entropy

MDPI

Article

Right-Censored Time Series Modeling by Modified Semi-Parametric A-Spline Estimator

Dursun Aydın [1], Syed Ejaz Ahmed [2] and Ersin Yılmaz [1,*]

[1] Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, Kotekli 48000, Turkey; duaydin@mu.edu.tr
[2] Department of Mathematics and Statistics, Faculty of Science, Brock University, 1812 Sir Isaac Brock Way, St. Catharines, ON L2S 3A1, Canada; sahmed@brocku.ca
* Correspondence: ersinyilmaz@mu.edu.tr

Abstract: This paper focuses on the adaptive spline (A-spline) fitting of the semiparametric regression model to time series data with right-censored observations. Typically, there are two main problems that need to be solved in such a case: dealing with censored data and obtaining a proper A-spline estimator for the components of the semiparametric model. The first problem is traditionally solved by the synthetic data approach based on the Kaplan–Meier estimator. In practice, although the synthetic data technique is one of the most widely used solutions for right-censored observations, the transformed data's structure is distorted, especially for heavily censored datasets, due to the nature of the approach. In this paper, we introduced a modified semiparametric estimator based on the A-spline approach to overcome data irregularity with minimum information loss and to resolve the second problem described above. In addition, the semiparametric B-spline estimator was used as a benchmark method to gauge the success of the A-spline estimator. To this end, a detailed Monte Carlo simulation study and a real data sample were carried out to evaluate the performance of the proposed estimator and to make a practical comparison.

Keywords: adaptive splines; B-splines; right-censored data; semiparametric regression; synthetic data transformation; time series

Citation: Aydın, D.; Ahmed, S.E.; Yılmaz, E. Right-Censored Time Series Modeling by Modified Semi-Parametric A-Spline Estimator. *Entropy* **2021**, *23*, 1586. https://doi.org/10.3390/e23121586

Academic Editor: Jan Mielniczuk

Received: 19 October 2021
Accepted: 22 November 2021
Published: 27 November 2021

1. Introduction

Time series datasets are censored from the right under specific conditions, such as a detection limit or an insufficient observation process. Consider a device which cannot measure values above a certain point, which is known as a detection limit. Since the device cannot determine the real value of an observation above its detection limit, such observations are recorded as right-censored data points. The hourly observed cloud ceiling heights data collected by the National Center for Atmospheric Research (NCAR) and modelled by [1,2] can be used as an example of a right-censored time series. Although right-censored time series are encountered frequently in the real world, in the literature, there are truly few studies completed on the estimation of right-censored time series. This may be because censorship is an unwanted data irregularity for the researchers, and it is therefore often ignored or solved by outdated techniques.

To solve the censorship problem before modelling the time series, reference [1] used the Gaussian imputation technique to estimate the series using modified ARMA models. In a similar manner, references [2,3] solved the censorship problem by using data imputation techniques. The common ground of these studies is the use of imputation and data augmentation methods to estimate the regression models with autoregressive errors for right-censored time series. On the other hand, there is an easier way to handle the censorship problem called synthetic data transformation. Although data imputation techniques have some merits, they are generally based on iterative algorithms and their calculations are costly. Reference [4] estimated the temporally correlated and right-censored

series by Nadaraya–Watson estimator nonparametrically, solving the censorship problem using a data transformation technique. Various data transformation (or synthetic data) methods have been proposed and studied in the literature for independent and identically distributed (i.i.d.) datasets; for example, see [5–7]. Because synthetic data transformation manipulates the data structure, which is disadvantageous, this solution method is no longer the preferred technique for right-censored time series. This paper aims to propose a method which can overcome the disadvantage of the synthetic data transformation method.

Note that the studies mentioned above consider the modeling of time series data using parametric or nonparametric methods. The data structure of a time series in the real world is generally not suitable for parametric modelling, because it requires rigid assumptions to reach reasonable estimates. Single-index nonparametric models, on the other hand, are very flexible, which is an important advantage over parametric methods and there are valuable studies on the subject [2,8,9]. However, nonparametric approaches lose their statistical efficiencies, when the number of covariates increases. In addition, it should be noted that, when a time series dataset is right-censored, the weaknesses of both methods are further increased.

Considering the issues mentioned above, this paper adopts semiparametric regression model for estimating right-censored time series. Although several researchers have introduced different types of semiparametric estimators for time series data, such as [10,11], there remains a significant gap in the research regarding the modelling of right-censored time series data. To address this absence, our paper proposes a modified semiparametric A-spline (AS) estimator based on synthetic data transformation. Thus, the bidirectional flexibility of the semiparametric model will be used, and the censorship problem will be effectively solved.

The paper is designed as follows: the methodology and fundamental ideas about right-censored semiparametric time series model with autoregressive errors and the synthetic data transformation method are given in Section 2. Section 3 introduces a modified AS estimator for parametric and nonparametric components of the right-censored time series model, and a semiparametric B-spline (BS) is given as a benchmark. Section 4 involves the statistical properties and evaluation criteria for both the modified AS and benchmark BS methods. Section 5 introduces some additional information about the penalty term of the semiparametric AS approach. Sections 6 and 7 contain a detailed Monte Carlo simulation study and a real-world data example, respectively. Conclusions are presented in Section 8.

2. Background

The classical semiparametric model can be defined as a hybrid model with a finite dimensional parametric component and a nonparametric component having an infinite dimensional nuisance parameter. See [12–15] for additional information. In both theory and practice, the semiparametric model brings a new perspective to data modeling, since it includes both parametric and nonparametric components. As mentioned in the previous section, it is well-suited to time series data, because it brings the advantages of the semiparametric model to time series analysis.

Suppose that a time series dataset $\{Z_t,\ \mathbf{x}_t,\ s_t,\ t\ =\ 1,2,\ldots,n\ \}$ satisfies an uncensored semiparametric time series model of the form:

$$Z_t\ =\ \mathbf{x}_t\boldsymbol{\beta} + f(s_t) + \varepsilon_t,\ \ a\ =\ s_1 < \ldots < s_n\ =\ b, \tag{1}$$

where Z_t's are the observations of stationary time series, $\mathbf{x}_t\ =\ \left(\mathbf{x}_{t1},\ldots,\mathbf{x}_{tp}\right)$ and $\mathbf{x}_1,\ldots,\mathbf{x}_n$ are known p-dimensional vectors of the explanatory variables, $\boldsymbol{\beta}\ =\ \left(\beta_1,\beta_2,\ldots,\beta_p\right)'$ is an unknown p-dimensional vector of the regression coefficients to be estimated, $f(.)$ is an unknown smooth function that describes the relationship between Z_t and a nonparametric temporal covariate s_t, and finally, ε_t's are the stationary autoregressive error terms generated by:

$$\varepsilon_t\ =\ \rho_1\varepsilon_{t-1} + \ldots + \rho_k\varepsilon_{t-k} + u_t, \tag{2}$$

where $\rho_1, \ldots, \rho_k$ are the autoregressive coefficients, and u_t denotes the independent and identically distributed random error terms with mean zero and a constant variance. Model (1) does not include lagged Z_t's and has auto-correlated errors. This expression makes it a suitable model for the semiparametric regression analysis of certain kinds of time series.

A common problem in practice is that dependent observations $Z_t's$ cannot be perfectly collected due to limitations including the detection limit of an evaluation tool or the end time for the study. To express this situation, we assume that $Z_t's$ are censored from the right by a non-negative random variable representing detection limit C_t. Therefore, instead of observing the values of Z_t, we now observe:

$$Y_t = \min(Z_t, C_t) \text{ and } \delta_t = \begin{cases} 1 \ if \ Z_t \leq C_t \ \text{(uncensored)} \\ 0 \ if \ Z_t > C_t \ \text{(censored)} \end{cases}, \tag{3}$$

where δ_t's denote the censoring information. Suppose that we are interested in estimating the mean semiparametric regression function. The distribution of the observable random variables does not identify the mean regression function uniquely. However, this problem can be solved as follows.

Let $F_Z(\alpha) = P(Z \leq \alpha)$, $G_C(\alpha) = P(C \leq \alpha)$, and $H_Y(\alpha) = P(Y \leq \alpha)$ for $\alpha \in \mathbb{R}$ be cumulative distribution functions of non-negative random variables Z_t, C_t, and Y_t, respectively. If random variables Z_t and C_t are independent, then the survival function $\overline{H}_Y(\alpha)$ for observed response variable Y_t can be defined from the basic relationship between F_Z and G_C:

$$\left\{\overline{H}_Y(\alpha) = 1 - H_Y(\alpha)\right\} = [(1 - F_Z(\alpha)) \cdot (1 - G_C(\alpha))]. \tag{4}$$

Given a random sample from the distribution of $(Y_t, X_t, s_t, \delta_t)$, it is of interest to examine the explanatory variables' effect on the observations of time series (i.e., response variable) by estimating the survival function $\overline{H}_Y(\alpha) = P(Y > \alpha)$, which is the regression function $E(Y_t|\mathbf{x}_t, s_t) = \mathbf{x}_t\boldsymbol{\beta} + f(s_t)$, the conditional mean of time series Y_t. However, because of the censoring, ordinary methods cannot be applied directly to estimate the regression function. To overcome censored observations, a data transformation technique should be used. One of the most widely used techniques is the synthetic data transformation, detailed in the section below.

Synthetic Data

To extend the penalized sum of squares approach to right-censored semiparametric regression analysis, we updated the synthetic data approach developed by [5]. The first step is to create an unbiased synthetic response variable of which the expectation is equal to the original and then to obtain the penalized squares estimator by means of this synthetic variable. The main goal of this transaction is to consider the censoring effect on the distribution of response variable. In the case of censored data, the authors of [16,17] used the synthetic data approach.

In the synthetic approach, we replace observed variable Y_t with transformed data Y_{tG}; a transformation maintains the conditional expectation of original variable Z_t. To describe this situation, it is easier to proceed directly using the cumulative distributions given in Lemma 1 below. Note also that if G_C is known then it is possible to transform observed data $\{(Y_t, \delta_t), t = 1, \ldots, n\}$ into unbiased synthetic data, given by:

$$Y_{tG} = \frac{\delta_t Y_t}{1 - G_C(Y_t)}, \tag{5}$$

where $G_C(.)$ is the distribution function of the censoring time C_t, as defined before. It should be noted that the distribution of G_C is rarely known. In this case, we use the Kaplan–Meier estimator defined by:

$$1 - \hat{G}_c(y) \;=\; \prod_{t=1}^{n} \left(\frac{n-t}{n-t+1} \right)^{\mathrm{I}[Y_{(t)} \leq y, \quad \delta_{(t)} = 0]}, \; y \geq 0, \tag{6}$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ are the sorted values of $Y_1, \ldots, Y_n$ and $\delta_{(t)}$ is the δ_t related to $Y_{(t)}$. Equation (5) has the following properties: (a) if distribution $\mathrm{G_C}$ is selected arbitrarily, some $Y_{(i)}$ can be identical. In this case, the ranking of $Y_1, \ldots, Y_n$ into $Y_{(1)} \ldots Y_{(n)}$ is not unique. However, the Kaplan–Meier estimator allows us to define the ranking of Y_t uniquely; (b) $\hat{G}_C(.)$ has jumps only at the censored observations of the time series (see [18]).

Substituting $\hat{G}_C(.)$ for $G_C(.)$ in Equation (5), we construct the following synthetic data, given by:

$$Y_{t\hat{G}} \;=\; \frac{\delta_t Y_t}{1 - \hat{G}_C(Y_t)}. \tag{7}$$

Then, one practical consequence of the following Lemma is that synthetic data $Y_{t\hat{G}}$ and completely observed response times Z_t have the same conditional expectations, as claimed in before.

Lemma 1. *Consider time series data Z_t denoted as a response variable. If the data is censored by random censoring variable C with distribution G_C, transform observed series $Y_t = min(Z_t, C_t)$ to $Y_{tG} = \delta_t Y_t / 1 - G_C(Y_t)$ in an unbiased form, as defined in Equation (4). Based on the information, it can be easily verified that $E[Y_{tG}|\boldsymbol{x}_t, s_t] = E[Z_t|\boldsymbol{x}_t, s_t] = \boldsymbol{x}_t\boldsymbol{\beta} + f(s_t)$. However, generally, G_C is unknown as mentioned before. Therefore, $Y_{t\hat{G}}$ is used which is defined in Equation (7), instead of Y_{tG}. Because of $\hat{G}_c \to G$ when $n \to \infty$, (see [5]), it is ensured that $E\left[Y_{t\hat{G}}|\boldsymbol{x}_t, s_t\right] \cong E[Y_{tG}|\boldsymbol{x}_t, s_t] = \boldsymbol{x}_t\boldsymbol{\beta} + f(s_t)$.*

Let us consider that $\tau_{H_Y} = \sup\{\alpha : H_Y(\alpha) < 1\}$, where $H_Y(.)$ is defined right after Equation (3). In the literature, the convergence rate of the Kaplan–Meier estimator is examined in two classes: (i) restriction of time-interval as $[0, \alpha]$ with $\alpha < \tau_{H_Y}$; (ii) extension of time-interval $\left[0, \tau_{H_Y}\right]$ (see [19] for more detailed discussions). Here, the convergence rate of the Kaplan–Meier estimator is inspected with regard to case (ii). However, $\left[0, \tau_{H_Y}\right]$ cannot be used without some strong conditions that can be given by:

(i) $G\left(\tau_{H_Y}\right) < 1 = F\left(\tau_{H_Y}\right)$;
(ii) $\tau_{H_Y} < \infty$;
(iii) $\int_0^{\tau_{H_Y}} \frac{1}{1-G(\alpha)} dF < \infty$.

Details about conditions (i)–(iii) were studied by [20]. The convergence of $\hat{G} \to G$ over the interval $\left[0, \tau_{H_Y}\right]$ can be provided. Reference [19] clearly shows both strong and weak convergences at the rate $n^{-\vartheta}$ where $0 \leq \vartheta \leq 1/2$.

The proof of Lemma 1 is given in Appendix A.

The major concern of this paper is to overcome the censoring problem and to estimate the semiparametric time series model efficiently. To achieve this goal, we used two different approaches, BS and modified AS estimators. In the following section, we applied these approaches to the transformed data to estimate time series observations under random right-censorship.

3. Estimating the Semiparametric Model Based on the BS Estimator

We first introduce the BS considered for estimating the components of model (1). A univariate B-spline is constructed by a piecewise polynomial function of degree q such that its derivatives up to order $(q-1)$ is continuous at each knot point $r_1, \ldots, r_k$. The set of BSs of degree q over the real numbers $(r_1, \ldots, r_k) = \boldsymbol{r}$ is a vector space of dimension $q + k + 1$. In addition, note that k denotes the number of interior knots, while $q \geq 0$ indicates the

polynomial order. For example, the polynomials of order $q = 0, 1, 2,$ and 3 are defined as constant, linear, quadratic, and cubic BS basis functions, respectively. If the knots are equally spaced (i.e., separated by same distance $h = (r_{k+1} - r_k)$), the knot points and the corresponding BSs are called uniform.

Definition 1. *Given an ordered knot vector* $\boldsymbol{r} = \{r_1 \leq r_2 \leq \ldots \leq r_k\}$ *in the domain of covariate* s_t*, then* i^{th} *BS basis functions* $\{B_{i,q}(s_t),\ i = 1, 2, \ldots, q+k+1\}$ *of degree* $q = 0$ *and* $q > 0$ *can be defined in recursive series, respectively, as:*

$$B_{i,0}(s) = \begin{cases} 1 & if \quad r_i \leq s \leq r_{i+1} \\ 0, & otherwise \end{cases}, \tag{8}$$

$$B_{i,q}(s) = \frac{s - r_i}{r_{i+q} - r_i} B_{i,q-1}(s) + \frac{r_{i+q+1} - s}{r_{i+q+1} - r_{i+1}} B_{i+1,q-1}(s). \tag{9}$$

Note that if the denominator of Equation (9) is equal to zero, then the BS basis function is assumed to be zero. From Equations (8) and (9), a set of $(q+k+1)$ *basis functions have the following important properties:*

(a) The BS basis functions form a partition of unity, $\sum_{i=1}^{q+k+1} B_{i,q}(s) = 1$*;*
(b) For all values of covariate s_t*,* $B_{i,q}(s) \geq 0$*; and*
(c) $B_{i,q}(s)$ *is realized in the interval* $[r_k,\ r_{k+q+1}]$*.*

Reference [21] proposes an algorithm to solve equation (9). See also the work of [22] for more detailed discussions on the BS approximation. Note also that the BS curve can be uniquely represented as a linear combination of the BSs basis functions in Equation (9), as given in the next section. Note that references [23,24] could be counted as recent studies about BSs.

3.1. BS Estimator

As previously noted, in this paper, we fit semiparametric time series model (1) with right-censored data. For this purpose, the BS estimator can be used as an approximation method. Using the synthetic data in Equation (7), we estimated the parametric and non-parametric components of model (1). Therefore, the sum of the squares of the differences between the censored time series values $Y_{t\hat{G}}$ and $(\mathbf{x}_t\boldsymbol{\beta} + f(s_t))$ are minimum. Assume that $f(.)$ is a smooth function that can be approximated by a linear combination of the BSs basis functions in Equations (8) and (9):

$$f(s) \cong \sum_{i=1}^{m=q+k+1} \alpha_i B_{i,q}(s) = \mathbf{B}\boldsymbol{\alpha}\,, \tag{10}$$

where $m = (q+k+1)$ is the total number of BS basis functions being used, $\hat{\alpha}_i{}'s$ are estimated coefficients (or control points) for each BS, $\mathbf{B}$ is an $(n \times m)$-dimensional matrix which includes BSs as defined by Equation (9) and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$ is a parameter vector of the BS function. Note also that the autoregressive errors in model (1) follow an n-dimensional multivariate normal distribution with a zero mean and stationary $(n \times n)$ covariance matrix $\boldsymbol{\Sigma}$, that is, $(\varepsilon_1, \ldots, \varepsilon_n)^T \sim N_n(\mathbf{0},\ \boldsymbol{\Sigma}\ )$, where the covariance matrix $\boldsymbol{\Sigma}$ is a symmetric and positive definite matrix with elements:

$$\boldsymbol{\Sigma} = \frac{\sigma_u^2}{1-\rho^2}\mathbf{R},\ \ R(t,j) = \rho^{|t-j|},\ \ 1 \leq (t,j) \leq n. \tag{11}$$

Throughout the paper, the notation is used as $\boldsymbol{\Sigma}^{-1} = \mathbf{V}$. Note that $\mathbf{V}$ is generally unknown. However, its elements can be obtained by the generalized least squares (GLS) based on an iterative process. Then, as in [25] which is a penalized BS study combining

BS and difference penalties, the estimates of the components of semiparametric model (1) were obtained by minimizing the penalized sum of squares (PSS) criterion:

$$PSS = \sum_{t=1}^{n} \mathbf{V}\Big\{Y_{t\hat{G}} - \sum_{j=1}^{p} x_{tj}\beta_j - \sum_{i=1}^{m} \alpha_i B_{i,q}(s)\Big\}^2 + \lambda \sum_{i=q+1}^{m} (\Delta^q \alpha_i)^2, \quad (12)$$

where $\Delta\alpha_i = (\alpha_i - \alpha_{i-1})$ is the first-order difference penalty on the coefficients of the BSs. The other differences can be defined as follows:

$$\Delta^2\alpha_i = \Delta(\Delta\alpha_i) = (\alpha_i - \alpha_{i-1}) - (\alpha_{i-1} - \alpha_{i-2}) = \alpha_i - 2\alpha_{i-1} + \alpha_{i-2}, \quad (13)$$

and similarly:

$$\Delta^q\alpha_i = \Delta\Big(\Delta^{q-1}\alpha_i\Big). \quad (14)$$

Note that if degree $q = 0$ in Equation (12), we obtain semiparametric ridge regression based on BSs. When $\lambda = 0$ in Equation (12), we have the minimization equation of ordinary least squares regression with a correlated error. If $\lambda > 0$, the penalty only influences the main diagonal and q sub-diagonals (on both sides of the main diagonal elements) of the banded structure system due to the limited overlap of the BSs.

We rewrite the minimization criterion described as Equation (12) in a matrix and vector notation:

$$PSS = \left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha}\right)' \mathbf{V} \left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha}\right) + \lambda\|\mathbf{D}\boldsymbol{\alpha}\|^2, \quad (15)$$

where $\|.\|$ denotes Euclidean norm, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, $\mathbf{Y}_{\hat{G}} = \left(Y_{1\hat{G}}, \ldots, Y_{t\hat{G}}\right)'$ is the synthetic response vector defined in Equation (7), $\lambda > 0$ is a smoothing parameter, and $\mathbf{D}$ denotes the matrix notation of the difference operator (Δ^q) defined in Equation (13). For example, $\mathbf{D}$ is an $(n-2) \times n$-dimensional banded matrix that corresponds to the second-order difference penalty, given by:

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -2 & 1 \end{bmatrix}. \quad (16)$$

From simple algebraic operations, it follows that the solution to the minimization problem in Equation (15) satisfies the following block matrix equation:

$$\begin{pmatrix} \mathbf{X}'\mathbf{V}\,\mathbf{X} & \mathbf{X}'\mathbf{V}\,\mathbf{B} \\ \mathbf{B}'\mathbf{V}\,\mathbf{X} & \left(\mathbf{B}'\mathbf{V}\,\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{X}' \\ \mathbf{B}' \end{pmatrix} \mathbf{V}\mathbf{Y}_{\hat{G}}. \quad (17)$$

Given a parameter $\lambda > 0$, the corresponding estimators based on BSs for vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be easily obtained by:

$$\hat{\boldsymbol{\alpha}}_{BS} = \left[\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right]^{-1} \mathbf{B}'\mathbf{V}\left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{BS}\right), \quad (18)$$

and:

$$\hat{\boldsymbol{\beta}}_{BS} = \left[\left(\mathbf{X}'\mathbf{V} - \boldsymbol{A}_{BS}\right)\mathbf{X}\right]^{-1} \left(\mathbf{X}' - \boldsymbol{A}_{BS}\right) \mathbf{V}\mathbf{Y}_{\hat{G}}, \quad (19)$$

where $\boldsymbol{A}_{BS} = \mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right]^{-1}\mathbf{B}'\mathbf{V}$. It should be noted that the estimates of the unknown regression function in a censored semiparametric model are obtained by:

$$\hat{\mathbf{f}}_{BS} = \mathbf{B}\hat{\boldsymbol{\alpha}}_{BS} = \left[\hat{f}(s_1), \ldots, \hat{f}(s_n)\right]'. \quad (20)$$

From Equations (19) and (20), we see that the fitted values of dependent time series data can be written as:

$$\hat{\boldsymbol{\mu}}_{BS} \;=\; \left(\mathbf{X}\hat{\boldsymbol{\beta}}_{BS} + \hat{\mathbf{f}}_{BS}\right) \;=\; \mathbf{H}_{BS}\mathbf{Y}_{\hat{G}} \;=\; E[Y \mid X,\, s], \tag{21}$$

where $\mathbf{H}_{BS}$ is a hat matrix for BSs and computed as follows:

$$\mathbf{H}_{BS} \;=\; \left[\mathbf{X}\left[\left(\mathbf{X}^{'}\mathbf{V} - \boldsymbol{A}_{BS}\right)\mathbf{X}\right]^{-1}\left(\mathbf{X}^{'} - \boldsymbol{A}_{BS}\right)\mathbf{V}(\mathbf{I} - \mathbf{M}_{BS}) + \mathbf{M}_{BS}\right], \tag{22}$$

where $\mathbf{M}_{BS} \;=\; \mathbf{B}\left[\mathbf{B}^{'}\mathbf{V}\mathbf{B} + \lambda\mathbf{D}^{'}\mathbf{D}\right]^{-1}\mathbf{B}^{'}\mathbf{V}$.

3.2. AS Estimator

The adaptive spline (AS) applies an adaptive ridge penalty to the BS method, which makes it more flexible for knot determination. The AS concept is explained in [26] in a nonparametric context. However, in this paper, we generalized this estimation concept to the semiparametric environment based on synthetic response observations. It should be noted that the location and number of knots have crucial importance in terms of synthetic data transformation. This issue is discussed in detail in Section 4.3. The point here is that a more efficient estimator based on synthetic responses is needed, as most of the existing smoothing techniques (spline smoothing, kernel smoothing, etc.) cannot properly handle synthetic data. This article aims to solve this issue with the AS estimator.

When a BS is defined on the knots $r_1 \leq r_2 \leq \ldots \leq r_k$ such that $\Delta^q\alpha_i \;=\; 0$ for some i^{th} knot, it may be reparametrized as a BS on the knots $r_1, r_2, \ldots, r_{i-1}, r_{i+1}, \ldots, r_k$. Accordingly, when $m \;=\; (q + k + 1)$, we want to put a penalty on the number of non-zero differences indicated as below:

$$\lambda \sum_{i\,=\,q+1}^{m} \left\|\Delta^q\alpha_j\right\|_0, \tag{23}$$

where $\Delta^q\alpha_i$ is the q^{th}-order difference operator and $\left\|\Delta^q\alpha_i\right\|_0$ is the L_0-norm of the differences, that is, $\left\|\Delta^q\alpha_i\right\|_0 = 0$ if $\Delta^q\alpha_j \;=\; 0$, otherwise, $\left\|\Delta^q\alpha_i\right\|_0 = 1$, and λ is a positive penalty parameter that ensures the tradeoff between the goodness of fit to the data and the smoothness of the fitted curve. This penalty enables us to remove knot r_i that is not related to the smoothing problem, to join the neighbor intervals $[r_{i-1}, r_i)$ and $[r_i,\ r_{i+1})$, and to carry on fitting with a BS described over the remaining knot points. Note also that when $\lambda \to 0$, the fitted curve becomes a BS with knots $r_i, i \;=\; 1,\ 2, \ldots, k$ and when $\lambda \to \infty,$ the fitted function becomes a polynomial of degree q.

It should be emphasized that one of the important points about the adaptive ridge penalty is that Equation (23) cannot be differentiated due to the L_0-norm. As a result, the fitting process is made numerically untraceable. An approximate solution to dealing with the L_0-norm is provided by [27,28]. Following the studies of these authors, we approximate the L_0-norm by using an iterative process referred to as an "adaptive ridge" based on synthetic data. The new criterion function is expressed by the following weighted penalized sum of squares:

$$WPSS \;=\; \left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha}\right)^{'}\mathbf{V}\left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha}\right) + \lambda\textstyle\sum_{i\,=\,q+1}^{m} w_i(\,\Delta^q\alpha_i)^2, \tag{24}$$

where w_i's denote the positive weights. It should be noted that the penalty is close to the L_0-norm of the differences when the weights are iteratively calculated from the parameter vector α of BS following the equation:

$$w_i \;=\; \left[(\Delta^q\alpha_i)^2 + \gamma^2\right]^{-1},\ \ \gamma > 0, \tag{25}$$

where γ is a constant properly determined by the researcher.

Remark 1. *There are a few important points to know about the selection of* γ. *If* $(\Delta^q \alpha_i) < \gamma$, *then the magnitudes of* w_i*'s might be quite large, resulting in* $(\Delta^q \alpha_i) \cong 0$ *and the penalty term turning into* $w_i(\Delta^q \alpha_i)^2 \cong 0$. *Furthermore, if* $(\Delta^q \alpha_i) \gg \gamma$, *then* $w_i(\Delta^q \alpha_i)^2 \cong \|\Delta^q \alpha_i\|_0$. *This convergence gives us a measure of how relevant the* i^{th} *knot point is. In practice, one possible choice, suggested by [28], is* $\gamma = 10^{-5}$. *They select the knots (denoted as* r_{i^*}*) with a weighted difference bigger than 0.99. The number of parameters of the chosen BS is* $m_\lambda = q + k_\lambda + 1$, *where* k_λ *denotes the number of selected knot points.*

Note that reference [28] provides a figure to show the effects of different norm degrees (q) on the quality of estimation. It is seen from that the performance of estimation does not change for different values of γ when norm degree is zero ($q = 0$). However, it affects the performance seriously if $q > 0$.

For some $\lambda > 0$ and non-negative weights, the $WPSS$ of Equation (26) can be rewritten as:

$$WPSS = (\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha})' \mathbf{V} (\mathbf{Y}_{\hat{G}} - \mathbf{X}\boldsymbol{\beta} - \mathbf{B}\boldsymbol{\alpha}) + \lambda \boldsymbol{\alpha}' \mathbf{K} \boldsymbol{\alpha}, \tag{26}$$

where $\mathbf{K}$ is a penalty matrix and written as $\mathbf{K} = \mathbf{D}'\mathbf{W}\mathbf{D}$, where $\mathbf{W} = \text{diag}(w_{q+1}, \ldots, w_m)$ and $\mathbf{D}$ is the matrix form of the difference operator Δ^q, as defined in Equation (13). Simple algebraic operations show that the solution to the minimization problem $WPSS$ in Equation (26) satisfies the block matrix equation:

$$\begin{pmatrix} \mathbf{X}'\mathbf{V}\mathbf{X} & \mathbf{X}'\mathbf{V}\mathbf{B} \\ \mathbf{B}'\mathbf{V}\mathbf{X} & \left(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda \mathbf{K}\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{X}' \\ \mathbf{B}' \end{pmatrix} \mathbf{V}\mathbf{Y}_{\hat{G}}. \tag{27}$$

By similar arguments as in the case of the BS approach, the corresponding estimators $\hat{\boldsymbol{\alpha}}_{AS}$ and $\hat{\boldsymbol{\beta}}_{AS}$ of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, based on the right-censored semiparametric time series model (1) with correlated data, can be easily obtained, respectively, as:

$$\hat{\boldsymbol{\alpha}}_{AS} = \left[\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda \mathbf{K}\right]^{-1} \mathbf{B}'\mathbf{V}' \left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{AS}\right), \tag{28}$$

and:

$$\hat{\boldsymbol{\beta}}_{AS} = \left((\mathbf{X}'\mathbf{V} - \boldsymbol{A}_{AS})\mathbf{X}\right)^{-1} \left(\mathbf{X}' - \boldsymbol{A}_{AS}\right) \mathbf{V}\mathbf{Y}_{\hat{G}}, \tag{29}$$

where $\boldsymbol{A}_{AS} = \mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{K}\right]^{-1}\mathbf{B}'\mathbf{V}'$. The proofs and derivations of Equations (28) and (29) are given in Appendix B. Notice that the estimates corresponding to the nonparametric part of the semiparametric model (1) are obtained using Equation (28) as described in the following equation:

$$\hat{\mathbf{f}}_{AS} = \mathbf{B}\hat{\boldsymbol{\alpha}}_{AS} = \left[\hat{f}(s_1), \ldots, \hat{f}(s_n)\right]'. \tag{30}$$

From Equations (29) and (30), we can see that the fitted values of the dependent time series data can be obtained as:

$$\hat{\boldsymbol{\mu}}_{AS} = \left(\mathbf{X}\hat{\boldsymbol{\beta}}_{AS} + \hat{\mathbf{f}}_{AS}\right) = \mathbf{H}_{AS}\mathbf{Y}_{\hat{G}} = E[Y|X, s], \tag{31}$$

where $\mathbf{H}_{AS}$ denotes the hat matrix, given by:

$$\mathbf{H}_{AS} = \left[\mathbf{X}\left[\left(\mathbf{X}'\mathbf{V} - \boldsymbol{A}_{AS}\right)\mathbf{X}\right]^{-1}\left(\mathbf{X}' - \boldsymbol{A}_{AS}\right)\mathbf{V}(\mathbf{I} - \mathbf{M}_{AS}) + \mathbf{M}_{AS}\right], \tag{32}$$

with $\mathbf{M}_{AS} = \mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{K}\right]^{-1}\mathbf{B}'\mathbf{V}'$.

To make the computation process efficient, all penalty terms $\left(\mathbf{D}^T\mathbf{W}\mathbf{D}\right)$ are calculated by using the iteration process instead of finding matrix $\mathbf{D}$ and knot set individually. The iterative algorithm is given in Algorithm 1 below.

Algorithm 1. Iterative algorithm process for the modified A-spline (AS) estimator $\hat{\alpha}_{AS}$.

Input: **X, s,** $\mathbf{Y}_{\hat{G}}$.

Output: $\hat{\boldsymbol{\beta}}_{\boldsymbol{AS}}^{(\boldsymbol{i})} = \left(\hat{\beta}_1^{(i)}, \hat{\beta}_2^{(i)}, \ldots, \hat{\beta}_p^{(i)}\right) \qquad \hat{\boldsymbol{\alpha}}_{\boldsymbol{AS}}^{(\boldsymbol{i})} = \left(\hat{\alpha}_1^{(i)}, \hat{\alpha}_2^{(i)}, \ldots, \hat{\alpha}_{q+k+1}^{(i)}\right)'$

1: **Begin**

2: Give initial values, $\boldsymbol{\beta}^{(0)} = 1_p$, $\boldsymbol{\alpha}^{(\mathbf{0})} = \mathbf{0}_{q+k+1}$ and $\mathbf{W}^{(0)} = \mathbf{I}$ to start iterative process

3: **do** until converges weighted differences to L_0-norm

4: $\hat{\boldsymbol{\beta}}_{\boldsymbol{AS}}^{(\boldsymbol{i})} = (\left(\mathbf{X}'\mathbf{V} - \boldsymbol{A}\right)\mathbf{X})^{-1}\left(\mathbf{X}' - \mathbf{A}\right)\mathbf{V}\mathbf{Y}_{\hat{G}}$

5: $\hat{\boldsymbol{\alpha}}_{\boldsymbol{AS}}^{(\boldsymbol{i})} = \left[\mathbf{B}'\mathbf{V}\mathbf{B} + \boldsymbol{\lambda}\boldsymbol{K}\right]^{-1}\mathbf{B}'\mathbf{V}'\left(\mathbf{Y}_{\hat{G}} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{AS}}^{(\boldsymbol{i})}\right)$

6: Determine $\gamma = 10^{-5}$

7: $w_i^{(i)} = \left[\left(\Delta^q \alpha_i^{(i)}\right)^2 + \gamma^2\right]^{-1}$

8: $\hat{\boldsymbol{\beta}}_{\boldsymbol{AS}} = \boldsymbol{\beta}_{\boldsymbol{AS}}^{(\boldsymbol{i})}$, $\hat{\boldsymbol{\alpha}}_{\boldsymbol{AS}} = \hat{\boldsymbol{\alpha}}_{\boldsymbol{AS}}^{(\boldsymbol{i})}$, $\mathbf{W} = diag\left(w_i^{(i)}\right)$

9: **end**

10: Calculate $\mathbf{r}_{(i^*)}$ by the criterion of $\left(\Delta^q \boldsymbol{\alpha}_{AS}^{(i)}\right)^2 \mathbf{W}^{(i)} > 0.99$

11: Return $\hat{\boldsymbol{\beta}}_{AS}^{(i)} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)$, $\hat{\boldsymbol{\alpha}}_{AS}^{(i)} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{q+k+1}\right)'$

12: **End**

Remark 2. *For the constant value of* $\gamma = 10^{-5}$, the iteration process repeats between step 3 and step 9 until the pre-determined tolerance value $\delta = 10^{-4}$ is obtained where $\delta = \sum_{i=1}^{n} n^{-1}\left|Y_i - \hat{Y}_{i\hat{G}}\right|$. *From our experience, the expected number of iterations is observed as* $no.iteration = 20$ *to achieve the convergence.*

Notice that the complexity and efficiency of Algorithm 1 is analyzed from different aspects that are given by:

(i) Number of local searches: algorithm does not involve a local search procedure which is an advantage for the speed of Algorithm 1;

(ii) Number of nested loops: due to the fact that there is only an iteration loop (without nested loops), if an algorithm does not include nested loops, its "order of growth" will be $O(n)$;

(iii) Asymptotic behaviors: as the former inference mentioned, Algorithm 1 has $O(n)$ which means that the limiting case of its convergence speed is considerable when it is compared with its alternative BS method on this issue.

As mentioned at the beginning of this section, the choice of an optimum smoothing parameter λ is required for both semiparametric BS and AS estimators. In this context, the improved Akaike information criterion (AIC_c) proposed by [29] is used, which is computed with the following equation:

$$AIC_c(\lambda) = \log\left(\hat{\sigma}^2\right) + 1 + \frac{2\{tr(\mathbf{H}) + 1\}}{n - tr(\mathbf{H}) - 2}, \tag{33}$$

where $\hat{\sigma}^2$ is the estimate of the model variance, which is estimated for both methods separately in the next section, and $\mathbf{H}$ denotes the hat matrix for any of two methods. It is replaced by $\mathbf{H}_{AS}$ for the AS method and $\mathbf{H}_{BS}$ for the BS method, respectively.

4. Statistical Properties of the Estimators

In this paper, we introduced the semiparametric AS and BS estimators for the estimation of the right-censored time series model. It should be noted that these two methods were used for the first time in the setting of a time series estimation procedure. Inferences were therefore carried out about their statistical properties. For example, among these, the error terms obtained from the estimates of both methods and the estimators of parametric and nonparametric components were inspected and their properties were extracted.

4.1. Properties of the Semiparametric BS Estimator

Firstly, the parametric component was inspected. As is known, in a parametric context, errors can be decomposed into the bias and the variance terms that provide the quality of the estimator. Accordingly, the estimator $\hat{\boldsymbol{\beta}}_{BS}$ of the parametric coefficients vector is expanded as follows:

$$\hat{\boldsymbol{\beta}}_{BS} = \left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{Y}_{\hat{G}} = \boldsymbol{\beta} + \left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{f}, \quad (34)$$

where $\mathbf{V}$, $\mathbf{A}_{BS}$ and $\mathbf{M}_{BS}$ matrices are as defined in Section 3.1 and $\mathbf{f} = [f(s_1), f(s_2), \ldots, f(s_n)]'$. From here, bias $B(\hat{\boldsymbol{\beta}}_{BS})$ and variance-covariance $V(\hat{\boldsymbol{\beta}}_{BS})$ of estimator $\hat{\boldsymbol{\beta}}_{BS}$ can be computed as follows:

$$B(\hat{\boldsymbol{\beta}}_{BS}) = E(\hat{\boldsymbol{\beta}}_{BS}) - \boldsymbol{\beta} = \left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{f}, \quad (35)$$

$$V(\hat{\boldsymbol{\beta}}_{BS}) = \sigma^2\left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}\right]^{-1}, \quad (36)$$

where σ^2 is the variance of the fitted semiparametric model. Since the variance is not generally known, instead of σ^2, the estimation (denoted by $\hat{\sigma}^2_{BS}$) based on the BS is used. It can be computed from the residuals sum of squares (RSS) using error terms:

$$\hat{\sigma}^2_{BS} = \frac{RSS}{tr(\mathbf{I} - \mathbf{H}_{BS})^2} = \frac{\|(\mathbf{I} - \mathbf{H}_{BS})\hat{Y}_{\hat{G}_{BS}}\|^2}{tr\left[(\mathbf{I} - \mathbf{H}_{BS})'(\mathbf{I} - \mathbf{H}_{BS})\right]}, \quad (37)$$

where $tr(\mathbf{I} - \mathbf{H}_{BS})^2 = n - 2tr(\mathbf{H}_{BS}) + tr(\mathbf{H}'_{BS}\mathbf{H}_{BS})$ denotes the degrees of freedom. In addition, $tr(\mathbf{H}'_{BS}\mathbf{H}_{BS})$ needs $O(n)$ algebraic operations. In the context of the BS, if the data have a normal distribution, $\hat{\sigma}^2_{BS}$ is asymptotically unbiased.

Secondly, the properties of estimated nonparametric component $\hat{\boldsymbol{\alpha}}_{BS} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{q+k+1}\right)'$ are given here. The bias of $\hat{\boldsymbol{\alpha}}$ is one of the quality measurements for the estimated model. The bias is denoted as conditional expectation $E[\hat{\boldsymbol{\alpha}}|s_t]$, given by:

$$E[\hat{\boldsymbol{\alpha}}_{BS}|s_t] = \left(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{B}'\mathbf{V}\mathbf{B}\boldsymbol{\alpha}. \quad (38)$$

From that, the bias is given by:

$$\begin{gathered} Bias(\hat{\boldsymbol{\alpha}}_{BS}) = E[\hat{\boldsymbol{\alpha}}_{BS}|s_t] - \boldsymbol{\alpha} = [(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{f} - [(\mathbf{B}'\mathbf{V}\mathbf{B} + \\ \lambda\mathbf{D}'\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS}) - [(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}' = \\ [(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{BS}). \end{gathered} \quad (39)$$

Accordingly, the covariance of $\hat{\boldsymbol{\alpha}}_{BS}$ can be computed as:

$$Cov(\hat{\boldsymbol{\alpha}}_{BS}) = \hat{\sigma}^2_{BS}\frac{1}{n}\left(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right)^{-1}\left(\mathbf{B}'\mathbf{V}\mathbf{B}\right)\left(\mathbf{B}'\mathbf{V}\mathbf{B} + \lambda\mathbf{D}'\mathbf{D}\right)^{-1}, \quad (40)$$

where $\hat{\sigma}^2_{BS}$ is defined by Equation (36). In addition, to reveal the performance of $\hat{\mathbf{f}}_{BS} = \mathbf{B}\hat{\boldsymbol{\alpha}}_{BS}$, the root square of mean squared error $RMSE\left(\mathbf{f}, \hat{\mathbf{f}}_{BS}\right)$ is used:

$$RMSE(\mathbf{f}, \hat{\mathbf{f}}_{BS}) = n^{-1}\sum_{t=1}^{n}\left[f(s_t) - \hat{f}_{BS}(s_t)\right]^2 = n^{-1}(\mathbf{f} - \hat{\mathbf{f}}_{BS})'(\mathbf{f} - \hat{\mathbf{f}}_{BS}). \quad (41)$$

4.2. Properties of the Semiparametric AS Estimator

Similar to in Section 4.1, the same properties for parametric and nonparametric components are given for the AS estimator here. The necessary expansion is written as follows to derivate the bias and variance of $\hat{\boldsymbol{\beta}}_{AS}$:

$$\hat{\boldsymbol{\beta}}_{AS} = [(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{Y}_{\hat{G}} = \boldsymbol{\beta}+[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{f}, \quad (42)$$

where $\mathbf{A}_{AS}$ and $\mathbf{M}_{AS}$ are given in Section 3.2. Now, the bias and the covariance matrix of the estimator $\hat{\boldsymbol{\beta}}_{AS}$ can be provided by:

$$B(\hat{\boldsymbol{\beta}}_{AS}) = E(\hat{\boldsymbol{\beta}}_{A})-\boldsymbol{\beta} = \left[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{f}, \quad (43)$$

$$V(\hat{\boldsymbol{\beta}}_{AS}) = \sigma^2\left[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}\left[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}\right]^{-1}, \quad (44)$$

where σ^2 is the variance of the fitted semiparametric model. Similar to Equation (40), instead of the model variance, $\hat{\sigma}^2_{AS}$ is obtained as follows:

$$\hat{\sigma}^2_{AS} = \frac{RSS}{tr(\mathbf{I}-\mathbf{H}_{AS})^2} = \frac{\|(\mathbf{I}-\mathbf{H}_{AS})\hat{\mathbf{Y}}_{\hat{G}_{AS}}\|^2}{tr\left[(\mathbf{I}-\mathbf{H}_{AS})'(\mathbf{I}-\mathbf{H}_{AS})\right]}. \quad (45)$$

The properties of estimated nonparametric component $\hat{\boldsymbol{\alpha}}_{AS} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{q+k+1}\right)'$ for the AS method are described below. The bias and the variance of the AS estimator $\hat{\boldsymbol{\alpha}}_{AS}$ can be given, respectively, as:

$$\begin{aligned} Bias(\hat{\boldsymbol{\alpha}}_{AS}) &= E[\hat{\boldsymbol{\alpha}}_{AS}|s_t]-\boldsymbol{\alpha} = [(\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda\mathbf{D}'\mathbf{W}\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{f}-[(\mathbf{B}'\mathbf{V}\mathbf{B}+ \\ &\lambda\mathbf{D}'\mathbf{W}\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})-[(\mathbf{B}'\mathbf{V}\mathbf{B}+ \\ &\lambda\mathbf{D}'\mathbf{W}\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{f} = [(\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda\mathbf{D}'\mathbf{W}\mathbf{D})]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}[(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V}-\mathbf{A}_{AS}), \end{aligned} \quad (46)$$

and

$$Cov(\hat{\boldsymbol{\alpha}}_{AS}) = \hat{\sigma}^2_{AS}\frac{1}{n}\left(\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda\mathbf{D}'\mathbf{W}\mathbf{D}\right)^{-1}\left(\mathbf{B}'\mathbf{V}\mathbf{B}\right)\left(\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda\mathbf{D}'\mathbf{W}\mathbf{D}\right)^{-1}. \quad (47)$$

Thus, the value of $RMSE\left(\mathbf{f}, \hat{\mathbf{f}}_{AS}\right)$ for $\hat{\mathbf{f}}_{AS} = \mathbf{B}\hat{\boldsymbol{\alpha}}_{AS}$, similar to Equation (41), is calculated as follows:

$$RMSE(\mathbf{f}, \hat{\mathbf{f}}_{AS}) = n^{-1}\sum_{t=1}^{n}\left[f(s_t)-\hat{f}_{AS}(s_t)\right]^2 = n^{-1}(\mathbf{f}-\hat{\mathbf{f}}_{AS})'(\mathbf{f}-\hat{\mathbf{f}}_{AS}). \quad (48)$$

4.3. Quality Measures for the Fitted Model

After assessing the parametric and nonparametric components of the model in Sections 4.1 and 4.2, several measurements are introduced in this section to evaluate the overall model performance. In the literature on time series modelling, mean absolute percentage error ($MAPE$), mean absolute error (MAE), and mean squared error (MSE) are the most commonly used performance criteria. To represent these criteria, $MAPE$ is preferred in this study. In addition, median absolute error ($MedAE$) was used, which allowed us to account for missing or censored data. Generalized MSE ($GMSE$) and the ratio of $GMSE$ ($RGMSE$) proposed by [30] and [2], respectively, were used to measure the quality of the fitted time series model. The aforementioned criteria can be defined as follows:

$$MAPE(Y_{t\hat{G}}, \hat{Y}_{t\hat{G}}) = \frac{n^{-1}\sum_{t=1}^{n}|Y_t-\hat{Y}_{t\hat{G}}|}{Y_{t\hat{G}}}, \quad MedAE(\mathbf{Y}_{\hat{G}}, \hat{\mathbf{Y}}_{\hat{G}}) = Median(|\mathbf{Y}_{\hat{G}}-\hat{\mathbf{Y}}_{\hat{G}}|),$$

$$GMSE(\mathbf{Y}_{\hat{G}}, \hat{\mathbf{Y}}_{\hat{G}}) = (\mathbf{Y}_{\hat{G}}-\hat{\mathbf{Y}}_{\hat{G}})'E(\mathbf{Y}_{\hat{G}}\mathbf{Y}'_{\hat{G}})(\mathbf{Y}_{\hat{G}}-\hat{\mathbf{Y}}_{\hat{G}}),$$

where $\hat{Y}_{t\hat{G}}$ and $\hat{\mathbf{Y}}_{\hat{G}}$ denote the fitted dependent variable values and vector for any estimation method. Here, $\hat{Y}_{t\hat{G}}$ and $\hat{\mathbf{Y}}_{\hat{G}}$ are replaced by $\hat{Y}_{t\hat{G}_{BS}}$ and $\hat{\mathbf{Y}}_{\hat{G}_{BS}}$ for the BS, and $\hat{Y}_{t\hat{G}_{A}}$ and $\hat{\mathbf{Y}}_{\hat{G}_{A}}$ for the AS. In addition, to make a more considerable comparison between the AS and BS estimators, $RGMSE$ is defined below.

Definition 2. *The ratio of GMSE can be defined as follows:*

$$RGMSE(\mathbf{Y}_{\hat{G}_{BS}}, \hat{\mathbf{Y}}_{\hat{G}_{AS}}) = \frac{GMSE(\hat{\mathbf{Y}}_{\hat{G}_{AS}})}{GMSE(\hat{\mathbf{Y}}_{\hat{G}_{BS}})}. \tag{49}$$

Regarding the RGMSE criterion, if RGMSE $(\mathbf{Y}_{\hat{G}_{BS}}, \hat{\mathbf{Y}}_{\hat{G}_{AS}}) < 1$, then it can be seen that the fitted model by the AS method shows better performance then the BS method.

5. Further Information for Adaptive-Ridge Penalty

The semiparametric AS estimator proposed for the right-censored time series model, with its adaptive nature, aims for qualified estimations despite the censorship. To approach the L_0-norm given in Equation (23), the most suitable knot locations can be chosen due to the weighted penalty term. Thus, the model avoids the disadvantages of synthetic data transformation, which gives higher magnitudes to uncensored observations.

This section is designed to inspect some of the large sample properties of the modified AS estimator under right-censored data. It should be noted that adaptive ridge penalty in the setting of regression has been studied by many authors; see for example [25,26,28]. However, the aforementioned studies consider adaptive ridge penalty individually, not as a part of a semiparametric time series model. This section provides basic information for the large sample properties of the proposed AS estimator in the context of a semiparametric time series model.

As previously stated, the AS approximation is a modified version of the P-splines (penalized BSs) estimator proposed by [31]. Note also that the AS method diverges from BSs with a significant difference of the L_0-norm in the penalty term. The AS estimator is obtained by an iterative process with determining weights, as expressed in Section 3.2. In addition, apart from the usage of the AS method in the literature, it is also used for modelling censored time series. For these reasons, we can make several important assumptions. The large sample properties are written based on the assumptions given below:

Assumption 1. *The minimization problem for the semiparametric AS is given in Equation (26). To make this expression more general, it can be rewritten as follows:*

$$PSS(\alpha;\lambda) = \sum_{t=1}^{n} \boldsymbol{V}\left\{ Y_{t\hat{G}} - \sum_{l=1}^{p} x_{tl}\beta_l - \sum_{j=1}^{v} \alpha_j B_{j,q}(s_t) \right\}^2 + \lambda \sum_{j=q+1}^{q+k+1} \|\Delta^q \alpha_j\|_\tau, \tag{50}$$

where $\|\Delta^q \alpha_j\|_\tau$ *represents the* τ*-norm of the penalty term. The first assumption is* $\tau \to 0$*, which allows approximation to the* L_0*-norm with the acquisition of weights via the iterative process. Otherwise, the* L_0*-norm needs overly complex calculations, which leads to the loss of practicality when using the method. From our knowledge of the literature, when* $\tau \to 0$*, such as in Equation (26), the minimization of Equation (50) works by penalizing the non-zero coefficients* α_j*'s, as shown by [32].*

Assumption 2. *When* $\hat{\boldsymbol{\alpha}}_{AS}$ *is examined asymptotically, the objective function of Equation (26) may not have a global minimum, since it is not clearly convex. However, if we assume that:*

$$\boldsymbol{R}_n = \frac{1}{r}\sum_i^r \boldsymbol{B}_i \boldsymbol{B}_i^{'} \to \boldsymbol{R}, \tag{51}$$

then it is possible to point out some important aspects of asymptotic consistency. Therefore, it should be presumed that $\boldsymbol{R}$ *is a non-negative definite matrix and:*

$$\frac{1}{q+k+1} \underset{1\le i\le r}{max} \boldsymbol{B}_i' \boldsymbol{B}_i \to 0, \tag{52}$$

where elements of $diag(R_i) = 1$.

Assumption 3. *$\mathbf{B}_j^T\mathbf{B}_j$, $\left(\mathbf{B}_j^T\mathbf{B}_j\right)^{-1}$, and $\mathbf{R}$ are assumed to be full rank matrices. Under the assumptions given above, to see asymptotic consistency of $\hat{\boldsymbol{\alpha}}_{AS}$ and $\hat{\boldsymbol{\beta}}_{AS}$, an equation can be obtained from Equation (50) as follows:*

$$M_n\left(\hat{\boldsymbol{\alpha}}_{ASn}, \hat{\boldsymbol{\beta}}_{ASn}\right) = \sum_{t=1}^{n} \begin{array}{c} \mathbf{V}\left\{Y_{t\hat{G}} - \sum_{l=1}^{p} x_{tl}\hat{\beta}_{ASnl} - \sum_{j=1}^{r} \hat{\boldsymbol{\alpha}}_{ASnj}B_{j,q}(s_t)\right\}^2 \\ +\lambda_n \sum_{i=q+1}^{q+k+1} \left\|\Delta^q \hat{\boldsymbol{\alpha}}_{ASn_i}\right\|_\tau \end{array}, \quad (53)$$

where $\left(\hat{\boldsymbol{\alpha}}_{ASn}, \hat{\boldsymbol{\beta}}_{ASn}\right)$ denote the limiting case of the estimators for $\lambda_n = O(n)$. Note that Equation (52) is ensured by following Theorem 1.

Theorem 1. *Based on Assumptions 1–3, and $\lambda_n \to \lambda \geq 0$, then $\left(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{\boldsymbol{\alpha}}_{AS_n}\right) \xrightarrow{d} argmin(M_n)$ where:*

$$M_n\left(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{\boldsymbol{\alpha}}_{AS_n}\right) = \left[\left(\hat{\boldsymbol{\beta}}_{AS_n} \ \hat{\boldsymbol{\alpha}}_{AS_n}\right)' - (\boldsymbol{\beta} \ \boldsymbol{\alpha})'\right]' \mathbf{R}\left[\left(\hat{\boldsymbol{\beta}}_{AS_n} \ \hat{\boldsymbol{\alpha}}_{AS_n}\right)' - (\boldsymbol{\beta} \ \boldsymbol{\alpha})'\right] + \lambda_n \sum_{i=q+1}^{m=q+k+1} \left\|\Delta^q \alpha_i\right\|_\tau. \quad (54)$$

Therefore, for optimal $\lambda_n = O(1)$, pair $(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{\boldsymbol{\alpha}}_{AS_n})$ can be counted as a consistent AS estimator of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. In this context, when $n \to \infty$ then $|\hat{\boldsymbol{\beta}}_{AS_n}, \hat{\boldsymbol{\alpha}}_{AS_n}| \to (\boldsymbol{\beta}, \boldsymbol{\alpha})$.

For the proof of Theorem 1, see Appendix C.

To clearly indicate the place of Assumptions 1–3 in the estimation process, the following explanations are given for each assumption.

- Assumption 1 is independent from the data. We assume that to provide a practical solution when minimizing Equation (50). Therefore, in both empirical and real data studies, this assumption does not impose anything to the dataset, but it is necessary to reduce the computational complexity.
- In real data studies, to ensure Assumption 2, **"B"** matrix obtained by using the non-parametric covariate needs to have independent columns. Because $\left(\mathbf{B}'\mathbf{B}\right)$ should be identifiable and avoid the ill-posed problem, $\left(\mathbf{B}'\mathbf{B}\right)$ must be a full-ranked matrix.
- Assumption 3 confirms Assumption 2. Thus, it can be seen that asymptotic consistency can be confirmed by Assumption 3. From that it can be said that Assumption 3 is indirectly depended on the dataset.

5.1. Asymptotic Distribution and Consistency of the Proposed Estimator

In this section, the estimate of parametric component $\hat{\boldsymbol{\beta}}_{AS}$ is inspected in terms of asymptotic consistency and asymptotic distribution.

Assume the following regularity conditions:

(i) $\mathbf{F}_n = n^{-1}\left(\mathbf{X}_i^T\mathbf{V} - \mathbf{A}\right)\mathbf{X}_i \to \mathbf{F}$ for non-negative matrix $\mathbf{F}$;

(ii) $n^{-1}\max_{1\leq t\leq n}\left(\mathbf{X}_i^T\mathbf{V} - \mathbf{A}\right)\mathbf{X}_i \to 0$;

(iii) Autoregressive errors ε_t's given in Equation (2) are stationary with independent and identically distributed random error terms u_t's that have zero mean and finite variance $0 < \sigma^2 < \infty$;

(iv) $\mathbf{F}_n^{-1} = n^{-1}\left[\left(\mathbf{X}_i^T\mathbf{V} - \mathbf{A}\right)\mathbf{X}_i\right]^{-1}$ exists.

Here, condition (ii) indicates that the diagonal elements of $\mathbf{F}$ and $\mathbf{F}_n$ are identical and one, because the covariates are scaled. To obtain the asymptotic distribution of $\hat{\boldsymbol{\beta}}_{AS}$, "nearly-singular" designs are performed due to $\tau \to 0$ for $\mathbf{F}_n$. Thus, it can be ensured that $\mathbf{F}_n \to \mathbf{F}$ asymptotically. On the other hand, $\mathbf{F}_n$ and $\mathbf{F}$ are assumed as non-singular in Section 5.1.

To show the consistency and asymptotic normality of the semiparametric AS estimator when conditions (i), (ii), and (iii) are ensured with non-singular $\mathbf{F}$, first the case of $\tau \geq 1$ is considered, followed by the case of $\tau < 1$.

Let $\hat{\boldsymbol{\beta}}_{AS_n}$ be an asymptotic estimator. The consistency of $\hat{\boldsymbol{\beta}}_{AS_n}$ can be reached by using following minimization function:

$$\psi_n\left(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{f}(s_t)\right) = n^{-1}\sum_{t=1}^{n}\left[Y_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_{AS_n} - \hat{f}(s_t)\right]^2 + \lambda_n n^{-1}\sum_{j=1}^{p}\left|\hat{\beta}_{(j)AS_n}\right|^{\tau}. \quad (55)$$

The following theorem shows the consistency of $\hat{\boldsymbol{\beta}}_{AS_n}$ for validated additional assumption $\lambda_n = O(n)$.

Theorem 2. *Assume that* $\boldsymbol{F}$ *is non-singular,* $\hat{f}(s_t)$ *behaves stable, and* $\lambda_n n^{-1} \to \lambda_0 \geq 0$. *It can then be said that as* $n \to \infty$*:*

$$\hat{\boldsymbol{\beta}}_{AS_n} \xrightarrow{d} \boldsymbol{\beta}, \quad (56)$$

where $\hat{\boldsymbol{\beta}}_{AS_n}$ *is a consistent estimator of* $\boldsymbol{\beta}$. *The proofs of this theorem are given in Appendix D. For* $\lambda_n = O(n)$, $argmin(\psi) = \boldsymbol{\beta}$ *and therefore* $\hat{\boldsymbol{\beta}}_{AS_n}$ *is a consistent estimator.*

It should be emphasized that the consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}}$ is sufficient to show that $\lambda_n = O(n)$. However, this depends on the magnitude of growth of λ_n. When λ_n grows more slowly, then a limiting distribution $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}} - \boldsymbol{\beta})$ exists. It is clear from Theorem 2 that the mean of the limiting distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}} - \boldsymbol{\beta})$ converges to zero for the consistency of $\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}}$. In addition, its asymptotic variance can be obtained based on conditions (i) and (iv) as $\sigma^2\mathbf{F}^{-1}$. Accordingly, the asymptotic distribution of the semiparametric AS estimator is written as:

$$\theta = \sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}} - \boldsymbol{\beta}) \xrightarrow{d} N\left[0, \sigma^2\mathbf{F}^{-1}\right]. \quad (57)$$

However, the limiting distribution depends on whether $\tau < 1$ or $\tau \geq 1$. In the context of this paper, Theorem 3 is given for the limiting distribution of $\hat{\boldsymbol{\beta}}_{\mathrm{AS_n}}$ when $\tau < 1$.

Theorem 3. *Assume that* $\tau < 1$ *if* $\lambda_n / n^{\frac{\tau}{2}} \to \lambda_0 \geq 0$. *Then:*

$$\boldsymbol{\theta} = \sqrt{n}\left(\hat{\boldsymbol{\beta}}_{AS_n} - \boldsymbol{\beta}\right) \xrightarrow{d} argmin(\xi), \quad (58)$$

where $\xi(\boldsymbol{\theta}) = -2\boldsymbol{\theta}^T F + \boldsymbol{\theta}^T\boldsymbol{F}\boldsymbol{\theta} + \lambda_0\sum_{j=1}^{p}\|\theta_j\|^{\tau} I(\beta_j = 0)$. *The proofs of Theorem 3 are given in Appendix E.*

6. Simulation Study

In this section, a simulation study was conducted to inspect the finite-sample behaviors and performances of the two semiparametric estimators $(\hat{\boldsymbol{\alpha}}_{BS}, \hat{\boldsymbol{\beta}}_{BS})$ and $(\hat{\boldsymbol{\alpha}}_{AS}, \hat{\boldsymbol{\beta}}_{AS})$ under right-censored time series. These estimators were then compared through the quality measurements given in Section 4. The simulation scenarios are designed as follows:

(a) We use the model $Z_t = \mathbf{X}_t\boldsymbol{\beta} + f(s_t) + \varepsilon_t$, $t = 1, 2, \ldots, n$ to generate datasets in the simulation experiments.

(b) The unknown smooth regression function $f(s_t)$ is constructed by combining the functions $\{S_j, j = 1, \ldots, 5\}$ that denote seasonal effects on the data, that is, $f(s_t) = U_{j=1}^{5} S_j(s_i)$, where $S_j(s_i) = s_i \sin^2(s_i)$ with $s_i = \frac{(i-0.5)}{\frac{n}{5}}, i = 1, \ldots, (n/5)$.

(c) The design matrix is generated from a normal distribution: $\mathbf{X}_t \sim N(\mu_x = 5, \sigma_x^2 = 1)$, where $\mathbf{X}_t$ is the $(n \times p)$ dimensional matrix for $p = 3$. Note also that the distribution may not be normal, and that one can thus consider a uniform or other distributions. The vectors of the regression coefficients are $\boldsymbol{\beta} = (3, 0.5, -1)$.

(d) The autoregressive error terms are generated from a one-lagged process $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$ with $\rho = 0.5$ and $u_t \sim N(0,1)$.

(e) Thus, as stated in (a), completely observed dependent time series Z_t's are generated from the sum of the parametric, nonparametric, and error terms using (b), (c), and (d).

(f) To produce the right censored variable Y_t, as specified in Equation (3), we generate the censoring variable C_t from the binomial distribution with proportions or censoring levels (CLs) at 5%, 20%, and 40%. The Algorithm 2, given below, demonstrates how the censoring variable is created.

Algorithm 2. Generation of censoring variable C_t.

Input: Completely observed Z_t
Output: Right-censored dependent variable Y_t
1: For given censoring level (CL), produce $\delta_t = I(Z_t \leq C_t)$ from the binomial distribution
2: **for** (t *in* 1 *to* n)
3: **If** ($\delta_t = 0$)
4: **while** ($Z_t \leq C_t$)
5: generate $C_t \sim N(\mu_Z, \sigma_Z^2)$
6: **Else**
7: $C_t = Z_t$
8: **end** (for loop in step 2)
9: **for** (t *in* 1 *to* n)
10: **If** ($Z_t \leq C_t$)
11: $Y_t = Z_t$
12: **Else**
13: $Y_t = C_t$
14: **end** (for loop in Step 9)

(a) To deal with censored observations in Y_t obtained with Algorithm 2, we use synthetic data values $Y_{t\hat{G}}$ obtained through the Kaplan and Meier estimator [18], as described in Equation (6).

(b) AR(1) model is used as a naïve model to estimate the right-censored time-series as in [1,2]. Thus, the finite sample performance of the introduced methods can be made.

For each CL in the simulation experiments, we generated 1000 random samples for size $n = 50, 100,$ and 200.

The results of the simulation study were divided into three parts for parametric components, nonparametric components, and overall model performance. Accordingly, the outcomes of the estimated models, comparative results, and corresponding comments are given together in the following tables and figures. To understand the simulated datasets and the scenarios, examples of some of the simulation configurations are given in Figure 1. Panel (a) shows the dataset for small sample size and low censorship. Panel (b) is drawn to show the case when the censoring level is really high. Panels (c)–(d) indicates the cases for medium and large sample sized data with censoring levels 20% and 40% respectively.

6.1. Assessing the Parametric Component

In this section, the performances of the two methods were compared in terms of the parametric components of the right-censored semiparametric linear models generated by the simulation. It should be also noted that in this simulation study, 54 different configurations were analyzed to provide a broad perspective of the adequacy of each method. The results from the parametric components in the simulation study are displayed in Table 1 and Figure 2. Note that bold colored scores indicate the best (minimum) scores.

From the careful inspection of Table 1, it can be demonstrated that the behaviors of the BS and AS change noticeably in different scenarios. Let us look at low and medium CLs for $n = 50$; under these conditions, the BS has remarkable superiority over the AS. This can be interpreted as the BS fitting the data better when the data's structure is distorted less by

censorship. However, for $CL = 40\%$, which means the data are heavily censored, the AS method gives better scores.

As the sample size increases, although the bias and variance values from the methods are obtained more closely, the AS provides more efficient performance in estimating the parametric component. Regarding the parametric component, it should be emphasized that the AS behaves as expected and gives the best scores for cases of heavy censorship.

In general, the best scores for each method can be evaluated in terms of bias and variance results. When we examined the bias results of the regression coefficients, the AS method gives the best score in only 12 out of 27 configurations while the BS method gives the best score in 15. However, regarding the variances, the AS gives the best score in 18 of 27 configurations, while the BS is superior in only 9 configurations. In Figure 2, Panels (a–c) shows the calculated biases for each simulation repetition for all cases when sample size is small, medium, and large.

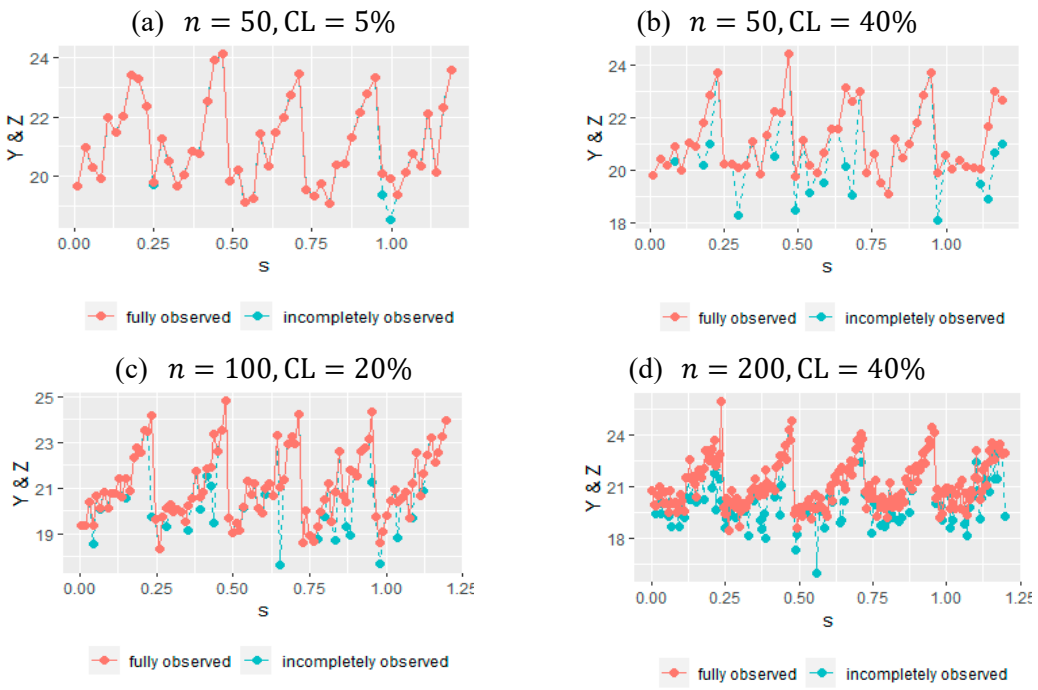

Figure 1. Some of the datasets generated using Algorithm 2 including both fully observed and censored data points for different censoring levels and sample sizes.

Table 1. Estimated regression coefficients from the AS and the B-spline (BS) with values of variance and bias.

		$\beta_1 = 3$				$\beta_2 = 0.5$				$\beta_3 = -1$			
		$Bias(\hat{\beta}_1)$		$Var(\hat{\beta}_1)$		$Bias(\hat{\beta}_2)$		$Var(\hat{\beta}_2)$		$Bias(\hat{\beta}_3)$		$Var(\hat{\beta}_3)$	
n	*C.L.*	**AS**	**BS**	**AS**	**BS**	**AS**	**BS**	**AS**	**BS**	**AS**	**BS**	**AS**	**BS**
50	5	0.887	**0.870**	0.936	**0.842**	0.809	**0.786**	0.922	**0.845**	0.867	**0.837**	0.884	**0.804**
	20	**0.852**	0.895	**1.180**	1.290	**0.888**	0.892	**1.210**	1.358	0.963	**0.949**	**1.191**	1.336
	40	**0.999**	1.172	**1.455**	1.641	**0.916**	1.108	**1.431**	1.657	**0.946**	1.145	**1.453**	1.674
100	5	0.510	**0.470**	0.440	**0.425**	0.539	**0.434**	0.433	**0.422**	0.515	**0.467**	0.439	**0.431**
	20	**0.514**	0.610	**0.583**	0.609	**0.538**	0.579	**0.583**	0.609	**0.527**	0.599	**0.590**	0.618
	40	0.535	**0.433**	**0.619**	0.689	**0.525**	0.622	**0.619**	0.689	**0.535**	0.610	**0.629**	0.692
200	5	0.285	**0.271**	0.260	**0.253**	0.290	**0.272**	0.255	0.255	0.294	**0.271**	**0.252**	0.254
	20	**0.310**	0.324	**0.333**	0.355	0.311	**0.300**	**0.325**	0.351	0.304	**0.296**	**0.328**	0.353
	40	**0.314**	0.333	**0.338**	0.352	**0.321**	0.337	**0.332**	0.356	**0.307**	0.336	**0.332**	0.363

The bolded values indicate the best scores.

6.2. Evaluating the Nonparametric Component

As in the case of parametric components, we constructed 1000 estimates of the regression function $f(.)$, which is the nonparametric component of model (1). For each method, 1000 replications were carried out, and the estimated bias, variance and *RMSE* values were computed for each estimator. This section is designed to show the simulated results related to the nonparametric component.

The results in Table 2 showed that the AS method proves its efficiency for the estimation of the nonparametric component when time series data are moderately to heavily censored. On the other hand, for $CL = 5\%$, the BS method gives better results for all sample sizes according to our evaluation metrics. One of the main reasons for this is that the BS adapted to the knots more than the AS. Consequently, when the data points are manipulated by censorship, these knots force the BS to make inefficient estimates. At this point, the knot determination of the AS based on the weights given in Equation (24) diminishes the effect of the censorship. That is why the AS method performs better under moderately and heavily censored time series data.

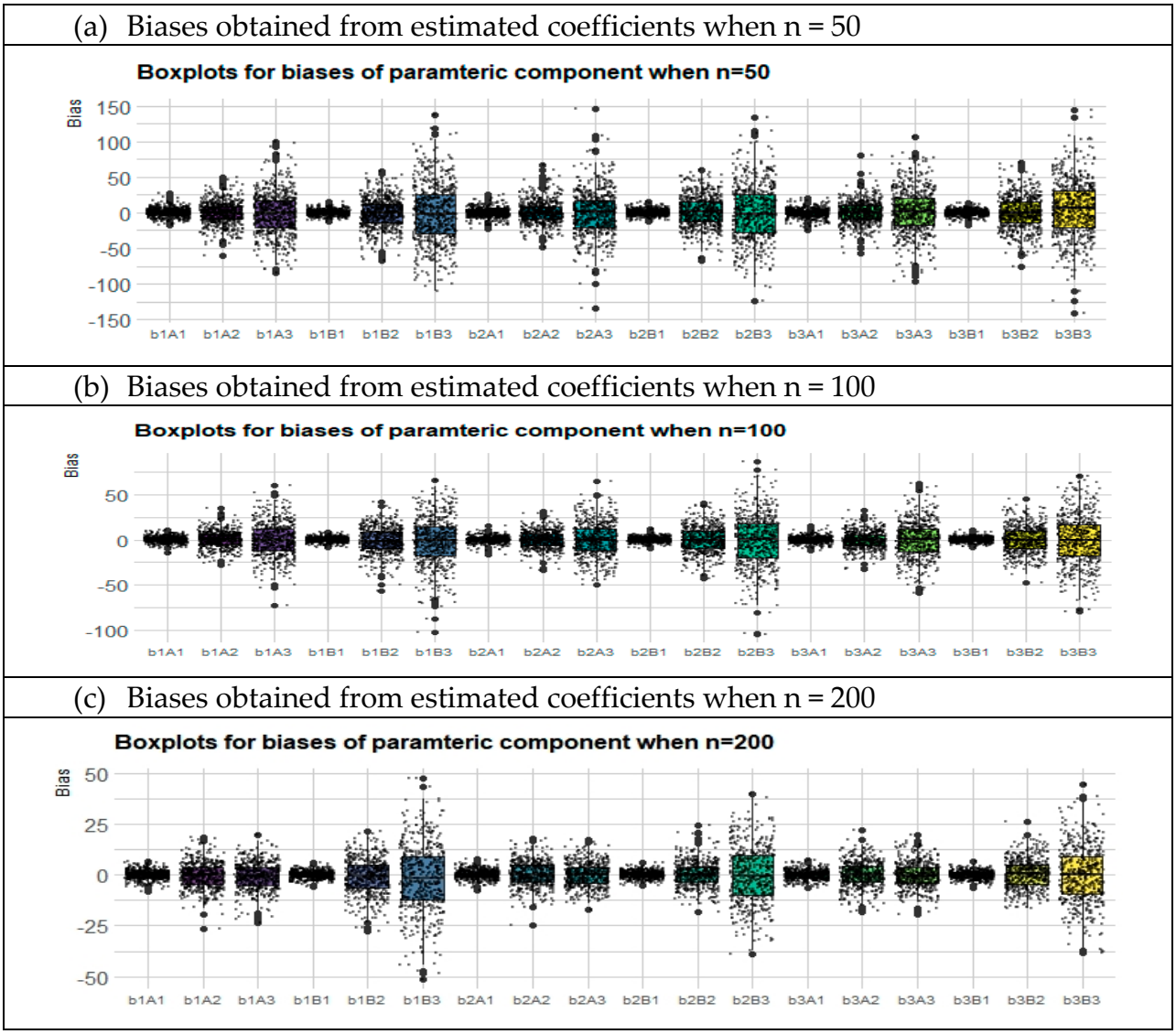

Figure 2. Boxplots of bias values for both the AS and BS methods for all configurations. In the x-axis, b1, b2, and b3 denote β_1, β_2, and β_3; A1, A2, and A3 denote biases obtained from the AS method for CLs of 5%, 20%, and 40%. Similarly, B1, B2, and B3 denote biases for the BS method, when CLs are 5%, 20%, and 40%.

Table 2. Outcomes from the fitted nonparametric components.

		$Bias(\hat{\alpha})$		$Var(\hat{\alpha})$		$RMSE(f,\hat{f})$	
n	***CLs***	**AS**	**BS**	**AS**	**BS**	**AS**	**BS**
50	5	1.085	**0.629**	0.048	**0.022**	1.135	**0.883**
	20	**1.128**	1.498	**0.066**	0.075	**1.099**	2.061
	40	**1.287**	2.510	**0.079**	0.095	**2.511**	3.127
100	5	0.961	**0.851**	**0.022**	0.025	0.824	**0.664**
	20	**1.040**	1.217	**0.030**	0.041	**1.255**	1.779
	40	**1.070**	1.302	**0.037**	0.070	**1.815**	2.331
200	5	0.891	**0.813**	0.009	**0.008**	0.670	**0.435**
	20	**0.928**	0.959	**0.013**	0.021	**1.547**	1.871
	40	**0.995**	1.070	**0.017**	0.028	**2.397**	2.882

The bolded values indicate the best scores.

Figure 3, consisting of four panels (a), (b), (c), and (d), is drawn to illustrate the performance of the AS and BS methods in nonparametric curve estimation and to present different simulation configurations. Panel (a) show the estimated curves for small sample size and medium censoring level. Similarly, Panel (b) shows the case when medium sample size and high censoring level. Panel (c) indicates the estimated curves for small sample size and low censoring level. Finally, Panel (d) shows the estimated curves when sample size is large and censoring level is medium. When panels (**a**) and (**c**) are analyzed comparatively, the effect of the censorship level can be seen. At the first glance, the distortion of both curves is noticeable. However, the BS method is insufficient to represent censored time series compared to the AS method. In addition, panel (b) shows that when data are heavily censored, the BS curve is drawn towards the x = 0 line, due to the presence of zero values in the synthetic response variable. Finally, panel (**d**) indicates that although the time series contains censored data points, the qualities of the estimates for both the AS and BS methods become better as the sample size increases.

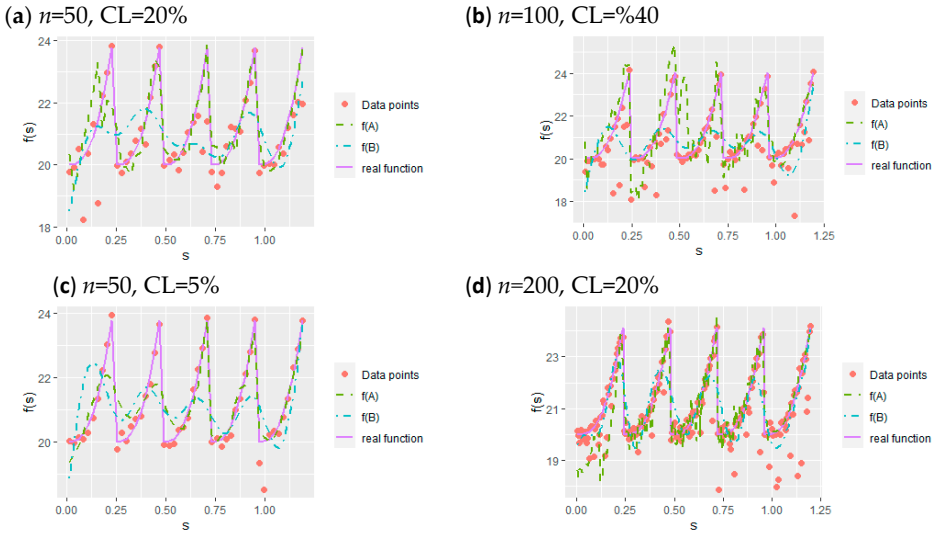

Figure 3. Data points, real regression functions, and curves fitted by two methods. In the legend of the plots, f(A) and f(B) represent function estimates obtained from the AS and BS methods, respectively.

6.3. Assessing the Performances of Methods

This section involves the results for overall model estimations obtained from the AS and BS methods. Although results are given for parametric and nonparametric components in the previous sections, a separate review for the whole model estimation is required for a healthy comparison. Accordingly, the performance scores for $MAPE$, $MedAE$, and $GMSE$ are given in Table 3, and Figure 4 is drawn to illustrate the $RGMSE$ values.

Table 3. The values of performances from the AS and BS methods.

		MAPE			*MedAE*			*GMSE*		
n	***CLs***	**AS**	**BS**	**AR(1)**	**AS**	**BS**	**AR(1)**	**AS**	**BS**	**AR(1)**
50	5	0.166	**0.157**	0.322	0.419	**0.383**	0.999	3.119	3.510	4.915
	20	0.358	**0.348**	0.388	**0.737**	0.896	1.052	4.468	4.920	5.142
	40	**0.584**	0.688	1.980	**1.030**	1.519	1.971	7.762	9.542	10.751
100	5	**0.154**	0.186	0.303	0.323	**0.320**	0.860	1.001	0.928	3.614
	20	**0.333**	0.336	0.365	**0.668**	0.750	0.914	1.870	1.988	4.147
	40	**0.514**	0.528	1.476	**1.025**	1.831	1.891	3.663	4.182	6.798
200	5	0.111	**0.096**	0.283	0.264	**0.251**	0.717	0.983	**0.761**	1.935
	20	**0.312**	0.332	0.364	**0.552**	0.606	0.847	**2.065**	2.497	3.411
	40	**0.499**	0.508	0.654	**1.008**	1.086	1.501	**2.759**	2.816	3.131

The bolded values indicated the best scores.

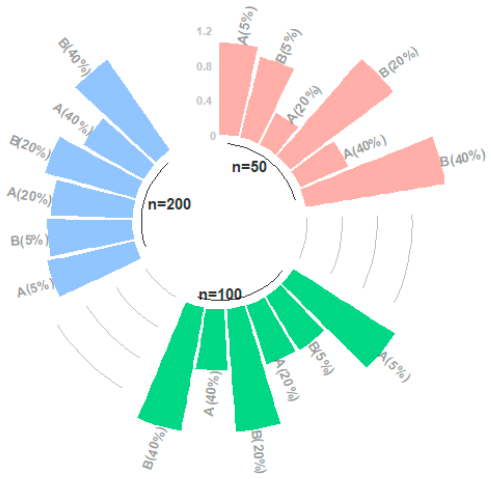

Figure 4. 360° bar chart for the *RGMSE*s of all simulation combinations.

When Table 3 is examined, it can be seen that the results obtained for the model estimates are slightly different from the previous results, as expected. The total error obtained from the estimation of parametric and nonparametric components is one of the reasons for this discrepancy. In addition, considering the situations where the two methods produce extremely similar scores, this difference can be understood better. Note that AR(1) model shows poor performance, which depends on its parametric and linear structure. However, for the large sample size (n = 200), the scores of models obtained are close to each other. However, it is clearly seen that the AS and BS methods are much better on the estimation of right-censored time series.

As can be seen from the bolded scores, the AS method generally performs better. From Table 3, it can be seen that the *MAPE* values obtained by BS are better for $n = 50$. However, as mentioned earlier, in this study, the *MedAE* criterion, which is not frequently used for time series data, is used to measure the durability of the predictions. When the scores of this criterion are examined, it is understood that, as stated from the beginning of the study, the BS method has more successful estimates under low censorship levels, but the AS method is superior for medium and high censorship levels.

Figure 4 includes the *RGMSE* scores for both the AS and BS methods that are formed by the ratio of the *GMSE* values of each method. In Figure 4, the difference between the qualities of the estimates is clearly very small for $CL = 5\%$. However, the difference becomes more significant for $CL = 20\%$ and $CL = 40\%$. Note that for $CL = 5\%$, the BS method gives smaller ratio values, which confirms the results given in Table 3. As stated before, the AS method is demonstrably superior at higher censorship levels, which can be seen in Figure 4 for all sample sizes.

7. Real-World Data

This section is designed to show how the newly introduced semiparametric estimator AS and benchmark BS method behave with a real right-censored time series dataset. For this purpose, we consider unemployment duration data involving the monthly unemployment period rates years between 2004 and 2019 for Turkey; this dataset is available at https://ec.europa.eu/eurostat/databrowser/view/UNE_RT_M__custom_1635127/default/table?lang=en. In the dataset, the last three months of 2004 and the last three months of 2019 cannot be observed correctly. Therefore, these data points can be censored from the right by the detection limit zero, because none of the data points are negative values. Accordingly, the introduced semiparametric methods, AS and BS, can be used for this time series analysis. In addition, as in the simulation study, the results of the AR model are given in the following tables. However, different from the simulation study, AR(2) model was used for the real data study, because the optimal lag values is determined as $lag = 2$ from Table 4. Before the modelling procedure, the stationarity of the time series data was tested with the augmented Dickey–Fuller (ADF) test, the suitable lag is determined under null hypothesis $H_0 : y_t \; is \; non - stationary$. The test results are given in Table 4 below:

Table 4. Augmented Dickey–Fuller (ADF) test results for the stationarity of time series data and the determination of the appropriate lag.

No. Lag	ADF Test Results	*p*-Value
0	−2.61	0.318
1	−3.27	0.077
2	**−3.52**	**0.041**
3	−3.33	0.066
4	−3.30	0.072

Bold scores are significant score for the 95% confidence level.

Table 4 shows that the second lag for this time series is suitable for the modelling. From this information, the semiparametric time series model can be given by:

$$UED_t = \beta_1 UED_{(t-1)} + \beta_2 UED_{(t-2)} + f(s_t) + \varepsilon_t, \;\; t = 1, \ldots, 186, \tag{59}$$

where UED_ts represent the dependent time series of the unemployment duration ratio, $UED_{(t-1)}$ and $UED_{(t-2)}$ denote the first and second lags of the dependent series UED_t that are used as covariates, respectively, $s_t = (1, \ldots, n)^T$ denotes the seasonality, and finally, ε_t's are the stationary autoregressive error terms as given in Equation (2). The estimation of model (6.1) is realized by both the AS and BS methods, and then, results are presented in Tables 5 and 6 and Figure 5.

Table 5. The performances of the BS and AS methods for the estimation of both parametric and nonparametric components.

Measurement	Bias		Variance	
	AS	**BS**	**AS**	**BS**
$\hat{\beta}_1$	**1.941**	2.682	**1.272**	1.703
$\hat{\beta}_2$	**0.915**	1.139	**1.562**	1.624
$\hat{\alpha}$	**3.628**	4.566	0.067	**0.058**

The bolded values indicate the best scores.

Table 6. Scores of performance measures for the AS and BS methods obtained from the whole model estimation.

Method	*MAPE*	*MedAE*	*GMSE*	*RGMSE*	$RMSE\left(\mathbf{f},\hat{\mathbf{f}}\right)$
AS	**0.623**	**0.510**	**1.275**	**0.824**	**1.154**
BS	1.315	1.166	1.546	1.212	1.385
AR(2)	1.856	4.506	3.702	2.775	-

The bolded values indicate the best scores.

Figure 5. Estimated curves for the seasonality $f(s_t)$ obtained from the AS and BS methods.

Table 5 involves the bias and variance values for estimated regression coefficients $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$ and $\hat{\boldsymbol{\alpha}} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_{q+k+1}\right)^T$. Accordingly, the AS method gives smaller bias and variance values than the BS method regarding $\hat{\boldsymbol{\beta}}$. Moreover, the AS method has better bias values for $\hat{\boldsymbol{\alpha}}$, but the BS method gives smaller variance values for $\hat{\boldsymbol{\alpha}}$ than the AS method. In overview, the AS and BS methods give similar values, because the data properties are $n = 186$ and $CL = 8.1\%$. Thus, it can be seen that the results of the unemployment duration data ensure the simulation outputs.

In addition, it should be noted that the outcomes obtained from estimated model (7.1) are given in Table 6 with $RMSE$ scores for the estimated nonparametric function $f(s_t)$. Upon close inspection, it is obviously seen from the results that the AS method produces the best scores. It should be emphasized that the largest difference between the methods regarding performance criteria is in $MedAE$, which indicates the strength of the AS method under censorship. Table 6 indicates the results of AR(2) model that are worse than the results of the other two as in the simulation study. Note that because of the sample size of the real data of $n = 186$ which is close to the simulation configurations when $n = 200$, scores are relatively close to each other. Figure 5 is given to compare the AS and BS methods in representing data under censorship.

As can be seen in Figure 5, the estimated curves are quite similar due to the data properties of a large sample size and a low CL. The effect of synthetic data manipulation is obvious in the figure with zero values. Like the simulation study, the BS method is affected by these zero values more than the AS method. The reason for this is that the knots of the AS method are determined by iteratively calculated weights. Therefore, the optimal knot sequence diminishes the effect of censorship.

8. Concluding Remarks

This paper demonstrated the estimation of right-censored time series data using a newly introduced semiparametric AS estimator and making a comparison with the BS method as a benchmark. The results obtained from both a simulation study and a real data example proved that the introduced method (AS) achieves the superior modelling of right-

censored time series data in a semiparametric context. Comparative outcomes also support that the AS method provides better performance scores over the BS method in most simulation configurations and the real data example. The most important factor in the success of the AS method is the adaptive nature of the method based on iteratively calculated weights. In the AS method, weights are responsible for determining and controlling the penalty term and for dependently obtaining the optimal knot points. Accordingly, our findings showed that the proposed method provides an advantage in modelling right-censored time series over the benchmark.

The simulation study examined the performance of the methods in three parts: the outcomes for the estimated parametric component (Table 1 and Figure 2), the nonparametric component (Table 2 and Figure 3), and the whole semiparametric model (Table 3 and Figure 4). The unemployment data estimation was evaluated for bias and variance (Table 5) using the criteria of $MAPE$, $MedAE$, $GMSE$, and $RGMSE$ (Table 6). Given the outcomes of the simulation study and the real data example, our general and detailed conclusions are as follows:

- As expected, the estimation qualities for both the AS and BS methods change for different CLs and sample sizes. The performances of the methods are affected negatively by increasing CLs, and they give better results for larger sample sizes. This claim is seen clearly from Tables 1–3.
- When unemployment duration data were analyzed, it can be seen that the results agreed with the corresponding configuration ($n = 200$; $CL = 20\%$) of the simulation study.
- One of the striking results of this paper is that, as Tables 1–3 demonstrate, while the AS method gives worse results at low censorship levels than the BS method, it provides significantly better results at medium and high censorship levels. This conclusion proves the claim of the paper, which is that using the AS method reduces the effect of the data manipulation of synthetic data transformation.
- When all the results obtained from simulation and real data studies were inspected, the AS method gives better results than the BS method, except in the configurations for low CLs, which supports the targeted conclusion.
- Unemployment duration data were modelled by the BS and AS methods using two lagged parametric components and the seasonal effect as a nonparametric component. Tables 5 and 6 show each method's scores using four evaluation metrics, which indicate the superiority of the AS method. Figure 5 shows the estimated curves for both methods, which are similar. However, the estimated curves show that the AS method is less affected by zero values of synthetic data and thus gives more satisfying estimates for the right-censored time series model than the BS method.

Finally, as can be understood from the whole paper, the AS method is superior for estimating right-censored time series over the BS method in both theory and practice.

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

Funding: Professor Ahmed research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We consider unemployment duration data involving the monthly unemployment period rates years between 2004 and 2019 for Turkey; this dataset is available at https://ec.europa.eu/eurostat/databrowser/view/UNE_RT_M__custom_1635127/default/table?lang=en.

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

Appendix A

Proof of Lemma 1. Lemma 1 can be ensured based on the common censorship assumption that Z_t and C_t are independent. From that, the proof can be written as follows:

$$\begin{aligned} E[Y_{tG}|x,s] &= E\left[\tfrac{\delta_t Z_t}{1-G(Z_t)}|\, x,s\right] = E\left[\tfrac{\delta_t Z_t}{\overline{G}(Z_t)}|x,s\right] = E\left[\tfrac{I(Z_t\le C_t)\min(Z_t,C_t)}{\overline{G}[\min(Z_t,C_t)]}|x,s\right] = \\ E\left[I(Z_t\le C_t)\tfrac{Z_t}{\overline{G}(Z_t)}|x,s\right] &= E\left[E\left[\tfrac{Z_t}{\overline{G}(Z_t)}I(Z_t\le C_t)|x,s\right]|x,s\right] = E\left[\tfrac{Z_t}{\overline{G}(Z_t)}\,\overline{G}(Z_t)|x,s\right] = \\ E[Z_t|x,s] &= \mathbf{x}_t\boldsymbol{\beta}+f(s_t) \end{aligned} \tag{A1}$$

Thus, Lemma 1 is proven. Here, $\overline{G}(.) = 1-G(.)$. Generally, distribution $G(.)$ is unknown. Therefore, its Kaplan–Meier estimator $\hat{G}(.)$ is used instead of $G(.)$, which is given in Equation (5). □

Appendix B

Derivations of Equations (29) and (30).

To show the derivations of Equations (29) and (30), two equations obtained from Equation (27) are written as:

$$\left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)\boldsymbol{\beta}+\mathbf{X}'\mathbf{V}\mathbf{B}\boldsymbol{\alpha} = \mathbf{X}'\mathbf{V}\mathbf{Y}_{\hat{G}} \quad \mathbf{B}'\mathbf{V}\mathbf{X}\boldsymbol{\beta}+\left(\mathbf{B}'\mathbf{V}\mathbf{B}+\boldsymbol{\lambda}\mathbf{K}\right)\boldsymbol{\alpha} = \mathbf{B}'\mathbf{V}\mathbf{Y}_{\hat{G}} \tag{A2}$$

From Equation (B1), $\hat{\boldsymbol{\alpha}}_{AS}$ can be acquired by the algebraic operations:

$$\left(\mathbf{B}'\mathbf{V}\mathbf{B}+\boldsymbol{\lambda}\mathbf{K}\right)\boldsymbol{\alpha} = \mathbf{B}'\mathbf{V}\mathbf{Y}_{\hat{G}}-\mathbf{B}'\mathbf{V}\mathbf{X}\boldsymbol{\beta} \quad \left(\mathbf{B}'\mathbf{V}\mathbf{B}+\boldsymbol{\lambda}\mathbf{K}\right)\boldsymbol{\alpha} = \mathbf{B}'\mathbf{V}\left(\mathbf{Y}_{\hat{G}}-\mathbf{X}\boldsymbol{\beta}\right). \tag{A3}$$

Thus, if $\boldsymbol{\beta}$ is replaced by $\hat{\boldsymbol{\beta}}_{AS}$, then $\hat{\boldsymbol{\alpha}}_{AS}$ can be written as:

$$\hat{\boldsymbol{\alpha}}_{AS} = \left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\left(\mathbf{Y}_{\hat{G}}-\mathbf{X}\hat{\boldsymbol{\beta}}_{AS}\right). \tag{A4}$$

Therefore, Equation (27) can be derived. Accordingly, the derivation of $\hat{\boldsymbol{\beta}}_{AS}$ can be obtained by using (B1):

$$\begin{aligned} \left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)\boldsymbol{\beta}+\mathbf{X}'\mathbf{V}\mathbf{B}\left[\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\left(\mathbf{Y}_{\hat{G}}-\mathbf{X}\boldsymbol{\beta}\right)\right] &= \mathbf{X}'\mathbf{V}\mathbf{Y}_{\hat{G}}, \\ \left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)\boldsymbol{\beta}+\mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{Y}_{\hat{G}}-\mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}\boldsymbol{\beta} &= \mathbf{X}'\mathbf{V}\mathbf{Y}_{\hat{G}}, \\ \left[\left(\mathbf{X}'\mathbf{V}\mathbf{X}\right)-\mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{X}\right]\boldsymbol{\beta} &= \mathbf{X}'\mathbf{V}\mathbf{Y}_{\hat{G}}-\mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\lambda K\right]^{-1}\mathbf{B}'\mathbf{V}'\mathbf{Y}_{\hat{G}}. \end{aligned} \tag{A5}$$

To simplify the calculations, let $\boldsymbol{A}_{AS} = \mathbf{X}'\mathbf{V}\mathbf{B}\left[\mathbf{B}'\mathbf{V}\mathbf{B}+\boldsymbol{\lambda}\mathbf{K}\right]^{-1}\mathbf{B}'\mathbf{V}'$. Therefore,

$$\left[\left(\mathbf{X}'\mathbf{V}-\boldsymbol{A}_{AS}\right)\mathbf{X}\right]\boldsymbol{\beta} = \left(\mathbf{X}'-\boldsymbol{A}_{AS}\right)\mathbf{V}\mathbf{Y}_{\hat{G}},\ \hat{\boldsymbol{\beta}}_{AS} = (\left(\mathbf{X}'\mathbf{V}-\boldsymbol{A}_{AS}\right)\mathbf{X}\right)^{-1}\left(\mathbf{X}'-\boldsymbol{A}_{AS}\right)\mathbf{V}\mathbf{Y}_{\hat{G}}. \tag{A6}$$

The derivations of Equations (29) and (30) are thus completed.

Appendix C

Proof of Theorem 1. To validate the Theorem 1, necessary equations are given by:

$$\sup_{\hat{\alpha}_{ASn} \in Q} \left| M_n(\hat{\alpha}_n) - M(\hat{\alpha}_{ASn}) - \sigma_\varepsilon^2 \right| \xrightarrow{p} 0, \tag{A7}$$

where σ_ε^2 is the variance of the model defined in Equation (7), Q is a compact set in a metric space and by using Equations (54)–(57), it can be seen that:

$$|\hat{\boldsymbol{\alpha}}_{ASn}| \to \boldsymbol{\alpha}, \text{ as } n \to \infty. \tag{A8}$$

See [33] for more details. □

Appendix D

Proof of Theorem 2. For ensured regularity conditions (i)–(iv), $plim(\hat{\boldsymbol{\beta}}_{AS_n})$ is written as follows:

$$\begin{aligned} plim(\hat{\boldsymbol{\beta}}_{AS_n}) &= \boldsymbol{\beta} + plim(n^{-1}[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})\mathbf{f}) + plim(n^{-1}[(\mathbf{X}'\mathbf{V} - \\ &\quad \mathbf{A}_{AS})\mathbf{X}]^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})\boldsymbol{\varepsilon}) \\ plim(\hat{\boldsymbol{\beta}}_{AS_n}) &= \boldsymbol{\beta} + plim\{n^{-1}\,[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})\mathbf{X}]^{-1}\}plim\{n^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})[\mathbf{f} + \boldsymbol{\varepsilon}]\}. \end{aligned} \tag{A9}$$

Because $\mathbf{f}$ can be counted as a nuisance parameter, and from assumptions (i) and (ii), $plim\left\{n^{-1}\,\left[(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})\mathbf{X}\right]^{-1}\right\} = \mathbf{F}_n^{-1}$ and $plim\{n^{-1}(\mathbf{X}'\mathbf{V} - \mathbf{A}_{AS})[\mathbf{f} + \boldsymbol{\varepsilon}]\} = o(1)$. Therefore, the expression at the right side in (D1) goes to zero. Thus, from that, it is obtained that:

$$\text{argmin}(\psi_n) \xrightarrow{p} \text{argmin}(\psi), \quad \hat{\boldsymbol{\beta}}_{AS_n} \xrightarrow{d} \boldsymbol{\beta}. \tag{A10}$$

Note that the results obtained above are for $\tau \geq 1$, which means ψ_n has a convex structure (see [34,35]). However, the proposed AS estimator includes the case of $\tau < 1$, so that ψ_n is not convex. In this matter, Equation (D2) is processed differently as:

$$\psi_n\left(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{f}(s_t)\right) > n^{-1} \sum_{t=1}^{n} \left[Y_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_{AS_n} - \hat{f}(s_t)\right]^2 = \psi_n^{(0)}\left(\hat{\boldsymbol{\beta}}_{AS_n}, \hat{f}(s_t)\right) \tag{A11}$$

Note that Equation (D3) is validated for all $\hat{\boldsymbol{\beta}}_{AS_n}$. Moreover, $argmin(\psi_n) = O_p(1)$, because $\left(\psi_n^{(0)}\right) = O_p(1)$. □

Appendix E

Proof of Theorem 3. To show the proof of Theorem 3, due to the non-convex structure of $\tau < 1$, some complex expressions are needed for minimization criterion ξ. These are given by:

$$\xi_n(\theta) = \sum_{t=1}^{n}\left[\left(\varepsilon_t - \frac{\theta^T\mathbf{X}_t}{n^{-1}}\right)^2 - \varepsilon_t\right] + \lambda_n \sum_{j=1}^{p}\left[\left|\beta_j + \frac{\theta_j}{n^{-1}}\right|^\tau - |\beta_j|^\tau\right]. \tag{A12}$$

Due to $\lambda_n = O\left(n^{\tau/2}\right) = o\left(\sqrt{n}\right)$, the following expression is obtained similar to Theorem 3:

$$\lambda_n \sum_{j=1}^{p}\left[\left|\beta_j + \frac{\theta_j}{n^{-1}}\right|^\tau - |\beta_j|^\tau\right] \xrightarrow{d} \lambda_0 \sum_{j=1}^{p}|\theta_j|^\tau \mathrm{I}(\beta_j = 0). \tag{A13}$$

Then the convergence is realized as follows:

$$argmin(\xi_n) \xrightarrow{d} argmin(\xi). \tag{A14}$$

Thus, the proof is finished. It is important to note that, for $\tau < 1$, the non-zero regression coefficients of the model can be estimated without asymptotic bias if zero ones are shrunk to the zero with a positive probability. □

References

1. Park, J.W.; Genton, M.G.; Ghosh, S.K. Censored time series analysis with autoregressive moving average models. *Can. J. Stat.* **2007**, *35*, 151–168. [CrossRef]
2. Aydin, D.; Yilmaz, E. Censored nonparametric time-series analysis with autoregressive error models. *Comput. Econ.* **2020**, *58*, 169–202. [CrossRef]
3. Hopke, P.K.; Liu, C.; Rubin, D.B. Multiple imputation for multivariate data with missing and below-threshold measurements: Time-series concentrations of pollutants in the arctic. *Biometrics* **2001**, *57*, 22–33. [CrossRef] [PubMed]
4. Ghouch, A.E.; Keilegom, I.V. Non-parametric Regression with Dependent Censored Data. *Scand. J. Stat.* **2008**, *35*, 228–247. [CrossRef]
5. Koul, H.; Susarla, V.; Van Ryzin, J. Regression Analysis with Randomly Right-Censored Data. *Ann. Stat.* **1981**, 1276–1285. [CrossRef]
6. Leurgans, S. Linear models, random censoring and synthetic data. *Biometrika* **1987**, *74*, 301–309. [CrossRef]
7. Zhou, M. Asymptotic Normality of the 'Synthetic Data' Regression Estimator for Censored Survival Data. *Ann. Stat.* **1992**, *20*, 1002–1021. [CrossRef]
8. Linton, O.; Nielsen, J.P.; Nielsen, S.F. Non-parametric regression with a latent time series. *Econom. J.* **2010**, *12*, 187–207. [CrossRef]
9. Vogt, M. Nonparametric regression for locally stationary time series. *Ann. Stat.* **2012**, *40*, 2601–2633. [CrossRef]
10. Gao, J. *Nonlinear Time Series: Semiparametric and Nonparametric Methods*; CRC Press: Boca Raton, FL, USA, 2007.
11. Chen, J.; Gao, J.; Li, D. Semiparametric trending panel data models with cross-sectional dependence. *J. Econom.* **2012**, *171*, 71–85. [CrossRef]
12. Engle, R.F.; Granger, C.W.J.; Rice, J.; Weiss, A. Semiparametric Estimates of the Relation between Weather and Electricity Sales. *J. Am. Stat. Assoc.* **1986**, *80*, 310–320. [CrossRef]
13. Härdle, W. *Applied Nonparametric Regression (No. 19)*; Cambridge University Press: Cambridge, UK, 1990.
14. Green, P.J.; Silverman, B.W. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*; CRC Press: Boca Raton, FL, USA, 1994.
15. Ruppert, D.; Wand, M.P.; Carroll, R.J. *Semiparametric Regression (No. 12)*; Cambridge University Press: Cambridge, UK, 2003.
16. Guessoum, Z.; Ould-Said, E. On nonparametric estimation of the regression function under random censorship model. *Stat. Decis.* **2009**, *26*, 159–177. [CrossRef]
17. Aydin, D.; Yilmaz, E. Modified estimators in semiparametric regression models with right-censored data. *J. Stat. Comput. Simul.* **2018**, *88*, 1470–1498. [CrossRef]
18. Kaplan, E.L.; Meier, P. Nonparametric Estimation from Incomplete Observations. *J. Am. Stat. Assoc.* **1958**, *53*, 457–481. [CrossRef]
19. Chen, K.; Lo, S.H. On the rate of uniform convergence of the product-limit estimator: Strong and weak laws. *Ann. Stat.* **1997**, *25*, 1050–1087. [CrossRef]
20. Gu, M.G.; Lai, T.L. Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation. *Ann. Probab.* **1990**, *18*, 160–189. [CrossRef]
21. De Boor, C. *A Practical Guide to Splines*; Springer: New York, NY, USA, 1978.
22. Lyche, T.; Manni, C.; Speleers, H. Foundations of spline theory: B-splines, spline approximation, and hierarchical refinement. In *Splines and PDEs: From Approximation Theory to Numerical Linear Algebra*; Springer: Cham, Switzerland, 2018; pp. 1–76.
23. Jin, H.; Guan, Y.; Yao, L. Minimum entropy active fault tolerant control of the non-Gaussian stochastic distribution system subjected to mean constraint. *Entropy* **2017**, *19*, 218. [CrossRef]
24. Havrylenko, Y.; Kholodniak, Y.; Halko, S.; Vershkov, O.; Miroshnyk, O.; Suprun, O.; Dereza, O.; Shchur, T.; Śrutek, M. Representation of a Monotone Curve by a Contour with Regular Change in Curvature. *Entropy* **2021**, *23*, 923. [CrossRef] [PubMed]
25. Eilers, P.; De Menezes, R. Quantile smoothing of array CGH data. *Bioinformatics* **2005**, *21*, 1146–1153. [CrossRef] [PubMed]
26. Ahmed, S.E.; Aydın, D.; Yılmaz, E. Imputation Method Based on Sliding Window for Right-Censored Data. In *International Conference on Management Science and Engineering Management*; Springer: Cham, Switzerland, 2020; pp. 433–446.
27. Rippe, R.C.; Meulman, J.J.; Eilers, P.H. Visualization of genomic changes by segmented smoothing using an L_0 penalty. *PLoS ONE* **2012**, *7*, e38230. [CrossRef]
28. Frommlet, F.; Nuel, G. An adaptive ridge procedure for l 0 regularization. *PLoS ONE* **2016**, *11*, e0148620. [CrossRef]
29. Hurvich, C.M.; Simonoff, J.S.; Tsai, C.L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B* **1998**, *60*, 271–293. [CrossRef]
30. Li, R.; Liang, H. Variable selection in semiparametric regression modeling. *Ann. Stat.* **2008**, *36*, 261–286. [CrossRef] [PubMed]
31. Eilers, P.H.; Marx, B.D. Flexible smoothing with B-splines and penalties. *Stat. Sci.* **1996**, *11*, 89–121. [CrossRef]
32. Frank, L.E.; Friedman, J.H. A statistical view of some chemometrics regression tools. *Technometrics* **1993**, *35*, 109–135. [CrossRef]

33. Fu, W.; Knight, K. Asymptotics for lasso-type estimators. *Ann. Stat.* **2000**, *28*, 1356–1378. [CrossRef]
34. Anderson, P.K.; Gill, R.D. Cox's regression model for counting processes: Large sample study. *Ann. Stat.* **1982**, *10*, 1100–1120. [CrossRef]
35. Pollard, D. Asymptotics for least absolute deviation regression estimators. *Econom. Theory* **1991**, *7*, 186–199. [CrossRef]