.

Table 4 summarizes the performance of the modified A-spline estimator. Note that the values of $\hat{\sigma}_{\varepsilon_{\hat{C}}}$, MSE and $Var(\hat{\alpha})$ are better than the results of Aydın and Yılmaz (2018), who previously used regression splines to model right-censored data. In addition, to provide a healthier comparison, the results of the RS and SS methods are given in the table. As can be seen, the results are pretty similar to the simulation results. Here, A-spline gives the best score, which proves the benefit of the introduce method. Additionally, when the shape of the dataset is inspected from Figure 6, it can be described as having an irregular shape, and it can be seen that this irregularity increases after synthetic data transformation, which is demonstrated by the blue dots in the figure. Despite this challenging case, the A-spline fit seems to represent the data well.

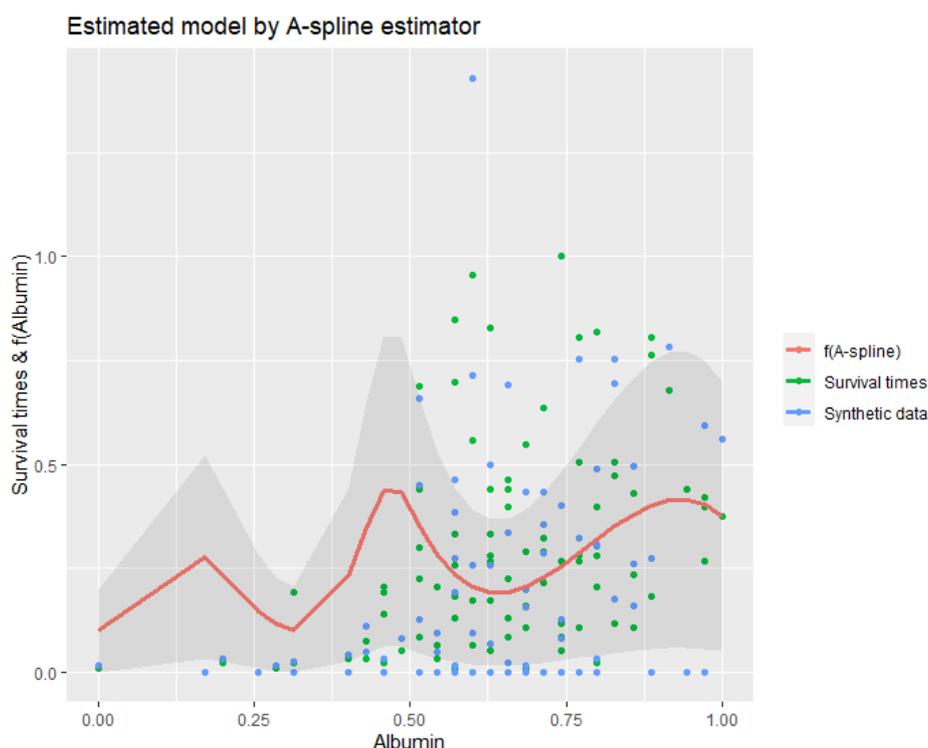


Figure 6. Estimated model for cancer data by the A-spline estimator.

Table 4. Outcomes for the estimated regression model for colon cancer data.

	$\hat{\sigma}_{\varepsilon\hat{\varepsilon}}$	MSE	Var($\hat{\alpha}$)	\mathcal{L}	n
$\hat{f}(Albumin)$	0.0051	0.0814	0.0037		
$\hat{f}_{RS}(Albumin)$	0.0540	0.0821	-	32.98%	97
$\hat{f}_{SS}(Albumin)$	0.0895	0.0863	-		

6. Concluding Remarks

This paper demonstrates that a modified A-spline estimator can be used to estimate the right-censored nonparametric regression model successfully. This is because it uses an adaptive procedure for determining the penalty term and works with only optimum knot points. A simulation study and real data example were carried out to demonstrate the performance of the method, and it can be seen from our findings that the modified A-spline estimator has merit for the estimation of right-censored data.

In the general frame of the numerical examples, incremental changes in the sample size affect the performance of the method, which gives closer results to real observations. This can be seen in Figures 3–5. Moreover, changes to the censoring level also influence the goodness of fit, and, as expected, when the censoring level increases, the performance of the method is negatively affected. However, there is an important difference here in terms of the modified A-spline. The main purpose of the usage of this method is to diminish the effect of censorship on the modeling process, and most of our results show that the introduced method achieves this purpose. For an example of these results, see Table 2. In the simulation study, two different function types are used to generate the model. f_1 is a cliché pattern of the sinus curve and is not difficult to estimate for any smoothing method. f_2 is a little bit more difficult to handle, especially by the smoothing techniques that use all data points as knots. In this paper, it can be seen that for almost all of the simulation configurations, the modified A-spline estimator gives really close values in terms of all evaluation metrics.

The real-world application uses the dataset of colon cancer patients. Their survival times are estimated by using albumin values in their blood. Figure 6 and Table 4 show the outcomes of this study. As mentioned above, our method does a good job despite the unsteadily scattered data points. The confidence interval given by the shaded region in Figure 6 seems wide because synthetic data transformation puts censored points (as zeros) far from the uncensored points. Considering the mentioned properties, it can be said that the modified A-spline estimator can be counted as a robust estimator for right-censored datasets. As a result of this study, we recommend that the modified A-spline estimator is appropriate for modeling clinical datasets.

Author Contributions: Conceptualization, D.A. and S.E.A.; methodology, D.A. and S.E.A.; software, E.Y.; validation, D.A., S.E.A. and E.Y.; formal analysis, E.Y.; investigation, D.A. and E.Y.; resources, D.A.; data curation, E.Y.; writing—original draft preparation, D.A. and E.Y.; writing—review and editing, S.E.A.; visualization, D.A. and E.Y.; supervision, S.E.A.; project administration, S.E.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: Thank the editors and reviewers for their objective assessments.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Appendix A.1 Proof of Lemma 1

Lemma 1 can be proven by using the common independency assumption between c_i and y_i , which is given in Section 1. From that, proof is given as follows:

$$\begin{aligned} E[t_{iG}|x_i] &= E\left[\frac{\delta_i t_i}{1-G(t_i)}|x_i\right] = E\left[\frac{\delta_i t_i}{\bar{G}(t_i)}|x_i\right] \\ &= E\left[\frac{I(y_i \leq c_i) \min(y_i, c_i)}{\bar{G}(\min(y_i, c_i))}|x_i\right] = E\left[I(y_i \leq c_i) \frac{y_i}{\bar{G}(t_i)}|x_i\right] \\ &= E\left[E\left[\frac{y_i}{\bar{G}(t_i)} I(y_i \leq c_i)|x_i, y_i\right]|x_i\right] = E\left[\frac{y_i}{\bar{G}(t_i)} \bar{G}(t_i)|x_i\right] = E(y_i|x_i) = f(x_i) \end{aligned}$$

Thus, proof of Lemma is completed. Note that because of distribution of censoring distribution G is unknown, it is replaced by its Kaplan–Meier estimator \hat{G} that is given in Equation (5).

Appendix A.2 Proof of Theorem 1

The equations given below need to be shown for validation of Theorem 1:

$$\sup_{\hat{\alpha}_n \in Q} |U_n(\hat{\alpha}_n) - U(\hat{\alpha}_n) - \sigma_\varepsilon^2| \xrightarrow{P} 0 \quad (\text{A1a})$$

where σ_ε^2 is the variance of the model defined in (Section 3.1), Q is a compact set in a metric space and using by Equations (28)–(32), it can be said that

$$|\hat{\alpha}_n| \rightarrow \alpha, \text{ as } n \rightarrow \infty \quad (\text{A1b})$$

(See [28], for more details).

References

1. Stute, W. Consistent Estimation Under Random Censorship When Covariables Are Present. *J. Multivar. Anal.* **1993**, *45*, 89–103. [\[CrossRef\]](#)
2. Kaplan, E.L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Statist. Assoc.* **1958**, *53*, 457–481.
3. Cox, D.R. Regression Models and Life-Tables. *J. R. Stat. Soc. Ser. B* **1972**, *34*, 187–202. [\[CrossRef\]](#)
4. Miller, R.G. Least squares regression with censored data. *Biometrika* **1976**, *63*, 449–464. [\[CrossRef\]](#)
5. Buckley, J.; James, I. Linear regression with censored data. *Biometrika* **1979**, *66*, 429–436. [\[CrossRef\]](#)
6. Miller, R.; Halpern, J. Regression with censored data. *Biometrika* **1982**, *69*, 521–531.
7. Jin, Z.; Lindgren, C.M.; Ying, Z. On least-squares regression with censored data. *Biometrika* **2006**, *93*, 147–161. [\[CrossRef\]](#)
8. Ritov, Y. Estimation in a linear regression model with censored data. *Ann. Stat.* **1990**, *18*, 303–328. [\[CrossRef\]](#)
9. Lai, T.L.; Ying, Z. Estimating a distribution function with truncated and censored data. *Ann. Stat.* **1991**, *19*, 417–442. [\[CrossRef\]](#)
10. Köhler, M.; Máthé, K.; Pinter, M. Prediction from randomly right censored data. *J. Multivar. Anal.* **2002**, *80*, 73–100. [\[CrossRef\]](#)
11. Winter, S. Smoothing Spline Regression Estimates for Randomly Right Censored Data. Ph.D. Thesis, University of Stuttgart, Stuttgart, Germany, 2013.
12. Aydin, D.; Yilmaz, E. Modified spline regression based on randomly right-censored data: A comparative study. *Commun. Stat.-Simul. Comput.* **2017**, 1–25. [\[CrossRef\]](#)
13. El Ghouch, A.; van Keilegom, I. Non-parametric regression with dependent censored data. *Scand. J. Stat.* **2008**, *35*, 228–247. [\[CrossRef\]](#)
14. Aydin, D.; Yilmaz, E. Nonparametric regression with randomly right-censored data. *Int. J. Math. Comput. Methods* **2016**, *1*, 186–189.

15. Kim, H.T.; Truong, Y.K. Nonparametric regression estimates with censored data: Local linear smoothers and their applications. *Biometrics* **1998**, *54*, 1434–1444. [[CrossRef](#)]
16. Peng, L.; Sun, S. Comparisons between local linear estimator and kernel smooth estimator for a smooth distribution based on MSE under right censoring. *Commun. Stat.-Theory Methods* **2007**, *36*, 297–312. [[CrossRef](#)]
17. Koul, H.; Susarla, V.; van Ryzin, J. Regression analysis with randomly right-censored data. *Ann. Stat.* **1981**, *9*, 1276–1288. [[CrossRef](#)]
18. Leurgans, S. Linear models, random censoring and synthetic data. *Biometrika* **1987**, *74*, 301–309. [[CrossRef](#)]
19. Goepp, V.; Bouaziz, O.; Nuel, G. Spline regression with automatic knot selection. *arXiv* **2018**, arXiv:1808.01770, preprint.
20. Frommlet, F.; Nuel, G. An adaptive ridge procedure for L0 regularization. *PLoS ONE* **2016**, *11*, e0148620. [[CrossRef](#)]
21. Rippe, R.C.A.; Meulman, J.J.; Eilers, P.H.C. Visualization of genomic changes by segmented smoothing using an L0 penalty. *PLoS ONE* **2012**, *7*, e38230. [[CrossRef](#)]
22. De Boor, C. *A Practical Guide to Splines*; Springer: New York, NY, USA, 1978.
23. Reinsch, C.H. Smoothing by spline functions. *Numer. Math.* **1967**, *10*, 177–183. [[CrossRef](#)]
24. Hurvich, C.M.; Simonoff, J.; Tsai, C. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B* **1998**, *60*, 271–293. [[CrossRef](#)]
25. Eilers, P.H.C.; De Menezes, R.X. Quantile smoothing of array CGH data. *Bioinformatics.* **2004**, *21*, 1146–1153. [[CrossRef](#)] [[PubMed](#)]
26. Eilers, P.H.; Marx, B.D. Flexible smoothing with B-splines and penalties. *Stat. Sci.* **1996**, *11*, 89–121. [[CrossRef](#)]
27. Frank, I.E.; Friedman, J.H. A statistical view of some chemometrics regression tools (with discussions). *Technometrics* **1993**, *35*, 109–148. [[CrossRef](#)]
28. Fu, W.; Knight, K. Asymptotics for lasso-type estimators. *Ann. Stat.* **2000**, *28*, 1356–1378. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).