Journal of Statistical Computation and Simulation, 2016
Vol. 86, No. 3, 431–442, http://dx.doi.org/10.1080/00949655.2015.1016432

Exponential regression for censored data with outliers

Jing Zhang, Yanyan Liu and Yuanshan Wu*

School of Mathematics and Statistics, Wuhan University, Wuhan, Hubei 430072, People's Republic of China

(*Received 11 May 2014; accepted 3 February 2015*)

We propose a penalized likelihood method to detect the possible outliers in the exponential regression model while it is utilized to fit the censored survival data. It features that the proposed method can simultaneously cope with outlier detection and estimation for the regression coefficient. We recast the outlier detection issue into a high-dimensional regularization regression and employ the coordinate descent algorithm to facilitate the computation. From both extensive simulation studies and an illustrative real example, it is shown that the proposed method works quite well in outlier detection as well as parameter estimation for the exponential regression model.

Keywords: censored data; exponential regression; outlier detection; penalty function; smoothly clipped absolute deviation

1. Introduction

Most real-world data sets contain outliers that have unusually larger or smaller values when compared with others in the data set. Outliers may cause a negative effect on data analysis, or may provide useful information about data when we look into an unusual response to a given study. Thus, outlier detection is an important part of data analysis. Grubbs [1] pointed out that an outlier is one that appears to deviate markedly from other members of the sample in which it occurs. Furthermore, Barnett and Lewis [2] considered that an observation which appears to be inconsistent with the remainder of that set of data. Outliers usually arise because of human error, instrument error, natural deviations in populations, fraudulent behaviour, changes in behaviour of systems or faults in systems. Hample et al. [3] estimated that a routine data set may contain about 1–10% (or more) outliers. In many cases, outliers may have serious effects in estimation, inference, and model selection. Weisberg [4] pointed out that although outliers have these bad effects, sometimes they often go unnoticed. Consequently, it is vital to detect the potential outliers for a given data set before conducting the routine statistical inference.

The outlier detection methods can be roughly classified into two categories: testing the discordancy and labelling the outliers. For the discordancy test method, it usually tests whether or not the target extreme value deviates from the assumed underlying well-behaving distribution. Iglewicz and Hoaglin [5] reviewed and compared five test statistics which are applied to the normal distribution. On the other hand, the outlier labelling method usually proposes an interval

*Corresponding author. Email: shan@whu.edu.cn or criterion, and any observation beyond the interval or criterion is considered as an outlier. The classical approach to labelling the data for outliers is the Z-score which is defined by the ratio of the sample mean and the sample standard deviation. However, the sample mean or standard deviation can be affected by the extreme values. Iglewicz and Hoaglin [5] proposed a robust version of the Z-score method, where the median and the median of the absolute deviation of the median are employed to replace the mean and the standard deviation, respectively.

It also has been appeared extensive literature on the outlier detection in the context of regression models. The observation with larger residual is usually more likely to be suspected as an outlier. Furthermore, the leave-one-out method [4] and Cook's distance [6] and their variants are frequently used in outlier detection for regression analysis. She and Owen [7] proposed a new outlier detection method from the view of regularization method by using the nonconvex penalized linear regression.

Although the methods for the outlier detection are relatively well developed for the completely observed data, such as the linear regression model and the general linear model,[2,8] the detection of outliers for the censored survival data has received little attention despite that the censored data are commonly encountered in clinical trial and medical science. The aforementioned methods cannot be applicable when response are subject to censoring. What is more, the censoring may mask the potential outliers. Some martingale-based residual methods were proposed by Gramhsch and Therneau [9] to suspect the possible outliers for censored data in terms of graphics. It is intuitive but not rigorous.

Motivated by She and Owen,[7] we propose a penalized likelihood method to detect the possible outliers in the exponential regression model while it is utilized to fit the censored survival data. It features that the proposed method can simultaneously cope with outlier detection and parameter estimation for the regression coefficient. We recast the outlier detection issue into a framework of high-dimensional regularization regression so that we can borrow the sophisticated techniques for the analysis of high-dimensional data to handle the outlier detection of interest. Furthermore, we extend the proposed method to the outlier detection for the high-dimensional exponential regression model.

The rest of the article is organized as follows. In Section 2, we present the outlier detection method for censored data in the framework of the exponential regression model. The iterative coordinate descent algorithm is also developed. We conduct extensive simulation studies to evaluate the proposed method in Section 3. An illustrative real example is analysed in Section 4. Some concluding remarks are relegated in Section 5.

2. Outlier detection and parameter estimation

Let T denote the failure time and C the censoring time. Correspondingly, let $Y = \min(T, C)$ be the observed time and $\Delta = I(T \leq C)$ be the failure indicator. Let $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}}$ be the associated covariates. Assume that the censoring mechanism is random, that is, the survival time T and the censoring time C are conditionally independent given $\mathbf{X}$.

For the ith subject $(i = 1, \ldots, n)$, assume that, given covariate $\mathbf{X}_i$, the conditional survival time T_i follows the exponential distribution with rate $r(\mathbf{X}_i)$. Further assume

$$r(\mathbf{X}_i) = \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i),$$

where $\boldsymbol{\beta}$ is the unknown regression parameter, and the unknown shifted parameter γ_i is incorporated into the exponential regression to indicate whether the ith observation is an outlier. Let $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$ and assume the sparsity on $\boldsymbol{\gamma}$. Consequently, the nonzero components of $\boldsymbol{\gamma}$

will correspond to the outliers. In what follows, we will propose a method to identify the outliers, or equivalently, to find the nonzero components of $\boldsymbol{\gamma}$, as well as to estimate the unknown regression coefficient $\boldsymbol{\beta}$.

Based on the independent observations $(Y_i, \Delta_i, \mathbf{X}_i)$, $i = 1, \dots, n$, the likelihood function can be written as

$$L_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \prod_{i=1}^{n} [\{\exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i)\}^{\Delta_i} \exp\{-Y_i \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i)\}].$$

Obviously, we need estimate $\boldsymbol{\gamma}$ as well as $\boldsymbol{\beta}$, but the dimension of the unknown parameters is $n+p$ which is greater than the sample size n, nevertheless p may be less than n. We recast the task of detecting outliers into a framework of high-dimensional variable selection so that we can employ the well-developed tools such as the regularization regression method to handle our concerns in this paper.

Fan and Li [10] advocated the smoothly clipped absolute deviation (SCAD) penalty, which satisfies $p_\lambda(0) = 0$ and has the first-order derivative

$$p'_\lambda(\theta) = \lambda \left\{ I(\theta \le \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda} I(\theta > \lambda) \right\},$$

for some $a > 2$ and $\theta > 0$, where $\lambda > 0$ is the tuning parameter. The SCAD penalty is continuous and differentiable on $(-\infty, 0) \cup (0, \infty)$, but not differentiable at 0. Its derivative vanishes outside $[-a\lambda, a\lambda]$. Hence, the SCAD penalty can produce estimators with continuity, sparsity, and unbiasedness for the large coefficients. More details can be found in [10]. Zou and Li [11] proposed a local linear approximation to the SCAD penalty, which maintains the same asymptotic properties and significantly improves the computational efficiency. As suggested by Fan and Li,[10] we fix $a = 3.7$. To emphasize the dependence of λ on n, we denote λ by λ_n hereafter.

As a result, the penalized log likelihood is given by

$$Q_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) - n \sum_{j=1}^{n} p'_{\lambda_n}(|\hat{\gamma}_j^{(0)}|)|\gamma_j|,$$

where

$$l_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \log L_n(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{i=1}^{n} \{\Delta_i(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i) - Y_i \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i)\}$$

and $\hat{\gamma}_j^{(0)}$ is an initial estimator.

Define

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}) = \operatorname{argmax}_{\boldsymbol{\beta}, \boldsymbol{\gamma}} Q_n(\boldsymbol{\beta}, \boldsymbol{\gamma}),$$

which are the proposed estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. However, the maximizing procedure is nontrivial. We propose an iterative algorithm between $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ to get estimators and meanwhile the coordinate descent algorithm [12] is utilized to obtain the estimator of $\boldsymbol{\gamma}$ for each iteration. The coordinate descent algorithm, which was also used by Wu and Lange [13] and Breheny and Huang,[14] is one of the most efficient algorithms in processing large-scale data for its simple operation and fast convergence. The procedure of our algorithm proceeds as follows:

(1) Set an initial estimate $\hat{\boldsymbol{\beta}}^{(0)}$, which is obtained by the exponential regression ignoring the possible outliers. Set $k = 1$.

(2) Let $\hat{\boldsymbol{\gamma}}^{(k)} = \text{argmax}_{\boldsymbol{\gamma}} Q_n(\hat{\boldsymbol{\beta}}^{(k-1)}, \boldsymbol{\gamma})$, where the coordinate descent algorithm is adopted to do the maximization. Specifically, for $j = 1, \ldots, n$,

$$\hat{\gamma}_j^{(k)} = \text{argmax}_{\gamma_j} Q_n(\hat{\boldsymbol{\beta}}^{(k-1)}, (\hat{\gamma}_1^{(k)}, \ldots, \hat{\gamma}_{j-1}^{(k)}, \gamma_j, \hat{\gamma}_{j+1}^{(k-1)}, \ldots, \hat{\gamma}_n^{(k-1)})^{\mathrm{T}})$$

and $\hat{\boldsymbol{\gamma}}^{(k)} = (\hat{\gamma}_1^{(k)}, \ldots, \hat{\gamma}_n^{(k)})^{\mathrm{T}}$.

(3) Let $\hat{\boldsymbol{\beta}}^{(k)} = \text{argmax}_{\boldsymbol{\beta}} Q_n(\boldsymbol{\beta}, \hat{\boldsymbol{\gamma}}^{(k)})$.

(4) Update $k = k + 1$ and repeat Steps 2 and 3 until the prespecified convergence criterion is met.

Obviously, the iteration sequence $(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{\gamma}}^{(k)})$ satisfies that

$$Q_n(\hat{\boldsymbol{\beta}}^{(k)}, \hat{\boldsymbol{\gamma}}^{(k)}) \geq Q_n(\hat{\boldsymbol{\beta}}^{(k-1)}, \hat{\boldsymbol{\gamma}}^{(k)}) \geq Q_n(\hat{\boldsymbol{\beta}}^{(k-1)}, \hat{\boldsymbol{\gamma}}^{(k-1)}),$$

which implies the iterative algorithm is convergent.

For practical implementation, it is desirable to have an automatically data-driven method for selecting the tuning parameter λ_n involved in $p'_{\lambda_n}(\cdot)$. Here, we select λ_n via the Bayesian information criterion (BIC), which was developed by Schwarz.[15] Specifically, define the BIC as follows:

$$\text{BIC}(\lambda_n) = -2\ln l_n(\hat{\boldsymbol{\beta}}(\lambda_n), \hat{\boldsymbol{\gamma}}(\lambda_n)) + k\ln(n),$$

where k is the number of nonzero elements of $\hat{\boldsymbol{\gamma}}(\lambda_n)$. Use

$$\hat{\lambda}_n = \text{argmin}_{\lambda_n} \text{BIC}(\lambda_n)$$

as the optimal tuning parameter.

3. Simulation studies

To assess the effectiveness of the proposed procedure, we employ three evaluation criteria.[7] We present the fraction of masking outliers (denoted by $\mathcal{M}$), the fraction of mislabelling normal points as outliers (denoted by $\mathcal{S}$), and the fraction of labelling normal points as normal ones and true outliers as outliers (denoted by $\mathcal{R}$). Obviously, the three quantities $\mathcal{M}$, $\mathcal{S}$, and $\mathcal{R}$, respectively, reflect the masking probability, the swamping probability, and the right detection probability for the outlier detection procedure. In outlier detection, masking is more serious than swamping. The former can cause gross distortions while the latter is often just a matter of lost efficiency. The criterion $\mathcal{R}$ is usually used as the overall index to evaluate the effectiveness of the outlier detection method. Ideally, $\mathcal{M} \approx 0$, $\mathcal{S} \approx 0$, and $\mathcal{R} \approx 100\%$.

We randomly chose d components of $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^{\mathrm{T}}$, half of which are set to be 5 and half -5. The remaining $n - d$ components of $\boldsymbol{\gamma}$ are set to be zero to indicate the normal points. We refer to such an outlier-generated mechanism as 'Case (a)' whereas another different manners were considered subsequently. We independently generated the survival time T_i from the exponential distribution with rate,

$$r(\mathbf{X}_i) = \exp(\mathbf{X}_i^{\mathrm{T}}\boldsymbol{\beta} + \gamma_i).$$

Set $\boldsymbol{\beta} = (1, 0.5)^{\mathrm{T}}$ and $\mathbf{X}_i = (X_{i1}, X_{i2})^{\mathrm{T}}$, where X_{i1} was randomly generated from the Unif$(0, 1)$ and X_{i2} from the Bernoulli distribution with success probability of 0.5. We took the censoring time $C = \tilde{C} \wedge L$, where $\tilde{C}$ was generated from Unif$(0, L + 2)$. The study duration L was chosen to yield the desirable censoring rate. Set sample size $n = 200$ and 400. We considered the

Table 1. The masking, swamping, and right detection proportions of the proposed outlier detection procedure among 500 simulations.

		Censoring rate of 20%			Censoring rate of 40%		
n	d/n	$\mathcal{M}$	$\mathcal{S}$	$\mathcal{R}$	$\mathcal{M}$	$\mathcal{S}$	$\mathcal{R}$
200	0%	–	0.0777	0.9223	–	0.0795	0.9205
	5%	0.0010	0.0757	0.9280	0.0010	0.0786	0.9253
	10%	0.0017	0.0761	0.9313	0.0015	0.0798	0.9280
	20%	0.0014	0.0776	0.9376	0.0013	0.0778	0.9375
400	0%	–	0.0712	0.9288	–	0.0731	0.9269
	5%	0.0020	0.0712	0.9323	0.0014	0.0739	0.9298
	10%	0.0018	0.0733	0.9339	0.0015	0.0743	0.9330
	20%	0.0014	0.0714	0.9426	0.0010	0.0745	0.9402

d/n, the proportion of the outliers; $\mathcal{M}$, $\mathcal{S}$, and $\mathcal{R}$, the proportions of masking, swamping, and right detection, respectively.

censoring rates of 20% and 40%, coupled with the proportions of the outliers, i.e. d/n, ranging from 0% to 20%. For each configuration, we repeated 500 simulations.

We denote by $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ the proposed estimator. We also considered the naive estimator, $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$, by ignoring the outliers and maximizing $L_n(\boldsymbol{\beta}, \mathbf{0})$ over $\boldsymbol{\beta}$; the oracle estimator, $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, by assuming that $\boldsymbol{\gamma}$ was known and maximizing $L_n(\boldsymbol{\beta}, \boldsymbol{\gamma})$ over $\boldsymbol{\beta}$; and the second oracle estimator, $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$, by assuming that $\boldsymbol{\gamma}$ was known, removing the d outliers, and maximizing $L_{n-d}(\boldsymbol{\beta}, \mathbf{0})$ over $\boldsymbol{\beta}$.

The simulation results for the proposed outlier detection procedure are presented in Table 1. In general, we can see that the values of $\mathcal{M}$ and $\mathcal{S}$ are very close to 0 and the value of $\mathcal{R}$ is close to 100%, which demonstrates the proposed outlier detection procedure enjoys low masking and swamping probabilities as well as high right detection probabilities. As a conclusion, it demonstrates that the proposed method is effective for outlier detection in censored exponential regression.

Table 2 summarizes estimators for $\boldsymbol{\beta}$ under sample size $n = 200$. We make the following observations: (i) As expected, in terms of bias or standard error, the oracle estimator $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$ is superior to the proposed estimator $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ and the naive estimator $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$. Moreover, $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$ also outperforms $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ because the later discards some data and thus leads to efficiency loss while the former is efficient. Nevertheless, these two estimators are not implementable in practice and just work as a benchmark for comparison; (ii) The proposed estimator $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ is essentially unbiased and performs well under different censoring rates or proportions of outliers. Furthermore, it is comparable with the second oracle estimator $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$; (iii) Not surprisingly, the naive estimator $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$ is seriously biased, especially when the proportion of the outliers d/n is increased; (iv) When d/n is 0%, the naive estimator $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$, the oracle estimator $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, and the second oracle estimator $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ coincide with each other. As a result, we only reported the results of $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$. The performances of $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$ are very close. In other words, conservatively viewing the ‘clear’ data as ‘unclear’, the proposed method still works well.

We further investigated performances of estimators for $\boldsymbol{\beta}$ under sample size $n = 400$. The corresponding simulation results are summarized in Table 3, from which we can draw similar conclusions. In particular, we do not observe the mitigation of the inherent bias for the naive estimator $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$ as sample size increases to 400. On the contrary, larger sample size yields more accurate estimators $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$, and $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$.

Also considered were another three avenues for generating outliers via setting different values for $\boldsymbol{\gamma}$. The number of nonzero elements of $\boldsymbol{\gamma}$ was fixed at d.

Case (b): half of nonzeroes were generated from Unif(2, 10) and half from Unif($-10, -2$).

Table 2. Simulation results for parameter estimation with sample size $n = 200$.

			$\beta_1 = 1$			$\beta_2 = 0.5$		
d/n	Cen. rate	Method	EST	SE	MSE	EST	SE	MSE
0%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0065	0.1641	0.0270	0.5001	0.1358	0.0184
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0152	0.1684	0.0286	0.4967	0.1382	0.0191
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0027	0.1914	0.0366	0.4968	0.1580	0.0250
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0314	0.1940	0.0386	0.5019	0.1589	0.0252
5%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0085	0.1613	0.0261	0.4985	0.1375	0.0189
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0076	0.1645	0.0271	0.4997	0.1385	0.0192
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9558	0.1664	0.0297	0.4703	0.1456	0.0221
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0144	0.1702	0.0292	0.4952	0.1410	0.0199
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0018	0.1906	0.0363	0.4982	0.1607	0.0258
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	0.9997	0.1963	0.0385	0.4997	0.1634	0.0267
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9975	0.1942	0.0377	0.4904	0.1656	0.0275
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0283	0.1988	0.0403	0.5053	0.1636	0.0268
10%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0095	0.1621	0.0264	0.4974	0.1373	0.0189
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0087	0.1681	0.0283	0.4985	0.1411	0.0199
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.8864	0.1796	0.0452	0.4340	0.1537	0.0280
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0123	0.1730	0.0301	0.4940	0.1433	0.0206
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0026	0.1913	0.0366	0.4982	0.1590	0.0253
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0005	0.2023	0.0409	0.4995	0.1656	0.0274
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9888	0.1959	0.0385	0.4811	0.1642	0.0273
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0295	0.2048	0.0428	0.5050	0.1661	0.0276
20%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0127	0.1705	0.0292	0.4983	0.1391	0.0194
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0129	0.1814	0.0331	0.4989	0.1469	0.0216
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.6608	0.2132	0.1605	0.3320	0.1771	0.0596
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0033	0.1862	0.0347	0.4903	0.1490	0.0223
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0073	0.1928	0.0372	0.4956	0.1569	0.0246
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0054	0.2131	0.0454	0.4960	0.1718	0.0295
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9713	0.2071	0.0437	0.4586	0.1664	0.0294
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0315	0.2149	0.0472	0.5001	0.1723	0.0297

Note: $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, the oracle estimators with or without removing outliers; $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$, the naive estimator; and $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$, the proposed estimator.

Case (c): all nonzeroes were generated from Unif$(2, 10)$.
Case (d): all nonzeroes were generated from Unif$(-10, -2)$.

The remaining set-up was kept the same as before. Table 4 reports the resulting outlier detection while Table 5 summarizes estimates for $\boldsymbol{\beta}$. It can be seen that the proposed method is immune to the considered outlier-generated mechanisms and performs equally well. On the other hand, the existence of outliers deteriorates the behaviours of the naive method.

It is interesting to evaluate the ability of our method to cope with high-dimensional covariate $\mathbf{X}$. Denote the dimension of $\mathbf{X}$ by p. We generated $\mathbf{X}$ from the multivariate normal distribution with mean $\mathbf{0}$ and correlation matrix $(0.5^{|i-j|})_{i=1,j=1}^{p}$. Set $n = 200$, $p = 250$, and $\boldsymbol{\beta} = (1, 0.5, 0, 2, 0, \ldots, 0)^{\mathrm{T}}$ while keeping the remaining settings as before. As $p > n$, we need add a penalty function to select significant variables and estimate the sparse regression coefficient $\boldsymbol{\beta}$.

Table 3. Simulation results for parameter estimation with sample size $n = 400$.

			$\beta_1 = 1$			$\beta_2 = 0.5$		
d/n	Cen. rate	Method	EST	SE	MSE	EST	SE	MSE
0%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0059	0.1212	0.0147	0.4975	0.0999	0.0100
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0140	0.1217	0.0150	0.4973	0.0991	0.0098
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0067	0.1451	0.0211	0.4977	0.1159	0.0134
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0311	0.1466	0.0225	0.5058	0.1186	0.0141
5%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0057	0.1225	0.0150	0.4984	0.1009	0.0102
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0056	0.1243	0.0155	0.4982	0.1017	0.0103
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9506	0.1273	0.0186	0.4674	0.1067	0.0125
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0100	0.1234	0.0153	0.4974	0.1006	0.0101
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0065	0.1465	0.0215	0.4982	0.1168	0.0137
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0064	0.1488	0.0222	0.4977	0.1176	0.0138
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	1.0021	0.1470	0.0216	0.4883	0.1180	0.0141
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0313	0.1500	0.0235	0.5048	0.1206	0.0146
10%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0063	0.1233	0.0152	0.4990	0.1012	0.0102
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0060	0.1265	0.0160	0.4992	0.1035	0.0107
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.8837	0.1373	0.0324	0.4367	0.1156	0.0174
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0076	0.1271	0.0162	0.4971	0.1026	0.0105
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0061	0.1474	0.0218	0.5001	0.1175	0.0138
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0055	0.1537	0.0236	0.5002	0.1212	0.0147
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9947	0.1512	0.0229	0.4834	0.1211	0.0149
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0303	0.1557	0.0252	0.5070	0.1241	0.0155
20%	20%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0087	0.1263	0.0160	0.4954	0.1022	0.0105
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0081	0.1338	0.0180	0.4965	0.1080	0.0117
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.6592	0.1461	0.1375	0.3311	0.1296	0.0453
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0006	0.1331	0.0177	0.4887	0.1075	0.0117
	40%	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0074	0.1463	0.0215	0.4969	0.1166	0.0136
		$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0054	0.1587	0.0252	0.4985	0.1259	0.0158
		$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.9741	0.1490	0.0229	0.4615	0.1237	0.0168
		$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0298	0.1614	0.0269	0.5043	0.1295	0.0168

Note: $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, the oracle estimators with or without removing outliers; $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$, the naive estimator; and $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$, the proposed estimator.

Without loss of generality, we chose the SCAD penalty function and shared the tuning parameter λ_n. In addition, the coordinate descent algorithm was also employed to obtain the estimate for $\boldsymbol{\beta}$ and the tuning parameter was selected via the BIC criterion. The resulting outlier detection and parameter estimation are summarized in Tables 6 and 7, respectively. The column 'Size' in Table 7 is the average number of nonzero components of $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$ while 'Prop' is the proportion of the true model being nested by the selected models. Apparently, the proposed method exhibits sound results even in the scenario of high-dimensional covariates.

It demonstrates in Figure 1 that the outliers indeed exert negative effects on the estimates of the survival functions whereas, after kicking out the suspected outliers using the proposed method, the resulting Kaplan–Meier curves agree well with the true ones. Note that these estimates were calculated based on one simulated data set without covariates in order to facilitate the comparison. We fixed $n = 200$ and both the censoring rate and the proportion of outliers at 20%.

Table 4. The masking, swamping, and right detection proportions of the proposed outlier detection procedure with sample size $n = 200$, censoring rate of 20%, and outlier proportion $d/n = 10\%$ under various outlier-generated scenarios.

Case	$\mathcal{M}$	$\mathcal{S}$	$\mathcal{R}$
(b)	0.0441	0.0783	0.9252
(c)	0.0318	0.0799	0.9249
(d)	0.0360	0.0774	0.9268

Note: $\mathcal{M}$, $\mathcal{S}$, and $\mathcal{R}$, the proportions of masking, swamping, and right detection, respectively.

Table 5. Simulation results for parameter estimation with sample size $n = 200$, censoring rate of 20%, and outlier proportion $d/n = 10\%$ under various outlier-generated scenarios.

		$\beta_1 = 1$			$\beta_2 = 0.5$		
Case	Method	EST	SE	MSE	EST	SE	MSE
(b)	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0094	0.1620	0.0263	0.4973	0.1377	0.0190
	$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0087	0.1681	0.0283	0.4985	0.1411	0.0199
	$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.8879	0.1790	0.0446	0.4353	0.1535	0.0277
	$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0132	0.1742	0.0305	0.4930	0.1428	0.0204
(c)	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0044	0.1677	0.0281	0.4992	0.1374	0.0189
	$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0031	0.1793	0.0322	0.5005	0.1459	0.0213
	$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	1.1507	0.1839	0.0565	0.5457	0.1501	0.0246
	$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	1.0243	0.1842	0.0345	0.5005	0.1483	0.0220
(d)	$\hat{\boldsymbol{\beta}}_{\mathrm{O}}$	1.0090	0.1683	0.0284	0.4984	0.1395	0.0195
	$\hat{\boldsymbol{\beta}}_{\mathrm{R}}$	1.0097	0.1685	0.0285	0.4986	0.1399	0.0196
	$\hat{\boldsymbol{\beta}}_{\mathrm{N}}$	0.5305	0.1968	0.2591	0.2869	0.1733	0.0754
	$\hat{\boldsymbol{\beta}}_{\mathrm{P}}$	0.9793	0.1710	0.0297	0.4800	0.1434	0.0210

Note: $\hat{\boldsymbol{\beta}}_{\mathrm{R}}$ and $\hat{\boldsymbol{\beta}}_{\mathrm{O}}$, the oracle estimators with or without removing outliers; $\hat{\boldsymbol{\beta}}_{\mathrm{N}}$, the naive estimator; and $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$, the proposed estimator.

Table 6. The masking, swamping, and right detection proportions of the proposed outlier detection procedure with sample size $n = 200$, the dimension of covariates $p = 250$, censoring rate of 20%, and outlier proportion $d/n = 10\%$ under various outlier-generated scenarios.

Case	$\mathcal{M}$	$\mathcal{S}$	$\mathcal{R}$
(a)	0.0087	0.0578	0.9471
(b)	0.0603	0.0579	0.9419
(c)	0.0412	0.0609	0.9411
(d)	0.0539	0.0570	0.9433

Note: $\mathcal{M}$, $\mathcal{S}$, and $\mathcal{R}$, the proportions of masking, swamping, and right detection, respectively.

4. Real example

As an illustration, we applied the proposed method to the German Breast Cancer (GBC) study. The data can be downloaded from http://www.umass.edu/statdata/statdata/data. In this study, a total of 686 patients with primary node positive breast cancer were recruited between July 1984 and December 1989. There were 440 patients treated with hormone therapy and 246 patients with chemotherapy. The primary endpoint of the study was the recurrence-free survival time (in days) whereas the corresponding censoring rate is 56.4%. Figure 2 shows the Kaplan–Meier curves for

Table 7. Simulation results for parameter estimation and variable selection of the proposed method with sample size $n = 200$, the dimension of variables $p = 250$, censoring rate of 20%, and outlier proportion $d/n = 10\%$ under various outlier-generated scenarios.

	$\beta_1 = 1$		$\beta_2 = 0.5$		$\beta_4 = 2$			
Case	EST	SE	EST	SE	EST	SE	Size	Prop
(a)	0.9827	0.1653	0.4124	0.2261	1.9455	0.1597	8.19	86.2%
(b)	0.9926	0.1631	0.4197	0.2284	1.9583	0.1546	7.98	85.6%
(c)	1.0175	0.1155	0.4883	0.1565	2.0384	0.1085	3.78	95.8%
(d)	0.9426	0.2035	0.3780	0.2575	1.8526	0.1748	13.70	79.4%

Note: Size, the average number of nonzero components of $\hat{\boldsymbol{\beta}}_{\mathrm{P}}$; Prop, the proportion of the true model being nested by the selected models.

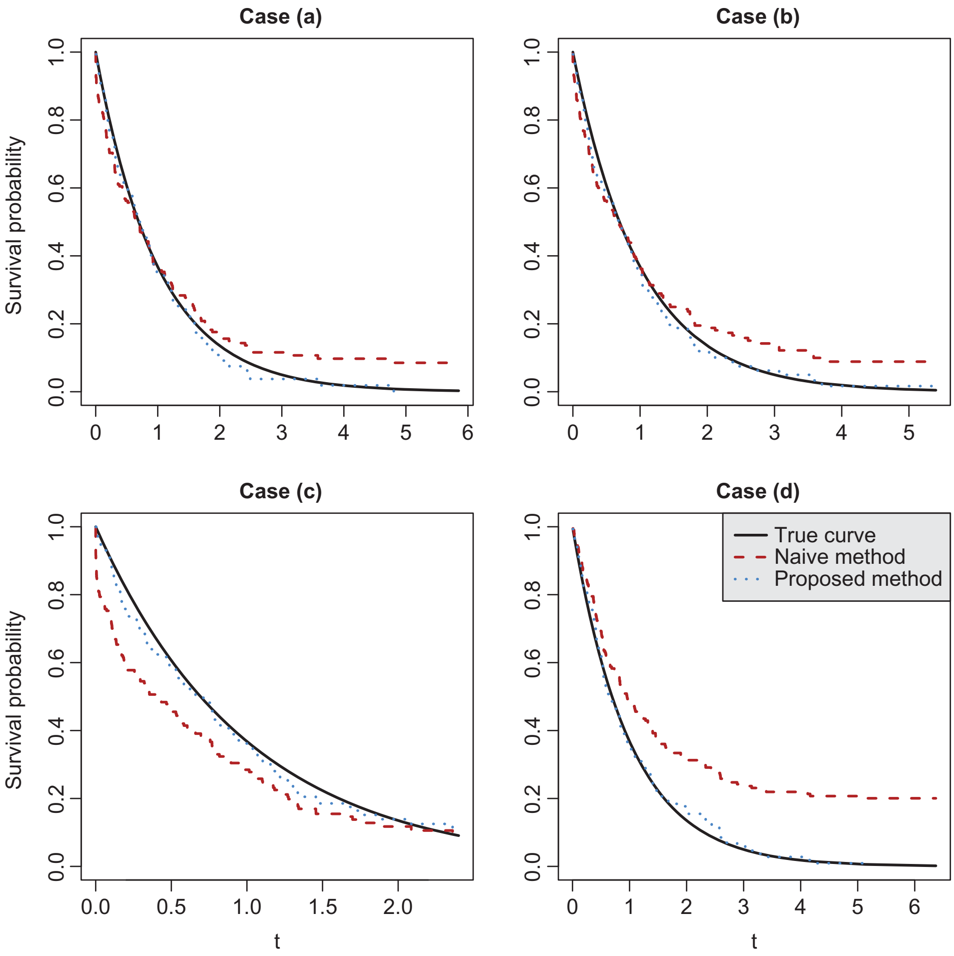

Figure 1. The Kaplan–Meier estimates of the naive method and the proposed method, respectively, under four different outlier generation manners.

the two different therapy arms, from which we can see that patients in the chemotherapy arm experienced longer recurrence-free survival time.

The covariates of interest included in this analysis were treatment (abbreviated as TRT, being 1 if treated with hormone therapy and 2 otherwise), age at diagnosis (AGE), menopausal status (MS, being 1 if menopaused and 2 otherwise), the tumour size (TS), the tumour grade

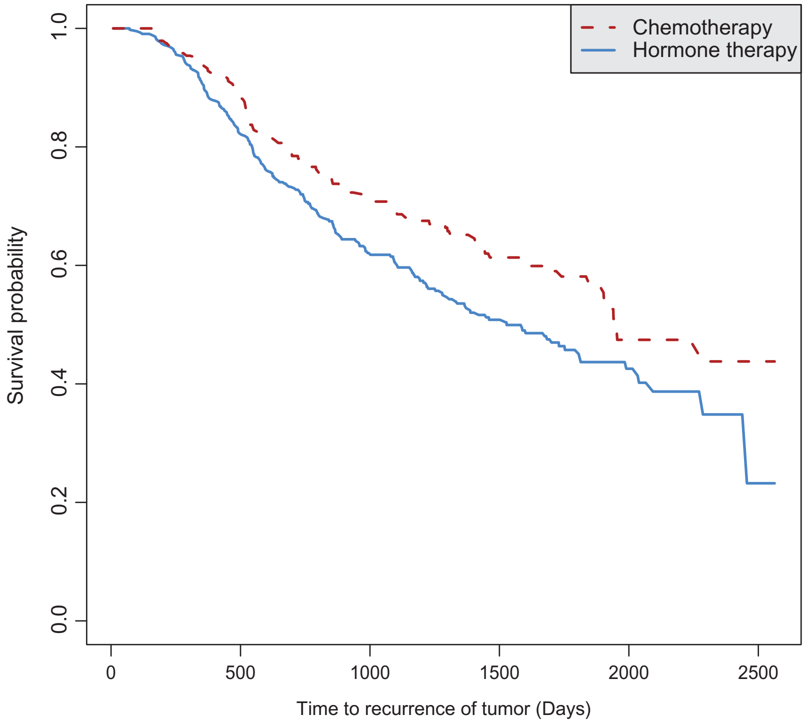

Figure 2. The Kaplan–Meier survival curves stratified by two different therapies in the GBC study.

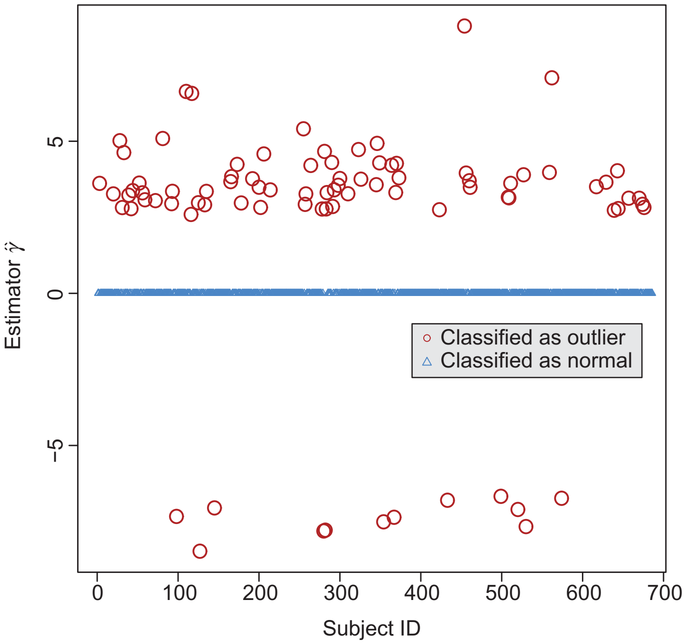

Figure 3. The scatter diagram of the estimator $\hat{\gamma}$ versus subject ID for the GBC study.

(TG), the number of nodes (NN), the number of progesterone receptors (NPR), and the number of oestrogen receptors (NER). For more detailed introduction of the GBC study, see Schmoor et al.[16]

Table 8. The analysis results of the covariates effects for the GBC study.

		Method				
Covariate		$\hat{\boldsymbol{\beta}}_{\text{Cox}}$	$\hat{\boldsymbol{\beta}}_{\text{N}}$	$\hat{\boldsymbol{\beta}}_{\text{P}}$	$\hat{\boldsymbol{\beta}}_{\text{C-P}}$	$\hat{\boldsymbol{\beta}}_{\text{C-Cox}}$
TRT	EST	– 0.3372	– 1.2027	– 1.3522	– 1.3752	– 0.8453
	SE	0.1290	0.2022	0.2918	0.1982	0.1629
	p-value	.0089	< .0001	< .0001	< .0001	< .0001
AGE	EST	– 0.0094	– 0.1042	– 0.1147	– 0.1141	– 0.0596
	SE	0.0093	0.0086	0.0112	0.0083	0.0116
	p-value	.3121	< .0001	< .0001	< .0001	< .0001
MS	EST	0.2673	0.7649	1.1061	1.1187	0.7663
	SE	0.1833	0.2260	0.2999	0.2137	0.2139
	p-value	.1448	.0007	.0002	< .0001	.0003
TS	EST	0.0077	– 0.0071	– 0.0022	– 0.0018	0.0066
	SE	0.0039	0.0060	0.0084	0.0058	0.0047
	p-value	.0507	.2368	.7934	.7563	.1629
TG	EST	0.2803	– 0.9582	– 1.0038	– 1.0178	– 0.2271
	SE	0.1061	0.1048	0.1361	0.1059	0.1274
	p-value	.0082	< .0001	< .0001	< .0001	.0747
NN	EST	0.0499	0.0517	0.0610	0.0611	0.0592
	SE	0.0074	0.0133	0.0237	0.0159	0.0079
	p-value	< .0001	.0001	.0100	.0001	< .0001
NPR	EST	– 0.0022	– 0.0058	– 0.0078	– 0.0079	– 0.0054
	SE	0.0006	0.0012	0.0025	0.0014	0.0009
	p-value	.0001	< .0001	.0018	< .0001	< .0001
NER	EST	0.0002	0.0016	0.0019	0.0019	0.0012
	SE	0.0004	0.0006	0.0009	0.0006	0.0005
	p-value	.7084	.0077	.0348	.0015	.0123

Note: $\hat{\boldsymbol{\beta}}_{\text{Cox}}$, the Cox regression estimator; $\hat{\boldsymbol{\beta}}_{\text{N}}$, the naive estimator; $\hat{\boldsymbol{\beta}}_{\text{P}}$, the proposed estimator; and $\hat{\boldsymbol{\beta}}_{\text{C-P}}$ and $\hat{\boldsymbol{\beta}}_{\text{C-Cox}}$, the exponential and Cox regression estimators after removing the suspected outliers.

Based on the proposed method, we obtained the estimator $\hat{\boldsymbol{\gamma}}$ and evaluated which observation is an outlier. The scatter points of $\hat{\boldsymbol{\gamma}}$ versus subject ID shown in Figure 3 clearly demonstrate that there are 85 possible outliers among the total 686 observations, which leads to the rate of outliers being of 12.39%.

We summarized the analysis results of the covariates effects in Table 8, where we also presented the results by fitting the data with the Cox proportional hazards regression model [17] and the resultant estimator was denoted by $\hat{\boldsymbol{\beta}}_{\text{Cox}}$. Also, we obtained a 'clear' data set by deleting the suspected outliers based on the nonzero components of the proposed estimator $\hat{\boldsymbol{\gamma}}$. As a comparison, we re-analysed the clear data using the exponential and Cox regression methods, and denoted the resulting estimators by $\hat{\boldsymbol{\beta}}_{\text{C-P}}$ and $\hat{\boldsymbol{\beta}}_{\text{C-Cox}}$, respectively. We adopt the bootstrap method to obtain the standard error estimates, although it lacks rigorous justification. Including or excluding the suspected outliers, the Cox regression method would get inconsistent conclusions for evaluating effects of covariates, such as AGE, MS, and TG. The proposed method is very effective to detect the possible outliers and estimate the regression parameter, resulting in similar conclusions from methods $\hat{\boldsymbol{\beta}}_{\text{C-P}}$ and $\hat{\boldsymbol{\beta}}_{\text{C-Cox}}$. The magnitude of the covariate effects differs between the proposed and naive methods. Furthermore, from all the considered methods, we can conclude that patients in the chemotherapy arm would possess longer survival time to resist the recurrence of the tumour, compared with ones in the hormone therapy arm. This coincides with Figure 2.

5. Conclusion

The main contribution of this paper is to propose a penalized likelihood method to detect the possible outliers in the exponential regression model as well as to estimate the unknown regression parameter. The method can be also extended to high-dimensional covariates. Numerical results demonstrate that the proposed method exhibits reasonable performance in practice.

It is surprised that the outlier detection procedure as a traditional topic in regression analysis is under-developed for censored data analysis. We conjecture that it may be the mixing of censoring and outlier that makes outlier detection and parameter estimation difficult. As an overture, our current work assumes that the failure time is from the exponential regression model. It warrants to further consider the outlier detection and parameter estimation in the classic Cox proportional hazards model.

Acknowledgments

We thank the editor and the reviewer for their many constructive and insightful comments that have led to significant improvements in the article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported in part by grants from the National Science Foundation of China [grant numbers 11171263, 41261087, 11201350, 61371126]; the Doctoral Fund of Ministry of Education of China [grant number 20110141120004], and the Fundamental Research Funds for the Central Universities [grant number 2042014kf0256].

References

[1] Grubbs FE. Procedures for detecting outlying observations in samples. Technometrics. 1969;11:1–21.
[2] Barnett V, Lewis T. Outliers in statistical data. New York: Wiley; 1984.
[3] Hampel FR, Ronchetti EM, Rousseeuw PJ, Stahel WA. Robust statistics: the approach based on influence functions. New York: Wiley; 1986.
[4] Weisberg S. Applied linear regression. New York: Wiley; 1985.
[5] Iglewicz B, Hoaglin D. How to detect and handle outliers. Milwaukee: American Society for Quality Control; 1993.
[6] Cook RD. Detection of influential observation in linear regression. Technometrics. 1977;42:65–68.
[7] She Y, Owen AB. Outlier detection using nonconvex penalized regression. J Am Statist Assoc. 2011;106:626–639.
[8] Hawkins DM. Identification of outliers. London: Chapman and Hall; 1980.
[9] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 1994;81:515–526.
[10] Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Statist Assoc. 2001;96:1348–1360.
[11] Zou H, Li R. One-step sparse estimates in nonconcave penalized likelihood models. Ann Stat. 2008;36:1509–1533.
[12] Friedman J, Hastie T, Höfling H, Tibshirani R. Pathwise coordinate optimization. Ann Appl Stat. 2007;1:302–332.
[13] Wu TT, Lange K. Coordinate descent algorithms for lasso penalized regression. Ann Appl Stat. 2008;2:224–244.
[14] Breheny P, Huang J. Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Ann Appl Stat. 2011;5:232–253.
[15] Schwarz G. Estimating the dimension of a model. Ann Stat. 1978;6:461–464.
[16] Schmoor C, Olschewski M, Schumacher M. Randomized and non-randomized patients in clinical trials: experiences with comprehensive cohort studies. Stat Med. 1996;15:263–271.
[17] Cox DR. Regression models and life tables. J R Statist Soc, Ser B. 1972;34:187–220.