Computational Statistics and Data Analysis 54 (2010) 1975–1982

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

On sufficient dimension reduction for proportional censorship model with covariates

Xuerong Meggie Wen

Department of Mathematics and Statistics, Missouri University of Science and Technology, MO 65409, USA

ARTICLE INFO

Article history:
Received 25 September 2009
Received in revised form 23 February 2010
Accepted 23 February 2010
Available online 6 March 2010

Keywords:
Conditional Koziol–Green model
Multiple covariates
Proportional hazards model
Sufficient dimension reduction

ABSTRACT

The requirement of constant censoring parameter β in Koziol–Green (KG) model is too restrictive. When covariates are present, the conditional KG model (Veraverbekea and Cadarso-Suárez, 2000) which allows β to be dependent on the covariates is more realistic. In this paper, using sufficient dimension reduction methods, we provide a model-free diagnostic tool to test if β is a function of the covariates. Our method also allows us to conduct a model-free selection of the related covariates. A simulation study and a real data analysis are also included to illustrate our approach.

1. Introduction

1.1. Sufficient dimension reduction

Thanks to modern technology, we are now overwhelmed by a huge amount of data. In most sciences, advances in data collection and storage capabilities have led to an information overload. Researchers are now facing larger and larger observations and simulations on a daily basis. Such large data sets pose new challenges in data analysis. Traditional methods developed for smaller data sets break down. Dimensionality is a major concern in analyzing large data sets. Many methods, such as principal component analysis, factor analysis, projection pursuit etc., were developed to reduce the dimension of the original data prior to any modeling. All these methods try to gasp the "important" features or patterns in the data. For regression problems, the notion of effective (sufficient) dimension reduction was first introduced by Li (1991).

For a typical regression problem with a univariate random response Y and a vector of random predictors $\mathbf{X} = (X_1, \ldots, X_p)^T \in \mathbb{R}^p$, the goal is to understand how the conditional distribution $Y|\mathbf{X}$ depends on the value of $\mathbf{X}$. The spirit of sufficient dimension reduction is to reduce the dimension of $\mathbf{X}$ without loss of information on the regression and without requiring a pre-specified parametric model. Assuming the following semiparametric regression model: $Y = g(\boldsymbol{\beta}_1^T\mathbf{X}, \boldsymbol{\beta}_2^T\mathbf{X}, \ldots, \boldsymbol{\beta}_K^T\mathbf{X}, \epsilon)$, where $g(.)$ is an unknown function and ϵ is an unknown random error independent of $\mathbf{X}$, we can see that the conditional distribution of $Y|(\boldsymbol{\beta}_1^T\mathbf{X}, \ldots, \boldsymbol{\beta}_K^T\mathbf{X})$ is the same as that of $Y|\mathbf{X}$ for all values of $\mathbf{X}$. Hence, these $\boldsymbol{\beta}$'s which provides a parsimonious characterization of the conditional distribution of $Y|\mathbf{X}$. We call these $\boldsymbol{\beta}$'s *effective directions* (Li, 1991). When K is small which is often the case in real applications, the original regression problem (data) can be effectively reduced by projecting X along these effective directions.

More formally, we search for subspaces $\mathcal{S} \subseteq \mathbb{R}^p$ such that

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X} | P_{\mathcal{S}}\mathbf{X}$$

E-mail address: wenx@mst.edu.

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.02.022

where $\perp\!\!\!\perp$ indicates independence, and $P_{(\cdot)}$ stands for a projection operator with respect to the standard inner product. The intersection of all such $\mathcal{S}$ is defined as the *central subspace* $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ (Cook, 1998), which almost always exists in practice under mild conditions. We assume the existence of the central subspaces through this article. Sufficient dimension reduction is concerned with making inference for the central subspace. Unlike other nonparametric approaches, sufficient dimension reduction can often avoid the curse of dimensionality. Many sufficient dimension reduction methods enjoy $\sqrt{n}$ convergence rates since they exploit the global features of the dependence of $\mathbf{Y}$ on $\mathbf{X}$.

One of the most popular sufficient dimension reduction methods is called sliced inverse regression (SIR) proposed by Li (1991). SIR is based on $E(\mathbf{X}|Y)$, the first conditional moment of the inverse regression where $\mathbf{X}$ is regressed on Y. Although SIR is easy to implement and has received considerable interests in applications, Cook and Ni (2005) showed that it is not asymptotically optimal. They proposed inverse regression estimation (IRE) which is asymptotically efficient among methods based on first conditional moments. However, both SIR and IRE fail to recover the whole central subspace when the response surface is symmetric about the origin. Methods based on the second conditional moment ($E(\mathbf{X}\mathbf{X}^T|Y)$) or the first two conditional moments, such as sliced average variance estimation (SAVE; (Cook and Weisberg, 1991), SIRII (Li, 1991) and combinations of SIR and SIRII (Gannoun and Saracco, 2003; Saracco, 2005), DR (Li and Wang, 2007) were developed in response to this limitation. In addition to those methods exploiting the global features of the dependency of Y on $\mathbf{X}$, Xia et al. (2002) proposed MAVE which makes use of the local features of the dependency of Y on $\mathbf{X}$. Wang and Xia (2008) recently proposed a method called sliced regression (SR) which combined SIR and MAVE.

Sufficient dimension reduction methods have been developed under various regression settings. To list a few, Li et al. (2003) studied estimation methods with a multivariate $\mathbf{Y}$; Zhong et al. (2005) used a method called regularized sliced inverse regression (RSIR) for motif discovery in microarray analysis; Wen and Cook (2007) and Liquet and Saracco (2007) investigated dimension reductions with a mix of continuous and categorical predictors. Li et al. (1999), Wen and Cook (2009), discussed dimension reduction methods that allow for censoring. In this paper, we will study dimension reduction methods in survival analysis data under the context of proportional censorship model.

1.2. Proportional censorship model with covariates

Right-censored data are common in many medical and epidemiological studies. Let $(T_1, C_1), \ldots, (T_n, C_n)$ be independent pairs of positive continuous random variables where T_j, $j = 1, \ldots, n$, represents the survival time (time to event) and C_j represents the right censoring time of the jth object under study. Let $Y_j = \min(T_j, C_j)$ and $\delta_j = I\{T_j \leq C_j\}$. the problem of interest is to estimate the distribution of the survival time T. However, due to the existence of right censoring, only (Y_j, δ_j) are fully observable. Many methods have been proposed in literature to tackle the problem under the assumption of non-informative censoring (Lagakos, 1979), i.e., a censored subject has the same risk of having an event as those who have complete follow-up. When the censoring is informative which is often the case in real applications, the so-called proportional hazards model of random censorship or Koziol–Green model (KG model: Koziol and Green, 1976; Faraway and Csörgö, 1998) has been used, in which we assume that there exists a constant $\beta > 0$ such that $\bar{G}(t) = \bar{F}(t)^{\beta}$ where $F(t)$ and $G(t)$ are the distribution functions of T and C respectively, $\bar{F}(t) = 1 - F(t)$, $\bar{G}(t) = 1 - G(t)$. Let $\Lambda(t)$ denote the cumulative hazard function, then this condition is equivalent to $\Lambda_G(\cdot) = \beta\Lambda_F(\cdot)$. The parameter β is interpreted as the censoring parameter since $P(C_i < T_i) = \beta/(1 + \beta)$.

Csörgö (1988) gave a detailed review on KG model and related estimation and testing procedures. Faraway and Csörgö (1998) argued that the assumption of constant β makes KG model impractical in reality (censoring is too informative, too good to be true in practice). In recognizing the limitations of KG model, there are several proposals on generalizing the KG model. Peña and Rohatgi (1989) proposed a generalized proportional hazard model with random censorship which allowed β to be a nonnegative random variable. Gather and Pawlitschko (1998) considered the partial KG model where some censorings are informative and the others are not. However, neither of those models considered the effects of covariates. As pointed out by Zeng (2004), in most studies, the covariates $\mathbf{X}$ contain not only subject demographic information and disease history, but also auxiliary information which is informative in predicting subjects failure time or explaining the censoring mechanism. Gaddah and Braekers (2009) and Veraverbekea and Cadarso-Suárez (2000) investigated estimation methods which incorporate covariate information. The conditional proportional model proposed by Veraverbekea and Cadarso-Suárez (2000) assumes that β is a function of covariate $\mathbf{X}$. Hence, whether a subject is being censored, also depends on the covariate $\mathbf{X}$. However, they only considered a single covariate, i.e., $X \in \mathbb{R}^1$.

In this paper, we will investigate the estimation method for conditional proportional model with many covariates, say, $\mathbf{X} \in \mathbb{R}^p$, $p > 1$. Specifically, we assume that

$$\bar{G}_{T|\mathbf{X}}(t|\mathbf{x}) = \mathrm{E}_B[\bar{F}_{T|\mathbf{X}}(t|\mathbf{x})]^{\beta(\mathbf{x})}, \tag{1.1}$$

where $\beta(\mathbf{X}) = h(\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_q^T\mathbf{X}, \epsilon)$, $q \leq p$, $B(b) = P\{\beta \leq b\}$, $\boldsymbol{\eta}_j \in \mathbb{R}^p$, $j = 1, \ldots, q$, and ϵ is a random error independent with $\mathbf{X}$. So $\beta(\mathbf{X})$ is a function of q linear combinations of the original covariates $\mathbf{X}$. Without assuming any parametric model, we first use sufficient dimension reduction methods to estimate q and $\boldsymbol{\eta}$, where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_q)$ is a $p \times q$ matrix consisting $\boldsymbol{\eta}_j$ as its columns. An asymptotic chi-squared test is developed to test whether the proportion of non-censoring should be modeled through the covariate. All our estimates are asymptotically consistent.

The rest of this paper is organized as follows. Section 2 is dedicated to the discussion of estimation methods for q and $\boldsymbol{\eta}$. If $q > 0$, we then fit a logit model similar to Yuan (2005) to obtain $\beta(\mathbf{X})$. In Section 3, we present simulation results. An

illustration of our method is also given via a real data analysis. We conclude our paper and discuss future research directions in Section 4. Proofs will be delayed to Appendix.

2. Estimation procedure via sufficient dimension reduction

Following the notation in Section 1, we first study estimation methods for q and $\boldsymbol{\eta}$ assuming model (1.1). If $q = 0$, then $\beta(\mathbf{X}) = h(\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_q^T\mathbf{X}, \epsilon)$ is then reduced to a constant and model (1.1) is reduced to the regular KG model. Let $\mathcal{S}_{T|\mathbf{X}}$, $\mathcal{S}_{C|\mathbf{X}}$, $\mathcal{S}_{\delta|\mathbf{X}}$, $\mathcal{S}_{Y|\mathbf{X}}$ and $\mathcal{S}_{(Y,\delta)|\mathbf{X}}$ denote the corresponding central spaces. The following theorem provides a connection among these spaces under model (1.1).

Theorem 1. *Assuming model (1.1) with $\beta(\mathbf{X}) = h(\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_q^T\mathbf{X}, \epsilon)$, where ϵ is a random error independent with $\mathbf{X}$, then*

1. $\mathcal{S}_{\delta|\mathbf{X}} = \mathrm{Span}\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_q\}$.
2. $\mathcal{S}_{(Y,\delta)|\mathbf{X}} = \mathcal{S}_{Y|\mathbf{X}} + \mathcal{S}_{\delta|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}} + \mathcal{S}_{C|\mathbf{X}}$.
3. $\mathcal{S}_{Y|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}} = \mathcal{S}_{C|\mathbf{X}}$ *if* $q = 0$, *where* $q = \dim(\mathcal{S}_{\delta|\mathbf{X}})$.

Since δ is a binary variable, most existing sufficient dimension reduction methods, such as SIR (Li, 1991), SAVE (Cook and Weisberg, 1991) or IRE (Cook and Ni, 2005), could be used directly to infer about q and $\mathcal{S}_{\delta|\mathbf{X}}$. To illustrate our procedure, a brief review of SIR is provided as following.

Let $\Sigma = \mathrm{Cov}(\mathbf{X})$ denote the marginal covariance matrix of $\mathbf{X}$. Sufficient dimension reduction can often be formulated as a generalized eigenvalue problem of the form $M\mathbf{v}_i = \lambda_i G\mathbf{v}_i$, where M is a method-specific symmetric kernel matrix and is nonnegative definite, G is a symmetric and positive definite matrix, $\mathbf{v}_1, \ldots, \mathbf{v}_p$ are eigenvectors satisfying $\mathbf{v}_i^T G\mathbf{v}_j = 1$ if $i = j$, and 0 otherwise, and $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ are the corresponding eigenvalues. Different choices of M and G would yield different dimension reduction methods. For example, $M = \Sigma$ and $G = I_p$ correspond to principal component analysis. Li's (1991) SIR method took $M = \mathrm{Cov}[\mathrm{E}(\mathbf{X}|\delta) - \mathrm{E}(\mathbf{X})]$ and $G = \Sigma$. Let $\hat{M}$ denote a consistent estimator of M. Then the sum of the smallest eigenvalues of $\hat{M}$ is utilized to construct test statistics for estimating q, the dimension of $\mathcal{S}_{\delta|\mathbf{X}}$. And the eigenvectors corresponding to the q largest eigenvalues are used to estimate $\boldsymbol{\eta}$. Specifically, a sequence of hypotheses $H_0 : q = m$ versus $H_a : q > m$, with m incremented by 1 until the hypothesis is not rejected. At which point $\hat{q}$ is the last value of m tested. Asymptotically, the test statistic of SIR is a linear combination of chi-square random variables.

In addition to the above spectral decomposition approach, there exists another approach called the minimum discrepancy approach (Cook and Ni, 2005) in the literature of sufficient dimension reduction. IRE (Cook and Ni, 2005) takes the minimum discrepancy approach and outperforms SIR asymptotically. It also provides an asymptotic chi-squared test for testing the dimension of $\mathcal{S}_{\delta|\mathbf{X}}$. Though with small sample size, SIR might perform as well as or even better than IRE. Interested readers could refer to Cook and Ni (2005) for a detailed discussion. We will apply both procedures in our simulation studies and data analysis. Both SIR and IRE on the regression of δ with respect to $\mathbf{X}$ could provide us with consistent estimators for $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_q)$, where $\mathcal{S}_{\delta|\mathbf{X}} = \mathrm{Span}\{\boldsymbol{\eta}\}$; and allowed us to do predictor selection without assuming any parametric model (Wen and Cook, 2007). In Theorem 2, we cited the results from Cook and Ni (2005) for reference. Furthermore, with a binary response δ, IRE and SIR actually provide the same results. Interested readers could refer to Appendix for detailed discussion on this issue. The proof of Theorem 2 is similar to that of Ni and Cook.

Theorem 2. *Assuming model (1.1) with $\beta(\mathbf{X}) = h(\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_q^T\mathbf{X}, \epsilon)$, where ϵ is a random error independent with $\mathbf{X}$. Assume that the data $(Y_i, \mathbf{X}_i, \delta_i)$, $i = 1, \ldots, n$, are a simple random sample of $(Y, \mathbf{X}, \delta)$ with finite fourth moments. Let $\hat{\boldsymbol{\eta}}$ be the estimator from SIR (IRE), then*

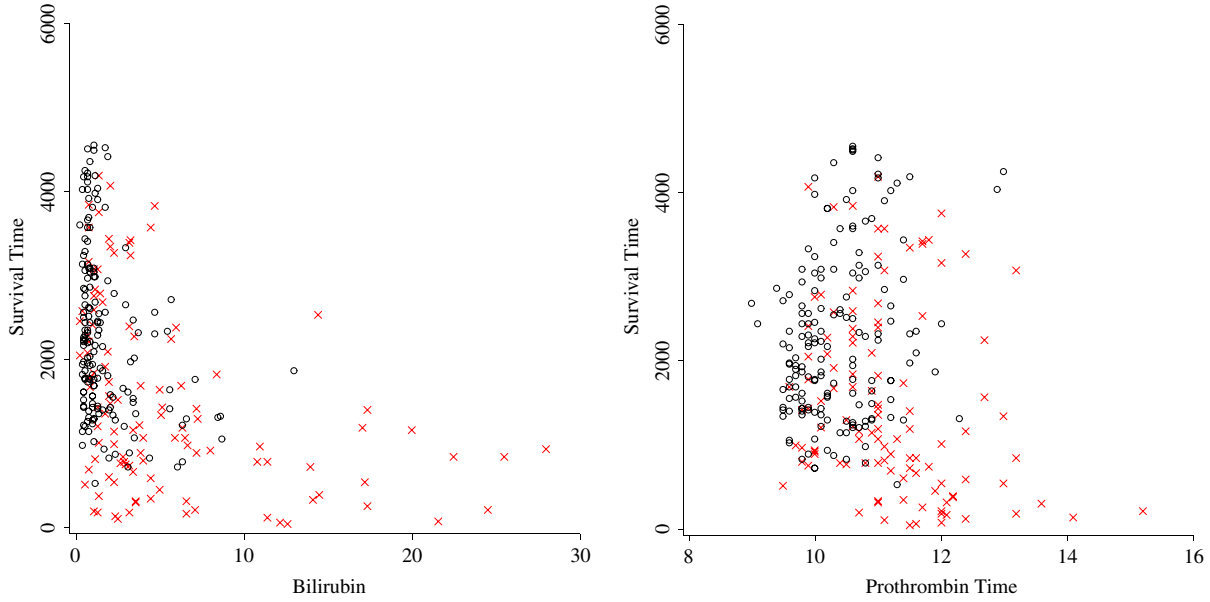
1. *The estimate* $\mathrm{vec}(\hat{\boldsymbol{\eta}})$ *is asymptotically efficient, and follows an asymptotic normal distribution, where* vec(.) *denotes the operator that constructs a vector from a matrix by stacking its columns.*
2. *The test statistic on testing* $\dim(\mathcal{S}_{\delta|\mathbf{X}}) = q = 0$ *has an asymptotic chi-squared distribution with degrees of freedom of p, where p is the number of covariates.*
3. $\mathrm{Span}(\hat{\boldsymbol{\eta}})$ *is a consistent estimator of* $\mathrm{Span}(\boldsymbol{\eta})$.

When $q(\hat{q}) = 0$, model (1.1) is reduced to the regular KG model, and β is a constant. Inference methods on KG model could be applied to the data directly. Or we can directly use sufficient dimension reduction methods to the survival analysis data to infer about the central subspace of $\mathcal{S}_{Y|\mathbf{X}}$, since $\mathcal{S}_{T|\mathbf{X}} = \mathcal{S}_{Y|\mathbf{X}}$ according to Theorem 1. So sufficient dimension reduction methods can serve as a diagnostic tool as a goodness fit test of regular KG model. It might outperform other approaches since it does not pre-assume any modeling assumptions on the survival function of T ($\bar{F}$). This can be a good starting point for further parametric and nonparametric modeling procedures.

When $q(\hat{q}) \neq 0$, model (1.1) is the so-called conditional proportional model (Veraverbekea and Cadarso-Suárez, 2000). Since an estimator of $\boldsymbol{\eta}$ is given by the sufficient dimension reduction method, and q is often 1 or 2, many traditional parametric modeling methods are applicable with $\boldsymbol{\eta}^T\mathbf{X}$ as the new covariates. Following Yuan (2005), we consider the following logistic regression:

$$\mathrm{logit}[P(\delta = 1|\mathbf{X})] = -\log(\beta(\mathbf{x})) = \gamma_0 + \boldsymbol{\gamma}^T(\boldsymbol{\eta}^T\mathbf{X}), \tag{2.2}$$

where $\gamma_0 \in \mathbb{R}^1$, and $\boldsymbol{\gamma} \in \mathbb{R}^q$ are the unknown coefficients. We may replace the above linear function of $\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_q^T\mathbf{X}$ with a polynomial function in case of necessity.



(a) Bilirubin level. (b) Prothrombin time.

Fig. 1. PBC data: Scatter plot of survival time (in days) versus bilirubin levels and prothrombin time. ×: Event; ○: Censored.

Assuming model (1.1), since $\Lambda_{T|x}(t|x) = \frac{1}{1+\beta(\mathbf{X})}\Lambda_{Y|x}(t|x)$, where $\Lambda_{T|x}$ and $\Lambda_{Y|x}$ are the cumulative hazard functions of T and Y respectively. We can always estimate $\Lambda_{T|x}(t|x)$ from $\Lambda_{Y|x}(y|x)$ by plugging in the value of $\beta(\mathbf{X})$ obtained from (2.2). Or we can take an approach similar to Abdushukurov–Cheng–Lin estimator (Cheng and Lin, 1987), since $1 - F_{T|x}(t|x) = (1 - H_{Y|x}(t|x))^{\frac{1}{1+\beta(\mathbf{X})}}$ where H is the cumulative distribution function of Y. Veraverbekea and Cadarso-Suárez (2000) have a detailed discussion on this type of estimator. $H(.)$ and $\beta(\mathbf{X})$ can be estimated independently since Y and δ are independent given the covariates $\mathbf{X}$ under model (1.1). Under this circumstance, the advantage of using sufficient dimension reduction methods for pre-processing is that now the covariates is reduced to fewer linear combinations of the original covariates.

3. Data illustration and simulation studies

In this section, we will first illustrate our method using a real data, the Primary Biliary Cirrhosis data set (PBC data; Fleming and Harrington, 1991). We then conduct several simulation studies.

3.1. Example: primary biliary cirrhosis data

The PBC data set came from the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between 1974 and 1984. The major goal of this double-blinded randomized placebo controlled trial is to assess the efficacy of a new drug, the D-penicillamine. This data set contains survival time and other information on 312 PBC patients participating in the trial. Fleming and Harrington (1991) provided a detailed description of this data set. They selected five predictors from the original seventeen covariates. Since previous studies have shown that there was no therapeutic differences between control and D-penicillamine-treated patients. The goal of their study is to explain the relationship between a patient's survival time and those covariates.

As suggested by Veraverbekea and Cadarso-Suárez (2000), the proportion of non-censoring is related strongly to the covariates. A conditional proportional model such as (1.1) was recommended. Though only one covariate, Serum bilirubin, was investigated in their study. Let Y denote the observed time, which is the number of days between registration and the earlier of the death or censoring; δ denote the censoring indicator. We would consider the following covariates:

Age = Age in years.
Edema = Presence of edema. 0 = no edema and no diuretic edema; 0.5 = edema present for which no diuretic therapy was given, or edema resolved with diuretic therapy; 1 = edema despite diuretic therapy.
Bilirubin = Serum bilirubin, in mg/dl.
Albumin = Albumin in gm/dl.
Triglycerides = Triglycerides, in mg/dl.
Platelet = Platelet count.
Prothrombin time = Prothrombin time, in s.

Table 1
PBC data: p-values from marginal predictor tests using IRE.

Predictor	Step 1	Step 2	Step 3	Step 4	Step 5
Age	0.002	0.002	0.002	0.002	0.001
Albumin	0.150	0.153	0.104	0.116	Deleted
Platelet	0.952	Deleted			
Edema	0.613	0.617	Deleted		
Prothrombin time	0.000	0.000	0.000	0.000	0.000
Bilirubin	0.000	0.000	0.000	0.000	0.000
Triglycerides	0.117	0.171	0.185	Deleted	

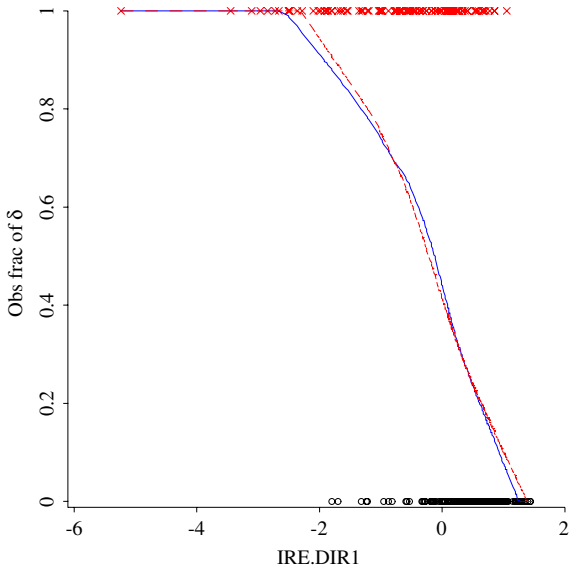


Fig. 2. PBC data: Model checking plot for Model (2.2). Solid blue line for response, dashed red line for fit. ×: Event; ○: Censored.

Cases with missing values were ignored. The remaining 278 cases were analyzed using SIR and IRE. The same outliers as Li et al. (1999) reported were found and removed in further analysis.

Fig. 1 showed the scatter plots of Y versus two covariates, Serum bilirubin and Prothrombin time respectively. As pointed by Veraverbekea and Cadarso-Suárez (2000), the proportion of non-censoring relates strongly to the bilirubin since censoring is much more prevalent for patients with low levels of bilirubin. From panel (b), we can see similar dependence of δ ($\beta(\mathbf{X})$) on the values of prothrombin time.

Since SIR and IRE provide same results with binary response, we will only show the analysis result using IRE. Started with a regression of δ versus the above eight covariates, we first did a model-free backward variable selection. As shown in Table 1, four predictors were screened out using 5% tests, leaving 3 predictors for further analysis. This screening requires only the standard independence condition $T \perp\!\!\!\perp C|\mathbf{X}$.

We then fit the logistic regression model (2.2) taking $\boldsymbol{\eta}^T\mathbf{X}$ as the first direction from IRE. Fig. 2 shows the model checking plot for PBC data. The solid blue line is the smooth of the observed fraction of non-censoring versus the first direction from IRE, while the dashed red line is obtained by smoothing the fitted probabilities of non-censoring versus the first direction from IRE. The two smooths are reasonably close which suggests that our model agrees with the data.

So we have the following model for the PBC data:

$$\text{logit}[P(\delta = 1|\mathbf{X})] = -\log(\beta(\mathbf{x})) = -0.349 + 1.774 \times \text{IRE}_1$$

where $\text{IRE}_1 = 0.054 \times \text{Age} + 0.977 \times \text{Prothrombin time} + 0.205 \times \text{Bilirubin}$.

3.2. Simulation studies

In this section, we conducted the following simulation via SIR. Both T and C were generated from Cox proportional hazards model. $\lambda_T(t|\mathbf{X}) = \lambda e^{X_1+X_2}$, and $\beta(\mathbf{X}) = e^{X_3+X_4} - 1$. Hence $q = \dim(\mathcal{S}_{\delta|\mathbf{X}}) = 1$, and $\boldsymbol{\eta}^T = (0, 0, 1, 1, 0, 0)$. We sampled X_3 and X_4 from a uniform distribution $U[0, 1]$, while X_i, $i = 1, 2, 5, 6$, are standard normal random variables. We ran 1000 simulations at four different sample sizes $n = 500, 200, 100, 50$. Table 2 reported the median p-values for testing $q = 0$, the average angles (the smaller the better) between the estimated direction $\hat{\boldsymbol{\eta}}^T\mathbf{X}$ and the true direction $\boldsymbol{\eta}^T\mathbf{X} = X_3 + X_4$, and the median number of significant predictors from the model-free variable selection procedure via SIR/IRE (Cook and Ni, 2005). With $n = 500$, SIR always rejected the null hypothesis of $q = 0$; about 85% of times, SIR was able to pick and only

Table 2
Results for non-constant β at 4 different sample sizes.

n	p-value for testing $q = 0$	Average angle	Significant predictors
500	1E−11	7.99	2
200	5E−5	12.83	2
100	0.008	18.60	2
50	0.055	27.96	1

Table 3
The percentage of right decisions when $\beta = 0.5$, 1.0 and 1.5.

n	0.5	1.0	1.5
500	95.2	95.5	95.3
200	95.7	94.8	94.7
100	96.1	94.9	94.6
50	94.6	96.2	97.0

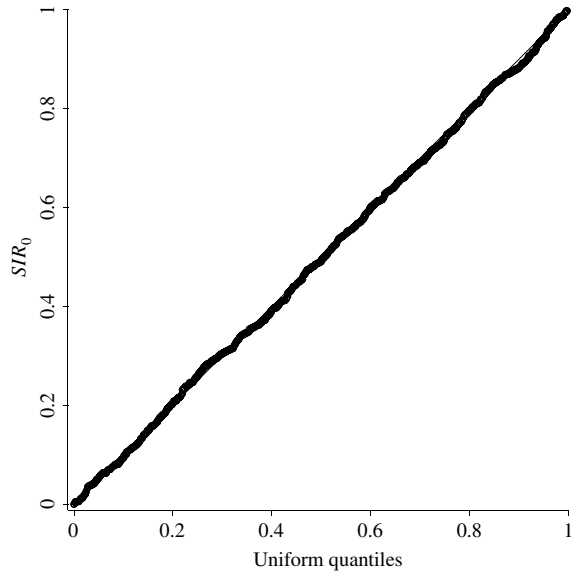


Fig. 3. Uniform quantile plot of p-values testing $H_0 : q = 0$ with $\beta = 0.5$ and $n = 500$.

pick the two truly effective predictors X_3 and X_4 out; the angles between the true direction and estimated direction are reasonably small. The performance of SIR deteriorates as n became smaller.

We also considered the cases with constant β under similar setup as above. The only difference is now that instead of generating β as a function of $\mathbf{X}$, we took $\beta = 0.5$, 1.0 and 1.5 respectively, which corresponded to 30%, 50% and 60% of censoring rate. Table 3 showed us the frequencies of accepting mull hypothesis $q = 0$ at significance level 0.05. About 95% of times, our tests made the right decision. It seems that our method is immune to the change of β. Fig. 3 showed the uniform quantile plot of p-values for testing $H_0 : q = 0$ at $n = 500$ and $\beta = 0.5$ with 1000 replications. As we can see, the sampling distribution of SIR's test statistics is close to the asymptotic one, suggesting a close agreement between the actual and nominal levels.

4. Conclusion and future research directions

We proposed a model-free approach to test whether the censoring parameter β depends on the covariates $\mathbf{X}$ under the context of conditional KG model. When β is indeed a function of $\mathbf{X}$, our approach also allows us to conduct a model-free variable selection procedure. Since our method does not require any modeling assumptions on the survival function of T ($\bar{F}$). It can serve as a good starting point for further parametric and nonparametric modeling procedures.

The adoption of sufficient dimension reduction to survival analysis data has shown great promise. Researches on inferring both the survival function and β simultaneously under the conditional KG model are also underway.

Another future research direction deserves serious investigations is dimension reduction in functional regressions where $\mathbf{X}$ and Y are random functions. Functional data analysis is now an important research field with many applications, see Ferraty and Vieu (2006) for an extensive review. Since both $\mathbf{X}$ and Y are infinite dimensional, parsimony is even more important therein. Ferre and Yao (2005), Amato et al. (2006) and Ait-Sidi et al. (2008) investigated dimension reduction methods assuming a multiple (single) index functional model. As Ferre and Yao (2005) pointed out, there are some technical

difficulties arising in inverting an estimator of Var(E(**X**|*Y*)) since it is ill-conditioned. Ferre and Yao (2005) introduced *functional* SIR, while Amato et al. extended both MAVE (Xia et al., 2002) and SIR to functional data. More attentions will be brought to this area as analysis of massive data become a routine in statistical analysis.

Acknowledgements

We thank the Associate Editor and two anonymous referees for their insightful suggestions and comments which led to great improvement of an earlier draft.

Appendix

Proof of Theorem 1. 1. Since for binary variable δ, under model (1.1), $P(\delta = 1|\mathbf{X}) = P(C_i < T_i|\mathbf{X}) = \beta(\mathbf{X})/(1+\beta(\mathbf{X}))$, hence it is obvious that $\mathcal{S}_{\delta|\mathbf{X}} = \text{Span}\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_q\}$.

2. Given **X**, both T and C, and Y and δ are pairwise independent under model (1.1), by Proposition 2 of Wen (2007), we have

$$\mathcal{S}_{(Y,\delta)|\mathbf{X}} = \mathcal{S}_{Y|\mathbf{X}} + \mathcal{S}_{\delta|\mathbf{X}},$$
$$\mathcal{S}_{(T,C)|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}} + \mathcal{S}_{C|\mathbf{X}}.$$

Also, note that (Y, δ) is a function of (T, C), therefore, $\mathcal{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathcal{S}_{(T,C)|\mathbf{X}}$.

Following the proof of Proposition 1 from Ebrahimi et al. (2003), let $S_T(t) = \bar{F}_{T|\mathbf{X}}(t|\mathbf{x}) = pr(T \geq t|\mathbf{X})$, $S_C(t) = \bar{G}_{T|\mathbf{X}}(t|\mathbf{x}) = pr(C \geq t|\mathbf{X})$, $S_Y(t) = pr(Y \geq t|\mathbf{X})$, and $S_\delta(t) = pr(Y \geq t, \delta = 1|X)$. Since $T \perp\!\!\!\perp C|\mathbf{X}$, we have

$$S_T(t) = S_Y(t) \exp^{-\int_0^t \frac{dS_\delta(u)}{S_Y(u)}}.$$

Therefore, $\mathcal{S}_{T|X} \subseteq \mathcal{S}_{(Y,\delta)|\mathbf{X}}$. In addition, it is straightforward that $S_Y(t) = S_T(t)S_C(t)$. Hence, $\mathcal{S}_{C|X} \subseteq \mathcal{S}_{(Y,\delta)|\mathbf{X}}$. So $\mathcal{S}_{T|\mathbf{X}} + \mathcal{S}_{C|\mathbf{X}} \subseteq \mathcal{S}_{(Y,\delta)|\mathbf{X}}$.

So $\mathcal{S}_{(Y,\delta)|\mathbf{X}} = \mathcal{S}_{Y|\mathbf{X}} + \mathcal{S}_{\delta|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}} + \mathcal{S}_{C|\mathbf{X}}$.

3. When $q = 0$, $\mathcal{S}_{\delta|\mathbf{X}} = \emptyset$, by result from (2) and (1.1), we have $\mathcal{S}_{Y|\mathbf{X}} = \mathcal{S}_{T|\mathbf{X}} = \mathcal{S}_{C|\mathbf{X}}$. □

The equivalence of SIR and IRE with binary response

When Y is continuous, Li (1991) proposed estimating E(**X**|$Y = y$) by replacing Y with a discrete version constructed by partitioning the range of Y into h fixed slices. Accordingly, we follow standard methodology and assume that Y takes values in $\{1, 2, \ldots, h\}$.

Cook (2004) showed that we can derive SIR from the following nonlinear least squares objective function:

$$F_d^{\text{sir}}(\mathbf{B}, \mathbf{C}) = \sum_{y=1}^{h} (\hat{f}_y \hat{\boldsymbol{\xi}}_y - \mathbf{B}\mathbf{C}_y)^T \hat{f}_y^{-1} \hat{\boldsymbol{\Sigma}} (\hat{f}_y \hat{\boldsymbol{\xi}}_y - \mathbf{B}\mathbf{C}_y)$$

where $\mathbf{B} \in \mathbb{R}^{p\times d}$, $\mathbf{C}_y \in \mathbb{R}^d$, $\hat{\boldsymbol{\xi}}_y = \hat{\boldsymbol{\Sigma}}^{-1}(\bar{\mathbf{X}}_{y\bullet} - \bar{\mathbf{X}}_\bullet) = \hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \hat{\mathbf{Z}}_{y\bullet}$, $y = 1, \ldots, h$, $\bar{\mathbf{X}}_{y\bullet}$ is the sample average of **X** in the yth slice, $\bar{\mathbf{X}}_\bullet$ is the sample average of **X**, and $\hat{f}_y = n_y/n$ is the fraction of sample points in slice y. With $h = 2$ and notice that $\hat{f}_1\hat{\boldsymbol{\xi}}_1 + \hat{f}_2\hat{\boldsymbol{\xi}}_2 = 0$, the above ionization problem is the same as minimizing the following quadratic discrepancy function which is exactly how IRE is derived (Cook and Ni, 2005):

$$F_d^{\text{ire}}(\mathbf{B}, \mathbf{C}) = (\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(\mathbf{B}\mathbf{C}))^T \mathbf{V}_n (\text{vec}(\hat{\boldsymbol{\xi}}) - \text{vec}(\mathbf{B}\mathbf{C})),$$

where $\hat{\mathbf{f}} = (\hat{f}_1, \ldots, \hat{f}_{h-1})^T$, $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\xi}}_1, \ldots, \hat{\boldsymbol{\xi}}_{h-1}) \in \mathbb{R}^{p\times(h-1)}$ and $\mathbf{V}_n^{-1}$ is the sample covariance matrix of the limiting normal distribution of $\sqrt{n}(\text{vec}(\hat{\boldsymbol{\xi}}) - \text{E}(\text{vec}(\hat{\boldsymbol{\xi}})))$. Since with $h = 2$, $V_n \propto \hat{f}_y^{-1}\hat{\boldsymbol{\Sigma}}$.

References

Ait-Sidi, A., Ferraty, F., Kassa, R., Vieu, P., 2008. Cross-validated estimation in the single functional index model. Statistics 42, 475–494.
Amato, A., Antoniadis, A., De Feis, I., 2006. Dimension reduction in functional regression with applications. Computational Statistics and Data Analysis 50, 2422–2446.
Cheng, P.E., Lin, G.D., 1987. Maximum likelihood estimation of a survival function under the Koziol–Green proportional hazards models. Statistics and Probability Letters 5, 75–80.
Cook, R.D., 1998. Regression Graphics. Wiley, New York.
Cook, R.D., 2004. Testing predictor contributions in sufficient dimension reduction. The Annals of Statistics 32, 1062–1092.
Cook, R.D., Ni, L., 2005. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. Journal of the American Statistical Association 100, 410–428.
Cook, R.D., Weisberg, S., 1991. Discussion of sliced inverse regression for dimension reduction. Journal of the American Statistical Association 86, 316–342.
Csörgö, S., 1988. Estimation in the proportional hazards model of random censorship. Statistics 19, 437–463.
Ebrahimi, N., Molefe, D., Ying, Z., 2003. Identifiability and censored data. Biometrika 90, 724–727.
Faraway, J., Csörgö, S., 1998. The paradoxical nature of the proportional hazards model of random censorship. Statistics 31, 67–78.
Ferraty, F., Vieu, P., 2006. Nonparametric Functional Data Analysis: Theory and Practice. In: Springer Series in Statistics, Springer, Berlin.
Ferre, L., Yao, A.F., 2005. Smoothed functional inverse regression. Statistica Sinica 15, 665–683.
Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York, NY.

Gaddah, A., Braekers, R., 2009. Weak convergence for the conditional distribution function in a Koziol–Green model under dependent censoring. Journal of Statistical Planning and Inference 139, 930–943.
Gannoun, A., Saracco, J., 2003. An asymptotic theory for SIR_α method. Statistica Sinica 13, 297–310.
Gather, U., Pawlitschko, J., 1998. Estimating the survival function under a generalized Koziol–Green model with partially informative censoring. Metrika 48, 189–207.
Koziol, J.A., Green, S.B., 1976. A Cramër-von mises statistic for randomly censored data. Biometrika 63, 465–474.
Lagakos, S.W., 1979. General right censoring and its impact on the analysis of survival data. Biometrics 35, 139–156.
Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–342.
Li, K.C., Aragon, Y., Shedden, K., Agnan, C.T., 2003. Dimension reduction for multivariate response data. Journal of the American Statistical Association 98, 99–109.
Liquet, B., Saracco, J., 2007. Pooled marginal slicing approach via SIR_α with discrete covariables. Computational Statistics 22, 599–617.
Li, B., Wang, S., 2007. On directional regression for dimension reduction. Journal of the American Statistical Association 102, 997–1008.
Li, K.C., Wang, J.L., Chen, C.H., 1999. Dimension reduction for censored regression data. The Annals of Statistics 27, 1–23.
Peña, E.A., Rohatgi, V.K., 1989. Survival function estimation for a generalized proportional hazard models of random censorship. Journal of Statistical Planning and Inference 22, 371–389.
Saracco, J., 2005. Asymptotics for pooled marginal slicing estimator based on SIR_α approach. Journal of Multivariate Analysis 96, 117–135.
Veraverbekea, N., Cadarso-Suárez, C., 2000. Estimation of the conditional distribution in a conditional Koziol–Green model. Test 9, 97–122.
Wang, H., Xia, Y., 2008. Sliced regression for dimension reduction. Journal of the American Statistical Association 103, 811–821.
Wen, X., 2007. A note on sufficient dimension reduction. Statistics and Probability Letters 77, 817–821.
Wen, X., Cook, R.D., 2007. Optimal sufficient dimension reduction in regressions with categorical predictors. Journal of Statistical Planning and Inference 137, 1961–1978.
Wen, X., Cook, R.D., 2009. New approaches to model-free dimension reduction for bivariate regression. Journal of Statistical Planning and Inference 139, 734–748.
Xia, Y., Tong, H., Li, W.K., Zhu, L.-X., 2002. An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society. Series B 64, 363–410.
Yuan, M., 2005. Semiparametric censorship model with covariates. Test 14, 489–514.
Zeng, D., 2004. Estimating marginal survival function by adjusting for dependent censoring using many covariates. The Annals of Statistics 32, 1533–1555.
Zhong, W., Zeng, P., Ma, P., Liu, J., Zhu, Y., 2005. RSIR: regularized sliced inverse regression for motif discovery. Bioinformatics 21, 4169–4175.