Computational Economics
https://doi.org/10.1007/s10614-020-10010-8

Censored Nonparametric Time-Series Analysis with Autoregressive Error Models

Dursun Aydin[1] · Ersin Yilmaz[1]

Accepted: 18 June 2020

Abstract
This paper focuses on nonparametric regression modeling of time-series observations with data irregularities, such as censoring due to a cutoff value. In general, researchers do not prefer to put up with censored cases in time-series analyses because their results are generally biased. In this paper, we present an imputation algorithm for handling auto-correlated censored data based on a class of autoregressive nonparametric time-series model. The algorithm provides an estimation of the parameters by imputing the censored values with the values from a truncated normal distribution, and it enables unobservable values of the response variable. In this sense, the censored time-series observations are analyzed by nonparametric smoothing techniques instead of the usual parametric methods to reduce modelling bias. Typically, the smoothing methods are updated for estimating the censored time-series observations. We use Monte Carlo simulations based on right-censored data to compare the performances and accuracy of the estimates from the smoothing methods. Finally, the smoothing methods are illustrated using a meteorological time- series and unemployment datasets, where the observations are subject to the detection limit of the recording tool.

Keywords Censored time series · Penalized spline · Smoothing spline · Auto-correlated data · Imputation method

✉ Ersin Yilmaz
ersinyilmaz@mu.edu.tr

Dursun Aydin
duaydin@mu.edu.tr

[1] Department of Statistics, Faculty of Science, Mugla Sitki Kocman University, 48000 Mugla, Turkey

Published online: 23 July 2020

1 Introduction

In the statistical literature, researchers use the term "right-censored observation" for a unit's failure time that is only known to exceed a detection limit. Generally, the measurements collected over time are observed with data irregularities, such as censoring due to a threshold value. Ordinary statistical methods cannot be applied directly to such observations, especially for time-series data. As is known, time-series measurements are often auto correlated and analyzed by modeling autocorrelations via their appropriate autoregressive structures. Box and Jenkins (1970) presented the first study dealing with time-series analysis within a parametric framework. However, although parametric approaches are highly useful for analyzing time-series data, they can produce biased estimates or lead to wrong conclusions, especially for censored data.

When using time series, we may encounter severe problems using data with censored or auto-correlated errors. During the last decade, many techniques have been proposed for dealing with such problems in which the dependent variable is subject to censoring. Our view is that these techniques are fundamentally divided into parametric and nonparametric methods, depending on the estimated autocorrelation function. Here, we focus only on the nonparametric approaches and try to discern which will provide a better estimation of auto-correlated censored data.

Suppose we consider a nonparametric time-series regression model with autoregressive errors for censored response observations at time t, given by

$$Y_t = f\left(X_t\right) + e_t, \quad t = 1, \dots, n \tag{1}$$

where Y_t represents a stationary time series, and its prediction depends on the explanatory variable X_t, and f is an unknown smooth function giving the conditional mean of Y_t given X_t. In addition, e_t, defined in (1), is a stationary autoregressive error term generated by

$$e_t = \phi e_{t-1} + \phi e_{t-2} + \cdots + \phi e_{t-p} + \varepsilon_t \tag{2}$$

where $\phi = \left(\phi_1, \dots, \phi_p\right)^\top$ is the vector of the autoregressive coefficients and ε_t represents independent and identically distributed random variables with zero mean and variance σ^2. Model (1) does not contain lagged $Y_t' s$ and has auto-correlated error terms. This makes it an appropriate model for the regression analysis of certain kinds of time-series data.

Our objective is to estimate both the unknown function f (.) and the autoregressive structure in (1) by nonparametric methods using censored time- series data. There are numerous studies suitable for the estimation approaches of f (.) in a nonparametric regression model based on censored data (Zheng 1984; Dabrowska 1992; Kim and Truong 1998; Yang 1999; Cai and Betensky 2003; and Aydin and Yilmaz 2017). As noted, while there are extensive studies on estimating nonparametric models with censored responses, the literature on censored time-series response data is limited. Examples of these works include Zeger and Brookmeyer (1986), Park et al. (2007), and Wang and Chan (2017).

Problems with censored time- series data are commonly solved using data augmentation techniques. Several researchers, including Robinson (1983), Parzen (1983), Tanner (1991), Hopke et al. (2001), and Park et al. (2007, (2009), have used this technique for regression models with autoregressive errors when censored response observations are considered. Note that both imputation and augmentation methods are addressed by several authors for missing values, excluding time- series observations. For example, see the studies of Rubin (1996), Dempster et al. (1977), Heitjan and Rubin (1990), and Meng (1994). In addition, there are several nonparametric estimation methods for obtaining the autocovariance function in the literature, such as Hall and Patil (1994), who suggested a nonparametric estimation method for estimating the autocovariance function based on kernel smoothing (*KS*). Elogne et al. (2008) estimated the autocovariance function with interpolation techniques, and there are several similar studies: Glasbey (1988), Shapiro and Botha (1991), Sampson and Guttorp (1992), Bjørnstad and Grenfell (2001), and Wu and Pourahmadi (2003).

In the literature, there are essentially two approaches to handling censoring. One is to eliminate the censored values, and the other is to use the censored data points as observed. However, the study of Park et al. (2007) demonstrates that both approaches yield biased and inefficient estimates. It is possible that the performance of the parameter estimation can be improved by using datasets that have a lower censoring rate (Helsel 1990). However, such methods are often not applicable, and outcomes depend heavily on rigid parametric model assumptions. As indicated above, even though some numerical solutions have been proposed in the literature to cope with the problem of censored responses in autoregressive error models, there are no studies making inferences for censored time-series models in terms of nonparametric approaches. By contrast, we propose nonparametric estimation procedures for time-series containing right-censored observations, instead of using parametric approaches. Thus, our study is remarkably different from other similar studies used in the statistical literature.

In this paper, we consider three nonparametric approaches, smoothing spline (*SS*), kernel smoothing (*KS*), and regression (penalized) spline (*RS*), for estimating an autoregressive time-series regression model with right- censored observations. Note that these nonparametric approaches cannot be applied directly to censored observations, and a data transformation is required to estimate the censored response observations. To overcome this problem and to get stable solutions, we used a data augmentation method, namely a the Gaussian imputation technique. This data transformation method, which is a modified extension of ordinary Gaussian imputation, is used to adjust the censoring response variable in the setting of a time series. We also compare the performances of the smoothing methods with the benchmark $AR(1)$ model. To the best of our knowledge, such a study has not been conducted. It should be noted that the best estimation of the censored lifetime observations Y_t that depend on an explanatory variable $X_t = x$ could be expressed as the conditional mean of the response $E(Y_t|X_t = x) = f(x)$, which minimizes the quantity

$$E\{Y_t - f(x)\}^2 = E\left\{Y_t - \hat{Y}_t\right\}^2 \tag{3}$$

Normally, it is not necessary that function f be linear, and the conditional variance is homoscedastic, but the error term indicated in model (1) is generally presented as follows

$$E(\varepsilon_t|X_t) = 0\,, Var(\varepsilon_t|X_t) = 1.$$

As indicated above, the minimization of Eq. (3) is carried out through three smoothing methods, the *SS*, *KS*, and *RS* methods. The primary purpose of this study is to estimate a right-censored time series non- parametrically and provide consistency in the estimation with the help of the data augmentation method for censored observations. One of the important points of data augmentation is to estimate the covariance function (Park et al. 2009). Here, this function is estimated with a nonparametric Nadaraya-Watson estimator. With this method, there is no need to describe the prior distribution of the time- series data because data augmentation is achieved using the nonparametric method.

The paper is organized as follows. Section 2 introduces the thory of the censored autoregressive model and algorithm of Gaussian imputation to make data augmentation. In Sect. 3, the three smoothing methods are explained, and their modifications are illustrated according to censored data. Section 4 involves the evaluation measurements for the three modified smoothing techniques. Furthermore, an estimation of the covariance functions of the estimators are expressed. To obtain empirical results, a detailed simulation study is done in Sect. 5. Also, two real-data applications with cloud and unemployement datasets are realized in Sect. 6. Finally, conclusions and discussion are presented in Sect. 7.

2 Censored Autoregressive Model and Preliminaries

Consider the nonparametric regression model defined in Eq. (1). In many applications, we may be unable to directly observe the response variable Y_t. Instead of attempting to observe all the values of the response variable, we may observe only Y_t when Y_t exceeds a constant value C_t that denotes a cut-off value or a detection limit. Time- series observations are often obtained with a detection limit. For example, a monitoring tool usually has a detection limit, and it records the limit value when the true value exceeds the detection limit. This case is often called censoring, which is also a type of missing data mechanism. Censoring of the response measurements occurs in different situations in the physical sciences, business, and economics. If censored observations are ignored, the resulting parameter estimates are usually biased.

Let Z_t be the value we observe instead of Y_t due to censoring. Then, in the right-censored cases, we consider the following type of censoring, given by

$$Z_t = min(Y_t, C_t), \delta_t = I(Y_t \le C_t) \tag{4}$$

where $I(.)$ is an indicator function,$\delta = 0$ denotes a censored observation, and Z_t and C_t are the failure times (or observed lifetimes) and the censoring times, respectively. In light of (4), we assume that only the response variable Y_t is censored on the right by censoring variable C_t, so our observations are the triples $\left\{(Z_t,\ \delta_t,\ X_t),\ t = 1, 2, \ldots, n\right\}$. Then, model (1) along with Eq. (4), reduce *the censored autoregressive nonparametric regression model* of order p. It should be noted that Y_t and Z_t have different distributions. Therefore, Z_t cannot be used directly to make inferences about model (1) described by Y_t. To use the Z_t series as a response variable, a conditional distribution is constructed through a truncated normal distribution.

Assume that $Y_t = (Y_1, \ldots, Y_n)^\top$ is a realization from a stationary time series described by model (1) with autoregressive error terms, which has a Gaussian disturbance ε_t, as defined in (2). Note that the autoregressive errors follow an n-dimensional multivariate normal distribution with a mean zero and stationary $n \times n$ covariance matrix $\mathbf{\Sigma}$. In other words, $e_t = (e_1, \ldots, e_n)^\top \backsim N_n(\mathbf{0}, \mathbf{\Sigma})$, where $\mathbf{\Sigma} = \sigma^2 R_n(\phi)$ is a $n \times n$ covariance matrix, given by

$$R_n(\phi) = \frac{1}{\sigma^2_{\varepsilon_t}} \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{pmatrix} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{pmatrix}$$

where $\gamma_0, \gamma_1, \ldots, \gamma_{n-1}$ are theoretical autocovariances of the process and $\rho_h = \left\{\gamma_h / \gamma_0,\ \ h = 1, 2, .., p\right\}$ are the theoretical autocorrelations of the process. It should be noted that $\gamma_h = E(\varepsilon_t \varepsilon_{t-h}^\top)$ denotes the autocovariance at lag h and $\gamma_0 = E\left(\varepsilon_t \varepsilon_t^\top\right) = \sigma^2$ gives the constant variance at lag zero.

In light of the equations and explanations outlined above, it is understood that $Y_t \sim N_n(\boldsymbol{\mu},\ \mathbf{\Sigma})$ for complete data. When we consider the responses with a censoring mechanism, as in (4), $Y_t \sim TN_n(\boldsymbol{\mu}, \mathbf{\Sigma}; R_c)$, where $TN_n(.\ ;\ R_c)$ denotes the truncated normal distribution on the interval R_c (see Vaida and Liu (2009) for a more detailed discussions). Note that the interval R_c depends on whether a data point is censored. Basically, the interval R_c is $\left(0, C_t\right)$ if $\delta_t = 1$(observed values) and R_c is $[C_t, \infty)$ if $\delta_t = 0$ (censored values). To calculate the unknown function in the censored autoregressive nonparametric regression model, the first task is to consider separately the observed and censored data points of the response variable at the beginning of the estimation procedure. In this context, by using permutation matrix P, which maps $(1, \ldots, k)^\top$ into the permutation vector $\boldsymbol{p} = (p_1, \ldots, p_k)^\top$, the order of the data can be rearranged as

$$PY_t = \begin{pmatrix} P_o \\ P_c \end{pmatrix} Y_t = \begin{pmatrix} \boldsymbol{Y}_o \\ \boldsymbol{Y}_c \end{pmatrix} \tag{5}$$

where $\boldsymbol{Y}_o$ denotes the observed part of the $Y_t' s$, whereas $\boldsymbol{Y}_c$ indicates the right- censored part of the same variable. Furthermore, PY_t provides a multivariate normal distribution defined by

$$PY_t \sim N_n\left(\boldsymbol{\mu}\begin{pmatrix} P_o \\ P_c \end{pmatrix}, \begin{bmatrix} P_o\boldsymbol{\Sigma}P_o^\top & P_o\boldsymbol{\Sigma}P_c^\top \\ P_c\boldsymbol{\Sigma}P_o^\top & P_c\boldsymbol{\Sigma}P_c^\top \end{bmatrix}\right) = N_n\left(\begin{pmatrix} \boldsymbol{\mu}_o \\ \boldsymbol{\mu}_c \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{oc} \\ \boldsymbol{\Sigma}_{co} & \boldsymbol{\Sigma}_{cc} \end{bmatrix}\right) \tag{6}$$

where $\boldsymbol{\Sigma}_{oo} = \boldsymbol{E}\left[(\boldsymbol{Y}_o - \boldsymbol{\mu}_o)(\boldsymbol{Y}_o - \boldsymbol{\mu}_o)^\top\right]$ shows the covariance matrix obtained by taking the observed part of the Y_t, while $\boldsymbol{\Sigma}_{cc} = \boldsymbol{E}\left[(\boldsymbol{Y}_c - \boldsymbol{\mu}_c)(\boldsymbol{Y}_c - \boldsymbol{\mu}_c)^\top\right]$ states the covariance matrix that corresponds to the right-censored data points of the $Y_t's$, and $\boldsymbol{\Sigma}_{oc} = \boldsymbol{\Sigma}_{co} = \left[(\boldsymbol{Y}_c - \boldsymbol{\mu}_c)(\boldsymbol{Y}_o - \boldsymbol{\mu}_o)^\top\right]$.

It should be noted that the conditional distribution of Y_c given Y_o is also a multivariate normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ (see, for example, Anderson 1984). In a similar manner to that used to obtain (5), the observed data vector Z_t can be portioned into sub-vectors, and is given by

$$PZ_t = \begin{pmatrix} P_o \\ P_c \end{pmatrix} Z_t = \begin{pmatrix} \boldsymbol{Z}_o \\ \boldsymbol{Z}_c \end{pmatrix} \tag{7}$$

One of the reasons for the study of conditional distribution derived from the multivariate normal distribution outlined above is to find an appropriate substitute for the right-censored data vector $\boldsymbol{Z}_c$. The basic idea is to replace the elements of the right-censored vector $\boldsymbol{Z}_c$ by sampling values obtained from the conditional distribution of the censored response vector $\boldsymbol{Y}_c$ given $\boldsymbol{Z}_o$ and $\boldsymbol{Z}_c$. This procedure is equivalent to applying the truncated multivariate normal distribution (Park et al. 2007):

$$\left(\boldsymbol{Y}_c | \boldsymbol{Z}_o, \boldsymbol{Z}_c \in R_c\right) \sim TN_{n_c}\left(\mathbf{M}, \boldsymbol{V}, R_c\right) \tag{8}$$

wheren_c denotes the number of censored data points, TN_{n_c}shows a truncated multivariate normal distribution with n_c-dimension, and R_c determines the region associated with the censoring of the response observations, as previously defined. The symbols $\mathbf{M}$ and $\boldsymbol{V}$ expressed in (8) are the parameters that correspond to the conditional mean and covariance of a non-truncated variant of a conditional multivariate normal distribution.

Now, suppose that we have a time-series vector $Y_t = (Y_1, \dots, Y_n)^\top$ that can be parametrized by the mean μ, variance σ^2, and autocorrelation ρ. As indicated above, these time series measurements are also considered a random vector from a multivariate normal distribution $Y_t \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu} = \mu\mathbf{1}_n$ and $\{\boldsymbol{\Sigma}\}_{i,j} = E\left(e_t e_t^\top\right) = \sigma_{e_t}^2 = \frac{\sigma_{\varepsilon_t}^2}{1-\rho^2}\rho^{|i-j|}, i,j = 1, \dots, n$. The conditional distribution of Z_t, given other observations, is a univariate normal distribution; if Z_t is censored, then the conditional distribution is a truncated univariate normal distribution. Many statistical approaches for obtaining the truncated normal distribution depend on a compacted simulation. Examples of such studies include Tanner and Wong (1987), Chen and Deely (1996), Gelfand et al. (1992), and so on. In general, the mentioned studies focus on the data augmentation method with a truncated normal distribution, Bayesian methods, and Gibbs samplers. In this study, we use a truncated normal distribution whose probability density function can be defined as

$$f\left(Z_t\right) = g\left(Z_t\right) I(\delta_t = 0) / \left[1 - F\left(C_t\right)\right] \tag{9}$$

where $I(.)$ denotes the indicator function, that is, $I = 1$ if $\delta_t = 0$, and $g(.)$ and $F(.)$ are the probability density function of the standard normal distribution and its cumulative distribution function, respectively. It should be noted that Gaussian density $g(.)$ is used for observations in the interval $[C_t, \infty]$ to obtain the distribution of the right-censored part of the data.

In this paper, we focus on the data imputation method to find a solution to the censored data problem. One should note that the basic idea in the imputation of the censored values is to replace every censored observation with a real value. To carry out this procedure, in the first stage, the parameters of the distributions outlined above must be estimated by iteratively applying an appropriate algorithm.

2.1 Imputation Algorithm

Let $\boldsymbol{\varphi} = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be the true parameter vector based on the distribution of Y_t with autocorrelation ρ and $\boldsymbol{\Phi} = (\mathbf{M}, V)$ be a parameter vector based on the conditional distribution of the censored response variable Y_c. By introducing the censored variable to the model (1), the imputation algorithm simplified to the generation of values from the truncated normal distribution defined in (9). Usually, all the censored observations are set equal to some constant value. The idea of the algorithm is to update the parameter estimates by filling (or imputing) the censored values with the values from a conditional sample. The iteratively updated algorithm is essentially divided into two components: data augmentation and parameter estimation. It should be emphasized that one needs to use a random sample generated from a truncated multivariate normal distribution for data augmentation, whereas any traditional method for the parameter estimation can be used.

The imputation algorithm consists of the following steps:

Step 1 Generate the truncated normal distribution and compute the density of the censored partition of the data.

Step 2 Obtain the initial parameter estimates $\hat{\mu}^{(0)}$, $\hat{\rho}^{(0)}$, and $\hat{\sigma}^{(0)}$ by using the equations

$$\hat{\mu}^{(0)} = n^{-1} \sum_{i=1}^{n} Z_t^{(0)},$$

$$\hat{\rho}^{(0)} = \left(\sum\nolimits_{t=2}^{n} \left(Z_{t-1}^{(0)} - \overline{Z}_{-n}^{(0)} \right)^2 \right)^{-1} \left(\sum\nolimits_{t=2}^{n} \left(Z_t^{(0)} - \overline{\overline{Z}}_{-1}^{(0)} \right) \left(Z_{t-1}^{(0)} - \overline{\overline{Z}}_{-n}^{(0)} \right) \right), \text{ and}$$

$$\left(\hat{\sigma}^{(0)} \right)^2 = (n-3)^{-1} \sum_{t=2}^{n} \left(Z_t^{(0)} - \hat{\mu}^{(0)} - \hat{\rho}^{(0)} \left(Z_{t-1}^{(0)} - \hat{\mu}^{(0)} \right) \right)^2$$

where notation $\hat{\mu}^{(0)}$ denotes the estimated mean for iteration zero, $\overline{Z}_{-n}^{(0)} = (n-1)^{-1} \sum_{t=1}^{n-1} Z_t$ and similarly $\overline{Z}_{-1}^{(0)} = (n-1)^{-1} \sum_{t=2}^{n} Z_t$. By using Step 1, determine the initial estimates of parameter vector $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\hat{\boldsymbol{\mu}}^{(0)} = \hat{\mu}^{(0)}\mathbf{1}_n,$$
$$\{\boldsymbol{\Sigma}\}_{i,j} = \left\{ \left(\hat{\sigma}^{(0)}\right)^2 \Big/ \left(1 - \left[\hat{\rho}^{(0)}\right]^2\right) \right\} \hat{\rho}^{(0)|i-j|}, i,j = 1, \ldots, n,$$

Step 3 Compute the conditional mean and variance $\hat{\mathbf{M}}^{(0)}$and $\hat{\boldsymbol{V}}^{(0)}$ based on the censored part of the parameter vector $\boldsymbol{\Phi} = (\mathbf{M}, \boldsymbol{V})$ according to

$$\hat{\mathbf{M}}^{(0)} = \hat{\mu}_c^{(0)} + \hat{\Sigma}_{co}^{(0)}\left(\hat{\Sigma}_{oo}^{(0)}\right)^{-1}\left(\boldsymbol{Z}_o - \hat{\mu}_o^{(0)}\right) \text{ and,}$$

$$\hat{\boldsymbol{V}}^{(0)} = \hat{\Sigma}_{cc}^{(0)} - \hat{\Sigma}_{co}^{(0)}\left(\hat{\Sigma}_{oo}^{(0)}\right)^{-1}\hat{\Sigma}_{oc}^{(0)}$$

where covariance matrices are described in (6).

Step 4 Generate the vector $\boldsymbol{Z}_c^{(1)}$ of the right-censored observations from truncated normal distribution, $TN_{n_c}\left(\hat{\mathbf{M}}^{(0)}, \hat{\boldsymbol{V}}^{(0)}, R_c\right)$, where R_c denotes the interval $[C_t, \infty)$.

Step 5 By applying the instructions expressed in Eqs. (5) and (7), construct the following augmented data from the observed part and the vector $\boldsymbol{Z}_c^{(1)}$ defined in the previous step:

$$\mathbf{Z}^{(1)} = P^{-1}\begin{bmatrix} \boldsymbol{Z}_o \\ \boldsymbol{Z}_c^{(1)} \end{bmatrix}.$$

Step 6 Re-compute the estimates of the parameter μ, σ^2, and ρ according to $\mathbf{Z}^{(1)}$ and update the parameters $\boldsymbol{\Sigma}$, $\mathbf{M}$, and $\boldsymbol{V}$, as defined in Steps 2 and 3.

Step 7 Repeat the algorithm from Step 2 to Step 6 under the condition $\psi < 0.005$ (for our simulations). Note that ψ indicates the convergence ratio of the parameter estimates at the kth iteration, given by

$$\psi = \left[\left(\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right)^{\top}\left(\hat{\boldsymbol{\theta}}^{(k+1)} - \hat{\boldsymbol{\theta}}^{(k)}\right)\right] \Big/ \left(\hat{\boldsymbol{\theta}}^{(k)}\right)^{\top}\hat{\boldsymbol{\theta}}^{(k)} \tag{10}$$

where $\boldsymbol{\theta} = (\mu, \sigma^2, \rho)^{\top}$ denotes the parameter vector.

We should recall, as stated about the imputation algorithm, we are not restricted to a specific method for the parameter estimation. For the autoregressive time-series models, once the augmented data is obtained, any suitable method (for example, Yule–Walker, least squares, or maximum likelihood) can be used to estimate the parameters (see, Park et al. 2007). In this paper, we update the *smoothing splines* (*SS*), *kernel smoothing* (*KS*), and *regression spline* (*RS*) methods to estimate the parameters of the nonparametric autoregressive time- series model with right-censored data.

3 Modified Estimation Methods

In this section, we introduce three different modified estimation procedures to estimate the unknown function in the autoregressive time- series model, which is defined in (1). The modified methods are based on a generalization of the ordinary

SS, *KS*, and *RS* methods or conventional censored nonparametric regression models. Examples studies using mentioned methods include Zheng (1998), Dabrowska (1992), Fan and Gijbels (1994), Guessoum and Ould Said (2008), Aydin and Yilmaz (2017).

3.1 Smoothing Splines

Now we consider ways to estimate the unknown regression function $f(x)$ stated in model (1), where $E\left(Y_t|X_t = x\right) = f(x)$. The first way to approach to the nonparametric regression is to fit a spline with knots at every data point. The main idea is to find a regression function $f(.)$ that minimizes the penalized residual sum of squares (*PRSS*) criterion

$$PRSS(f,\lambda) = \sum_{t=1}^{n}\left\{Z_t^{(k)} - f\left(X_t\right)\right\}^2 + \lambda\int_a^b \left\{f''(x)\right\}^2 dx,\quad a \le x_1 \le \cdots \le x_n \le b \tag{11}$$

where $Z_t^{(k)}$is the response variable that provides the criterion (10) at the kth iteration, $f \in C^2[a,b]$ is a unknown smooth function, and λ is a positive smoothing parameter, controlling the tradeoff between the closeness of the estimate to the data and roughness of the function estimate. If $\lambda \to \infty$, the roughness penalty term dominates and the the parameter λ forces $f''(x) \to 0$, yielding the linear least squares estimate. If $\lambda \to 0$, the penalty term becomes negligible, and the solution tends to an interpolating estimate. Therefore, the choice of the smoothing parameter is an important problem and the GCV method has been used to address this problem in this paper. Note that minimizing the *PRSS* in (11) over the space of all continuous differentiable functions leads to a unique solution, and this solution is a natural cubic spline with knots at the unique values $x_1, \ldots, x_n$ for a fixed smoothing parameter λ.

Suppose that f is a natural cubic spline (NCS) with knot points $x_1 < \cdots < x_n$. By giving its value and second derivative at each knot points we can determine the following NCS vectors

$$f\left(X_t\right) = \left(f\left(x_1\right), \ldots, f(x_n)\right)^\top = (f_1, \ldots, f_n)^\top = \mathbf{f}$$

and

$$f''\left(X_t\right) = \left(f''\left(x_2\right), \ldots, f''\left(x_{n-1}\right)\right)^\top = \left(\vartheta_2, \ldots, \vartheta_{n-1}\right)^\top = \boldsymbol{\vartheta}$$

that specify the curve f completely. However, not all possible vectors $\mathbf{f}$ and $\boldsymbol{\vartheta}$ represent NCSs. In this sense the following theorem, discussed by Green and Silverman (1994), provides a condition for the vectors to be an NCS on the given knot points.

Theorem 3.1 (Green and Silverman 1994). *The vectors* $\mathbf{f}$ *and* $\boldsymbol{\vartheta}$ *determine an NCS f if and only if the following criterion*

$$\mathrm{Q}^\top\mathbf{f} = \mathrm{R}\boldsymbol{\vartheta} \tag{12}$$

is satisfied. If Eq. (12) *is satisfied, then the roughness penalty term will provide*

$$\int_a^b \{f''(x)\}^2 dx = \boldsymbol{\vartheta}^\top \mathrm{R} \boldsymbol{\vartheta} = \mathbf{f}^\top \mathbf{K} \mathbf{f} \tag{13}$$

A useful algebraic result from Theorem 3.1 *is that the penalty term in* (11) *may be written in quadratic form. In matrix and vector form, the criterion in* (11) *can be rewritten as*

$$PRSS(\mathbf{f}, \lambda) = \left\| \mathbf{Z}^{(k)} - \mathbf{f} \right\|_2^2 + \lambda \mathbf{f}^\top \mathbf{K} \mathbf{f}. \tag{14}$$

Notice that $\mathbf{K}$*in Eqs.* (13) *and* (14) *is a symmetric* $n \times n$ *positive definite penalty matrix that can be decomposed to*

$$\mathbf{K} = \mathbf{Q}^\top \mathbf{R}^{-1} \mathbf{Q}$$

where $\mathbf{Q}$ *is a tri-diagonal* $(n-2) \times n$ *matrix with elements* $\mathbf{Q}_{ii} = 1/h_i$, $\mathbf{Q}_{i,\,(i+1)} = \left(\frac{1}{h_i} + \frac{1}{h_{i+1}} \right)$, *and* $Q_{i,(i+2)} = 1/h_{i+1}$, $\mathbf{R}$ *is a symmetric tri-diagonal matrix of order* $(n-2)$ *with* $\mathbf{R}_{(i-1),i} = \mathbf{R}_{i,(i-1)} = h_i/6$, $\mathbf{R}_{ii} = \left(h_i + h_{i+1}\right)/3$, *and* $h_i = x_{i+1} - x_i$ *denotes the distance between successive knot points.*

For some constant $\lambda > 0$, the corresponding solution based on the smoothing splines (*SS*)for the vector $\mathbf{f}$, which specify the unknown smooth function f in (1), can be obtained by

$$\widehat{\mathbf{f}}_\lambda^{SS} = (\boldsymbol{\Sigma} + \lambda \mathbf{K})^{-1} \boldsymbol{\Sigma} \mathbf{Z}^{(k)} = \mathbf{S}_\lambda^{SS} \mathbf{Z}^{(k)} \tag{15}$$

where $\boldsymbol{\Sigma}$ is the n-dimensional regularity matrix for autocorrelated errors and $\mathbf{S}_\lambda^{SS} = (\boldsymbol{\Sigma} + \lambda \mathbf{K})^{-1} \boldsymbol{\Sigma}$ is a positive definite spline smoother matrix that depends on parameter λ. Details on the derivation of Eq. (15) can be found in the "Appendix 1".

3.2 Kernel Smoothing

As explained in the imputation algorithm, the idea is that, instead of Y_t, we consider the response variable $Z_t^{(k)}$, which is employed to estimate the censored observations. In this case, model (1) reduces to the following nonparametric regression model

$$Z_t^{(k)} = f\left(X_t\right) + e_t, \quad t = 1, \dots, n \tag{16}$$

This model can be considered as equivalent to the regression model with autoregressive errors. Consequently, kernel smoothing can be used as an alternative nonparametric approach to (*SS*) to get a suitable estimate of function f. Analogous to (15), this leads to the kernel regression estimator, in other words, (*KS*) introduced by Nadaraya (1964) and Watson (1964). Accordingly, the estimate of $f(.)$ at fixed x can be computed by

$$\widehat{\mathbf{f}}_{\lambda}^{KS}(x) = \sum_{t=1}^{n} \Sigma_t w_{\lambda}\left(x, X_t\right) Z_t^{(k)} = \mathbf{S}_{\lambda}^{KS} \mathbf{Z}^{(k)} \tag{17}$$

where $\mathbf{S}_{\lambda}^{KS}$ is a kernel smoother matrix defined by weights

$$s_{\lambda}\left(x, X_t\right) = K\left(\frac{X_t - x}{\lambda}\right) / \sum_{t=1}^{n} K\left(\frac{X_t - x}{\lambda}\right) = K(u) / \sum K(u) \tag{18}$$

with a bandwidth λ (also called a smoothing parameter), which is a nonnegative number determining the degree smoothness of $\widehat{\mathrm{f}}_{\lambda}^{KS}(x)$, as in ($SS$). In fact, it is possible to write the weighted sum (17) in a more general form

$$\begin{aligned} \widehat{f}_{\lambda}^{KS}(x) &= \frac{n^{-1} \sum_{t=1}^{n} K(u) Z_t^{(k)}}{n^{-1} \sum_{t=1}^{n} K(u)} \\ &= \frac{1}{n} \sum_{t=1}^{n} \left(\frac{K\left(\frac{X_t - x}{\lambda}\right)}{n^{-1} \sum_{t=1}^{n} K\left(\frac{X_t - x}{\lambda}\right)} \right) Z_t^{(k)} = \frac{1}{n} \sum_{t=1}^{n} \Sigma_t s_{\lambda}\left(x, X_t\right) Z_t^{(k)} \end{aligned} \tag{19}$$

In the expressions above, $K(u)$ is a real-valued kernel function assigning weights to each data point, and the kernel K satisfies the following conditions: $(u) \geq 0$, $\int K(u) du = 1$, and $K(u) = K(-u)$ for all $u \in R$. Moreover, Eq. (19) shows that the kernel regression estimator is a weighted average of the response observations $Z_t^{(k)}$. Also, as expressed above, the λ is a critical parameter in kernel regression estimation. A large bandwidth λ provides an over-smooth estimate, whereas a small bandwidth λ produces a wiggly function curve. For example, when $\lambda \to 0$, then $w_{\lambda}\left(x, X_t\right) \to n$, and hence, the estimator (17) reproduces the response observations (i.e., $\widehat{f}_{\lambda}^{KS}(x) \to Z_t^{(k)}$). When $\lambda \to \infty$, then $s_{\lambda}\left(x, X_t\right) \to 1$, and the estimator (17) convergences to the mean of the response observations (i.e., $\widehat{f}_{\lambda}^{KS}(x) \to \overline{Z}_t^{(k)}$). In this context, the GCV method is used in to determine the amount of smoothness required, as with the SS method.

3.3 Regression Spline

One alternative approach to estimating a function non-parametrically is to fit a qth-order RS (or penalized spline), which can be computed in terms of truncated power functions. To be specific, it is assumed that the unknown univariate function $f(.)$ can be estimated by a penalized spline with a truncated polynomial basis

$$f\left(X_t\right) = b_0 + b_1 X_t + b_2 X_t^2 + \cdots + b_q X_t^q + \lambda \sum_{r=1}^{m} b_{q+r} \left(X_t - \kappa_r\right)_+^q \tag{20}$$

where λ is a positive smoothing parameter, as in the other two approaches, q is the degree of spline, $\left\{\kappa_r,\ r = 1, \ldots, m\right\}$ are spline knot points,

$$\{b_0, b_1, \dots, b_q, b_{q+1}, \dots, b_{q+m}\}$$

is a set of regression coefficients, $(X_t - \kappa_r)_+^q = (X_t - \kappa_r)^q$ when $X_t > \kappa_r$ and otherwise $(X_t - \kappa_r)_+^q = 0$. It is also assumed that $\min(X_t) < \kappa_1 < \cdots < \kappa_m < \max(X_t)$, such that spline knot points $\{\kappa_1, \dots, \kappa_m\}$ represent a subset of $\{X_1, \dots, X_n\}$ (see, Ruppert 2002 for knot selection).

It follows from Eq. (20) that, for a truncated polynomial basis, the nonparametric time-series model (1) is

$$Z_t^{(k)} = b_0 + b_1 X_t + b_2 X_t^2 + \cdots + b_q X_t^q + \lambda \sum_{r=1}^{m} b_{q+r} (X_t - \kappa_r)_+^q + e_t, \quad t = 1, \dots, n \tag{21}$$

where $Z_t^{(k)}$ is the observed response variable, as with the previous two approaches. As displayed in the Eq. (21), the truncated qth-degree power basis with knots κ_r has basis vectors. In a matrix and vector form, model (21) can be rewritten as

$$\mathbf{Z}^{(k)} = \mathbf{X}\mathbf{b} + \boldsymbol{e} \tag{22}$$

where

$$\mathbf{Z}^{(k)} = \begin{bmatrix} Z_1^{(k)} \\ \vdots \\ Z_n^{(k)} \end{bmatrix}, \; = \begin{bmatrix} 1 & X_1 & \dots & X_1^q & (X_1 - \kappa_1)_+^q & \dots & (X_1 - \kappa_m)_+^q \\ & & & \vdots & & & \\ 1 & X_n & \dots & X_n^q & (X_n - \kappa_1)_+^q & \dots & (X_n - \kappa_m)_+^q \end{bmatrix}, \; \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{q+m} \end{bmatrix}, \text{ and}$$

$$\boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}.$$

For the penalized- spline- fitting problem, the key is to choose the vector $\mathbf{b}$ that minimizes

$$PRSS(\mathbf{b}, \lambda) = \sum_{t=1}^{n} \Sigma_t \left(Z_t^{(k)} - f(X_t) \right)^2 + \lambda \sum_{r=1}^{m} b_{q+r}^2 = \boldsymbol{\Sigma} \left\| \mathbf{Z}^{(k)} - \mathbf{X}\mathbf{b} \right\|_2^2 + \lambda \mathbf{b}^\top \mathbf{D}\mathbf{b} \tag{23}$$

where $\mathbf{D} = diag(\mathbf{0}_{(q+1)}, \mathbf{1}_m)$ is a diagonal penalty matrix whose first (q+1) elements are 0 and whose other elements are 1. Note that $\lambda \sum_{r=1}^{m} b_{q+r}^2 = \lambda \mathbf{b}^\top \mathbf{D}\mathbf{b}$ in (23) is referred to as a penalty term whose magnitude is determined by a positive smoothing parameter λ. It is clear that λ plays an important role in estimating the regression model. As with the (SS) and the (KS) expressed in the previous sections, λ is selected by a GCV criterion.

Minimization of the Eq. (23) yields to estimates of the vector $\mathbf{b}$, where

$$\hat{\mathbf{b}} = \left(\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X} + \lambda \mathbf{D} \right)^{-1} \mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{Z}^{(k)} \tag{24}$$

For (24) the fitted values $\hat{\mathbf{f}}$ based on the (RS) can be stated as

$$\hat{\mathbf{f}}_\lambda^{RS} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}\left(\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X} + \lambda \mathbf{D} \right)^{-1} \mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{Z}^{(k)} = \mathbf{S}_\lambda^{RS} \mathbf{Z}^{(k)} \tag{25}$$

where $\mathbf{S}_\lambda^{RS} = \mathbf{X}\left(\mathbf{X}^\top \boldsymbol{\Sigma} \mathbf{X} + \lambda \mathbf{D}\right)^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}$ is the smoother matrix based on the penalized spline that is dependent on a smoothing parameter λ, as in the other smoother matrices. Details on the derivation of Eq. (25) can be found in "Appendix 2".

The penalized spline method uses polynomial functions to best fit the data and includes a penalty term to prevent the overfitting problem. Thus, the method can produce more accurate estimates. As noted, the penalized spline fits the model with the help of specified nodes. These nodes need to be optimally selected. In this paper, the full search algorithm is used for selecting the nodes defined by Ruppert et al. (2003). For more details about knot selection using censored data, see the study of Aydin and Yilmaz (2017).

4 Evaluating the Performance of the Estimators

For a given smoothing parameter λ the regression estimators studied here (such as smoothing spline, kernel and regression spline) can be written in the form

$$\widehat{f}_\lambda(X_t) = \left(\widehat{f}_\lambda(x_1), \dots, \widehat{f}_\lambda(x_n)\right)^\top = \left(\left(\widehat{f}_\lambda\right)_1, \dots, \left(\widehat{f}_\lambda\right)_n\right)^\top = \widehat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \mathbf{Z}^{(k)} = \widehat{\mathbf{Y}} \tag{26}$$

where $\mathbf{Z}^{(k)} = (Z_1^{(k)}, \dots, Z_n^{(k)})'$ and $\mathbf{S}_\lambda$ is a $n \times n$ dimensional smoother matrix, as mentioned in previous sections. Note that this smoother matrix depends on $x_1, \dots, x_n$ and a parameter λ, but not on response variable $\mathbf{Z}^{(k)}$. It should also be expressed that the estimators defined in (26) is referred to as linear smoothers. The considered smoothers in this paper need a choice of smoothing parameter λ. In this sense, the mentioned parameter λ is chosen by minimizing the generalized cross validation (GCV) score (see Craven and Wahba 1979), given by

$$\mathrm{GCV}(\lambda) = \frac{n^{-1}\left\|\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\mathbf{Z}^{(k)}\right\|^2}{\left[n^{-1} tr\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\right]^2} = n \times \frac{Residual\ sum\ of\ squares(RSS(\lambda))}{(Equavalent\ degrees\ of\ freedom(EDF))^2} \tag{27}$$

When addressing the problem of smoothing parameter selection, an important issue is to have a good idea into bias and variance of the estimators, since a balance between these two quantities constitutes the core of selection criteria (27). As in the usual parametric inference, there are two main matters: systematic error that occurs in the form of bias and random error that occurs in the form of variance. The parameter λ that provides a balance between these two aspects is also obtained by minimizing the mean square error (MSE) of an estimation:

$$\begin{aligned} MSE(\lambda) &= E\left\|RSS(\lambda) = \left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\mathbf{Z}^{(k)}\right\|^2 \\ &= \left\|\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\mathbf{f}\right\|^2 + \sigma^2\left[n - 2\left(\mathbf{S}_\lambda\right) + \left(\mathbf{S}_\lambda \top \mathbf{S}_\lambda\right)\right] \end{aligned} \tag{28}$$

One often refers to the first term in (28) as the squared bias and to the second term as the variance. To include both of these components, we consider the MSE values of the estimators. Details on the derivation of the Eq. (28) can be found in the "Appendix 3".

Due to the optimal$MSE(\lambda)$ in (28) depend on unknown quantity of σ^2, and it is not directly applicable in practice. In this case, just as in the ordinary linear regression model, the variance σ^2 can be estimated by the residual sum of squares (27):

$$\hat{\sigma}^2 = \frac{RSS(\lambda)}{EDF} = \frac{(\mathbf{Z}^{(k)})^\top \left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)^2 (\mathbf{Z}^{(k)})}{tr\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)^{\mathbf{2}}} = \frac{(\mathbf{Z}^{(k)})^\top \left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)^2 (\mathbf{Z}^{(k)})}{n-p} \tag{29}$$

where

$$EDF = tr\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)^2 = (n-2) \times [tr\left(\mathbf{S}_\lambda\right)\ ]+ tr(\mathbf{S}_\lambda \top \mathbf{S}_\lambda) \tag{30}$$

Supposing that the bias term $\left\| \left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\mathbf{f} \right\|^2$ given in (28) is negligible, it turns out that $RSS(\lambda)/EDF$ is an unbiased estimate of σ^2. Hence, the EDF for residuals,$(n-p)$, is used to correct for bias, as in a linear model.

In general, a comparison of the estimators can be made by using a quadratic risk function that measures the expected loss of a vector $\hat{\mathbf{f}}_\lambda$. This so-called quadratic risk is given in the Definition 4.1. Our application of the results of the simulation experiments is to approximate the risk in the nonparametric autoregressive time series model with right censored. Such approximates have the advantage of being simpler to optimize the practical selection of smoothing parameters. For convenience, we will consider with the scalar valued mean square error.

Remark 4.1 The quadratic risk is closely related to the MSE matrix of an estimator $\hat{\mathbf{f}}_\lambda$. The scalar valued version of this MSE matrix is

$$SMSE\left(\hat{\mathbf{f}}_\lambda, \mathbf{f}\right) = E\left(\hat{\mathbf{f}}_\lambda - \mathbf{f}\right)\left(\hat{\mathbf{f}}_\lambda - \mathbf{f}\right)^\top = tr\left(MSE(\hat{\mathbf{f}}_\lambda, \mathbf{f})\right) \tag{31}$$

Remark 4.2 The covariance matrix for the fitted vector such as $\hat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \mathbf{Z}^{(k)}$ is specified as

$$Cov\left(\hat{\mathbf{f}}_\lambda\right) = \sigma^2(\ \mathbf{S}_\lambda \mathbf{S}_\lambda^\top) \tag{32}$$

Lemma 4.1 Consider different estimators $\hat{\mathbf{f}}_\lambda$. The scalar valued MSE values of these estimators can be defined as the sum of the covariance matrix and the squared bias vector:

$$\begin{aligned} SMSE\left(\hat{\mathbf{f}}_\lambda, \mathbf{f}\right) &= E\sum_{i=1}^{n} \left(\hat{f}_{\lambda i}(X_t) - f_i(X_t)\right)^2 = E\left\|\hat{\mathbf{f}}_\lambda - \mathbf{f}\right\|^2 \\ &= \left\|\left(\mathbf{I}\text{-}\mathbf{S}_\lambda\right)\mathbf{f}\right\|^2 + \sigma^2 tr(\ \mathbf{S}_\lambda \mathbf{S}_\lambda^\top) \end{aligned} \tag{33}$$

See "Appendix 4" for Proof of the Lemma 4.1

Note also that when we adopt the smoothing spline, the computation of $\mathbf{S}_{\lambda}^{SS} = (\mathbf{I} + \lambda\mathbf{K})^{-1}$ in (15) instead of $\mathbf{S}_{\lambda}$ as stated in Eqs. (from (26) to (33)) above is needed. In a similar fashion, for kernel smoothing and penalized spline methods, we have to calculate the $\mathbf{S}_{\lambda}^{KS}$ in (17) and $\mathbf{S}_{\lambda}^{RS}$ in (25) matrices, respectively. The traces of the matrices, $tr(\mathbf{S}_{\lambda}^{SS})$ $tr(\mathbf{S}_{\lambda}^{KS})$ and $tr(\mathbf{S}_{\lambda}^{RS})$ can be found in O(n) algebraic operations, and hence, these matrices can be calculated in only a linear time.

The first term in (33) measures squared bias while the second term measures variance. Hence, we can compare the quality of two estimators by looking at the ratio of their *SMSE*. This ratio gives the following definition concerning the superiority of any two estimators.

Definition 4.1 The relative efficiency of an estimator $\hat{\mathbf{f}}_{\lambda}^{E1}$ compared to another estimator $\hat{\mathbf{f}}_{\lambda}^{E2}$ is defined by

$$RE = \frac{R\left(\hat{\mathbf{f}}_{\lambda}^{E1}, \mathbf{f}\right)}{R\left(\hat{\mathbf{f}}_{\lambda}^{E2}, \mathbf{f}\right)} = \frac{SMSE\left(\hat{\mathbf{f}}_{\lambda}^{E1}, \mathbf{f}\right)}{SMSE\left(\hat{\mathbf{f}}_{\lambda}^{E1}, \mathbf{f}\right)} \tag{34}$$

where $R(.)$ denotes the scalar risk that is equivalent to the Eq. (33). $\hat{\mathbf{f}}_{\lambda}^{E2}$ is said to be more efficient than $\hat{\mathbf{f}}_{\lambda}^{E1}$ if $RE < 1$.

Additional to evaluating the scale-dependent measures based on squared errors expressed above, we have also examined the alternative accuracy measures based on prediction errors to compare the smoothing methods. The most commonly used measures are briefly defined in the following way (see study of Chen and Yang 2004; Gooijer and Hyndman 2006; Chen et al. 2017 for more detailed discussion):

The mean absolute percentage error, *MAPE* is described by

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \frac{\left| \mathbf{Z}_t^{(k)} - \hat{Y}_t \right|}{\mathbf{Z}_t^{(k)}} \tag{35}$$

KL-N measure is computed based on the Kullback–Leibler (*KL*) divergence. It corresponds to the quadratic loss function scaled with variance estimate, and is given by the formula

$$KL - N = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \frac{\left(\hat{Y}_t - \mathbf{Z}_t^{(k)}\right)}{S_t^2}}, \quad \text{with} \;\; S_t^2 = \left(\frac{1}{t-1}\right) \sum_{j=1}^{t-1} \left(\mathbf{Z}_j^{(k)} - \overline{\mathbf{Z}}_{t-1}^{(k)}\right)^2 \tag{36}$$

where $\overline{\mathbf{Z}}_{t-1}^{(k)}$ is the mean of the first (t-1) value of variable $\mathbf{Z}_t^{(k)}$.

Inter quartile range, *IQR* is calculated by the formula

$$IQR = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \frac{\mathbf{Z}_t^{(k)} - \hat{Y}_t}{Iqr^2}} \tag{37}$$

where *Iqr* is the inter-quantile range of the vector $\boldsymbol{Z}_t^{(k)}$ defined as the difference between the third quantile and first quantile of the data.

The relative squared error, *RSE* is defined by

$$RSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{\left(\boldsymbol{Z}_t^{(k)} - \widehat{Y}_t\right)^2}{\left(\boldsymbol{Z}_t^{(k)} - \boldsymbol{Z}_{t-1}^{(k)}\right)^2}} \tag{38}$$

The relative absolute error, *RAE* is described by the formula

$$RAE = \sqrt{\sum_{t=1}^{n}\frac{\left|\boldsymbol{Z}_t^{(k)} - \widehat{Y}_t\right|}{\left|\boldsymbol{Z}_t^{(k)} - \boldsymbol{Z}_{t-1}^{(k)}\right|}} \tag{39}$$

It should be noted that the response variable $\boldsymbol{Z}_t^{(k)}$ obtained iteratively and their estimation values ($\widehat{Y}_t$) as expressed in Eqs. (from (35) to (39)) above are needed to calculate based on each smoothing methods, *SS*, *KS*, and *RS*.

5 Simulation Study

As indicated, the simulation studies were conducted to compare the estimation performances of the updated smoothing methods *SS*, *KS*, and *RS* defined in Sect. 3. For all the simulation studies, we considered a right-censored autoregressive model of order $p = 1$ (i.e. the, $AR(1)$- model) with one explanatory variable. The data- generating process from the model defined in (1) is as follows.

Step 1 The explanatory variable X_t is generated from Uniform distribution $X_t \sim \alpha U[0, 1]$, where α is the multiplier that determines the spatial variation for function $f(.)$. Note that the values of α are $\alpha_1 = 6.4$ (meaning this function has two peaks) and similarly $\alpha_2 = 12.8$ (meaning this function has four-peaks).

Step 2 Unknown smooth function $f(.)$ is determined as $f\left(X_t\right) = \sin\left(X_t\right)^2 + \omega,\ \ \omega > 1$, where ω is a constant that prevents the unidentifiability caused by log -transformation.

Step 3 Random error terms are generated as $e_t = \phi e_{t-1} + u_t$, where $u_t \sim N\left(\mu = 0, \sigma^2 = 1\right)$ and $\phi = 0.7$.

Step 4 The completely observed response values are generated using the expressions defined in Steps 1, 2 and 3: $Y_t = f\left(X_t\right) + e_t$.

Step 5 In this simulation setup, the censoring procedure of the response variable is stated in the following way:

1. Three censoring levels are considered, $\eta = 5\%,\ 20\%,\ 40\%$;.
2. For this simulation's purposes, the cutoff value c is determined by (see, Park et al. 2007):

$$c = \mu + \sigma \frac{F^{-1}(1-\eta)}{\sqrt{1-\phi^2}} \sqrt{1-\phi^{2(n+1)}}$$

where η is the censoring probability expressed as $\eta = P(Y_t > c)$, $F(.)$ represents the standard normal distribution function, ϕ is the autocorrelation parameter, and $\sqrt{1-\phi^{2(n+1)}}$ is the correction term for the finite sample sizes;.

1. After deciding cutoff point *c*, censored time-series C_t can be generated as

$$C_t = Y_t\left(1 - I\left(Y_t > c\right)\right) + c.I\left(Y_t > c\right), \quad t = 1, \ldots, n;$$

2. Hence, the new incompletely observed response measurements Z_t are constructed by means of Eq. (4). However, because of the censoring, these measurements cannot be applied directly. To overcome this problem, we used the variable $Z_t^{(k)}$ that provides the criterion (10) at the $k^t h$ iteration, as described in the imputation algorithm.

For each simulation setting, we generated 1000 random samples of sizes $n = 25, 50, 100,$ and 300 based on censoring levels and the Steps expressed above. Notice that the censored variable $Z_t^{(k)}$is modeled by $AR(1)$, which is considered a naïve model here, , in order to compare the finite sample performances of the proposed methods.

5.1 Results from the Simulations

In the simulation study, many simulated configurations were implemented to provide perspective on the adequacy of the smoothing methods examined in Sect. 3. Because 36 different configurations were analyzed, it was not possible to display the details of each configuration. Therefore, the main outcomes from the Monte Carlo simulation study, performed under different conditions, are summarized numerically and graphically. In general, the tables and figures are based on the comparison of fits from the methods under varying sample sizes and censoring levels.

Each panel in Fig. 1 presents a single realization of simulated censored and uncensored observations and therefore, two different curves. A cutoff point is determined for each simulation setting, and response observations are then censored according to this cutoff value. Hence, a nonparametric estimation of the regression functions in the case of right-censored data can be obtained via smoothing methods using the iterated response variable stated in the imputation algorithm. The outcomes from this simulation study are presented in the following tables and figures.

Table 1 displays the results for a minimal censoring level (5%). When the performance measures are inspected, it is shown that the three smoothing methods have considerable results, which is expected at a low censoring level. In addition, Table 1 shows that the proposed methods work well compared to the $AR(1)$ model, even at the lowest censoring levels. The best scores are indicated in bold color. Table 1 can

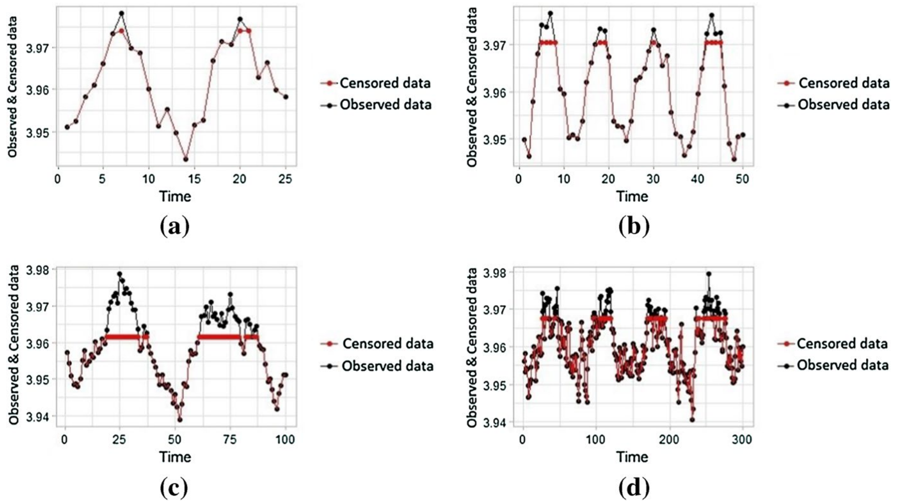

Fig. 1 The panels show the censored and uncensored time-series observations from functions simulated under varying conditions: **a** for n = 25 and $\eta = 5\%$; **b** for n = 50 and $\eta = 20\%$; **c** for n = 100 and $\eta = 40\%$; **d** for n = 300 and $\eta = 20\%$

be inspected from two perspectives. One is by the shapes of the function, and the other is by the sample size. In this study, to measure the strength of the smoothing methods against the fluctuations in the function, two and four-peaked functions are used, and these peaks are censored by a determined detection limit. In this case, all of the obtained outcomes are presented in Tables 1, 2 and 3. When the censoring level is low and all the results are considered, *KS*has the highest number of colored scores (20/40), followed by. *RS*(19/40) and *SS*(15/40). Numbers in parentheses represent the number of bold scores out of the total combinations. Note that usually, the values of *KS*and *SS* are close or equal in all three tables. When results are examined in terms of sample sizes, for small samples, *RS* appears more efficient than the others. For medium samples ($n = 50$ and $n = 100$), *KS* and *SS* have reasonable results, and in large samples, *SS* performs the best. When the results are analyzed in terms of spatial variation, it is clear that most of the methods' performance decreases when fluctuation is increased, but the results are still satisfactory. The methods can be interpreted as being resistant to these kinds of fluctuations. However, the *RS* shows a difference in modeling the multi-peak function. This result is not unexpected because *RS*has some advantages in knot selection. These results can easily be observed in Figs. 2 and 3.

In Table 2, the effect of the censoring level can be seen distinctly when compared with Table 1. All of the performance measures increased. In general, *RS* (22/40) showed better results than *KS* and *SS*. The results for the others were (18/40), (16/40), and (1/40) for *KS*, *SS* , *and AR*(1), follow *RS* respectively. Surprisingly, although *RS* is last at a low censoring level, it is the best when the censoring level increases. This result might be worth examining further. *RS* is possibly better in estimating censored time-series in comparison with *KS* and *SS*. As expected, *RS* is at

Table 1 Results based on 1000 simulated samples with all sample sizes (n) and regression functions for censoring level $\eta = 5\%$

η =5%			Two-peak ($\alpha = 6.4$)						Four-peak ($\alpha = 12.8$)			
n	*Meth.*	MSE	MAPE	KL-N	IQR	RSE	RAE	MSE	MAPE	KL-N	IQR	RSE	RAE
25	*SS*	0.0619	0.0233	0.6473	4.3347	4.9430	1.4773	0.0795	0.0244	0.9499	4.3744	5.9363	1.9842
	KS	0.0569	0.0233	0.6473	4.3346	4.9421	1.4573	0.0806	0.0244	0.9499	4.3744	5.9363	1.9842
	RS	0.0561	0.0225	0.6329	4.3332	4.9546	1.4656	0.0766	0.0213	0.9510	4.3495	5.9286	1.8591
	AR(1)	0.0997	0.0299	0.8160	4.6913	5.1160	2.2279	0.1967	0.0738	1.0730	4.5342	6.1395	2.1820
50	*SS*	0.0390	0.0201	0.8849	3.2515	4.5260	0.9229	0.0463	0.0225	0.9017	3.3343	4.8751	1.6415
	KS	0.0388	0.1701	0.8851	3.2515	4.5258	0.9138	0.0412	0.0225	0.9015	3.3343	4.8576	1.6402
	RS	0.0384	0.0202	0.8849	3.2545	4.4417	0.9108	0.0463	0.0227	0.0966	3.3683	4.8705	1.6451
	AR(1)	0.0629	0.0263	0.9140	3.5882	4.8246	1.6942	0.1794	0.0449	1.0222	0.7765	0.2014	0.4758
100	*SS*	0.0318	0.0201	0.8110	3.2237	3.4531	0.8448	0.0426	0.0209	0.6699	3.3041	4.5336	1.5548
	KS	0.0317	0.0201	0.8092	3.2229	3.4387	0.7814	0.0375	0.0209	0.6702	3.3041	4.5336	1.5543
	RS	0.0319	0.0203	0.8060	3.2244	3.4596	0.8144	0.0425	0.0192	0.5789	3.2975	4.5420	1.7252
	AR(1)	0.0623	0.0264	1.4669	3.5937	3.8448	1.1077	0.0668	0.0265	0.6594	3.6561	4.7912	1.8374
300	*SS*	0.0253	0.0193	0.7263	3.1228	3.3888	0.7387	0.0357	0.0207	0.5219	3.2990	4.4518	1.3033
	KS	0.0255	0.0198	0.7228	3.1333	3.3356	0.7612	0.0349	0.0209	0.5220	3.2990	4.4851	1.4740
	RS	0.0265	0.0196	0.7559	3.1349	3.3733	0.7571	0.0362	0.0209	0.5543	3.2990	4.4211	1.3033
	AR(1)	0.0421	0.0204	0.8728	3.4943	3.4780	0.8286	0.0494	0.0220	0.6981	3.5167	4.6407	1.6874

Table 2 Outcomes from the simulated examples for all sample sizes and regression functions under the censoring level $\eta = 20\%$

η =20%		Two-peak ($\alpha = 6.4$)						Four-peak ($\alpha = 12.8$)					
n	*Meth.*	MSE	MAPE	KL-N	IQR	RSE	RAE	MSE	MAPE	KL-N	IQR	RSE	RAE
25	*SS*	0.0954	0.0319	1.3247	5.4668	4.2788	2.4400	0.1145	0.0327	1.4425	5.5434	4.3633	2.4741
	KS	0.0936	0.0319	1.3247	5.4668	4.2788	2.4400	0.0973	0.0326	1.4426	5.5434	4.3633	2.4553
	RS	0.0954	0.0309	1.3413	5.4593	4.2533	2.4611	0.1148	0.0321	1.4450	5.5404	4.2676	2.4501
	AR(1)	0.1533	0.0402	1.4999	5.7280	4.3829	2.5312	0.2061	0.0549	1.4795	6.0372	5.1098	3.1600
50	*SS*	0.0821	0.0319	1.2459	5.3353	4.1207	2.2754	0.0917	0.0309	1.2740	5.3616	4.3292	2.3003
	KS	0.0842	0.0319	1.2459	5.3353	4.1206	2.2752	0.0863	0.0309	1.2740	5.3616	4.3294	2.3003
	RS	0.0821	0.0322	1.2443	5.3370	4.1045	2.2539	0.0847	0.0312	1.3996	5.3565	4.2918	2.3001
	AR(1)	0.0854	0.0360	1.6316	5.6642	4.3660	2.4191	0.1245	0.0512	1.8805	5.7260	4.6003	2.7809
100	*SS*	0.0781	0.0282	1.1382	4.3238	3.9762	1.3078	0.0811	0.0283	1.2148	4.4045	4.2123	1.6202
	KS	0.0770	0.0282	1.1213	4.3143	3.9771	1.3078	0.0810	0.0283	1.2155	4.4045	4.2121	1.6129
	RS	0.0786	0.0286	1.1654	4.3462	3.9647	1.3024	0.0801	0.0294	1.1840	4.4152	4.6194	1.6335
	AR(1)	0.0820	0.0319	1.3224	4.5928	4.0322	1.6298	0.0920	0.0309	1.2609	4.7009	4.7711	1.8807
300	*SS*	0.0663	0.0262	1.0550	4.2410	3.8468	0.9462	0.0772	0.0270	1.1886	4.3952	4.1156	0.9994
	KS	0.0510	0.0271	1.1938	4.2466	3.7708	0.9137	0.0747	0.0274	1.2050	4.3955	4.1175	1.0093
	RS	0.0584	0.0312	1.1715	4.2719	3.8171	0.8229	0.0701	0.0269	1.1394	4.3946	4.1657	1.2175
	AR(1)	0.0621	0.0315	1.1521	4.5691	3.8447	1.2020	0.0776	0.0289	0.9774	4.5594	4.2181	1.2360

Table 3 Results from the simulations for all sample sizes and all functions under the censoring level $\eta = 40\%$

η =40%		Two-peak ($\alpha = 6.4$)						Four-peak ($\alpha = 12.8$)					
n	*Meth.*	MSE	MAPE	KL-N	IQR	RSE	RAE	MSE	MAPE	KL-N	IQR	RSE	RAE
25	*SS*	0.6103	0.0401	1.5129	5.6074	4.4741	2.7991	0.7524	0.0522	1.7860	5.6842	4.5460	3.3245
	KS	0.6103	0.0401	1.5129	5.6074	4.4740	2.7990	0.7380	0.0522	1.7860	5.6842	4.5460	3.3245
	RS	0.5052	0.0360	1.5088	5.5108	4.4671	2.7433	0.7152	0.0520	1.5669	5.6821	4.6201	3.3855
	AR(1)	1.1668	0.0461	1.8998	5.6721	5.1503	3.2283	1.2367	0.0770	1.7596	5.8682	5.9158	4.6549
50	*SS*	0.2369	0.0383	1.4236	5.4201	4.3462	2.6952	0.4980	0.0466	1.4295	5.6210	4.4362	2.8977
	KS	0.2459	0.0383	1.4138	5.4201	4.3462	2.6952	0.4794	0.0466	1.4296	5.6210	4.4362	2.8977
	RS	0.2322	0.0391	1.4005	5.4209	4.3496	2.8020	0.4608	0.0399	1.2427	5.6222	4.4948	3.0864
	AR(1)	0.7055	0.0466	2.6540	5.5936	4.6044	3.5285	0.7878	0.0426	1.4004	5.7232	5.1689	3.4166
100	*SS*	0.1572	0.0308	1.1952	5.3809	4.1388	1.4055	0.1422	0.0351	1.2690	5.5673	4.3855	1.9796
	KS	0.1256	0.0309	1.1520	5.3811	4.1406	1.4056	0.1236	0.0351	1.2690	5.5672	4.3855	1.9794
	RS	0.1087	0.0305	1.1076	5.3790	4.1089	1.3417	0.1105	0.0375	1.3153	5.5584	4.3643	1.9401
	AR(1)	0.5591	0.0345	1.4197	5.5859	4.9682	2.3095	0.4870	0.0418	1.4081	5.7046	4.6060	2.2233
300	*SS*	0.0928	0.0271	1.0827	5.2767	4.1212	1.0666	0.0986	0.0273	1.1769	5.5246	4.1895	1.4907
	KS	0.0912	0.0299	0.9848	5.2921	4.1216	0.9719	0.0968	0.0273	1.2139	5.5246	4.1840	1.4488
	RS	0.0917	0.0294	0.9495	5.3022	4.0699	0.9535	0.0949	0.0275	1.2971	5.5154	4.1733	1.5250
	AR(1)	0.3258	0.0354	1.6633	5.5989	4.3519	1.4004	0.3948	0.0288	1.5992	5.6029	4.5226	1.8259

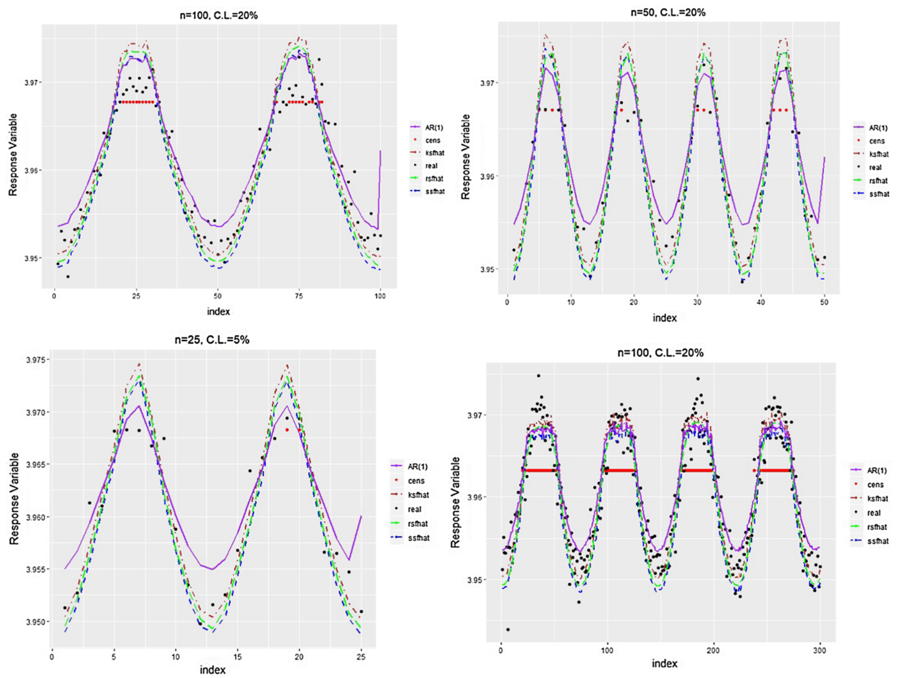

Fig. 2 The plots of simulated observations, right-censored data points, and fitted curves from an autoregressive nonparametric time-series model using a classical AR(1) model and three smoothing methods for different simulation configurations, with red points (.) denoting the censored points

the front in modelling the four-peaked function. Moreover, *KS* and *SS* have estimates for small and medium sample sizes that are more efficient, but for large samples, *KS* cannot match *RS*and *SS*.

The results of Table 3 support the inferences gathered in examining Table 2. Namely, *RS* can model the censored time series better than the other two methods in the scope of this simulation study. It is observed that, when the censoring level increases, the performance of *RS*also increases. Similar to other censoring levels, *RS* and *SS* produce better results for large samples, and for medium samples, *KS* and *SS* both perform reasonably well. Note that $AR(1)$ has the worst performance for almost all of the simulation configurations and that the, performance of $AR(1)$ is unsatisfying under heavy censoring levels, which is due to it being a linear and parametric method, as displayed in Tables 1, 2 and 3 and Figs. 2 and 3. Also note that Figs. 2 and 3 show the fits from the smoothing methods and an the $AR(1)$ model for some selected combinations, as defined in Tables 1, 2 and 3, and the graphically summarized results support the outcomes expressed in the tables.

Figure 2 The plots of simulated observations, right-censored data points, and fitted curves from an autoregressive nonparametric time- series model using a classical $AR(1)$ model and three smoothing methods for different simulation configurations. Red points (.) denote the censored points..

Figure 3 the same as Fig. 2 but for different combinations

Figs. 2 and 3 prove that the augmented method works well in estimating right-censored time series under different scenarios. The three smoothing methods perform satisfactorily, which is further illustrated in Tables4 and 5. It is worth noting that even if the censoring rate increases, the imputed values are still closer to real observations. To inspect this simulation study, see the web application available at https://ey13.shinyapps.io/right-cens-timeseries/, which dynamically reacts to different combinations of sample size, censoring level, and spatial variation.

As indicated in Sect. 4, we considered four different criteria for evaluating the performance of the proposed methods in this paper. However, especially for the simulation experiments, it was shown that the scores obtained from *IQR* and *MSE* criteria are very similar. Consequently, paired Wilcoxon tests were used to determine whether the difference between the median values of *MSE*s (or median values of *IQR*s) of any two criteria was significant at the 5% level. For example, if the median *MSE* value of a method is significantly less than the remaining three, it will be ranked first. If the median*MSE* value of a method is significantly larger than one of the methods but less than the remaining two, it will be ranked second, and so on ranked third. Methods having non-significantly different median values will share the same averaged rank. Note also that the method or methods having the smallest rank are superior. Similar interpretations can be made for the *IQR* criterion.

The Wilcoxon test results for *IQR* and *MSE* are given in Tables4 and 5. Note that Tables 4 and 5 are constructed from the averaged ranking values of *IQR*s and *MSE*s, respectively, according to Wilcoxon tests for each censoring level and sample size.

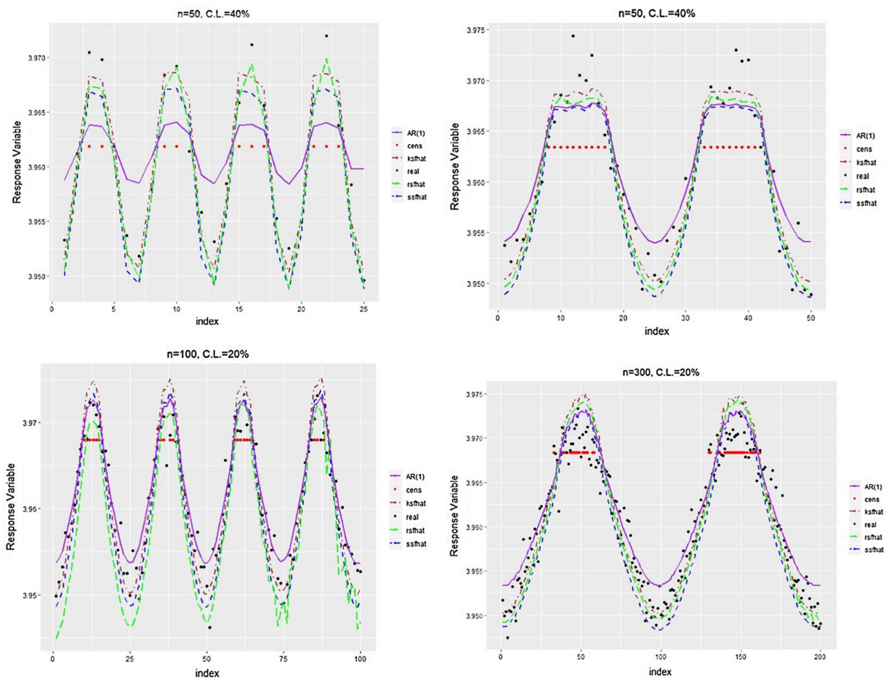

Fig. 3 The same as Fig. 2 but for different combinations

Table 4 Averaged Wilcoxon test ranking values based on *IQR* criterion for the four methods under each sample size and censoring level

Meth./n	Two-peak (α = 6.4)				Four-peak (α = 12.8)				Overall
	25	50	100	300	25	50	100	300	
5%									
SS	2.50	3.50	3.50	3.00	2.50	3.00	2.00	2.00	2.7500
KS	2.50	2.00	2.00	1.50	2.50	3.00	1.00	1.00	1.9375
RS	1.00	1.00	1.00	1.50	1.00	1.00	3.50	3.00	1.6250*
AR(1)	4.00	3.50	3.50	3.00	4.00	3.00	3.50	4.00	3.5625
20%									
SS	2.50	3.00	2.50	3.00	2.50	1.00	3.00	3.00	2.5625
KS	2.50	2.00	1.00	1.50	2.50	2.00	2.00	2.00	1.9375
RS	1.00	1.00	2.50	1.50	1.00	3.00	1.00	1.00	1.5000*
AR(1)	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.0000
40%									
SS	2.50	3.00	2.50	3.00	2.50	2.00	2.50	3.00	2.6250
KS	2.50	2.00	2.50	2.00	2.50	2.00	2.50	1.50	2.1875
RS	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.50	1.1875*
AR(1)	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.0000

* denotes the best score

Wilcoxon test results for the remaining criteria ($MAPE$, $KL-N$, RSE and RAE) are not provided here to save space. For given simulation configuration, *RS* appears to be better than *KS* and *SS* according to the results obtained by 1000 repetitions. The overall scores of the tables further show that *RS* has the smallest median values in terms of *IQR* and *MSE* criteria. The *SS* and *KS* give similar ranking points. Moreover, as the level of censorship increases, the difference between the points of the models increases. Note also that*KS* and *SS* generally have similar scores, but *KS* performs better. Lastly, according to, the overall Wilcoxon test rankings in Tables 4 and 5, $AR(1)$ has performed the worst.

5.2 Efficiency Comparisons

In order to illustrate and compare the relative efficiencies of the methods based on censored data, the bar charts of relative efficiencies computed from *RE* values are displayed in Fig.4. It is clearly shown that efficiencies ensure the outcomes in Tables 1, 2 and 3. As explained above, smaller efficiency values indicate a more efficient method. If Fig. 4 is inspected carefully, one can conclude that the *KS* has smaller values for small samples and low censoring levels, but when the censoring level begins to increase, the efficiency values of the *RS* decrease, which means the *RS*'s performance improves. The *SS* provides stable performance throughout the simulation experiments.

As seen in Fig. 4, efficiency values of the *KS* method are smaller than 1 for all censoring levels and sample sizes. Hence, *KS* method is more efficient than the other

Table 5 Averaged Wilcoxon test ranking values based on MSE criterion for the four methods under each sample size and censoring level

Meth./n	Two-peak (α = 6.4)				Four-peak (α = 12.8)				Overall
	25	50	100	300	25	50	100	300	
5%									
SS	2.50	2.50	3.00	3.00	2.50	3.00	2.50	3.00	2.7500
KS	2.50	2.50	3.00	1.50	2.50	3.00	2.50	3.00	2.5625
RS	1.00	1.00	1.00	1.50	1.00	1.00	2.50	1.00	1.2500*
AR(1)	4.00	4.00	3.00	3.00	4.00	4.00	2.50	3.00	3.4375
20%									
SS	2.50	2.00	2.00	2.00	2.50	2.00	2.00	2.50	2.1875
KS	2.50	3.50	3.00	3.00	2.50	3.50	3.00	2.50	2.9375
RS	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.0000*
AR(1)	4.00	3.50	4.00	4.00	4.00	3.50	4.00	4.00	3.8750
40%									
SS	2.50	2.00	2.50	2.00	2.50	2.50	2.00	1.50	2.1875
KS	2.50	3.00	2.50	3.00	2.50	2.50	2.00	2.50	2.5625
RS	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00	1.1250*
AR(1)	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.0000

* denotes the best score

methods. When *KS*, *RS,* and *SS* are compared to the benchmark *AR*(1), it is shown that the *RE* values obtained from *AR*(1) increase. This means that *AR*(1) has the worst efficiency in terms of *SMSE* values.

6 Real-Data Studies

6.1 Cloud-Ceiling Data

This section more explicitly examines the estimation performances of the three smoothing methods for right-censored time-series data by modeling cloud-ceiling data. Cloud-ceiling data were collected hourly by the National Center for Atmospheric Research (NCAR) in San Francisco throughout March 1989. A dataset of

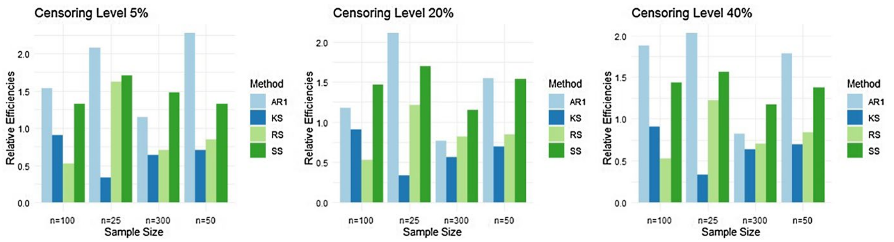

Fig. 4 The column chart provides the relative efficiencies from the smoothing methods *SS*, *KS*, *RS* and *AR*(1)

716 observations was formed. The scale of the data was originally 100 feet, but here, a logarithmic transformation was made. It is visualized in Fig. 5. As mentioned at the beginning of this paper, sometimes a measurement tool has a detection limit. For this data, this limit is 12,000 feet (4.791 in log-transformed data); therefore, this time series is considered to be right censored. This means that each observation that is larger than the detection limit is censored. In the dataset, 293 observations are censored, and consequently, the censoring rate is 41.62%.

In Fig. 5, the points in red represent the censored observations. The following tables and figures summarize the findings of the three smoothing methods for estimating right-censored time series. Note that the imputation method can also provide an advantage in estimating censored data when there is no detection limit in the process of obtaining the data.

The performance measures for the three smoothing methods and the naïve $AR(1)$ model, are presented in Table 6. The method with the best performance is indicated in bold. In this case, the KS had reasonable results according to almost all performance measures, and the KS is likely the best method for this type of censored time-series data. However, one must note that the RS shows its superiority for both absolute and relative performance measures, but in the general framework, the KS dominated the other two methods, while the $AR(1)$ gave the highest scores for all evaluation metrics. As previously mentioned, the first three criteria for the estimated models measure the absolute error and the last two illustrate the relative error. In Tabl 6, one clearly sees that the RS performs better than the SS according to the first three criteria, which are $MAPE$, $KL-N$, and IQR, and, when RSE andRAE are inspected, it is clear that the performance of the RS is better than the SS.

Figure 6 represents the estimated values for the nonparametric regression model for the three smoothing methods, with the bottom right of the figure showing the observed real data versus augmented data points. The figure shows that the three methods all perform satisfactorily. When comparing the methods using

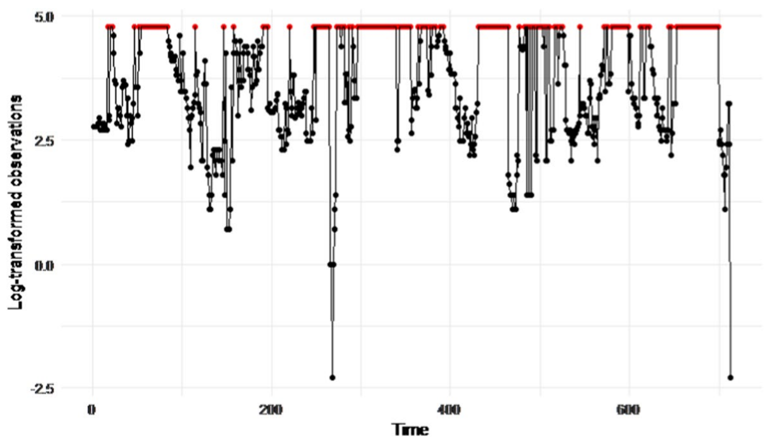

Fig. 5 Time plots of the log-transformed hourly cloud-ceiling height values with right-censored time series

this cloud-ceiling dataset, Table 6 is more useful than Fig. 6; therefore, as mentioned, the *KS* is superior in estimating right-censored time series.

With the real data, the weakness of the *RS* was that it uses determined knots. For this kind of complex data, optimizing the number of knots can be difficult. In this study, a full search algorithm was used to select the number of knots, which was proposed by Ruppert et al. (2003). Furthermore, the *KS*gives weight to observations, which is one of the reasons for its success with this dataset.

6.2 Unemployment Duration Data

A second real dataset, this time an econometric time-series, was used to analyze the performances of the three methods The unemployment rates in Turkey were estimated using the three smoothing methods and the $AR(1)$ model. This data set comprises monthly unemployment rates between 2004 and 2019 (see https: //ec.europa.eu/eurostat/data/database). Note that the data from 2004 and the last three months of 2019 cannot be observed correctly. Since these data cannot be negative, they can be censored from the right with zero as detection limit. Therefore, the nonparametric autoregressive model applied to the data is as follows:

$$UE_t = f(UE_{(t-1)}) + \varepsilon_t, \quad t = 1, \dots, 186 \tag{40}$$

where UE_t's represent the ratios of unemployment duration depend on time, $UE_{(t-1)}$ is the first lag of the response variable (UE_t) and ε_t's are the random error terms with $\varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right)$. The outcomes from the estimated model (40) are given in Table 7 and Fig. 7, in the same manner as the cloud- ceiling data example in Sect. 6.1.

The first three panels of Fig. 7 includes estimated curves obtained by the three smoothing methods and at the bottom-right panel includes censored data (black) with augmented data points (red) similar to Fig. 6. It is clearly shown that the curve estimated by the *RS* represents the data better than the other two. This case can be explained by the wiggly structure of the dataset. Because the *RS* is working with determined knot points, it can adapt the data structure more easily than *SS*and *KS*. Note that the, data are are estimated by the $AR(1)$ model, but it cannot give comparable results with introduced nonparametric techniques, which can be seen in Table 7.

The scores given in Table 7 indicate that the *RS* is the best qualified method for econometric time-series. For this dataset, the *KS* performed the second best. By contrast, the *SS* and $AR(1)$ provide unsatisfactory results with respect to all evaluation criteria. Lastly, one should note that the *KS* and the *RS* performed similarly, which demonstrates that the two real- data examples give compatible outcomes.

Table 6 The outcomes from the smoothing methods and AR(1) model

	MSE	MAPE	KL-N	IQR	RSE	RAE
SS	0.0849	0.0797	27.3869	0.1831	1.2113	64.1795
KS	0.0722	0.0502	24.8250	0.1174	0.9361	30.1658
RS	0.0761	0.0586	23.9862	0.1471	0.8693	43.3855
AR(1)	0.1052	0.2070	81.2845	0.2784	1.7426	69.5116

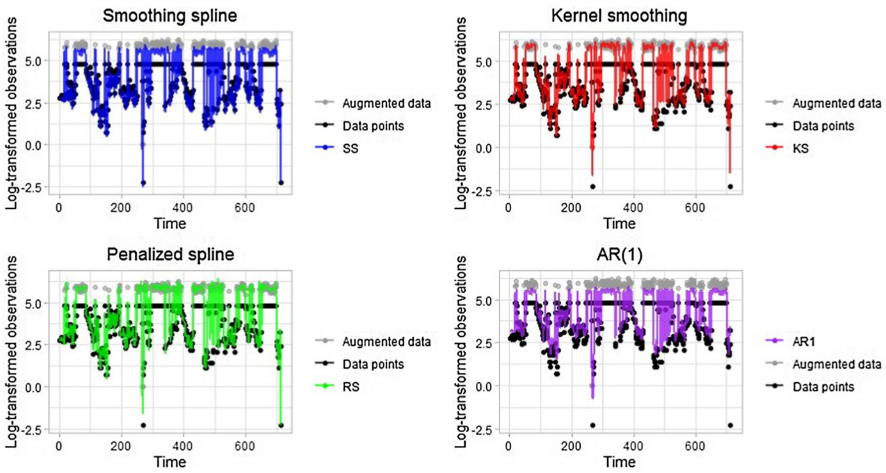

Fig. 6 Time plots of the log-transformed observations, including right-censored time series, and the augmented observations and fitted curves from an autoregressive nonparametric time-series model using three smoothing methods

7 Conclusions and Recommendations

This paper used the imputation method , to estimate right-censored time series with a nonparametric regression model, as initially proposed by Park et al. (2007). Three different smoothing methods were to estimate the nonparametric regression model: *SS*, *KS*, and *RS*. To realize our aim, simulation experiments were performed and a real-data examples were presented. The results obtained from the simulation study and the real-data example are presented in the tables and figures. The nonparametric regression model was found to be quite a useful for estimating right-censored time series. This paper compared the three smoothing methods. Therefore, the results show which method was the best under different conditions.

The empirical results confirmed that the *KS* and the *RS* perform reasonably well for different combinations of sample sizes, function types, and censoring rates. As expected, when the censoring rate increases, the quality of the estimation worsens. In general, the three smoothing methods provided satisfactory results, but there are some advantages and disadvantages for each methods. One of the most remarkable results was the success of the *KS*and *RS* methods for low and high censoring levels. The *KS* and the *SS* provided similar results for almost all combinations, but the *KS* performed better than the *SS* for lower censoring rates. Because the *RS* method works with determined knots, it was much more successful than the *SS* and the *KS* for numerical applications in this paper. In summary, the following suggestions and conclusions are offered based on the simulation study and the real- data example:

- In the simulation study, the *RS*method generally gave satisfying results, as shown in Tables 1, 2 and 3. Furthermore, the *KS* method is better than the other two

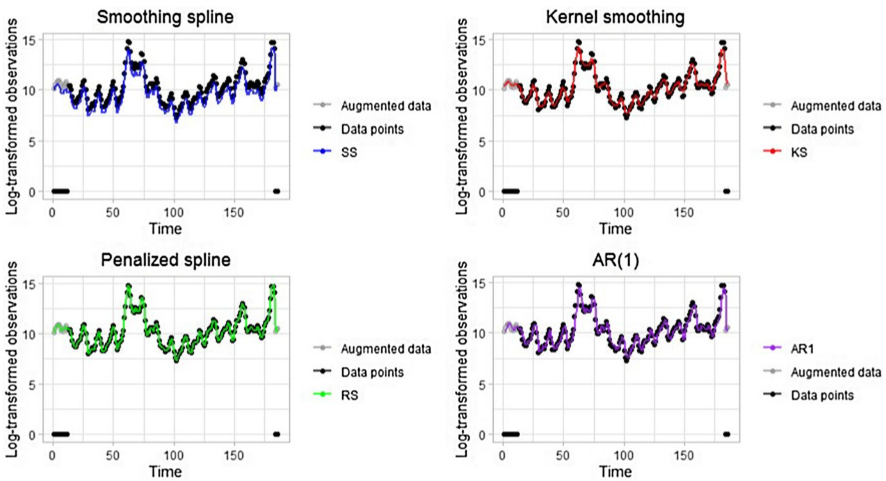

Fig. 7 Estimated curves of unemployment duration data represented by three smoothing methods

Table 7 Results of four models for the unemployment dataset

	MSE	MAPE	KL-N	IQR	RSE	RAE
SS	0.4753	0.0211	4.5128	0.2724	0.8969	25.7752
KS	0.0722	0.0055	1.4019	0.1407	0.2594	12.4707
RS	0.0707	0.0048	0.7308	0.1324	0.2146	11.4410
AR(1)	0.3256	0.0393	8.1460	0.3774	0.9331	23.0301

methods at a lower censoring level, as shown in Table 1. When the censoring level begins to increase, the *RS*method begins to stand out, and it gave the best results at both medium and high censoring levels. Although the quality of the estimation decreases with increases in the censoring level for all methods, these three smoothing methods still produce reasonable outcomes, proving that right-censored time series can be successfully modelled with nonparametric methods.

- One other important finding from the simulation concerns the performance of the methods with respect to spatial variation in the regression function. As explained above, the regression function was produced with two different shapes, one, two-peaked and one four-peaked. It was shown that each peak exceeded the detection limit, , so data was censored. The performances of the methods were measured according to this scenario. As expected, the performance criteria values changed for the worse for all methods when the fluctuation of the function increased. However, Tables 1, 2 and 3 reveal that the *RS* seems to be more durable than the others.
- The real-data examples, are illustrated in Figs. 5, 6 and 7 and their outcomes are presented in Tables 6 and 7.The censoring rate of the real data was 41.62%, which matches the high censoring level in the simulation study. Here, the expectation was that the *RS* would have the best performance since it did in the simulation results. However, because the cloud-ceiling dataset is highly complex and

wiggly, the *RS* cannot follow the data completely because it uses determined knots. Although the *KS* had the best performance using real data, the *RS* was a close second.

- For both studies, fitted curves for the estimated nonparametric regression models are illustrated in Figs. 2 and 3, 6 and 7. The figures demonstrate that the three smoothing methods for estimating right-censored time series have similar outcomes, and all are successful in representing the data.
- As shown in the simulation and real- data examples, the *KS*, the *RS* and the *SS* approximations had the best empirical performance. By comparison, the classical $AR(1)$ produced the worst performance at all censoring levels and sample sizes.

Finally, all of the results of the two numerical studies illustrated that for lower censoring levels and complicated datasets, the *KS* method performs best, and the *RS* method provides the most satisfactory estimates for nonparametric regression modeling of right-censored time-series data.

Acknowledgements We would like to thank the editor, the associate editor, and the anonymous referees for beneficial comments and suggestions.

Appendices: Supplemental Technical Materials

Appendix 1: Derivation of Equation (15)

Let $\mathbf{Z}_t^{(k)}$ be the vector $(Z_1^{(k)}, \dots, Z_n^{(k)})^\top$. It is easily seen that the residual sum of squares about f can be rewritten as

$$\sum_{t=1}^{n} \left\{ Z_t^{(k)} - f\left(X_t\right) \right\}^2 = \left\| \mathbf{Z}^{(k)} - \mathbf{f} \right\|_2^2 = \left(\mathbf{Z}^{(k)} - \mathbf{f} \right)^\top \left(\mathbf{Z}^{(k)} - \mathbf{f} \right) \tag{41}$$

since the vector $\mathbf{f}$ is definitely the vector of the values $f\left(X_t\right)$. Determine the penalty term $\int_a^b \left\{f''(x)\right\}^2 dx$ in (11) as $\mathbf{f}^\mathsf{T}\mathbf{K}\mathbf{f}$ from (13) to obtain the penalized residuals sum of squares criterion (14), given by

$$PRSS(\mathbf{f}, \lambda) = \left(\mathbf{Z}^{(k)} - \mathbf{f} \right)^\top \left(\mathbf{Z}^{(k)} - \mathbf{f} \right) + \lambda \mathbf{f}^\top \mathbf{K} \mathbf{f} = \mathbf{Z}^{(k)^\top} \mathbf{Z}^{(k)} - 2\mathbf{Z}^{(k)}\mathbf{f} + \mathbf{f}^\top (\mathbf{I} + \lambda \mathbf{K})\mathbf{f} \tag{42}$$

Taking the derivative with respect to $\mathbf{f}$ and setting it to zero:

$$\frac{\partial PRSS(\mathbf{f}, \lambda)}{\partial \mathbf{f}} = -2\mathbf{Z}^{(k)} + 2\mathbf{f}(\mathbf{I} + \lambda \mathbf{K}) = 0$$

Setting Eq. (42) equal to zero and replacing $\mathbf{f}$ by $\hat{\mathbf{f}}_\lambda^{SS}$, we obtain

$$\hat{\mathbf{f}}_\lambda^{SS} (\mathbf{I} + \lambda \mathbf{K}) = \mathbf{Z}^{(k)} \tag{43}$$

Hence, by multiplying the term $(\mathbf{I} + \lambda\mathbf{K})^{-1}$on the both sides of the equation above, we have the following solution based on smoothing spline for the vector

$$\hat{\mathbf{f}}_{\lambda}^{SS} = (\mathbf{I} + \lambda\mathbf{K})^{-1}\mathbf{Z}^{(k)}$$

as expressed in the Eq. (15).

Appendix 2: Derivation of Equation (25)

Consider the model $\mathbf{Z}^{(k)} = \mathbf{Xb} + \boldsymbol{e}$, where $\mathbf{b} = \left(b_0, b_1, \dots, b_q, b_{q+r}, r = 1, 2, \dots, m\right)$, with $b_{q+r}$the coefficient of the r^{th} knot. The vector of ordinary least squares residuals can be written as $\boldsymbol{e} = \mathbf{Z}^{(k)} - \mathbf{Xb}$ and hence

$$\boldsymbol{e}^{\top}\boldsymbol{e} = \left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right)^{\top}\left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right) \tag{44}$$

The primary interest is to obtain the estimate $\hat{\mathbf{b}}$ that minimizes the least squares residuals defined above. It should be noted that such an unrestricted estimation of the b_{q+r} leads to a wiggly fit. To overcome this problem, we impose the constraint $\sum_{r=1}^{m} b_{q+r}^2 < C$(where > 0) on b_{q+r} the coefficients. Also, we assume that $\mathbf{S}_{\lambda}^{PS}$ is a positive definite and symmetric smoother matrix based on penalized spline, and $\mathbf{D}$ is a $(m + 2) \times (m + 2)$ dimensional diagonal penalty matrix whose first $(q + 1)$ entries are $\mathbf{0}$, and the remaining entries are $\mathbf{1}$. Then, the Eq. (23) expressed in section (11) can be written as

$$minimum \ \left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right)^{\top}\left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right) \ subject \ \ to \ \ \lambda\mathbf{b}^{\top}\mathbf{Db} < C.$$

By the Lagrange multiplier method, minimizing this optimization problem subject to the constraint is equivalent to the following penalized residuals sum of squares

$$PRSS(\mathbf{b}, \lambda) = \left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right)^{\top}\left(\mathbf{Z}^{(k)} - \mathbf{Xb}\right) + \lambda\mathbf{b}^{\top}\mathbf{Db} = \mathbf{Z}^{(k)}\mathbf{T}\mathbf{Z}^{(k)} - 2(\mathbf{X}\mathbf{T}\mathbf{Z}^{(k)})\mathbf{b}+\mathbf{b'}(\mathbf{X}\mathbf{T}\mathbf{X})\mathbf{b} + \lambda\mathbf{b}^{\top}\mathbf{Db} \tag{45}$$

Similar to the ideas in (42), by taking the derivative with respect to $\mathbf{b}$ in (45) and setting it to zero, we obtain

$$\frac{\partial PRSS(\mathbf{b}, \lambda)}{\partial \mathbf{b}} = -2\mathbf{X}^{\top}\mathbf{Z}^{(k)} + 2\mathbf{b}\left(\mathbf{X}^{\top}\mathbf{X}\right) + \lambda 2\mathbf{bD} = 0 \tag{46}$$

Setting (46) equal to zero and replacing $\mathbf{b}$ by $\hat{\mathbf{b}}$, it is easily seen that we obtain the penalized least squares normal equations

$$\left(\mathbf{X}^{\top}\mathbf{X}+\lambda\mathbf{D}\right)\hat{\mathbf{b}} = \mathbf{X}^{\top}\mathbf{Z}^{(k)} \tag{47}$$

From (47), the estimated regression coefficients are simply

$$\hat{\mathbf{b}} = \left(\mathbf{X}^{\top}\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^{\top}\mathbf{Z}^{(k)} \tag{48}$$

Hence, the fitted values $\hat{\mathbf{f}}$ based on RS can be given by

$$\widehat{\mathbf{f}}_{\lambda}^{RS} = \mathbf{X}\widehat{\mathbf{b}} = \mathbf{X}\left(\mathbf{X}^{\top}\mathbf{X} + \lambda\mathbf{D}\right)^{-1}\mathbf{X}^{\top}\mathbf{Z}^{(k)} = \mathbf{S}_{\lambda}^{RS}\mathbf{Z}^{(k)} \tag{49}$$

as claimed.

Appendix 3: Derivation of Equation (28)

Consider the $RSS(\lambda)$ defined in Eq. (27). The $RSS(\lambda)$ can be rewritten as

$$RSS(\lambda) = \left(\widehat{\mathbf{f}}_{\lambda} - \mathbf{Z}^{(k)}\right)^{\top}\left(\widehat{\mathbf{f}}_{\lambda} - \mathbf{Z}^{(k)}\right) = \mathbf{Z}^{(k)}\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2}\mathbf{Z}^{(k)}$$

It is easily seen that $RSS(\lambda)$ can be expressed in a quadratic form. When we take the expected value of this form, we obtain

$$\begin{aligned} E[RSS(\lambda)] &= E\left\|\mathbf{Z}^{(k)}\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2}\mathbf{Z}^{(k)}\right\| = MSE(\lambda) \\ &= \widehat{\mathbf{f}}_{\lambda}^{\top}\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2}\widehat{\mathbf{f}}_{\lambda} + \sigma^{2} tr\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2} \\ &= \widehat{\mathbf{f}}_{\lambda}^{\top}\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2}\widehat{\mathbf{f}}_{\lambda} + n\sigma^{2} - 2\sigma^{2} tr\left(\mathbf{S}_{\lambda}\right) + \sigma^{2} tr\left(\mathbf{S}_{\lambda}^{\top}\mathbf{S}_{\lambda}\right) \\ &= \widehat{\mathbf{f}}_{\lambda}^{\top}\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)^{2}\widehat{\mathbf{f}}_{\lambda} + \sigma^{2}\left\{n - 2tr\left(\mathbf{S}_{\lambda}\right) + tr\left(\mathbf{S}_{\lambda}^{\top}\mathbf{S}_{\lambda}\right)\right\} \end{aligned}$$

as defined in the Eq. (28).

Appendix 4: Proof of the Lemma 4.1

$SMSE\left(\widehat{\mathbf{f}}_{\lambda}, \mathbf{f}\right) = E\left\|\widehat{\mathbf{f}}_{\lambda} - \mathbf{f}\right\|^{2}$, where as $\widehat{\mathbf{f}}_{\lambda} = \mathbf{S}_{\lambda}\mathbf{Z}^{(k)}$ Then the *SMSE* matrix can be written as the sum of variance and squared bias

$$\begin{aligned} SMSE\left(\widehat{\mathbf{f}}_{\lambda}, \mathbf{f}\right) &= \sum_{t=1}^{n}\left\{E\left(\widehat{f}_{\lambda}(X_t)\right) - f(X_t)\right\}^{2} + Var\left(\widehat{f}_{\lambda}(X_t)\right) \\ &= \sum_{t=1}^{n} bias^{2}\left(\widehat{f}_{\lambda}(X_t)\right) + \sum_{t=1}^{n} Var\left(\widehat{f}_{\lambda}(X_t)\right) \end{aligned} \tag{50}$$

As in the usual linear smoother, the *bias* and *Var* (i.e, variance) terms in (50) can be written, respectively, as

$$\sum_{t=1}^{n} bias^{2}\left(\widehat{f}_{\lambda}(X_t)\right) = \left(E\left(\widehat{\mathbf{f}}_{\lambda}\right) - \mathbf{f}\right)'\left(E\left(\widehat{\mathbf{f}}_{\lambda}\right) - \mathbf{f}\right) = \left(E\left(\mathbf{S}_{\lambda}\mathbf{Z}^{(k)}\right) - \mathbf{f}\right)^{\top}\left(E\left(\mathbf{S}_{\lambda}\mathbf{Z}^{(k)}\right) - \mathbf{f}\right) = \left\|\left(\mathbf{I} - \mathbf{S}_{\lambda}\right)\mathbf{f}\right\|^{2} \tag{51}$$

and

$$\sum_{t=1}^{n} Var\left(\hat{f}_{\lambda}(X_t)\right) = tr\left\{ Cov\left(\hat{f}_{\lambda}(X_t)\right)\right\} = tr\left\{ Cov\left(\hat{\mathbf{f}}_{\lambda} = \mathbf{S}_{\lambda}\mathbf{Z}^{(k)}\right)\right\} \\ = tr\left\{\mathbf{S}_{\lambda}\, Cov\left(\mathbf{Z}^{(k)}\right)\mathbf{S}_{\lambda}^{\top}\right\} \tag{52}$$

Assume that $Cov\left(\mathbf{Z}^{(k)}\right) = \sigma^2\mathbf{I}$ in (51) and hence, bringing (51) and (52) into (50) yields that

$$SMSE\left(\hat{\mathbf{f}}_{\lambda}, \mathbf{f}\right) = E\left\|\hat{\mathbf{f}}_{\lambda} - \mathbf{f}\right\|^2 = \left\|(\mathbf{I}\text{-}\mathbf{S}_{\lambda})\mathbf{f}\right\|^2 + \sigma^2 tr(\mathbf{S}_{\lambda}\mathbf{S}_{\lambda}^{\top})$$

This completes the proof of Lemma 4.1.

References

Anderson, T. W. (1984). *An introduction to multivariate statistical analysis*. Hoboken: Wiley.

Aydin, D., & Yilmaz, E. (2017). Modified spline regression based on randomly right-censored data: a comparison study. *Communications in Statistics-Simulation and Computation*,. https://doi.org/10.1080/03610918.2017.1353615.

Bjørnstad, N., & Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *American Association for the Advancement of Science*, *293*(5530), 638–643.

Box, G., & Jenkins, G. (1970). *Time series analysis: Forecasting and control*. San Francisco: Holden-Day.

Cai, T., & Betensky, R. A. (2003). Hazard regression for interval-censored data with penalized spline. *Biometrics*, *59*(3), 570–579.

Chen, M.-H., & Deely, J. J. (1996). Bayesian analysis for a constrained linear multiple regression problem for predicting the new crop apples. *Journal of Agricultural, Biological and Environmental Statistics*, *1*, 467–489.

Chen, C., Twycross, J., & Garibaldi, J. M. (2017). A new accuracy measure based on bounded relative error for time series forecasting. *PLos ONE*, *12*(3), 0174202.

Chen, Z., & Yang, Y. (2004). *Assessing forecast accuracy measures*. Ames: Iowa State University.

Craven, P., & Wahba, G. (1979). Smoothing noisy data with spline functions. *Numerical Mathematics*, *31*, 377–403.

Dabrowska, D. M. (1992). Nonparametric quantile regression with censored data. *Sankhya: The Indian Journal of Statistics Series A*, *54*(2), 252–259.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–38.

Elogne, S. N., Hristopulos, D. T., & Varouchakis, E. (2008). An application of Spartan spatial random fields in environmental mapping: Focus on automatic mapping capabilities. *Stochastic Environmental Research and Risk Assessment*,. https://doi.org/10.1007/s00477-007-0167-5.

Fan, J., & Gijbels, I. (1994). Censored regression: Local linear approximations and their applications. *Journal of the American Statistical Association*, *89*(426), 560–570.

Gelfand, A. E., Smith, A. F. M., & Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, *87*, 523–532.

Glasbey, C. A. (1988). Examples of regression with serially correlated errors. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *37*(3), 277–291.

Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, *22*(3), 443–473.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear model*. Boca Raton: Chapman & Hall.

Guessoum, Z., & Ould Said, E. (2008). On the nonparametric estimation of the regression function under censorship model. *Statistics & Decisions*, *26*, 159–77.

Hall, P., & Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields*, *99*(3), 399–424.
Heitjan, D. F., & Rubin, D. B. (1990). Inference from coarse data via multiple imputation with application to age heaping. *Journal of the American Statistical Association*, *85*(410), 304–314.
Helsel, D. R. (1990). Less than obvious: Statistical treatment of data below detection limit. *Environmental Science & Technology*, *24*, 1767–1774.
Hopke, P. K., Liu, C., & Rubin, D. B. (2001). Multiple imputation for multivariate data with missing and below threshold measurements: Time-series concentrations of pollutants in the arctic. *Biometrics*, *57*(1), 22–33.
Kim, H. T., & Truong, Y. K. (1998). Nonparametric regression estimates with censored data: Local linear smoothers and their applications. *Biometrics*, *54*, 1434–1444.
Meng, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science*, *9*(4), 538–558.
Nadaraya, E. A. (1964). On Estimating Regression. *Theory of Probability and Its Applications*, *10*, 186–190.
Park, J. W., Genton, M. G., & Ghosh, S. K. (2007). Censored time series analysis with autoregressive moving average models. *Canadian Journal of Statistics*, *35*(1), 151–168.
Park, J. W., Genton, M. G., & Ghosh, S. K. (2009). Nonparametric auto-covariance estimation from censored time series by Gaussian imputation. *Journal of nonparametric statistics*, *21*(2), 241–259.
Parzen, E. (1983). Time series analysis of irregularly observed data. In *Proceedings of a symposium held at Texas A&M University*, College Station, Texas.
Robinson, P. M. (1983). Nonparametric estimators for time-series. *Journal of Time Series Analysis*, *4*(3), 185–207.
Rubin, D. (1996). Multiple imputation after 18+ years (with discussion). *Journal of the American Statistical Association*, *91*, 473–489.
Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, *11*(4), 735–757.
Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression*. New York: Cambridge University Press.
Sampson, P. D., & Guttorp, P. (1992). Nonparametric estimation of non-stationary spatial covariance structure. *Journal of the American Statistical Association*, *87*(417), 108–119.
Shapiro, A., & Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics & Data Analysis*, *11*, 87–96.
Tanner, M. A. (1991). *Tools for statistical inference*. New York: Springer.
Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, *82*(398), 528–540.
Vaida, F., & Liu, L. (2009). Fast implementation for normal mixed effects models with censored response. *Journal of Computational and Graphical Statistics*, *18*(4), 797–817.
Wang, C., & Chan, K. S. (2017). Carx: An R package to estimate censored autoregressive time series with exogenous covariates. *R Journal*, *9*(2), 213–231.
Watson, G. S. (1964). Smooth regression analysis. *Sankhya, Series A*, *26*, 359–72.
Wu, W. B., & Pourahmadi, M. (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, *90*(4), 831–844.
Yang, S. (1999). Censored median regression using weighted empirical survival and hazard function. *Journal of the American Statistical Association*, *94*(445), 137–145.
Zeger, S. L., & Brookmeyer, R. (1986). Regression analysis with censored auto-correlated data. *Journal of the American Statistical Association*, *81*(395), 722–729.
Zheng, Z. K. (1984). Regression analysis with censored data. Ph.D. Dissertation, University of Colombia.
Zheng, J. X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric theory*, *14*, 123–138.