

Article

Variable Selection for Generalized Linear Models with Interval-Censored Failure Time Data

Rong Liu ^{1,2}, Shishun Zhao ^{1,2}, Tao Hu ^{3,*}  and Jianguo Sun ⁴

¹ Center for Applied Statistical Research, College of Mathematics, Jilin University, Changchun 130012, China; liurong18@mails.jlu.edu.cn (R.L.); zhaoss@jlu.edu.cn (S.Z.)

² National Applied Mathematical Center (Jilin), Changchun 130012, China

³ School of Mathematical Sciences, Capital Normal University, Beijing 100048, China

⁴ Department of Statistics, University of Missouri, Columbia, MO 65211, USA; sunj@missouri.edu

* Correspondence: hutaomath@foxmail.com

Abstract: Variable selection is often needed in many fields and has been discussed by many authors in various situations. This is especially the case under linear models and when one observes complete data. Among others, one common situation where variable selection is required is to identify important risk factors from a large number of covariates. In this paper, we consider the problem when one observes interval-censored failure time data arising from generalized linear models, for which there does not seem to exist an established method. To address this, we propose a penalized least squares method with the use of an unbiased transformation and the oracle property of the method is established along with the asymptotic normality of the resulting estimators of regression parameters. Simulation studies were conducted and demonstrated that the proposed method performed well for practical situations. In addition, the method was applied to a motivating example about children's mortality data of Nigeria.

Keywords: interval-censored data; unbiased transformation; linear model; variable selection; large sample properties

MSC: 62J12; 62N02; 62J07; 62G20



Citation: Liu, R.; Zhao, S.; Hu, T.; Sun, J. Variable Selection for Generalized Linear Models with Interval-Censored Failure Time Data. *Mathematics* **2022**, *10*, 763. <https://doi.org/10.3390/math10050763>

Academic Editor: Gaorong Li

Received: 18 January 2022

Accepted: 24 February 2022

Published: 27 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Variable selection is an important field in statistics, and there is a lot of literature on variable selection, especially in the context of linear models for complete data, such as stepwise regression, ridge regression, Bayesian variable selection, least absolute shrinkage and selection operator (LASSO), model averaging, smoothly clipped absolute deviation (SCAD), elastic net, adaptive LASSO (ALASSO), minimax concave penalty (MCP), seamless- L_0 (SELO) and broken adaptive ridge (BAR) (Goldberger et al., 1961 [1]; Hoerl and Kennard, 1970 [2]; Mitchell and Beauchamp, 1988 [3]; Tibshirani, 1996 [4]; Raftery et al., 1997 [5]; Zou and Hastie, 2005 [6]; Zou, 2006 [7]; Fan and Li, 2001 [8]; Zhang, 2010 [9]; Dicker et al., 2013 [10]; Liu and Li, 2016 [11]; Dai et al., 2018 [12]; Zheng et al., 2021 [13]). Among others, one type of methods that have recently attracted a lot of attention is the penalized method, for which various penalty functions have been proposed. LASSO by Tibshirani (1996) [4], SCAD by Fan and Li (2001) [8], elastic net by Zou and Hastie (2005), ALASSO by Zou (2006) [7], MCP by Zhang (2010) [9], SELO by Dicker et al. (2013) [10], and BAR by Liu and Li (2016) [11] are included for penalty functions.

Variable selection has been investigated by many authors for incomplete data such as right-censored and interval-censored failure time data (Cai et al., 2005 [14]; Fan and Li, 2002 [15]; Tibshirani, 1997 [16]; Zhang and Lu, 2007 [17]; Nasrullah Khan et al., 2018 [18]; Zhao et al., 2020 [19]; Li et al., 2020 [20]; Du and Sun 2021 [21]; Ali et al., 2021 [22]). By interval censored data, we mean that the failure time of interest is known or observed

only to belong to an interval instead of being observed exactly. It is easy to see that this is usually the case for periodic follow-up studies such as clinical trials and they include right-censored failure time data as a special case. Notably, the analysis of interval-censored data is much more challenging than the analysis of right-censored data. In terms of the well-known Cox model, the classical partial likelihood method for right-censored data is no longer available with interval censoring. This is because for the case of interval censoring, we have to estimate not only regression parameters of interest but also the nuisance parameters simultaneously. Recently, many authors have discussed the analysis of interval-censored data in various situations. For example, Zhao et al. (2015) [23], Wang et al. (2016) [24] and Li et al. (2018) [25] studied inference procedures for the Cox model, the additive hazard model, and the linear transformation model, respectively. Sun (2006) [26] provided a relatively complete review of the literature for the interval-censored failure time data analysis.

As mentioned above, several authors have discussed variable selection for interval-censored failure time data ([19,20]). However, existing methods cannot be applied directly to linear models or generalized linear models. In the variable selection procedure proposed below, we employ the unbiased transformation approach and the main idea behind it is to transfer two variables representing interval-censored data to a new, single variable that has the same conditional expectation. One of the early applications of the approach was given by Buckley and James (1979) [27] in dealing with right-censored data. Among others, one advantage of the proposed method over the existing methods is that it can be relatively easily implemented since one can make use of the existing variable selection programs for complete data with simple modifications.

Deng (2004) [28] and Deng et al. (2012) [29] discussed the use of the unbiased transformation approach for the analysis of interval-censored data. In particular, the latter considered the situation similar to that discussed here, but not on variable selection, under the assumption that the joint density function of the two variables representing the observed data is known. It is easy to see that this may not be true in many applications. To address this, we will adopt the kernel estimation approach [30,31] to estimate the needed density function and develop a unified approach to variable selection with different penalty functions for generalized linear models based on interval-censored data.

The rest of the article is organized as follows. In Section 2, we will first introduce some notations and assumptions to be used throughout the paper and present the proposed variable selection procedure. For the implementation of the presented method, a coordinate descent algorithm is developed in Section 3. In Section 4, the asymptotic properties of the proposed method under the BAR penalty are established, and Section 5 gives some numerical results obtained from a simulation study, which suggest that the method works well in practical situations. In Section 6, the method is applied to children's mortality data of Nigeria that motivated this study, and Section 7 contains some discussion and concluding remarks.

2. Unbiased Transformation Variable Selection Procedure

Consider a failure time study that consists of n independent subjects and for subject i , let T_i denote the failure time of interest and X_i a p -dimensional vector of covariates. Suppose that for each T_i , two observations are available at the observation times U_i and V_i such that they divide the axis $(0, \infty)$ into three parts $[0, U_i)$, $[U_i, V_i]$, (V_i, ∞) and we know which part T_i falls in. Thus, the observed data on subject i have the form $O_i = \{U_i, V_i, \delta_{1i}, \delta_{2i}, X_i\}$, where $\delta_{1i} = I(T_i \leq U_i)$ and $\delta_{2i} = I(U_i < T_i \leq V_i)$ with $I(\cdot)$ being the indicator function,

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

$i = 1, \dots, n$. In the following, we will assume that T_i is independent of U_i and V_i given X_i .

To describe the covariate effects, we will assume that given X_i , T_i has the form

$$H(T_i) = \beta_0 + \mathbf{X}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad (1)$$

where $H(\cdot)$ is a known function, ϵ_i is a zero mean random error with distribution unknown, and β_0 and $\boldsymbol{\beta}$ are unknown parameters. Note that for the estimation of the model above, if T_i was exactly observed, a simple method would be to take $H(T_i)$ as a new response variable and transform model (1) into a linear model. For the situation here where T_i is interval-censored, however, the above method cannot be carried out directly.

To overcome the problem, we adopt the unbiased transformation approach to first convert $H(T_i)$ into the variable

$$h_i^* = \phi_1(U_i, V_i)\delta_{1i} + \phi_2(U_i, V_i)\delta_{2i} + \phi_3(U_i, V_i)(1 - \delta_{1i} - \delta_{2i}) + H(0),$$

where $\phi_1(\cdot, \cdot)$, $\phi_2(\cdot, \cdot)$ and $\phi_3(\cdot, \cdot)$ are some continuous functions with finite continuous partial derivatives and also independent of the distribution of T_i . Let $g(\cdot, \cdot)$ denotes the joint density function of U_i and V_i and assume that $\phi_1(u, v)$, $\phi_2(u, v)$ and $\phi_3(u, v)$ satisfy the following conditions

$$\begin{cases} \int_{v=0}^{\infty} \int_{u=0}^v \phi_1(u, v)g(u, v)dudv = 0, \\ \int_y^{\infty} [\phi_2(y, v) - \phi_1(y, v)]g(y, v)dv \\ + \int_0^y [\phi_3(u, y) - \phi_2(u, y)]g(u, y)du = H'(y). \end{cases} \quad (2)$$

Then according to Deng et al. (2012) [29], we have that

$$E(h_i^*(U_i, V_i, \delta_{1i}, \delta_{2i})) = E(H(T_i))$$

for $i = 1, \dots, n$.

Under Equation (2), one can see that $\phi_1(U_i, V_i)$, $\phi_2(U_i, V_i)$ and $\phi_3(U_i, V_i)$ not only depend on U_i and V_i but also depend on $g(U_i, V_i)$. More specifically, one can rewrite h_i^* as

$$\begin{aligned} h_i^*(U_i, V_i, g(U_i, V_i), \delta_{1i}, \delta_{2i}) &= \phi_1(U_i, V_i, g(U_i, V_i))\delta_{1i} + \phi_2(U_i, V_i, g(U_i, V_i))\delta_{2i} \\ &+ \phi_3(U_i, V_i, g(U_i, V_i))(1 - \delta_{1i} - \delta_{2i}) + H(0). \end{aligned}$$

Thus, for estimating model (1) or $\boldsymbol{\beta}$, it is natural to consider the least squares method to minimize the mean square of the residual after the unbiased transformation

$$n^{-1} \sum_{i=1}^n (h_i^* - \beta_0 - \mathbf{X}_i^\top \boldsymbol{\beta})^2 \quad (3)$$

if $g(U_i, V_i)$ was known. Of course, in practice, $g(U_i, V_i)$ is unknown and for this, we propose to first estimate $g(U_i, V_i)$ by the kernel density estimator $\hat{g}(U_i, V_i)$ and then consider

$$\begin{aligned} \hat{h}_i^* &= \phi_1(U_i, V_i, \hat{g}(U_i, V_i))\delta_{1i} + \phi_2(U_i, V_i, \hat{g}(U_i, V_i))\delta_{2i} \\ &+ \phi_3(U_i, V_i, \hat{g}(U_i, V_i))(1 - \delta_{1i} - \delta_{2i}) + H(0). \end{aligned}$$

Note that $\hat{h}_i^*$ given above involves the estimation of the two-dimensional function g and according to Section 5 of Deng (2004) [28], one could equivalently replace it by

$$\begin{aligned} \hat{h}_i^* &= \phi_1(U_i, V_i, \hat{g}_u(U_i), \hat{g}_v(V_i))\delta_{1i} + \phi_2(U_i, V_i, \hat{g}_u(U_i), \hat{g}_v(V_i))\delta_{2i} \\ &+ \phi_3(U_i, V_i, \hat{g}_u(U_i), \hat{g}_v(V_i))(1 - \delta_{1i} - \delta_{2i}) + H(0), \end{aligned}$$

where $\hat{g}_u(U_i)$ and $\hat{g}_v(V_i)$ denote the kernel estimators of the marginal density functions $g_u(U_i)$ and $g_v(V_i)$. In Lemma A1 given in Appendix A, we show that $\hat{h}_i^*$ converge to h_i^* in probability. Since U_i and V_i are positive variables, we propose to adopt the log-transformation technique and employ the log kernel density estimator

$$\hat{g}_u(u) = (nh)^{-1} \sum_{i=1}^n u^{-1} K((\log(u) - \log(U_i))/h),$$

and

$$\hat{g}_v(v) = (nh)^{-1} \sum_{i=1}^n v^{-1} K((\log(v) - \log(V_i))/h)$$

(Parzen, 1962 [30]). Then by following Deng (2004) [28], one can estimate β by minimizing the mean squared residual after kernel estimation and unbiased transformation

$$n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - \mathbf{X}_i^\top \beta)^2. \quad (4)$$

Now we consider the variable selection or the selection of important covariates. For this, and motivated by (4), we propose to use the penalized least squares estimation method and to minimize the mean squared residual after kernel estimation and unbiased transformation with penalized criteria

$$n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - \mathbf{X}_i^\top \beta)^2 + \sum_{j=1}^p p_{a,\lambda}(|\beta_j|), \quad (5)$$

where $p_{a,\lambda}(|\beta_j|)$ denotes a penalty function with a and λ (some comments are given below). In the following, several commonly used penalty functions are considered, including the LASSO penalty $p_\lambda(|\beta_j|) = \lambda|\beta_j|$ proposed by Tibshirani (1996) [4] and the SCAD penalty

$$p_\lambda(|\beta_j|; a) = \begin{cases} \lambda|\beta_j| & \text{if } |\beta_j| \leq \lambda \\ -\frac{|\beta_j|^2 - 2a\lambda|\beta_j| + \lambda^2}{2(a-1)} & \text{if } \lambda < |\beta_j| \leq a\lambda \\ \frac{(a+1)\lambda^2}{2} & \text{if } |\beta_j| > a\lambda \end{cases}$$

with $a > 2$ by Fan and Li (2001) [8]. For a in the SCAD penalty, we set $a = 3.7$ suggested by Fan and Li (2001) [8]. Furthermore, we investigate the use of the MCP penalty

$$p_\lambda(|\beta_j|; a) = \lambda \int_0^{|\beta_j|} \frac{(a\lambda - x)_+}{a\lambda} dx$$

with $a > 1$ given in Zhang (2010) [9] and the BAR penalty $p_\lambda(|\beta_j|) = \lambda\beta_j^2 / \tilde{\beta}_j^2$ discussed in Zhao et al. (2020) [19], where $\tilde{\beta}_j (j = 1, \dots, p)$ denotes a nonzero “good” estimator of β_j .

Note that for the application of the method described above, one needs to choose the functions ϕ_1 , ϕ_2 , and ϕ_3 satisfying Equation (2). For this, many functions can be used and for simplicity, we suggest to employ (I) $\phi_1(U_i, V_i) = 0$, $\phi_2(U_i, V_i) = 0$, $\phi_3(U_i, V_i) = H'(V_i)/\hat{g}_v(V_i)$; (II) $\phi_1(U_i, V_i) = 0$, $\phi_2(U_i, V_i) = H'(U_i)/\hat{g}_u(U_i)$, $\phi_3(U_i, V_i) = H'(U_i)/\hat{g}_u(U_i)$; or (III) $\phi_1(U_i, V_i) = 0$, $\phi_2(U_i, V_i) = H'(U_i)/2\hat{g}_u(U_i)$, $\phi_3(U_i, V_i) = H'(U_i)/2\hat{g}_u(U_i) + H'(V_i)/2\hat{g}_v(V_i)$. The numerical study below indicates that they give satisfactory and robust results and more discussion on this can be found in Deng et al. (2012) [29].

3. Penalized Least Squares Coordinate Descent Algorithm

Let $\hat{\beta}$ denote the estimator of β given by minimizing the penalized criterion function in (5). In this section, we discuss the determination of β and investigate a coordinate descent algorithm that takes turn to update each element β_j of β while making all other elements of β fixed at their current estimates except β_j .

Define

$$M(\beta_j) = n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - \mathbf{X}_i^\top \beta)^2.$$

Then, at the k th iteration, the value of β_j is derived by minimizing $Q(\beta_j) = M(\beta_j) + p_\lambda(|\beta_j|)$ to determine $\hat{\beta}_j^{(k)}$. Note that by drawing the locally quadratic approximation idea discussed in Fan and Li (2001) [8], the locally quadratic approximation can be presented as

$$[p_\lambda(|\beta_j|)]' = p'_\lambda(|\beta_j|) \text{sgn}(\beta_j) \approx \{p'_\lambda(|\hat{\beta}_j^{(k-1)}|)/|\hat{\beta}_j^{(k-1)}|\} \beta_j$$

when $\beta_j \neq 0$. In other words, a quadratic function can locally approximate $p(|\beta_j|; \lambda)$ at $|\hat{\beta}_j^{(k-1)}|$ as

$$p_\lambda(|\beta_j|) \approx p_\lambda(|\hat{\beta}_j^{(k-1)}|) + 1/2 \{ p'_\lambda(|\hat{\beta}_j^{(k-1)}|) / |\hat{\beta}_j^{(k-1)}| \} [\beta_j^2 - (\hat{\beta}_j^{(k-1)})^2].$$

Meanwhile, $M(\beta_j)$ can be obtained by the second-order Taylor expansion,

$$M(\beta_j) \approx M(\hat{\beta}_j^{(k-1)}) + M'(\hat{\beta}_j^{(k-1)}) (\beta_j - \hat{\beta}_j^{(k-1)}) + 1/2 M''(\hat{\beta}_j^{(k-1)}) (\beta_j - \hat{\beta}_j^{(k-1)})^2,$$

where M' and M'' represent the first and second derivatives of $M(\cdot)$, respectively. Therefore, minimizing $Q(\beta_j)$ is equivalent to minimizing the function

$$\begin{aligned} & M(\hat{\beta}_j^{(k-1)}) + M'(\hat{\beta}_j^{(k-1)}) (\beta_j - \hat{\beta}_j^{(k-1)}) + \frac{1}{2} M''(\hat{\beta}_j^{(k-1)}) (\beta_j - \hat{\beta}_j^{(k-1)})^2 \\ & + p_\lambda(|\hat{\beta}_j^{(k-1)}|) + 1/2 \{ p'_\lambda(|\hat{\beta}_j^{(k-1)}|) / |\hat{\beta}_j^{(k-1)}| \} [\beta_j^2 - (\hat{\beta}_j^{(k-1)})^2] \end{aligned}$$

with respect to β_j . A closed form solution is given as

$$\hat{\beta}_j^{(k)} = \frac{\hat{\beta}_j^{(k-1)} M''(\hat{\beta}_j^{(k-1)}) - M'(\hat{\beta}_j^{(k-1)})}{M''(\hat{\beta}_j^{(k-1)}) + p'_\lambda(|\hat{\beta}_j^{(k-1)}|) / |\hat{\beta}_j^{(k-1)}|}. \quad (6)$$

Note that it is easy to see that the resulting solution (6) and the approximation used above for the penalty function apply to any penalty function. However, the BAR penalty is not necessary to do a locally quadratic approximation, due to the fact that it is already a quadratic function of coefficients. For that situation, we can obtain the closed-form iterative solution proposed by Wu et al., 2020 [32] as

$$\hat{\beta}_j^{(k)} = \hat{\beta}_j^{(k-1)} - \frac{Q'(\hat{\beta}_j^{(k-1)})}{Q''(\hat{\beta}_j^{(k-1)})}, \quad (7)$$

where $Q'(\hat{\beta}_j^{(k-1)})$ and $Q''(\hat{\beta}_j^{(k-1)})$ are the first and second derivatives of $Q(\beta_j)$ with respect to β_j evaluated at $\hat{\beta}_j^{(k-1)}$, respectively. By combining the discussion above, the algorithm can be implemented as follows:

Step 1: Set $k = 0$ and choose the initial estimate $\hat{\beta}^{(0)}$

$$\hat{\beta}^{(0)} = \arg \min_{\beta} n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - \mathbf{X}_i^\top \beta)^2.$$

Step 2: Use the coordinate descent algorithm to determine $\hat{\beta}^{(k)}$ by (6) for LASSO, SCAD and MCP, respectively. Meanwhile, we determine $\hat{\beta}^{(k)}$ by (7) for BAR.

Step 3: Repeat Step 2 until the convergence or k exceeds a given large number.

There are many various criteria to check the convergence in Step 3 above. In the simulation studies below, the proposed algorithm is declared to achieve convergence if the maximum of the absolute differences of the estimates between two successive iterations is less than 10^{-5} (Sun et al., 2019) [33]. To implement the algorithm above, one also needs to choose the tuning parameter λ_n . For the results given below, we use the Bayesian information criterion (BIC) proposed by Schwarz (1978) [34] which is data-dependent and defined as

$$BIC_\lambda = 2 \cdot n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - \mathbf{X}_i^\top \hat{\beta})^2 + df_\lambda \log(n).$$

In the above, $\hat{\beta}$ denotes the final estimator of β and df_{λ} represents the total number of nonzero estimates in $\hat{\beta}$, which serves as the degree of freedom. Alternatively, one could employ other methods such as a K-fold cross-validation (CV) (Verweij and Houwelingen, 1993) [35]. For given λ_n , we define BIC as above. Then, one can choose the value of λ_n that minimize BIC_{λ} . For variance estimation of the proposed estimators, we suggest using the nonparametric bootstrap method, which seems to work well as the numerical study indicates.

4. Asymptotic Properties

In this section, we establish the asymptotic properties of the variable selection procedure or the estimator proposed above with the use of the BAR penalty function. For this, let $\beta_0 = (\beta_{01}, \dots, \beta_{0p_n})^\top$ denote the true parameter value. Note that here, we replace p by p_n to emphasize the dependence of p on n and assume that p_n can diverge to infinity but $p_n < n$. For simplicity, we write $\beta_0 = (\beta_{01}^\top, \beta_{02}^\top)^\top$ and assume $\beta_{01} \neq 0$ and $\beta_{02} = 0$, where $\beta_{01} \in \mathbb{R}_n^{q_n}$ and $\beta_{02} \in \mathbb{R}_n^{p_n - q_n}$. Let $\hat{\beta}^* = (\hat{\beta}_1^{*\top}, \hat{\beta}_2^{*\top})^\top$ denote the BAR estimator of β corresponding to the same participation. Set $X_\alpha = (x_1, \dots, x_{q_n})$, $X_\gamma = (x_{q_n+1}, \dots, x_{p_n})$, $\Sigma_{n1} = X_\alpha^\top X_\alpha / n$ and $\Sigma_n = X^\top X / n$, where x_j is j th column of X for $j = 1, \dots, p_n$. Let $\hat{Y}_i^* = \hat{h}_i^* - n^{-1} \sum_{i=1}^n \hat{h}_i^*$, $\hat{Y}^* = (\hat{Y}_1^*, \dots, \hat{Y}_n^*)$, $Y_i^* = h_i^* - n^{-1} \sum_{i=1}^n h_i^*$, $Y^* = (Y_1^*, \dots, Y_n^*)$ for $i = 1, \dots, n$. For the asymptotic properties, we need the following regularity conditions.

C1. For every $t \in (0, \infty)$, $H'(t)$ exists, where $H'(t)$ is the derivative of $H(t)$ and $H(0) < \infty$.

C2. U_i and V_i are positive i.i.d random vectors with uniformly continuous density functions $g_u(u)$ and $g_v(v)$, respectively.

C3. T_i and (U_i, V_i) are independent given X_i .

C4. $\text{Var}(h^*) < \infty$.

C5. $C_n = n^{-1} \sum_{i=1}^n X_i^\top X_i \rightarrow C$ with probability 1 as n tends to ∞ , where C is a positive definite matrix ([19]).

C6. $K(t)$ is uniformly continuous and of bound variation, $\int |K(t)| dt < \infty$ and $K(t) \rightarrow 0$ as $|t| \rightarrow \infty$. $\int K(t) dt = 1$ and $\int |x \log |x||^{1/2} |dK(t)| < \infty$. $\lim_{n \rightarrow \infty} h_n = 0$, and $\lim_{n \rightarrow \infty} nh_n / \log n = \infty$.

C7. There exists a constant $E > 1$ such that $0 < 1/E < \lambda_{\min}(C_n) \leq \lambda_{\max}(C_n) < E < \infty$ for every integer n .

C8. $a_{0n} = \min_{1 \leq j \leq q_n} |\beta_{0j}|$ and $a_{1n} = \max_{1 \leq j \leq q_n} |\beta_{0j}|$. As $n \rightarrow \infty$, $p_n q_n / \sqrt{n} \rightarrow 0$, $(p_n/n)^{1/2} / a_{0n} \rightarrow 0$, $p_n / \lambda_n \rightarrow 0$ and $\lambda_n a_{1n} (q_n/n)^{1/2} / a_{0n}^2 \rightarrow 0$.

Note that Conditions **C1–C5** are necessary to obtain an unbiased transformation ([29]), where uniformly continuous density functions $g_u(u)$ and $g_v(v)$ are needed to make kernel estimation $\hat{g}_u(u)$ and $\hat{g}_v(v)$ converging to density functions $g_u(u)$ and $g_v(v)$ almost surely and Condition **C6** guarantees that $\hat{h}_i^*$ converges to h_i^* in probability ([31]). Condition **C7** assumes that C_n is positive definite almost surely and its eigenvalues are bounded away from zero and infinity. Condition **C8** gives some sufficient, but not necessary, conditions needed to prove the convergence and asymptotic properties of the BAR estimator and nonzero coefficients are assumed to be uniformly bounded away from zero and infinity ([12]). Define $\beta = (\alpha^\top, \gamma^\top)^\top$, where α and γ are $q_n \times 1$ and $(p_n - q_n) \times 1$ vectors, respectively. The following theorem gives the asymptotic properties.

Theorem 1. Assume that Conditions **C1–C8** given above hold and $\phi_1(\cdot, \cdot)$, $\phi_2(\cdot, \cdot)$, and $\phi_3(\cdot, \cdot)$ are continuous functions with finite continuous partial derivatives. Then, we have that:

(i) The fixed point of $f(\alpha) = \{X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha)\}^{-1} X_\alpha^\top \hat{Y}^*$ exists and is unique, where $D_1(\alpha) = \text{diag}(\alpha_1^{-2}, \dots, \alpha_{q_n}^{-2})$.

- (ii) (Oracle property) the BAR estimator $\hat{\beta}^* = (\hat{\beta}_1^{*\top}, \hat{\beta}_2^{*\top})^\top$ exists and is unique, where $\hat{\beta}_2^* = \mathbf{0}$ and $\hat{\beta}_1^*$ is the unique fixed point of $f(\alpha)$.
- (iii) (Asymptotic normality) $\sqrt{n}(\hat{\beta}_1^* - \beta_{01}) \rightarrow N(\mathbf{0}, \Sigma_{(1)})$ in distribution, where $\Sigma_{(1)}$ is defined in Appendix A.

5. Simulation Study

Now we present some results from a simulation study to assess the finite sample performance of the variable selection procedure presented in the previous sections. In the study, we generated the vector of covariates $\mathbf{X}_i$ from $\mathcal{N}(\mathbf{0}, \Sigma_X)$ and the covariance matrix Σ_X with the (l, m) element being $0.5^{|l-m|}$. Given the $\mathbf{X}_i$'s, the true failure times were generated based on model (1) with $H(T) = \log(T + 1)$ and the ε_i 's following the standard normal distribution. For the generation of the observed interval-censored data or the two observation times for each subject, they were set to be times from the homogeneous Poisson process with the interexamination times being independently and identically distributed in the exponential distribution with a mean of 0.4 (Li et al., 2019 [20]).

For the application of the proposed variable selection procedure, we set

$$h^* = \delta_1 * 0 + \delta_2 * 0 + \delta_3 * \frac{1}{1+v} / g_v(v) \quad (8)$$

by following Deng et al. (2012) [29]. Meanwhile, we set

$$h^* = \delta_1 * 0 + \delta_2 * \frac{1}{1+u} / g_u(u) + \delta_3 * \frac{1}{1+u} / g_u(u) \quad (9)$$

and

$$h^* = \delta_1 * 0 + \delta_2 * \frac{1}{1+u} / 2g_u(u) + \delta_3 * (\frac{1}{1+u} / 2g_u(u) + \frac{1}{1+v} / 2g_v(v)). \quad (10)$$

For the kernel estimators $\hat{g}(u)$ and $\hat{g}(v)$, we considered the following biquadratic kernel function

$$K(t) = \begin{cases} 3/4(1-t^2), & |t| \leq 1, \\ 0, & \text{else,} \end{cases}$$

and several different bandwidths. They include (a) $n^{-1/5}$, (b) $1.06 \cdot \hat{\sigma} \cdot n^{-1/5}$, (c) $1.06 \cdot \min(\hat{\sigma}, \hat{R}/1.34) \cdot n^{-1/5}$ with $\hat{R}$ being the 0.75 quantile minus the 0.25 quantile, and (d) $c_1 \cdot n^{-1/5}$, where $\hat{\sigma}$ denotes the sample standard deviation and c_1 is selected by the CV method over 20 numbers from (0.5, 1.5) with equal distance. The simulation results of above four circumstances are listed in Supplemental Materials. We also considered the selection of the tuning parameter λ_n and the bandwidth choice (d) together, based on the BIC described above with λ_n over 50 numbers from 0.001 to 0.01 with equal distance and c_1 over 10 numbers from (0.5, 1.5) with equal distance. The results given below are based on the sample size $n = 300$ or $n = 500$ with 100 replications and $p = 10, 30, 50$, or 100.

Tables 1, 2 and 3 are based on (8), (9), (10), respectively. Tables 1–3 present the results on the covariate selection with $\beta_0 = (0.5, 0.5, \mathbf{0}_{p-2}^\top)^\top$ and $\beta_0 = (0.5, 0.7, \mathbf{0}_{p-2}^\top)^\top$, corresponding to relatively moderate and weak signals, respectively, as well as $\beta_0 = (0.5, 0.5, 0.5, 0.5, \mathbf{0}_{p-4}^\top)^\top$. They include the average number of nonzero estimates of the parameters whose true values are not zero (TP) and the average number of nonzero estimates of the parameters whose true values are zero (FP). In addition, we calculated and included in the table the median mean squared errors (MMSE) given by $(\hat{\beta} - \beta_0)^\top \Sigma_X (\hat{\beta} - \beta_0)$, measuring the prediction accuracy, and the standard deviation of MSE (SD), where Σ_X denotes the population covariance matrix of the covariates. In the table, in addition to the BAR penalty function, we also considered the LASSO, MCP, and SCAD penalty functions, and the joint selection of the tuning parameter λ_n and the bandwidth based on the BIC was used. We also added backward stepwise variable selection based on BIC and for $\beta_0 = (0.5, 0.5, \mathbf{0}_{p-2}^\top)^\top$ in Table 1.

Table 1. Simulation results based on the joint selection of the tuning parameter and bandwidth for (I).

		$n = 300$			$n = 500$		
		$\beta = (0.5, 0.5, 0, \dots, 0)$					
p	Method	TP	FP	MMSE (SD)	TP	FP	MMSE (SD)
10	Stepwise	1.49	0.12	0.179 (0.097)	1.96	1.58	0.144 (0.060)
	LASSO	2.00	1.03	0.174 (0.071)	2.00	1.21	0.164 (0.065)
	SCAD	1.95	0.92	0.145 (0.070)	1.97	0.90	0.132 (0.063)
	MCP	1.92	0.77	0.139 (0.064)	1.95	0.76	0.131 (0.061)
	BAR	1.89	0.49	0.242 (0.090)	1.95	0.48	0.134 (0.058)
30	Stepwise	1.90	5.65	0.279 (0.115)	1.98	5.05	0.208 (0.069)
	LASSO	1.99	3.17	0.216 (0.145)	2.00	2.86	0.186 (0.056)
	SCAD	1.93	2.97	0.187 (0.085)	1.99	2.51	0.145 (0.059)
	MCP	1.90	2.68	0.187 (0.092)	1.98	2.01	0.137 (0.056)
	BAR	1.98	1.14	0.143 (0.055)	1.98	1.14	0.143 (0.054)
50	Stepwise	1.84	9.90	0.400 (0.140)	1.96	9.69	0.292 (0.083)
	LASSO	1.97	4.86	0.233 (0.063)	2.00	4.66	0.193 (0.054)
	SCAD	1.90	4.54	0.206 (0.079)	1.96	4.25	0.151 (0.063)
	MCP	1.88	4.32	0.220 (0.099)	1.95	3.78	0.155 (0.068)
	BAR	1.73	1.45	0.205 (0.079)	1.93	1.76	0.158 (0.066)
100	Stepwise	1.87	25.96	0.875 (0.277)	1.96	21.61	0.542 (0.137)
	LASSO	1.99	9.83	0.255 (0.079)	2.00	8.67	0.211 (0.064)
	SCAD	1.86	9.55	0.228 (0.103)	1.97	8.12	0.173 (0.081)
	MCP	1.87	8.85	0.299 (0.133)	1.95	7.32	0.189 (0.095)
	BAR	1.75	2.99	0.253 (0.101)	1.91	3.27	0.196 (0.082)
$\beta = (0.5, 0.7, 0, \dots, 0)$							
10	LASSO	2.00	1.13	0.278 (0.086)	2.00	1.25	0.225 (0.074)
	SCAD	1.94	0.85	0.225 (0.080)	1.97	0.86	0.209 (0.070)
	MCP	1.94	0.84	0.220 (0.076)	1.97	0.77	0.204 (0.067)
	BAR	1.89	0.29	0.229 (0.082)	1.95	0.45	0.214 (0.067)
30	LASSO	2.00	3.06	0.311 (0.096)	2.00	2.85	0.276 (0.069)
	SCAD	1.90	2.80	0.248 (0.099)	1.99	2.53	0.209 (0.069)
	MCP	1.91	2.69	0.258 (0.109)	1.99	2.12	0.207 (0.063)
	BAR	1.85	0.91	0.254 (0.102)	1.98	1.08	0.214 (0.065)
50	LASSO	1.98	4.78	0.337 (0.082)	2.00	4.13	0.278 (0.059)
	SCAD	1.90	4.64	0.284 (0.092)	2.00	3.55	0.214 (0.069)
	MCP	1.93	4.32	0.292 (0.115)	1.98	3.25	0.214 (0.067)
	BAR	1.82	2.07	0.373 (0.578)	1.97	1.48	0.221 (0.063)
100	LASSO	1.99	9.44	0.343 (0.092)	1.99	8.14	0.312 (0.069)
	SCAD	1.94	9.50	0.298 (0.121)	1.98	8.04	0.241 (0.079)
	MCP	1.90	8.45	0.346 (0.137)	1.99	7.19	0.249 (0.082)
	BAR	1.83	2.72	0.309 (0.113)	1.99	3.12	0.262 (0.079)
$\beta = (0.5, 0.5, 0.5, 0.5, 0, \dots, 0)$							
10	LASSO	3.92	1.03	0.649 (0.134)	3.97	1.10	0.588 (0.109)
	SCAD	3.60	0.81	0.591 (0.132)	3.81	0.87	0.530 (0.120)
	MCP	3.39	0.58	0.586 (0.137)	3.67	0.64	0.533 (0.125)
	BAR	3.24	0.25	0.615 (0.138)	3.61	0.37	0.552 (0.126)

Table 1. *Cont.*

$\beta = (0.5, 0.5, 0.5, 0.5, 0, \dots, 0)$							
30	LASSO	3.89	2.79	0.705 (0.127)	3.99	2.58	0.643 (0.112)
	SCAD	3.56	2.62	0.630 (0.139)	3.76	2.59	0.557 (0.121)
	MCP	3.38	2.27	0.614 (0.136)	3.67	2.07	0.540 (0.110)
	BAR	3.15	0.75	0.630 (0.140)	3.64	0.96	0.547 (0.109)
50	LASSO	3.90	4.49	0.754 (0.119)	3.98	4.40	0.663 (0.108)
	SCAD	3.51	4.30	0.679 (0.132)	3.70	4.38	0.578 (0.122)
	MCP	3.45	3.82	0.660 (0.144)	3.57	3.60	0.555 (0.123)
	BAR	3.08	1.19	0.673 (0.138)	3.57	1.68	0.566 (0.111)
100	LASSO	3.84	9.20	0.793 (0.139)	3.97	7.62	0.695 (0.086)
	SCAD	3.38	9.15	0.724 (0.157)	3.64	7.80	0.616 (0.094)
	MCP	3.05	8.04	0.742 (0.191)	3.52	6.50	0.600 (0.100)
	BAR	3.00	2.57	0.735 (0.172)	3.53	2.99	0.607 (0.101)

Table 2. Simulation results based on the joint selection of the tuning parameter and bandwidth $n = 300$ for (II).

$\beta = (0.5, 0.5, 0, \dots, 0)$				
p	Method	TP	FP	MMSE (SD)
10	LASSO	2.00	0.52	0.174 (0.047)
	SCAD	1.99	0.38	0.129 (0.048)
	MCP	1.99	0.32	0.127 (0.044)
	BAR	1.96	0.08	0.132 (0.047)
30	LASSO	2.00	1.47	0.197 (0.053)
	SCAD	1.99	1.20	0.151 (0.055)
	MCP	1.99	1.12	0.133 (0.049)
	BAR	1.92	0.31	0.150 (0.057)
50	LASSO	2.00	1.96	0.204 (0.048)
	SCAD	2.00	1.99	0.159 (0.054)
	MCP	1.96	1.57	0.140 (0.050)
	BAR	1.96	0.29	0.144 (0.048)
100	LASSO	2.00	4.01	0.215 (0.058)
	SCAD	2.00	1.99	0.159 (0.054)
	MCP	1.98	3.60	0.149 (0.057)
	BAR	1.97	0.70	0.159 (0.053)
$\beta = (0.5, 0.7, 0, \dots, 0)$				
10	LASSO	2.00	0.55	0.277 (0.055)
	SCAD	1.99	0.35	0.218 (0.053)
	MCP	1.99	0.28	0.213 (0.050)
	BAR	1.95	0.10	0.222 (0.055)
30	LASSO	2.00	1.40	0.304 (0.070)
	SCAD	1.98	1.07	0.232 (0.064)
	MCP	1.99	1.14	0.216 (0.062)
	BAR	1.95	0.29	0.231 (0.057)
50	LASSO	2.00	2.25	0.320 (0.061)
	SCAD	2.00	1.94	0.238 (0.062)
	MCP	2.00	1.85	0.227 (0.057)
	BAR	1.98	0.34	0.233 (0.056)

Table 2. *Cont.*

$\beta = (0.5, 0.7, 0, \dots, 0)$				
100	LASSO	2.00	3.76	0.325 (0.060)
	SCAD	1.99	3.82	0.251 (0.067)
	MCP	1.98	3.42	0.236 (0.062)
	BAR	1.94	0.68	0.249 (0.061)
$\beta = (0.5, 0.5, 0.5, 0.5, 0, \dots, 0)$				
10	LASSO	3.97	0.40	0.673 (0.102)
	SCAD	3.71	0.33	0.599 (0.100)
	MCP	3.58	0.23	0.594 (0.090)
	BAR	3.32	0.06	0.634 (0.106)
30	LASSO	3.99	1.43	0.731 (0.106)
	SCAD	3.77	1.20	0.645 (0.105)
	MCP	3.61	1.30	0.600 (0.107)
	BAR	3.44	0.26	0.626 (0.106)
50	LASSO	3.99	2.05	0.751 (0.092)
	SCAD	3.74	2.22	0.648 (0.103)
	MCP	3.63	1.81	0.618 (0.103)
	BAR	3.41	0.38	0.624 (0.112)
100	LASSO	3.96	3.55	0.775 (0.081)
	SCAD	3.76	3.77	0.692 (0.102)
	MCP	3.64	3.25	0.627 (0.118)
	BAR	3.38	0.55	0.644 (0.112)

Table 3. Simulation results based on the joint selection of the tuning parameter and bandwidth $n = 300$ for (III).

$\beta = (0.5, 0.5, 0, \dots, 0)$				
p	Method	TP	FP	MMSE (SD)
10	LASSO	2.00	0.63	0.165 (0.057)
	SCAD	1.98	0.42	0.125 (0.054)
	MCP	1.92	0.77	0.139 (0.064)
	BAR	1.95	0.14	0.128 (0.053)
30	LASSO	2.00	1.82	0.191 (0.057)
	SCAD	1.98	1.44	0.142 (0.058)
	MCP	1.90	2.68	0.187 (0.092)
	BAR	1.94	0.37	0.139 (0.055)
50	LASSO	2.00	2.83	0.200 (0.046)
	SCAD	1.99	2.31	0.153 (0.058)
	MCP	1.88	4.32	0.220 (0.099)
	BAR	1.96	0.42	0.137 (0.050)
100	LASSO	2.00	5.25	0.214 (0.055)
	SCAD	1.95	5.45	0.171 (0.071)
	MCP	1.96	1.57	0.140 (0.050)
	BAR	1.92	1.04	0.161 (0.066)
$\beta = (0.5, 0.7, 0, \dots, 0)$				
10	LASSO	2.00	0.78	0.261 (0.068)
	SCAD	1.98	0.53	0.204 (0.060)
	MCP	1.98	0.47	0.202 (0.056)
	BAR	1.94	0.13	0.211 (0.063)

Table 3. Cont.

$\beta = (0.5, 0.7, 0, \dots, 0)$				
30	LASSO	2.00	1.80	0.275 (0.067)
	SCAD	2.00	1.20	0.217 (0.064)
	MCP	2.00	1.42	0.201 (0.065)
	BAR	1.95	0.31	0.222 (0.064)
50	LASSO	2.00	2.54	0.305 (0.062)
	SCAD	1.95	2.36	0.223 (0.066)
	MCP	1.97	2.25	0.220 (0.061)
	BAR	1.96	0.58	0.225 (0.063)
100	LASSO	2.00	4.8	0.231 (0.081)
	SCAD	1.97	4.45	0.231 (0.066)
	MCP	1.97	4.70	0.158 (0.063)
	BAR	1.91	0.90	0.232 (0.070)
$\beta = (0.5, 0.5, 0.5, 0.5, 0, \dots, 0)$				
10	LASSO	3.96	0.58	0.618 (0.110)
	SCAD	3.74	0.40	0.553 (0.108)
	MCP	3.63	0.32	0.544 (0.104)
	BAR	3.43	0.11	0.578 (0.114)
30	LASSO	3.96	1.36	0.694 (0.106)
	SCAD	3.77	1.79	0.588 (0.104)
	MCP	3.59	1.07	0.571 (0.109)
	BAR	3.42	0.32	0.594 (0.113)
50	LASSO	3.97	2.19	0.702 (0.096)
	SCAD	3.69	2.54	0.627 (0.095)
	MCP	3.63	1.94	0.576 (0.102)
	BAR	3.39	0.58	0.600 (0.106)
100	LASSO	3.96	4.93	0.738 (0.093)
	SCAD	3.63	4.79	0.642 (0.102)
	MCP	3.57	4.70	0.612 (0.099)
	BAR	3.35	1.17	0.615 (0.117)

Furthermore, we added lower and greater sample sizes $n = 100$ and $n = 5000$ with $p = 5$ in Table 4 with $\beta_0 = (0.5, 0.5, \mathbf{0}_{p-2}^\top)^\top$. Furthermore, we added an extra simulation to demonstrate how this method works in the presence of noncontinuous covariates, i.e., the last covariates are generated from a Bernoulli distribution with a 0.5 success probability. The simulation results of this setup are presented in Table 5 with $n = 300$, $p = 10, 30, 50$, and $\beta_0 = (0.5, 0.5, \mathbf{0}^\top)^\top$. Finally, we considered a toy example in which the left endpoint imputation is considered (Sun, 2006). We illustrated the results in Table 6 to show the error that would be made if the data were considered as uncensored with $n = 100$, $p = 5$ and $\beta_0 = (0.5, 0.5, \mathbf{0}_{p-2}^\top)^\top$.

Table 4. Simulation results based on the joint selection of the tuning parameter and bandwidth for $n = 100$ or $n = 5000$ with $p = 5$.

$n = 100$					$n = 5000$		
$\beta = (0.5, 0.5, 0, \dots, 0)$							
p	Method	TP	FP	MMSE (SD)	TP	FP	MMSE (SD)
5	LASSO	1.81	0.68	0.234 (0.104)	2.00	0.74	0.084 (0.065)
	SCAD	1.60	0.58	0.220 (0.103)	2.00	0.36	0.077 (0.018)
	MCP	1.55	0.50	0.218 (0.107)	2.00	0.30	0.076 (0.019)
	BAR	1.21	0.14	0.267 (0.117)	2.00	0.03	0.079 (0.021)

Table 5. Simulation results based on the joint selection of the tuning parameter and bandwidth $n = 100$ for existing noncontinuous covariate.

$\beta = (0.5, 0.5, 0, \dots, 0)$				
p	Method	TP	FP	MMSE (SD)
10	LASSO	1.98	1.25	0.188 (0.074)
	SCAD	1.93	1.04	0.156 (0.076)
	MCP	1.90	0.94	0.155 (0.078)
	BAR	1.81	0.38	0.162 (0.085)
30	LASSO	1.99	3.42	0.229 (0.077)
	SCAD	1.93	3.07	0.199 (0.087)
	MCP	1.89	2.69	0.199 (0.093)
	BAR	1.82	1.26	0.200 (0.090)
50	LASSO	1.99	5.60	0.232 (0.074)
	SCAD	1.92	5.32	0.202 (0.086)
	MCP	1.90	4.83	0.218 (0.092)
	BAR	1.79	1.61	0.206 (0.087)

Table 6. Comparison of the proposed method to the left endpoint method.

Proposed Method					Left Endpoint Imputation		
$\beta = (0.5, 0.5, 0, \dots, 0)$							
p	Method	TP	FP	MMSE (SD)	TP	FP	MMSE (SD)
5	LASSO	1.81	0.68	0.234 (0.104)	1.89	0.10	0.313 (0.086)
	SCAD	1.60	0.58	0.220 (0.103)	1.58	0.06	0.305 (0.084)
	MCP	1.55	0.50	0.218 (0.107)	1.52	0.13	0.313 (0.071)
	BAR	1.21	0.14	0.268 (0.117)	0.97	0.02	0.385 (0.089)

One can see from Tables 1–3 that the proposed approach seems to perform well with all penalty functions considered and in general, the method with the BAR penalty function gave smaller FP and thus parsimonious models. Meanwhile, results are similar for (I), (II), and (III). As expected, the method with the LASSO penalty function gave slightly larger FP and tended to select more noises than other penalty functions, and the method gave better results on both variable selection and prediction accuracy when the sample size increased. Furthermore, as expected, the important covariates with weak effects were more difficult to be identified than moderate effect covariates. The stepwise variable selection gave the largest MMSE and SD. One can see from Table 4 that the results with a larger sample have lower MMSE and SD. The proposed method only needed 17 s, 15 s, 20 s, and 5 s for LASSO, SCAD, MCP and BAR, respectively, on average in Table 4. One can see from Table 5 that the proposed procedure works in the presence of noncontinuous covariates. The unbiased transformation method can improve accuracy as shown when comparing the proposed method to the left endpoint imputation in Table 6.

6. An Application

In this section, we apply the methodology proposed in the previous sections to a set of children's mortality data arising from the 2003 Nigeria Demographic and Health Survey (Kneib, 2006) [36]. The data set consists of 5730 children with their survival information and six covariates. They are AGE (the age of the children's mother when giving birth), BMI (the mother's body mass index at birth), HOSP (1 if the baby was delivered in a hospital and 0 otherwise), GENDER (1 for boys and 0 otherwise), EDU (1 if the mother received higher education and 0 otherwise), and URBAN (1 if the family lived in an urban area and 0 otherwise). Among others, one of the objectives of the study was to identify the covariates that had a significant effect on children's mortality in Nigeria.

In the study, for each subject, if death occurred within the first two months after birth, the failure time could be observed exactly, and otherwise, only interval-censored observa-

tions were obtained based on the interview times of their mothers. Among the 5730 children, 233 gave exact failure times, 430 interval-censored observations, and the others provided right-censored observations. To apply the proposed approach, as in the simulation study, we set $H(T) = \log(T + 1)$ and used the same four penalty functions. For the three functions $\phi_1(\cdot, \cdot)$, $\phi_2(\cdot, \cdot)$ and $\phi_3(\cdot, \cdot)$, we considered three choices and they are (I) $\phi_1(U_i, V_i) = 0, \phi_2(U_i, V_i) = 0, \phi_3(U_i, V_i) = H'(V_i)/\hat{g}_v(V_i)$; (II) $\phi_1(U_i, V_i) = 0, \phi_2(U_i, V_i) = H'(U_i)/\hat{g}_u(U_i), \phi_3(U_i, V_i) = H'(U_i)/\hat{g}_u(U_i)$; (III) $\phi_1(U_i, V_i) = 0, \phi_2(U_i, V_i) = H'(U_i)/2\hat{g}_u(U_i), \phi_3(U_i, V_i) = H'(U_i)/2\hat{g}_u(U_i) + H'(V_i)/2\hat{g}_v(V_i)$, which will be referred to as choices (I), (II), and (III) below, respectively. Note that for exact failure times, we used $H(T)$ as the response and made no unbiased transformation.

The results on the covariate selection and the estimated covariate effects are presented in Tables 7–9 with the use of choices (I)–(III), respectively. In addition, the estimated standard errors were obtained by using the nonparametric bootstrap method with 100 bootstrap samples and are included in the tables. One can see from these tables that the results seem to be robust with respect to the three choices, and all methods selected the factor AGE, suggesting that the age of the mother giving birth has a significant effect on the mortality risk of the children. The factor EDU was also selected by the LASSO, SCAD, and MCP penalty functions, and the results indicate that the level of the mother’s education seem to have a significant effect on the mortality risk. Xu et al. (2020) [37] have analyzed this application. They also found that EDU had a significant effect on the mortality risk of the children by the LASSO, ALASSO, and SCAD penalty functions. In contrast, the results suggest that the mother’s body mass index, the child’s gender, the baby’s birth location, and the family’s location did not appear to have significant effects on the children’s mortality or death rate.

Table 7. Analysis results of children’s mortality data based on choice (I).

Factor	LASSO	SCAD	MCP	BAR
Age	0.234 _(0.110)	0.248 _(0.110)	0.248 _(0.132)	0.162 _(0.134)
BMI	0.000 _(0.132)	0.000 _(0.111)	0.000 _(0.154)	0.000 _(0.110)
HOPS	0.000 _(0.123)	0.000 _(0.130)	0.000 _(0.118)	0.000 _(0.094)
GENDER	0.000 _(0.109)	0.000 _(0.108)	0.000 _(0.102)	0.000 _(0.088)
EDU	0.092 _(0.119)	0.115 _(0.143)	0.098 _(0.146)	0.000 _(0.115)
URBAN	0.000 _(0.126)	0.000 _(0.122)	0.000 _(0.101)	0.000 _(0.120)

Table 8. Analysis results of children’s mortality data based on choice (II).

Factor	LASSO	SCAD	MCP	BAR
Age	0.238 _(0.109)	0.252 _(0.115)	0.252 _(0.105)	0.167 _(0.090)
BMI	0.000 _(0.108)	0.000 _(0.123)	0.000 _(0.118)	0.000 _(0.100)
HOPS	0.000 _(0.131)	0.000 _(0.107)	0.000 _(0.144)	0.000 _(0.109)
GENDER	0.000 _(0.121)	0.000 _(0.119)	0.000 _(0.122)	0.000 _(0.119)
EDU	0.110 _(0.102)	0.133 _(0.161)	0.133 _(0.118)	0.000 _(0.105)
URBAN	0.000 _(0.127)	0.000 _(0.114)	0.000 _(0.104)	0.000 _(0.091)

Table 9. Analysis results of children’s mortality data based on choice (III).

Factor	LASSO	SCAD	MCP	BAR
Age	0.243 _(0.093)	0.257 _(0.111)	0.257 _(0.105)	0.162 _(0.134)
BMI	0.000 _(0.119)	0.000 _(0.127)	0.000 _(0.154)	0.000 _(0.110)
HOPS	0.000 _(0.145)	0.000 _(0.148)	0.000 _(0.118)	0.000 _(0.094)
GENDER	0.000 _(0.123)	0.000 _(0.101)	0.000 _(0.102)	0.000 _(0.088)
EDU	0.128 _(0.097)	0.149 _(0.109)	0.149 _(0.119)	0.000 _(0.115)
URBAN	0.000 _(0.117)	0.000 _(0.118)	0.000 _(0.105)	0.000 _(0.120)

7. Discussion and Concluding Remarks

This paper discussed the variable selection for generalized linear models when only interval-censored failure time data are available, and for that problem, a new unbiased-transformation-based approach was proposed. One advantage for the use of the unbiased transformation is that it allows one to employ the simple penalized least squares approach for estimation. The proposed approach can accommodate any penalty function such as LASSO, SCAD, MCP, and BAR, and a coordinate descent algorithm was developed for the implementation of the proposed procedure. In addition, the asymptotic properties of the resulting estimators were established, and the simulation study indicated that the proposed methodology works well for practical situations.

There exist several directions for future research. One is that in the proposed method, it was assumed that $H(T) < \infty$ and it is clear that this may not hold in practice. One such example is the accelerated failure time (AFT) model, in which case the proposed method would not be valid. In other words, one needs to modify the proposed unbiased transformation or develop some new methods that can be applied to the AFT model. Another direction is that in practical applications, one may encounter multivariate failure time data; in that scenario, Cai et al. (2005) [14] proposed a penalized marginal likelihood method for right-censored data with a mass of covariates. It would be helpful to develop some flexible and reliable methods to handle other types of multivariate failure time data including interval-censored data. A third direction is that it has been assumed in the previous sections that the dimension of covariates p_n can diverge to infinity but is smaller than the sample size n . Obviously, there may be case where p_n is greater than n such as in genetic or biomarker studies. In other words, some new methods that allow for $p_n > n$ need to be developed.

As one anonymous reviewer pointed out, neutrosophic statistics (Smarandache, 1998 [38]; Smarandache, 2013 [39]) is the extension of classical statistics and is applied when the data is coming from a complex process or from an uncertain environment. The current study can be extended using neutrosophic statistics as future research.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/math10050763/s1>, Table S1: Simulation results based on the bandwidth selection method (a), Table S2: Simulation results based on the bandwidth selection method (b), Table S3: Simulation results based on the bandwidth selection method (c), Table S4: Simulation results based on the bandwidth selection method (d).

Author Contributions: Formal analysis, R.L.; Investigation, S.Z.; Methodology, T.H.; Software, R.L.; Supervision, T.H. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Beijing Natural Science Foundation Z210003 and National Nature Science Foundation of China (Grant Nos: 12171328, 11971064). Shishun Zhao's study was supported by National Natural Science Foundation of China (NSFC) (12071176), Science and Technology Developing Plan of Jilin Province (No. 20200201258JC, 20190902012TC, 20190304129YY).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. Interested researchers can obtain the data in the application section of this paper on the official website of Demographic and Health Surveys (DHS) Program: <https://dhsprogram.com/data/available-datasets.cfm> (accessed on 19 February 2022). One can registered for a download account and apply for the data.

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

Appendix A

In this appendix, we sketch the proof of Theorem 1. For completeness, we introduce a few lemmas and notations, which are useful for the proof of Theorem 1. Let $C_n = X^\top X/n$. When we consider BAR penalty, i.e.,

$$\arg \min_{\beta} n^{-1} \sum_{i=1}^n (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - X_i^\top \beta)^2 + \lambda \sum_{j=1}^n \beta_j^2 / \tilde{\beta}_j^2. \quad (A1)$$

In Section 2 we can get

$$\beta^k = g(\beta^{k-1}) = \left\{ X^\top X + \lambda_n D(\beta^{k-1}) \right\}^{-1} X^\top \hat{Y}^* = \left(\alpha^*(\beta^{k-1})^\top, \gamma^*(\beta^{k-1})^\top \right)^\top. \quad (A2)$$

For simplicity, we write $\alpha^*(\beta)$ and $\gamma^*(\beta)$ as α^* and γ^* . C_n^{-1} can be partitioned as

$$C_n^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}^\top & A_{22} \end{pmatrix},$$

where the A_{11} is a $q \times q$ matrix. Multiplying $(X^\top X)^{-1} (X^\top X + \lambda_n D(\beta))$ to Equation (A2) gives

$$\begin{pmatrix} \alpha^* - \beta_{01} \\ \gamma^* \end{pmatrix} + \lambda_n/n \begin{pmatrix} A_{11} D_1(\alpha) \alpha^* + A_{12} D_2(\gamma) \gamma^* \\ A_{12}^\top D_1(\alpha) \alpha^* + A_{22} D_2(\gamma) \gamma^* \end{pmatrix} = (X^\top X)^{-1} X^\top \hat{\epsilon}^* = \hat{\beta}_{least} - \beta_0,$$

where $\hat{\epsilon}_i^* = \epsilon_i^* + (\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^*)$, $\hat{\beta}_{least} = (X^\top X)^{-1} X^\top \hat{Y}^*$, $D_1(\alpha) = \text{diag}(\alpha_1^{-2}, \dots, \alpha_{q_n}^{-2})$, $D_2(\gamma) = \text{diag}(\gamma_1^{-2}, \dots, \gamma_{p_n-q_n}^{-2})$, and $\epsilon_i^* = h_i^* - n^{-1} \sum_{q=1}^n h_q^* - X_i^\top \beta$.

Lemma A1. Assume that the conditions C2, C6 hold and $\phi_1(\cdot, \cdot)$, $\phi_2(\cdot, \cdot)$, and $\phi_3(\cdot, \cdot)$ are continuous functions with finite continuous partial derivatives. We have

$$\sup_{u,v} |\hat{h}^*(\hat{g}_u(u), \hat{g}_v(v)) - h^*(g_u(u), g_v(v))| \rightarrow 0 \quad a.s. \quad as \quad n \rightarrow \infty.$$

Proof. Under the conditions C2 and C6, we have that $\sup_u |\hat{g}_u(u) - g_u(u)| \rightarrow 0 \quad a.s.$ and $\sup_v |\hat{g}_v(v) - g_v(v)| \rightarrow 0 \quad a.s.$ as $n \rightarrow \infty$ according to Theorem A in Silverman (1978) [31]. Using Taylor's expansion, $\phi_1(\cdot, \cdot)$, $\phi_2(\cdot, \cdot)$, and $\phi_3(\cdot, \cdot)$ are continuous functions with finite continuous partial derivatives yielding

$$\sup_{u,v} |\hat{h}^*(\hat{g}_u(u), \hat{g}_v(v)) - h^*(g_u(u), g_v(v))| \rightarrow 0 \quad a.s. \quad as \quad n \rightarrow \infty.$$

The proof of Lemma A1 is completed. $\square$

Lemma A2. Let $\hat{\beta}_{least}$ denote the least squares estimator defined in (3), β_0 is a true value of β and suppose that the conditions C1 to C8 hold. Then, we have that

$$\|\hat{\beta}_{least} - \beta_0\| = O_p(\sqrt{p_n/n}),$$

Proof. When we consider (A1), according to C4, we get $\text{Var}(\epsilon) < \infty$. According to (3), we get $n^{-1} \sum_{i=1}^n (h_i^* - n^{-1} \sum_{q=1}^n h_q^* - X_i^\top \beta + \hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^*)^2$. After some simple algebraic manipulations, we get

$$(\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^* + \epsilon^*)^2 \leq 2(\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^*)^2 + 2\epsilon^{*2}.$$

Under the conditions C1 to C8, we have that $2(\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^*)^2 = o_p(1)$. According to Lemma A1, $nE(\hat{h}_i^* - n^{-1} \sum_{q=1}^n \hat{h}_q^* - h_i^* + n^{-1} \sum_{q=1}^n h_q^* + \epsilon_i^*)^2 = nE(\epsilon_i^{*2}) +$

$no_p(1)$. $nE\|\hat{\beta}_{least} - \beta_0\| = nE(\|(X^\top X)^{-1}X^\top(\hat{Y}^* - Y^* + \epsilon^*)\|^2) = (E(\epsilon^{*\top}\epsilon^*) + o_p(1)) \cdot \text{trace}(C_n) = O_p(p_n)$. We get that $\|\hat{\beta}_{least} - \beta_0\|^2 = O_p(p_n/n)$. $\square$

Lemma A3. Suppose that the conditions C2, C4, and C6 hold. Then, we have that

$$\lim_{n \rightarrow \infty} P(|\text{Var}(\hat{Y}^*) - \text{Var}(Y^*)| > \epsilon) = 0.$$

Proof. According to C4, we know $\text{Var}(Y^*) = E(\epsilon^{*\top}\epsilon^*) < \infty$. According to Lemma A1, $\hat{h}_i^*$ converges to h_i^* in probability for $i = 1, \dots, n$. Then, $\text{Var}(\hat{Y}^*) = E\{(\hat{Y}^* - Y^* + \epsilon^*)^\top(\hat{Y}^* - Y^* + \epsilon^*)\} = E((\epsilon^{*\top}\epsilon^*) + \epsilon^*(\hat{Y}^* - Y^*) + (\hat{Y}^* - Y^*)^2) = E(\epsilon^{*\top}\epsilon^*) + o_p(1)$. So Lemma A3 is proved. $\square$

Proof of Theorem 1 (Oracle Property). For a zero component, according to Lemma 1 in [12] conclusion (i), let δ_n be a sequence of positive real numbers satisfying $\delta_n \rightarrow \infty$ and $\delta_n^2 p_n / \lambda_n \rightarrow 0$. Define $H_n = \{\beta \in \mathbb{R}^{p_n} : \|\beta - \beta_0\| \leq \delta_n \sqrt{p_n/n}\}$, $H_{n1} = \{\alpha \in \mathbb{R}^{q_n} : \|\alpha - \beta_{01}\| \leq \delta_n \sqrt{p_n/n}\}$. Then, with probability tending to 1, we have

$$\|\gamma^*\| < \|\gamma\| \leq \delta_n \sqrt{p_n/n} \rightarrow 0. \quad (\text{A3})$$

For a nonzero component, according to Lemma 2 in [12], $f(\alpha)$ is a contraction mapping from H_{n1} to itself, where

$$f(\alpha) = \{X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha)\}^{-1} X_\alpha^\top \hat{Y}^*.$$

Let $\hat{\alpha}^\circ$ is the unique fixed point of $f(\alpha)$ defined by $\hat{\alpha}^\circ = \{X_\alpha^\top X_\alpha + \lambda_n D_1(\hat{\alpha}^\circ)\}^{-1} X_\alpha^\top \hat{Y}^*$.

Hence, to prove the consistency of nonzero parts in Theorem 1, it is sufficient to show that

$$\Pr\left(\lim_{k \rightarrow \infty} \|\hat{\alpha}^{(k)} - \hat{\alpha}^\circ\| = 0\right) \rightarrow 1. \quad (\text{A4})$$

Define $\gamma^* = 0$ if $\gamma = 0$. It is easy to see from (A3) that for any $\alpha \in H_{n1}$,

$$\lim_{\gamma \rightarrow 0} \gamma^*(\alpha, \gamma) = 0. \quad (\text{A5})$$

Combining (A5) with the fact that

$$\begin{pmatrix} X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha) & X_\alpha^\top X_\gamma \\ X_\gamma^\top X_\alpha & X_\gamma^\top X_\gamma + \lambda_n D_2(\gamma) \end{pmatrix} \begin{pmatrix} \alpha^* \\ \gamma^* \end{pmatrix} = \begin{pmatrix} X_\alpha^\top \hat{Y}^* \\ X_\gamma^\top \hat{Y}^* \end{pmatrix},$$

we find that, for any $\alpha \in H_n$,

$$\lim_{\gamma \rightarrow 0} \alpha^*(\alpha, \gamma) = \{X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha)\}^{-1} X_\alpha^\top \hat{Y}^* = f(\alpha). \quad (\text{A6})$$

According to conclusion b in Lemma 1 of Dai et al. (2018) [12], g in (A2) is a mapping from H_n to itself. This, together with (A3) and (A6), implies that, as $k \rightarrow \infty$,

$$\eta_k \equiv \sup_{\alpha \in H_n} \|f(\alpha) - \alpha^*(\alpha, \hat{\gamma}^{(k)})\| \rightarrow 0. \quad (\text{A7})$$

with probability tending to 1. Note that

$$\begin{aligned} \|\hat{\alpha}^{(k+1)} - \hat{\alpha}^\circ\| &= \|\alpha^*(\hat{\beta}^{(k)}) - \hat{\alpha}^\circ\| \leq \|\alpha^*(\hat{\beta}^{(k)}) - f(\hat{\alpha}^{(k)})\| + \|f(\hat{\alpha}^{(k)}) - \hat{\alpha}^\circ\| \\ &\leq \eta_k + \frac{1}{E} \|\hat{\alpha}^{(k)} - \hat{\alpha}^\circ\|, \end{aligned}$$

where the last step is on the basis of $\|f(\hat{\alpha}^{(k)}) - \hat{\alpha}^\circ\| = \|f(\hat{\alpha}^{(k)}) - f(\hat{\alpha}^\circ)\| \leq \|\hat{\alpha}^{(k)} - \hat{\alpha}^\circ\|/E$.

Let $a_k = \|\hat{\alpha}^{(k)} - \hat{\alpha}^\circ\|$, for every integer $k \geq 0$. From (A7), for any $\epsilon > 0$, there exists a positive integer N such that for every integer $k > N$, $|\eta_k| < \epsilon$

$$\begin{aligned} a_{k+1} &\leq \frac{a_{k-1}}{E^2} + \frac{\eta_{k-1}}{E} + \eta_k \leq \frac{a_1}{E^k} + \frac{\eta_1}{E^{k-1}} + \dots + \frac{\eta_N}{E^{k-N}} + \left(\frac{\eta_{N+1}}{E^{k-N-1}} + \dots + \frac{\eta_{k-1}}{E} + \eta_k\right) \\ &\leq (a_1 + \eta_1 + \dots + \eta_N) \frac{1}{E^{k-N}} + \frac{1 - (1/E)^{k-N}}{1 - 1/E} \epsilon \end{aligned}$$

with probability 1 as n tends to ∞ and the right-hand term tends to 0 as $k \rightarrow \infty$. This proves (A4). Therefore, $\lim_{k \rightarrow \infty} \beta^{(k)} = (\hat{\alpha}^{\circ\top}, 0)^\top$. $\square$

Proof of Theorem 1 (Asymptotic Normality). We next prove the asymptotic normality of the nonzero components in the BAR regression. Finally, based on (A6), we have

$$\sqrt{n}(\hat{\alpha}^* - \alpha) = \Pi_1 + \Pi_2,$$

where

$$\Pi_1 = \sqrt{n}[(X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha))^{-1} X_\alpha X_\alpha^\top - I_{q_n}] \alpha,$$

$$\Pi_2 = \sqrt{n}[(X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha))^{-1} (X_\alpha^\top \hat{Y}^* - X_\alpha^\top X_\alpha \alpha)].$$

According to the first-order resolvent expansion formula, we obtain

$$\begin{aligned} (X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha))^{-1} &= (X_\alpha^\top X_\alpha)^{-1} - \lambda_n (X_\alpha^\top X_\alpha)^{-1} D_1(\alpha) (X_\alpha^\top X_\alpha \\ &\quad + \lambda_n D_1(\alpha))^{-1}. \end{aligned}$$

This yields

$$\begin{aligned} \Pi_1 &= -\lambda_n / \sqrt{n} (X_\alpha^\top X_\alpha / n)^{-1} D_1(\hat{\alpha}) (X_\alpha^\top X_\alpha + \lambda_n D_1(\alpha))^{-1} \\ &\quad \times X_\alpha X_\alpha^\top \alpha. \end{aligned}$$

With conditions C7 and C8, we can get

$$\|\Pi_1\| = O_p(\lambda_n \sqrt{q_n/n}) \rightarrow 0,$$

$$\Pi_2 = \sqrt{n}[(n^{-1} X_\alpha^\top X_\alpha + o_p(n^{-1/2}))^{-1} (n^{-1} X_\alpha^\top \hat{Y}^* - n^{-1} X_\alpha^\top X_\alpha \alpha)].$$

Therefore, we can get

$$\sqrt{n}(\hat{\alpha}^* - \alpha) = E(X_\alpha^\top X_\alpha)^{-1} n^{-1/2} (X_\alpha^\top \hat{Y}^* - X_\alpha^\top X_\alpha \alpha) + o_p(1).$$

According to Lemma A3, we know that $\text{Var}(\hat{Y}^*)$ converge to $\text{Var}(Y^*)$ in probability. Then, we get

$$\sqrt{n} \Sigma_{(1)}^{(-1/2)} (\hat{\alpha}^* - \alpha^*) \rightarrow N(0, I_{q \times q})$$

in distribution. Denote $E(X_\alpha^\top X_\alpha)^{-1} \text{Var}(X_\alpha^\top Y^*) E(X_\alpha^\top X_\alpha)^{-1}$ as $\Sigma_{(1)}$. This complete the proof. $\square$

References

- Goldberger, A.S.; Jochems, D.B. Note on stepwise least squares. *J. Am. Stat. Assoc.* **1961**, *56*, 105–110. [\[CrossRef\]](#)
- Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [\[CrossRef\]](#)
- Mitchell, T.J.; Beauchamp, J.J. Bayesian Variable Selection in Linear Regression. *J. Am. Stat. Assoc.* **1988**, *83*, 1023–1041. [\[CrossRef\]](#)
- Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [\[CrossRef\]](#)
- Raftery, A.E.; Madigan, D.; Hoeting, J.A. Bayesian Variable model averaig for Linear Regression Models. *J. Am. Stat. Assoc.* **1997**, *92*, 179–191. [\[CrossRef\]](#)
- Zou, H.; Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B* **2005**, *67*, 301–320. [\[CrossRef\]](#)
- Zou, H. The Adaptive Lasso and Its Oracle Properties. *J. Am. Stat. Assoc.* **2006**, *101*, 1418–1429. [\[CrossRef\]](#)
- Fan, J.; Li, R. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [\[CrossRef\]](#)
- Zhang, C.H. Nearly Unbiased Variable Selection Under Minimax Concave Penalty. *Ann. Stat.* **2010**, *38*, 894–942. [\[CrossRef\]](#)
- Dicker, L.; Huang, B.; Lin, X. Variable Selection and Estimation with the Seamless- L_0 penalty. *Stata Sin.* **2013**, *23*, 929–962. [\[CrossRef\]](#)
- Liu, Z.; Li, G. Efficient Regularized Regression with L_0 Penalty for Variable Selection and Network Construction. *Comput. Math. Methods Med.* **2016**, *2016*, 3456153. [\[CrossRef\]](#)
- Dai, L.; Chen, K.; Sun, Z.; Liu, Z.; Li, G. Broken Adaptive Ridge Regression and Its Asymptotic Properties. *J. Multivar. Anal.* **2018**, *168*, 334–351. [\[CrossRef\]](#) [\[PubMed\]](#)

13. Zheng, X.; Rong, Y.; Liu, L.; Cheng, W. A More Accurate Estimation of Semiparametric Logistic Regression. *Mathematics* **2021**, *9*, 2376. [\[CrossRef\]](#)
14. Cai, J.; Fan, J.; Li, R.; Zhou, H. Variable Selection for Multivariate Failure Time Data. *Biometrika* **2005**, *92*, 303–316. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Fan, J.; Li, R. Variable Selection for Cox's Proportional Hazards Model and Frailty Model. *Ann. Stat.* **2002**, *30*, 74–99. [\[CrossRef\]](#)
16. Tibshirani, R. The Lasso Method for Variable Selection in the Cox Model. *Stat. Med.* **1997**, *16*, 385–395. [\[CrossRef\]](#)
17. Zhang, H.H.; Lu, W. Adaptive Lasso for Cox's proportional hazards model. *Biometrika* **2007**, *94*, 691–703. [\[CrossRef\]](#)
18. Khan, N.; Aslam, M.; Raza, S.M.; Jun, C. A New Variable Control Chart under Failure-censored Reliability Tests for Weibull distribution. *Qual. Reliab. Eng. Int.* **2018**, *35*, 572–581. [\[CrossRef\]](#)
19. Zhao, H.; Wu, Q.; Li, G.; Sun, J. Simultaneous Estimation and Variable Selection for Interval-Censored Data With Broken Adaptive Ridge Regression. *J. Am. Stat. Assoc.* **2020**, *115*, 204–216. [\[CrossRef\]](#)
20. Li, S.; Wei, Q.; Sun, J. Penalized estimation of semiparametric transformation models with interval-censored data and application to Alzheimer's disease. *Stat. Methods Med. Res.* **2019**, *29*, 2151–2166. [\[CrossRef\]](#)
21. Du, M.; Sun, J. Variable selection for interval-censored failure time data. *Int. Stat. Rev.* **2021**, *accepted*. [\[CrossRef\]](#)
22. Ali, S.; Raza, S.M.; Aslam, M.; Butt, M.M. CEV-Hybrid DEWMA charts for censored data Using Weibull distribution. *Commun. Stat.—Simul. Comput.* **2021**, *50*, 446–461. [\[CrossRef\]](#)
23. Zhao, S.; Hu, T.; Ma, L.; Wang, P.; Sun, J. Regression analysis of informative current status data with the additive hazards model. *Lifetime Data Anal.* **2015**, *21*, 241–258. [\[CrossRef\]](#)
24. Wang, P.; Zhao, H.; Sun, J. Regression Analysis of Case K Interval Censored Failure Time Data in the Presence of Informative Censoring. *Biometrics* **2016**, *72*, 1103–1112. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Li, S.; Hu, T.; Wang, P.; Sun, J. A Class of Semiparametric Transformation Models for Doubly Censored Failure Time Data. *Scand. J. Stat.* **2018**, *45*, 682–698. [\[CrossRef\]](#)
26. Sun, J. *The Statistical Analysis of Interval-Censored Failure Time Data*; Springer: New York, NY, USA, 2006.
27. Buckley, J.; James, I. Linear regression with censored data. *Biometrika* **1979**, *66*, 429–436. [\[CrossRef\]](#)
28. Deng, W. Some Issues on Interval Censored Data. Ph.D. Dissertation, Fudan University, Shanghai, China, 2004.
29. Deng, W.; Tian, Y.; Lv, Q. Parametric Estimator of Linear Model with Interval-Censored Data. *Commun. Stat.—Simul. Comput.* **2012**, *41*, 1794–1804. [\[CrossRef\]](#)
30. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [\[CrossRef\]](#)
31. Silverman, B.W. Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives. *Ann. Statist.* **1978**, *6*, 177–184. [\[CrossRef\]](#)
32. Wu, Q.; Zhao, H.; Zhu, L.; Sun, J. Variable Selection for High-dimensional Partly Linear Additive Cox model with application to Alzheimer's Disease. *Stat. Med.* **2020**, *39*, 3120–3134. [\[CrossRef\]](#)
33. Sun, L.; Li, S.; Wang, L.; Song, X. Variable Selection in semiparametric nonmixture cure model with interval-censored failure time data: An application to the prostate cancer screening study. *Stat. Med.* **2019**, *38*, 3026–3039. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Schwarz, G. Estimating the Dimension for a Model. *Ann. Stat.* **1978**, *6*, 461–464. [\[CrossRef\]](#)
35. Verweij, P.; Houwelingen, H. Cross-validation in survival analysis. *Stat. Med.* **2010**, *12*, 2305–2314. [\[CrossRef\]](#)
36. Kneib, T. Mixed Model-Based Inference in Geoaddditive Hazard Regression for Interval-Censored Survival Times. *Comput. Stat. Data Anal.* **2006**, *51*, 777–792. [\[CrossRef\]](#)
37. Xu, Y.; Zhao, S.; Tao, H.; Sun, J. Variable Selection for generalized odds rate mixture cure models with interval-censored failure time data. *Comput. Stat. Data Anal.* **2020**, *156*, 107115. [\[CrossRef\]](#)
38. Smarandache, F. *Neutrosophic Probability, Set, and Logic*; American Research Press: Rehoboth, NM, USA, 1998.
39. Smarandache, F. *Introduction to Neutrosophic Measure, Neutrosophic Integral, and Neutrosophic Probability*; Sitech-Education Publisher: Craiova, Romania, 2013.