Published in final edited form as:
J Am Stat Assoc. 2011 March 1; 106(493): 178–190. doi:10.1198/jasa.2011.tm08250.

Inverse regression estimation for censored data

Nivedita V. Nadkarni[*], **Yingqi Zhao**[†], and **Michael R. Kosorok**[‡]

[†]Yingqi Zhao is Ph.D. student, Department of Biostatistics, University of North Carolina at Chapel Hill, NC 27599 (yqzhao@email.unc.edu)

[‡]Michael R. Kosorok is Professor and Chair, Department of Biostatistics, and Professor, Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 (kosorok@unc.edu).

Abstract

An inverse regression methodology for assessing predictor performance in the censored data setup is developed along with inference procedures and a computational algorithm. The technique developed here allows for conditioning on the unobserved failure time along with a weighting mechanism that accounts for the censoring. The implementation is nonparametric and computationally fast. This provides an efficient methodological tool that can be used especially in cases where the usual modeling assumptions are not applicable to the data under consideration. It can also be a good diagnostic tool that can be used in the model selection process. We have provided theoretical justification of consistency and asymptotic normality of the methodology. Simulation studies and two data analyses are provided to illustrate the practical utility of the procedure.

Keywords

right censored data; accelerated failure time; sufficient dimension reduction

1 Introduction

An objective of analyzing survival data via regression is to develop a predictive model given covariates. Often this is done under semiparametric considerations when the covariate effects are summarized in a linear manner as in the Cox (1972) model. An important step in formulating the model involves variable selection. Most of the variable selection techniques used for analyzing censored data are extensions of the regression methodology for uncensored data. Stepwise deletion and best subset selection are the most popular ones in this context. Selection of the influential predictors is critical and becomes complicated if the data has many high dimensional covariates, as is often the case in clinical trials and more recently in microarray studies. In addition to selection, assessment of predictor performance is also crucial. It is therefore very beneficial to efficiently select a subset of significant variables which is sufficient for inference on the response and then to model those variables effectively.

A variety of variable and model selection procedures have been proposed to address these issues in the censored setup. Tibshirani (1997) suggested the Lasso for variable selection in the Cox model. This approach minimizes the log partial likelihood subject to the sum of the absolute values of the parameters being bounded by a constant. The nature of the constraint

[*] niveditan@gmail.com .

shrinks coefficients and produces some coefficients that are exactly zero. Tibshirani gives the example of the veteran's lung cancer data set, but the assumption of proportional hazards is unreasonable for nominal covariates such as cell type and Karnofsky score. Hence, the Lasso is not applicable when the proportional hazards assumption is not valid. Fan and Li (2002) proposed variable selection via penalized likelihood for Cox's proportional hazards and frailty models. Selection of significant variables and estimation of regression coefficients is done simultaneously in this method. As in the case of the Lasso, this procedure is applicable only for variable selection in Cox models. Keles et al. (2004) developed a model selection method to select among predictors of right censored outcomes in the context of prediction and density/hazard estimation problems. This procedure is applicable for estimating data-based parameters such as the conditional mean, conditional density, etc.

In many applications the assumptions made for model based inference may not be valid, and consequently the results can be biased. As a result, nonparametric methods are becoming increasingly popular. Recently, there have been several nonparametric alternatives for uncensored data that address the issue of variable selection without assuming an underlying model. Li (1991) introduced sliced inverse regression (SIR) and Cook (2004) developed a procedure for testing predictor contributions via SIR. In addition to these approaches, there have also been Bayesian based techniques in variable and model selection.

Li et al. (1999) extended SIR for censored data. They proposed methods of finding low dimensional projections of the data for visually examining the censoring pattern. A double slicing procedure that requires dimension reduction for both T, the failure time, and the censoring time C using principal component analysis was introduced. The example used to illustrate the procedure is the primary biliary cirrhosis of the liver (PBC) data collected at the Mayo clinic between 1974 and 1986. In the example, the authors use only 6 of the original 17 predictors for their analysis and the justification for the proposed method is via a comparison with the parametric analysis done by Fleming and Harrington (1991). Li's paper provides a background on implementing SIR for censored data and opens up avenues for further research in the area.

Cook (2004) formulated a methodology for testing predictor contributions using SIR. He introduced tests of hypothesis of no effect for selected predictors in regression for uncensored data, without assuming a model for the conditional distribution of the response given the predictors. The sufficient dimension reduction approach (hereafter SDR) via inverse regression was subsequently introduced by Cook and Ni (2005). They improve on the methodology developed by Cook (2004) using a more efficient approach. In their paper, a family of dimension reduction methods, the inverse regression family, is developed by minimizing a quadratic objective function. An optimal member of this family, the inverse regression estimator (IRE) is proposed, along with inference methods and a computational algorithm. An example on lean body mass regression is provided as also simulation studies which show the effectiveness of the method. A simulation comparison between SIR and IRE and theory supports the claim that SIR is a suboptimal member of the inverse regression family.

The purpose of this paper is the development of SDR for censored data without requiring semiparametric restrictions on the form of the censoring distribution. Let T be the failure time and let Z denote the $p \times 1$ vector of covariates. We are interested in inferring about $log(T)|Z$. The conditional distribution of $T|Z$ does not need to be modeled explicitly in order to identify a low dimensional representation of the covariate effect. We incorporate the inverse probability of censoring in our procedure which ensures that censoring is accounted for and also ensures computational ease.

SDR is based on a population meta-parameter, the central subspace (CS) (Cook (1996)). We represent it by $S_{T|Z}$ and define it as the intersection of all subspaces $S \subseteq \mathcal{R}^p$ having the property $T \perp Z|P_S Z$ where $\perp$ indicates independence and P_S is the orthogonal projection onto S in the usual inner product. Therefore, the statement translates as T is independent of Z given $P_S Z$. The CS is a uniquely defined subspace of $\mathcal{R}^p$ when it exists (Cook (1998)). If the central subspace exists, the statement

$$\log(T) \perp Z|\eta' Z \tag{1}$$

can be thought of as a dimension reduction model, where η is a $p \times dim(S_{T|Z})$ basis for the CS. The CS allows reduction of the predictor from Z to $\eta'Z$ without loss of information. $\eta'Z$ is therefore referred to as a "sufficient" predictor.

Our contribution to SDR for censored data is twofold. Firstly, we introduce inverse regression (IR hereafter) for censored data using inverse regression estimators with a quadratic objective function. Secondly, we utilize the inverse probability of censored weighting so that inference is based on the variable of interest T after adjusting for the censoring variable C. See Rotnitzky and Robbins (2003) for a reference on inverse probability of weighting. This ensures a simpler implementation than the one described in Li et al. (1999) in SIR for censored data since it bypasses the need to take the two variables' structure into account. For this approach, no underlying model assumptions are required for T or C except for some weak nonparametric smoothness assumptions on the density of C to be described shortly. This provides flexibility in assessing the variable contribution based purely on the data driven technique developed herein. The procedure is easy to implement and computationally fast. We use bootstrap methods to obtain the structural dimension of the regression. Therefore, we address the issue of variable selection in a nonparametric context, thus augmenting the literature beyond Fan and Li's and Tibshirani's papers.

The data setup and assumptions that are required for obtaining the model given in equation (1) are presented in Section 2. The assumptions are mainly needed to ensure proper inference on the meta-parameter. The proposed estimation procedure and the sample estimators are discussed in Section 3. A nonparametric Kaplan-Meier estimator is utilized to address the issue of nonparametrically estimating the distribution function of C. This facilitates computing the inverse probability of censored weighting. A minimum discrepancy approach is utilized for inverse regression, and bootstrap methods are developed for dimension selection and predictor testing. Theoretical properties of proposed methods are discussed in Section 4. The proofs of the theorems and lemmas in Section 4 are provided in the appendix. Simulation studies and data analyses demonstrate the applicability of the method in Section 5. The simulation studies look at dimension reduction for data drawn from the Cox model and the accelerated failure time model. The method is illustrated on the diffuse large B-cell lymphoma (DLBCL) data. We also provide an illustration on the PBC data to compare with Li et al. (1999). Finally, we discuss future research and open questions in Section 6.

2 The data setup and structure

2.1 Data assumptions

The observed data $(X_i, \delta_i, Z_i, i = 1, \ldots, n)$, consist of n i.i.d. realizations of (X, δ, Z), where $X = \min(T, C)$ and $\delta = I(T \leq C)$, T being the failure time and C the right censoring time. Z is the $p \times 1$ vector of covariates and is assumed to be restricted to a known, compact subset $\mathcal{Z} \subset \mathcal{R}^p$. Let $Y = \log(X)$ for notational convenience.

Let F_Z and G_Z denote the conditional distribution functions of T and C given Z respectively. We denote the respective conditional survival functions by S_Z and L_Z. We make the following additional assumptions:

(A1) $P[C = 0] = 0$, $P[C \geq \tau|Z] = P[C = \tau|Z] > 0$, almost surely, and censoring is independent of T given Z.

(A2) C is either discrete or continuous w.r.t a Lebesgue measure.

(A3) The vector of covariates Z is assumed to be time independent.

(A4) $L_Z(t) > 0$ for all $-\infty < t \leq \tau$ and $L_Z(t) = 0$ for $t > \tau$.

(A5) Assume that $\{TI(T \leq \tau), TI(T=\tau)\} \perp Z|\eta' Z$. More specifically, we require,

$$\begin{aligned} h_z(t) &= g_{\eta' z}(t), \forall t \in (0, \tau] \\ h_z^+ &= g_{\eta' z}^+, \end{aligned} \tag{2}$$

where $h_z(t)$ is the density of $(T|Z = z)$ and $h_z^+ = P(T > \tau|Z = z)$ where g and g^+ are some functions. We also assume h is Lipschitz continuous uniformly over Z, i.e., $\sup_{z \in \mathcal{Z}} |h_z(t_1) - h_z(t_2)| \leq K_0 |t_1 - t_2|$, for some $K_0 < \infty$.

2.2 Additional assumptions for dimension reduction

The most important assumption for dimension reduction is that the central subspace exists. For our setting, the dimension of the CS may be smaller than the dimension of the CS if $\log(T)$ were fully known. Inverse regression relies on an assumption about the marginal distribution of Z. The linearity condition requires that $E(Z|\eta' Z = u)$ is linear in u, where the columns of η form a basis for $S_{\log(T)|Z}$ (Cook 1998, Proposition 4.2). This condition connects the central subspace (CS) with inverse regression of Z on $\log(T)$. When it holds, $E[Z|\log(T)] \in S_{\log(T)|Z}$ and hence $Span(Cov(E(Z|\log(T)))) \subseteq S_{\log(T)|Z}$. This condition has been discussed in several places and is required for SIR as well. However, the performance of any of the dimension reduction methods is not sensitive to this condition. In view of the fact that most low-dimensional projections of high-dimensional data often appear like normal distributions (Diaconis and Freedman (1984)), Hall and Li (1993) argue for the generality of this condition in high-dimensional situations. On the other hand, reweighting and subsampling methods can also be applied to obtain this condition. This condition allows us to infer about a proper subset of the CS.

In order to guarantee the existence of the CS, we need to make assumptions on the predictors. We can make the assumption of elliptically contoured predictors for which the linearity condition holds. However, since this condition is more restrictive, we can relax the assumption and instead assume that the marginal distribution of the Z's has convex support. In this case, the CS is unique when it exists (Cook (1998)).

Therefore, we need to make just the following two assumptions:

(B1) The marginal distribution of the vector of covariates Z has convex support.

(B2) $E(Z|\eta' Z = u)$ is linear in u.

2.3 Assumptions needed for asymptotic properties of the basis estimator

In order for sufficient dimension reduction to be applicable for censored data, we outline more conditions required as part of the assumptions needed for the methodology to be effective.

We are dealing with a data structure of the form (X, δ) to make inference on $log(T)|Z$. To adjust for the censoring variable C, we use inverse probability of censoring weighting. This inverse weighting approach is incorporated in the nonparametric estimation of the weighted Kaplan-Meier estimator for the censored time, the Kaplan-Meier estimator for the failure time, and also in the estimation of the sample estimators. To ensure that this inverse weighting preserves the inherent nature of the methodology, we need the following conditions:

We define a collection of sets and related assumptions that will be necessary for the theoretical explanation of the construction of the weighted Kaplan-Meier estimator of the censoring time. We make the following assumptions:

(C1) For some $\gamma \in (0, 1]$ and some $K_1 < \infty$, the probability function $P(T > C, C \leq t|Z = z) = f(z, t)$ satisfies $\sup_{t\in(0,\tau]}|f(z_1, t) - f(z_2, t)| \leq K_1\|z_1 - z_2\|^{\gamma}$.

(C2) For the same γ as in (C1) and some $K_2 < \infty$, the probability function $P(T > t, C \geq t|Z = z) = g(z, t)$ satisfies $\sup_{t\in(0,\tau]}|g(z_1, t) - g(z_2, t)| \leq K_2\|z_1 - z_2\|^{\gamma}$.

(C3) We also assume that the conditional survival function for the censoring time is Lipschitz continuous uniformly over Z, i.e., $\sup_{z\in\mathcal{Z}}|L_z(u_1) - L_z(u_2)| \leq K_3|u_1 - u_2|$, for some $K_3 < \infty$.

(C1)–(C3) are needed to ensure asymptotic consistency of the weighted Kaplan-Meier estimator of the conditional censoring distribution and for establishing the convergence rate.

3 Methodology

3.1 Inverse regression

In this section, we discuss inverse regression and the minimum discrepancy approach. We begin by outlining the idea of inverse regression for censored data. The primary variables of interest are the failure time, T , and the vector of covariates, Z. We want to infer about $\log(T)|Z$ using inverse regression. First, we begin by defining some of the main terms of interest. Since inverse regression is based on constructing sample versions of $E(Z|\log(T))$, we proceed by partitioning the log of the failure time T into equal non-overlapping intervals $u_j = (t_j, t_{j+1}], j = 1, \ldots, h$, where $t_h = \tau < \infty$. This partition is one of many possible partitions and as n increases, the partition is allowed but not required to become finer. Σ is the covariance matrix of the predictor vector Z.

Define the working meta parameter,

$$S_{\xi}=\sum_{j=1}^{h} S\,pan\left(\xi_{u_j}\right),$$

where,

$$\xi_{u_j}=\Sigma^{-1}\left(E\left[Z|\log(T)\in u_j\right]-E[Z]\right)$$

Let $d = \dim(S_\xi)$ and let $\beta \in R^{p\times d}$ be a basis of S_ξ. We also define a vector γ_t^* such that $\xi_t = \beta\gamma_t^*$ for each t. An estimate of β provides an estimate of the basis of S_ξ under linearity, but inference about S_ξ itself does not require linearity. Define

$$\xi = (\xi_{u_1}, \ldots, \xi_{u_h}) = \beta\gamma^*,$$

where $\gamma^* = (\gamma_1^*, \ldots, \gamma_{u_h}^*)$. Let $f = (f_{u1}, \ldots, f_{uh})'$, where $f_{ut} = P(\log(T) \in u_t)$. The intrinsic location constraint gives $\xi f = \beta\gamma^* f = 0$.

Following Cook and Ni (2005), we obtain the basis estimate first and then link it with a testing procedure to select d, the structural dimension of the regression. The structural dimension of the regression is defined as the smallest number of distinct linear combinations of the predictors required to characterize the conditional distribution of the response given the predictors.

In this paragraph, we give a brief idea of the minimum discrepancy approach that we will be using. It is natural to estimate S_ξ with a d-dimensional subspace that is closest to the columns of the sample estimator of ξ. There are many ways to define "closeness". Letting vec(·) denote the operator that constructs a vector from a matrix by stacking its columns, we consider quadratic discrepancy functions of the form

$$F_d(B, K) = \left(\mathrm{vec}\left(\widehat{\xi}R_n\right) - \mathrm{vec}(BK)\right)' V_n \quad \left(\mathrm{vec}\left(\widehat{\xi}R_n\right) - \mathrm{vec}(BK)\right), \tag{3}$$

where $V_n \in \mathcal{R}^{pl\times pl}$ is a positive definite matrix. The columns of $B \in \mathcal{R}^{p\times d}$ represent a basis for Span(ξR_n); and $K \in \mathcal{R}^{d\times l}$, which is used only in fitting, represents the coordinates of ξR_n relative to B. The matrix $R_n \in \mathcal{R}^{h\times l}$ decides how we organize the columns of $\widehat{\xi}$. The subspace of $\mathcal{R}^p$ spanned by a value of B that minimizes F_d provides an estimate of a subset of S_ξ, depending on (R_n, V_n). One such pair corresponds to a dimension reduction method. These methods are called the IR family. Given (R_n, V_n), solutions of this minimization are not unique due to overparametrization, however this nonindentifiability is not an issue, because any complete basis suffices to specify S_ξ. It is possible to impose constraints to make the parametrization unique, but the overparametrized setting is more intuitive and generally easier to treat analytically.

Now we move on to obtaining the sample estimators for dimension reduction.

3.2 Estimators required for inverse regression

In this section, we obtain the estimators required to carry out inverse regression based on the observed data . We need to obtain a basis for S_ξ as well as a way to determine the dimension d of the basis. In order to do this, we first need to describe the sample estimates that will be required before we proceed to the actual basis estimation.

An important thing to note here is that since T is not observed we make use of the inverse probability of censored weighting to incorporate the information from the censored observations. We use the notation $Y = \log(X)$ to denote the transformed variable.

Since the failure time is not observed, we partition Y as enumerated earlier. Let u_y denote the interval $(t_j, t_{j+1}]$ which contains y and let Z_{yj} denote the j^{th} observation on Z in interval u_y, $j = 1, \ldots, n_y$, $y = 1, \ldots, h$, and $\sum_y n_y = n$. The mesh size should be fine enough to capture the dependency structure (as a function of $\beta'Z$), but it need not converge to zero. We therefore

assume hereafter that the mesh size is fine enough to capture the needed structure. Let $\bar{Z}_{..}$ be the overall average of Z, and $\bar{Z}_{y.}$ denote the average of the n_y points with $Y \in u_y$. We estimate $E[Z|\log(T) \in u_y]$ by $\bar{Z}_{y.}$ such that the missing information from censoring is incorporated. The theoretical justification is given in detail in Section 4.

In order to estimate the conditional expectation such that it is accurate and unbiased, we weight the sum in each interval by the inverse of the estimated probability $\hat{P}(C > T|Z)$. This probability is estimated using a kernel conditional Kaplan-Meier estimator (Dabrowska (1989)).

Therefore, the estimator of $E[Z|\log(T) \in u_y]$ can be expressed as,

$$\bar{Z}_{y.} = \frac{\mathbb{P}_n\left[\frac{\delta Z I(Y \in u_y)}{\hat{P}(C>T|Z)}\right]}{\mathbb{P}_n\left[\frac{\delta I(Y \in u_y)}{\hat{P}(C>T|Z)}\right]}. \tag{4}$$

The weighted Nelson-Aalen estimator for the cumulative hazard of the censoring time is defined as:

$$\widehat{\Lambda}^*_Z(t) = \int_0^t d\bar{N}^*_Z / \bar{Y}^*_Z, \tag{5}$$

where Y^*_Z and N^*_Z denote weighted processes of number at risk and events for censoring. Let $N^c_i(t)$ and $Y^c_i(t)$ denote the counting process and at risk process respectively for the i^{th} observation: $N^c_i(t) = 1(Y_i \le t, \delta_i = 0)$, $Y^c_i(t) = 1(Y_i > t)$. Then,

$$\bar{N}^*_Z = \frac{n^{-1}h^{-d}\sum_{i=1}^n K(\|z - z_i\|/h)\, N^c_i(t)}{n^{-1}h^{-d}\sum_{i=1}^n K(\|z - z_i\|/h)}; \tag{6}$$

$$\bar{Y}^*_Z = \frac{n^{-1}h^{-d}\sum_{i=1}^n K(\|z - z_i\|/h)\, Y^c_i(t)}{n^{-1}h^{-d}\sum_{i=1}^n K(\|z - z_i\|/h)}. \tag{7}$$

Here K is a kernel function and h is the bandwidth. p is the dimension of z, d is the number of covariates that are continuous, with $p - d$ being the number of discrete-valued covariates. Different types of kernel functions can be used with little difference in the results. When the dimension of the covariate space is high, we propose a simplification of the kernel function as $h^{-r}K(\|x - x_i\|/h)$, with x defined as the first $r \le d$ principal component of the d-dimensional covariates. Consequently, the weighted Kaplan-Meier estimator can be written as:

$$\widehat{L}_Z(t) = \phi\left(-\int_0^t \frac{n^{-1}\sum_{i=1}^n h^{-d}K(\|z - z_i\|/h)\ \mathrm{d}N^c_i(s)}{n^{-1}\sum_{i=1}^n h^{-d}K(\|z - z_i\|/h)\ Y^c_i(s)}\right), \tag{8}$$

where ϕ is the product integral functional.

Let $\hat{f}_{uj} = \hat{S}_Z(t_{j+1}) - \hat{S}_Z(t_j)$, where

$$\widehat{\Lambda}_Z(t) = \int_0^t \frac{\sum \frac{dN_i(s)}{\bar{L}_Z(s-)}}{\sum \frac{Y_i(s)}{\bar{L}_Z(s-)}} \tag{9}$$

is the estimate of the cumulative hazard for the failure time and $\hat{S}_Z$ is the resulting survival function estimate of the failure time. Let $\widehat{\Sigma}_{>0}$ be the usual sample covariance matrix for Z. Then, the sample version of ξ_{u_t} is $\widehat{\xi}_{u_y} = \widehat{\Sigma}^{-1}\left(\bar{Z}_{y\cdot} - \bar{Z}_{\cdot\cdot}\right)$, which ensures that $\widehat{\xi}_{u_y} \in \mathcal{R}^{p\times h}$.

We compute the survival function for T by inversely weighting the Kaplan-Meier with the corresponding probability $\hat{P}(C > T|Z)$ in the algorithm. After these probabilities have been computed, $\bar{Z}_{y\cdot}$ can be obtained easily.

We would like to mention here that Dabrowska (1989) has shown uniform consistency of a kernel conditional Kaplan-Meier estimate. This estimate is similar to ours, but is structured as a proper kernel estimate and requires more stringent conditions than the ones we specify for proof and implementation.

3.3 Basis estimation

We now discuss basis estimation. We consider inverse regression using a quadratic discrepancy function as outlined earlier. The basis for S_ξ is estimated with a d-dimensional subspace that is closest to the columns of $\widehat{\xi}$.

The choice of an optimal discrepancy function depends on the choices of R_n and V_n. We choose R_n to be nonsingular which, when incorporated into the discrepancy function, simplifies to:

$$\operatorname{vec}\left(\widehat{\xi}R_n\right) - \operatorname{vec}(BK) = R_n' \otimes I_p \quad \left(\operatorname{vec}\left(\widehat{\xi}\right) - \operatorname{vec}\left(BKR_n^{-1}\right)\right).$$

Because we will be eventually minimizing $F_d(B, K)$, K is redefined as KR_n^{-1} without loss of generality.

Let D_v denote a diagonal matrix with the elements of the vector v on the diagonal and construct a nonstochastic matrix $A \in \mathcal{R}^{h\times(h-1)}$ such that $A'A = I_{h-1}$ and $A'1_h = 0$. Then $D_{\widehat{f}}(A, 1_h) \in \mathcal{R}^{h\times h}$ is nonsingular and can be used as the choice for R_n. However, $\widehat{\xi}D_{\widehat{f}}1_h = 0$ due to the intrinsic location constraint and, consequently $\widehat{\xi}D_{\widehat{f}}(A, 1_h) = \left(\widehat{\xi}D_{\widehat{f}}A, 0\right)$. Since the last column is always zero, we will lose no generality by using the reduced data matrix $\widehat{\zeta} \equiv \widehat{\xi}D_{\widehat{f}}A$ in the construction of the discrepancy functions,

$$F_d(B, K) = \left(\operatorname{vec}\left(\widehat{\zeta}\right) - \operatorname{vec}(BK)\right)' V_n \left(\operatorname{vec}\left(\widehat{\zeta}\right) - \operatorname{vec}(BK)\right),$$

where $B \in \mathcal{R}^{p\times d}$, $K \in \mathcal{R}^{d\times(h-1)}$, and V_n has yet to be specified. The optimal choice of V_n in this version of the discrepancy function depends upon the asymptotic distribution of $\text{vec}(\hat{\zeta})$. We verify later that $\hat{\zeta}$ converges in probability to $\zeta \equiv \beta\gamma^* D_f A = \beta v$, where $v = \gamma^* D_f A$.

We now suggest an estimate for V_n that seems reasonable since the asymptotic variance of the basis estimate is diffcult to compute. Define h random variables J_y such that J_y equals the probability of falling in u_y if an observation is in u_y and 0 otherwise, $y = 1, \ldots, h$. Then, $E(J_y) = f_y$. Also define the random vector $\epsilon^* = (\epsilon_1^*, \ldots, \epsilon_h^*)'$, where its elements, ϵ_y^*, are the population residuals from the ordinary least squares fit of J_y on $\tilde{Z}$, where $\tilde{Z}$ is the standardized version of Z. We will use $\left(Cov\left(\text{vec}\left(\hat{\Sigma}^{-1/2}\,\tilde{Z}\,\epsilon^*\right)\right)\right)^{-1}$ as our sample estimate of V_n.

Now we consider minimization of the discrepancy function given V_n. This can be done by using the alternating least squares algorithm (Cook and Ni (2005)) to obtain basis estimates.

3.4 Dimension selection using the bootstrap

In order to test hypotheses of the form $d = d_0$ versus $d > d_0$, we utilize the limiting distribution of $n\hat{F}_d$, where $\hat{F}_d$ is the minimum value of $F_d(B, K)$. If $n\hat{F}_m$ exceeds a selected quantile of the asymptotic distribution of $n\hat{F}_d$ under the null, then the hypothesis is rejected.

It is difficult to derive this limiting distribution in our case. However, the limiting distribution of $n\hat{F}_d$ under the null hypothesis can be approximated using the bootstrap. Let Y^*, δ^*, Z^* denote a resampling of Y, δ, Z drawn randomly. Recall that $F_d(B, K) = \left(\text{vec}(\hat{\zeta}) - \text{vec}(BK)\right)' V_n \left(\text{vec}(\hat{\zeta}) - \text{vec}(BK)\right)$. The bootstrap estimate $\text{vec}(\zeta^*) - \text{vec}(BK)$, denoted as U^*, is computed based on the resample. Bootstrap estimates are centered by subtracting their mean $\bar{U}^*$ to reflect the null hypothesis. We then obtain the critical value from the bootstrap value of $n\hat{F}_d^*$ under the null, which can be calculated as $n(U^* - \bar{U}^*)'V_n(U^* - \bar{U}^*)$. The proof of this centered bootstrap approach follows along the lines of the proofs of Theorem 7 and 8 in Kosorok and Song (2007), after incorporating the results for kernel type estimates as described in Hall (1991). The details of the proof are omitted.

A series of such tests can be used to estimate d as follows. First, starting with $d_0 = 1$, test the hypothesis $d = d_0$. If the hypothesis is rejected, then increase d_0 by one and test again, stopping when the first non-significant result is obtained. Note that we start testing with $d_0 = 1$. Consequently, failing to reject $d_0 = 1$ does not necessarily imply that the one predictor contributes to the regression, because the predictor may be independent of the failure time. However, testing of full independence is beyond the scope of this paper, although this issue is an important one for future research.

3.5 Predictor testing using the bootstrap

The main hypothesis tests of interest would be those for which dimension is not specified yet the predictor contribution is tested robustly. More precisely, we wish to deal with tests of conditional independence,

$$\tilde{T} \perp P_{\mathcal{H}} Z | Q_{\mathcal{H}} Z,$$

where $\mathcal{H}$ is an r-dimensional user-specified subspace of the predictor space. We require $r \leq p$-dim($S_{\tilde{T}|Z}$). This can be accomplished by partitioning $Z' = \left(Z_r', Z_{-r}'\right)$, where we wish to test the hypothesis that r selected predictors do not contribute to the regression. In this case, $\mathcal{H}$=Span (H), with basis H= $\left(I_r', 0\right)$.

For the case of censored data, we are interested in developing the following Marginal Predictor tests:

Marginal Predictor Hypotheses: $P_{\mathcal{H}} S_{\tilde{T}|Z} = O_p$ versus $P_{\mathcal{H}} S_{\tilde{T}|Z} \neq O_p$.

The marginal predictor hypothesis is equivalent to the hypothesis $\mathbf{H}^T \zeta = 0$, where $\mathbf{H}$ is a $p \times r$ basis for $\mathcal{H}$. The test statistic,

$$T(\mathcal{H}) = n\text{vec}\left(\mathbf{H}'\widehat{\zeta}\right)' \left\{\left(\mathbf{I}_{h-1} \otimes \mathbf{H}'\right)\widehat{\Gamma}_{\widehat{\zeta}}(\mathbf{I}_{h-1} \otimes \mathbf{H})\right\}^{-1} \text{vec}\left(\mathbf{H}'\widehat{\zeta}\right),$$

can be used for this procedure. To determine if a predictor is significant, we can choose $\mathbf{H}$ to be $\mathbf{e_k}$, where $\mathbf{e_k}$ is the $p \times 1$ vector with 1 in the kth entry and 0 elsewhere. Then the test statistic is

$$T_k = n\mathbf{e_k}^T\widehat{\xi}\left\{\left(\mathbf{I}_{h-1} \otimes \mathbf{e_k}'\right)\widehat{\Gamma}_{\widehat{\zeta}}(\mathbf{I}_{h-1} \otimes \mathbf{e_k})\right\}^{-1}\widehat{\xi}\,\mathbf{e_k}.$$

Cook and Ni (2005) have used backward selection based on the chi-squared tests in order to select the variables for testing. To elaborate, marginal predictor tests were first carried out and p-values for each test obtained. In the second step, backward elimination is used with the variable having the most insignificant p-value in the marginal test being eliminated first and so on. However, in our case, it is hard to derive the null distributions for the above statistics. Fortunately, as we did previously, we can apply the bootstrap to center the test statistics to reflect the null hypothesis and to obtain critical values. In the marginal test setting, we compute ξ^* from resampling and then subtract the ξ^*s' mean. The T_k^*s are then calculated using these centered quantities. Critical value are obtained from the bootstrap quantiles of T_k^*.

4 Asymptotic properties

In this section, we will mainly discuss the theoretical background that is required for the methodology. To obtain a consistent estimate of the basis of the central subspace, we have to ensure that all of the sample estimators are consistent for their population counterparts. In our derivations, we have shown consistency of all of the estimators. We also use some earlier results from Cook and Ni (2005) and Shapiro (1986) to prove that the basis estimate is a consistent estimator for the basis of the underlying central subspace.

4.1 Consistency of the estimators

We show that the consistency of the weighted Kaplan-Meier estimator holds under the assumptions we have already outlined in Section 2. In addition, we impose certain assumptions on the kernel function, including that the kernel function $K(\cdot) \geq 0$ has a support on [0, 1], and range [0, 1]. It also satisfies $\int K(u)du = 1$, $\int uK(u)du = 0$.

Theorem 1: The weighted Kaplan-Meier estimator for the censoring distribution is consistent for $G_Z(t)$ under the assumptions outlined and achieves an optimal convergence rate

$O_p(n^{-\gamma/2(d+\gamma)})$ when $h = \tilde{O}_P(n^{-1/2(d+\gamma)})$, where $\tilde{O}_P(1)$ is a quantity bounded above and below in probability in the limit.

Lemma 1: The inversely weighted estimator of the survival function of T is consistent for S_Z with the same rate of convergence as the weighted Kaplan-Meier estimator.

The sample covariance matrix $\widehat{\Sigma}$ of the vector of covariates Z is $\sqrt{n}$ consistent for its population counterpart Σ. The overall average of the Z's is also $\sqrt{n}$ consistent for the true value by the law of large numbers.

We have proved consistency of both the weighted estimators for the survival distributions of the censoring time and the failure time. Since the weighted Kaplan-Meier estimator of the conditional censoring time is incorporated in the calculation of $\bar{Z}_{y.}$, we need to prove that this estimator is also consistent.

Lemma 2: The sample estimator $\bar{Z}_{y.}$ is consistent for $E(Z|Y \in u_y)$ with rate $O_P(n^{-\gamma/2(d+\gamma)})$.

Since all the sample estimators are consistent now we need to prove the consistency of the basis estimate. In the implementation of the alternating least squares algorithm, the inverse probability of the censored weighting scheme is utilized to adjust for the loss in information due to censoring.

Since A is a constant matrix, we consider only $\left(\text{vec}\left(\widehat{\xi}D_{\widehat{f}}\right) - \text{vec}\left(\beta\nu D_f\right)\right)$. In order to prove consistency, we need to incorporate the results in Shapiro (1986) on asymptotics of overparametrized discrepancy functions and two other supplemental results that need to be derived based on his main results. We also utilize results from Cook and Ni (2005) to conclusively prove consistency of the basis estimate. The proofs are given in the appendix.

Theorem 2: The first term of the discrepancy function $\text{vec}\left(\widehat{\xi}D_{\widehat{f}}\right)$ is asymptotically normal with rate $O_p(n^{-\gamma/2(d+\gamma)})$ and with mean=$\beta\gamma^* D_f$ and some variance covariance matrix $\Gamma_{\widehat{\zeta}}$.

Theorem 3: The estimate of the basis using the discrepancy function is consistent.

4.2 Validity of the bootstrap

We develop a measure to assess the accuracy of the estimation in data analysis via the bootstrap. Hall (1991) shows that the bootstrap approximation is valid for kernel density estimators. In our setting, the source of variation mainly comes from the kernel-type Kaplan-Meier estimate. Though this kernel type estimator does not achieve root-n consistency, the bootstrap can be shown to consistently approximate the limiting distribution of the discrepancy function, using arguments such as those given in Hall (1991). In particular, the bootstrap method is asymptotically valid for obtaining critical values in structural dimension determination and predictor selection, once we center the bootstrap estimates to reflect the null hypothesis.

5 Simulation studies and data analysis

Simulation studies are carried out to assess the performance of the estimator. For this section, we first report simulation studies to illustrate how our approach works in estimation and testing. Then we apply our method on the Diffuse large B-cell lymphoma data and the PBC data.

5.1 Basis estimation given *d*

We aim to compare performance between SIR using the double slicing estimator and our estimator of S_ξ when d is known. Both accelerated failure (AFT) models and Cox regression models for the failure time are considered for the simulations.

Model 1. First, we take $p = 6$ and generate $\mathbf{z} = (z_1, \cdots, z_6)$ from the standard normal distribution. The true survival time Y^0 is generated from

$$Y^0 = \exp(z_1 + z_3)\,\epsilon_1, \tag{10}$$

where ϵ_1 follows the exponential distributions with parameter 1. Censoring time C_1 is generated from

$$C_1 \sim \exp(z_1 + z_2 + z_3) \wedge 4, \tag{11}$$

which is a constant conditional on regressors. The censoring percentages is 45% approximately.

We vary the sample size from 50 to 100, 200, 400 and 800 to study the effect of sample size on estimation. Also, we study the performance of two estimators as the regressor dimension p gets larger. We increase p from 6 to 10, 15 and 20, and keep the same sample size of $n = 200$. The added predictors follow a standard normal distribution. For each simulation run, we compute the angle between S_ξ and its estimate. The angle between two vectors $\mathbf{a}$ and $\mathbf{c}$ is computed as $180 \cos^{-1}(|\mathbf{a}^{\mathbf{T}}\mathbf{c}|/\|\mathbf{a}\|\|\mathbf{c}\|)\pi$. In Model 1, the basis of the true central subspace is $(1, 0, 1, 0, 0, 0)'$. The leading direction obtained from the SIR method is set to be the SIR estimate, and $\hat{b}_1$ is our estimate using the method described in Section 3 by fixing the dimension of B to be 1.

Figures 1(a) and 1(b) show mean angles from 100 simulation runs in each case for different sample size or different numbers of parameters. As anticipated, we obtain biased estimates when the sample size is small, and the average angle converges to 0 as sample size grows. Our procedure did better than SIR with the double slicing procedure. Both estimators deteriorate gradually as p increases and are close. Increasing the number of covariates does not seem to have a significant effect on angle estimation.

Model 2. Similar to Model 1, the failure time follows (10) and the censoring time follows (14). For the covariates z_1, z_2 and z_3, we draw one of them from a Rademacher distribution and the remaining two from a normal distribution. A Rademacher random variable X satisfies $P(X = -1) = P(X = 1) = 0.5$ and is equivalent to a Bernoulli random variable with success probability 0.5 but standardized to have mean 0 and variance 1, corresponding to the first two moments of a standard normal distribution. We apply this to three different scenarios: failure time dependent on the binary variable, censoring time dependent on the binary variable or neither failure nor censoring time dependent on the binary variable. The purpose of these simulations is to evaluate the influence of a binary variable on the estimation of the basis. The simulation results suggest that our estimators have a better performance compared to SIR estimators from moderate to large sample sizes, although the SIR estimators can have better small-sample behavior. We also compare the two estimators' performance for different numbers of parameters. The sample size n is kept at 500, and the number of parameters is increased from 6 to 10, 15, and 20. According to the simulations, our estimators have less bias compared to SIR estimators in all scenarios, see Figures 2 to 4.

Model 3. We take $p = 6$ and generate $\mathbf{z} = (z_1, \cdots, z_6)$ from the standard normal distribution. The true survival time Y^0 is generated from

$$Y^0 = (-\log(\epsilon_2)/\exp(z_1+z_3)),$$

where ϵ_2 follows the uniform distribution on [0,1]. Note that the Cox model holds for this model. The censoring time C_1 is generated from

$$C_1 \sim \exp(z_1+z_2+|z_3|) \wedge 2. \tag{12}$$

We then compare two estimators for different sample sizes and different numbers of parameters. As shown in Figure 5, our estimators do not perform as well as the SIR estimators for the Cox regression model, although our estimators improve with increasing sample size.

5.2 Estimation of *d*

Using the methodology described in Section 3.4, we consider the following example for $n = 400$: let $z_1, \cdots, z_6$ be generated from the standard normal distribution.

$$Y^0 = \exp((z_1+z_4)\exp(z_3+z_5))\epsilon_1; \quad C \sim \exp(z_2) \wedge 4,$$

In this case, the basis for the central subspace is (1,0,0,1,0,0) and (0,0,1,0,1,0) with the true dimension $d = 2$.

Here is how to execute our procedure in this setting:

- Beginning with $d = 1$, the test statistic $n\hat{F}_1$ is 119.9 Using 1000 bootstraps of the centered $n\hat{F}_1$, we obtain that the 95% quantile is 54.8. Therefore, the hypothesis that $d = 1$ is rejected.
- Increasing to $d = 2$, we obtain $n\hat{F}_2$=1033.2. Using 1000 bootstraps of the centered $n\hat{F}_2$, we obtain that the 95% quantile is 1884.9. The result is not significant and we do not reject the hypothesis that $d = 2$.

Simulating this process 100 times, the hypothesis $d = 1$ is rejected 92 times. When $d = 1$ is rejected, we proceed to test the hypothesis $d = 2$. It is then rejected 29 times. In other words, the procedure identifies the true dimension 63 out of 100 times. The power improves as we increase the samplesize. When we repeat the procedure for $n = 600$, it identifies the true demension 97 out of 100 times.

5.3 Predictor test

We numerically compare the performance of the smoothly clipped absolute deviation (SCAD thereafter, Fan and Li (2002)), adaptive Lasso (ALASSO thereafter, Zhang and Lu (2007)) and the proposed method, where SCAD and ALASSO are existing model selection methods applied to survival data via penalized likelihood. These two methods are, however, based on the Cox proportional hazards model (frailty model) assumption. Under any other conditions, they might not be optimal. We have the following questions for investigation: how does the sample size or model sparsity affect performance of the methods? What is the influence of correlation between predictors in the selection results? Intuitively, tests are more powerful with larger sample size or with less sparse models, and the effectiveness

might deteriorate from increased correlation between covariates. We perform several simulations to evaluate these issues.

We test the significance of the predictors from 100 simulated data sets with the true survival time from

$$Y^0=\exp(z_1+z_3+\exp(2(z_1+z_3)))\epsilon_1, \tag{13}$$

where ϵ_1 follows an exponential distribution with mean 1, $\alpha_0 = 1$. $\mathbf{z} = (z_1, \cdots, z_6)$ is generated from the standard normal distribution. The censoring time C_1 is generated from

$$C_1 \tilde{} \exp(z_1+z_2+z_3) \wedge 4. \tag{14}$$

Note here that neither the proportional hazard or frailty model assumptions are satisfied. The sample size is varied from 100, 200, 400 to 800. Percentages of selecting important variables (z_1 or z_3) out of 100 simulation runs versus selecting non-important ones are shown in Figure 6 under different scenarios, for each variable selection method. Note that the proposed approach performs comparatively the best, in the sense that it generally yeilds lower percentage of non-important variables while also selecting significant covariates more frequently.

Keeping the simulation mechanism the same for failure time and censoring time, we increase the total number of parameters to 10, 15, 20 respectively, where additional predictors are simulated from the standard normal distribution independently. Figure 5.3 shows simulation results as the number of parameters p varies. As anticipated, increasing p leads to an increase in proportions of falsely selected non-significant variables. While the other two methods failed to select significant predictors correctly most of the time, the power of the proposed method stays above 0.8 under all scenarios. To investigate how the three methods perform when covariates are correlated, we let the correlation between different predictors ρ range from 0.2, 0.5 or 0.9. The sample size is set to be 200. The proposed method again outperforms SCAD and ALASSO, as seen in Figure 5.3. Though the power of the proposed test is reduced as ρ increases, it is overall higher than the other two. Also, the proposed approach is less likely to select non-important variables. The other two methods, however, do not work reliably in this context. We conclude this section by reporting simulation results from varying the coefficient of variation (CV), where $c_v = \sigma/\mu$, σ and μ refer to standard deviation and mean respectively. Figure 5.3 suggests that SCAD and ALASSO perform better as CV increases, whereas varying the CV has little impact on the proposed method.

5.4 Data analysis

For the analyses done in this paper, we handle categorical variables by introducing dummy variables as in regular regression. First we illustrate the method on the diffuse large B-cell lymphoma data and then consider the PBC data for comparison with Li et al. (1999).

5.4.1 Diffuse large B-cell lymphoma

The diffuse large B-cell lymphoma (DLBCL) data was first analyzed by Rosenwald et al. (2002). This data set consists of 240 patients with DLBCL including 138 patient deaths during the follow-up. For our analysis purposes, we have excluded those observations for which the time to death is zero. That leaves us with 235 observations. The other variables in the data set include the three gene expression subgroups of DLBCL, gene expression signatures (i.e., germinal center B-cell signature, major-histocompatibility-complex (MHC)

class II signature, lymph node signature and the proliferation signature), value for the BMP6 gene (a member of the transforming growth factor β superfamily of genes), the outcome predictor score, and the international-prognostic-index component (IPI) subgroup. We have excluded the IPI subgroup variable because there are a lot of missing values for it. Since the gene expression sub group is categorical, we use two dummy variables instead of the variable itself. Thus, there are eight covariates.

The marginal predictor test suggests that the gene expression subgroups are important predictors. This is consistent with the view of Rosenwald et al. (2002) that the overall survival after chemotherapy differed significantly among the three subgroups. According to the dimension test of $d = d_0$, we obtain that the central subspace dimension is one, since the F value under the null $d = 1$ is smaller than bootstrap critical value. The estimates, however, suggest that aside from the gene expression subgroups, some gene-expression signatures, especially the outcome predictor score, which is a linear combination of the different signatures and the value of the BMP6 gene as taken from the analysis by Rosenwald et al. (2002), also contribute to the linear combinations. This validates the premise of Rosenwald et al. that the outcome predictor score is a good indicator of the outcome of chemotherapy. See Table 1 for the estimates and bootstrap standard errors.

5.4.2 Primary biliary cirrhosis of the liver

The following briefly describes data collected for the Mayo Clinic trial in primary biliary cirrhosis (PBC) of the liver conducted between January 1974 and May 1984 comparing the drug D-penicillamine (DPCA) with a placebo. The first 312 cases participated in the randomized trial of D-penicillamine versus placebo, and contain largely complete data. The variables in the data set include case number, the number of days between registration and the earlier of death or study analysis time in 1986, censoring indicator, treatment code (1= DPCA, 2=placebo), age in years, sex (0=male, 1=female), presence of ascites (0=no, 1=yes), presence of hepatomegaly (0=no, 1=yes), presence of spiders (0=no, 1=yes), presence of edema, serum bilirubin, serum cholesterol, albumin, urine copper, alkaline phosphatase, SGOT, triglycerides, platelet count, prothrombine time, and histologic state of disease. We first make log transformations of the covariates serum bilirubin, albumin, serum cholesterol, prothrombine time following original publications (Fleming and Harrington (1991)). For the sake of simplicity, we will be considering the histologic state of disease to be numerically valued.

Two sets of analysis are carried out on the data. One is with only 6 covariates as in Li et al. (1999) and the other one with all 17 covariates.

We conduct the analysis with the 6 covariates first. Observations with missing data are discarded, leaving 308 observations. These covariates are z_1 =age, z_2 =presence of edema, z_3 =serum bilirubin, z_4 =albumin, z_5 =platelet count and z_6 =prothrombin time. Fleming and Harrington (1991) concluded that five baseline covariates—age, albumin, serum bilirubin, presence of edema and prothrombin time—are significant, and the true lifetime depends on x through the variable $\mathbf{Q} = 0.0333z_1 + 0.7847z_2 + 0.8792 \log z_3 - 3.0553 \log z_4 + 3.0157 \log z_6$. Using our proposed marginal predictor test, we identify covariates age, albumin, presence of edema and prothrombin time to be important. Survival time is independent of serum bilirubin (platelet count) given the other covariates. Different from previous results, serum bilirubin is not significant after adjusting for other variables. The dimension tests indicate that the central subspace dimension is two. Specifically, starting from $d = 1$, the test statistic is larger than the bootstrap critical value and we reject the null hypothesis. We do not reject the null when testing $d = 2$. Li et al. (1999) performed SIR separately for the failure time and the censoring time under the assumption that both the failure time and the censoring time are functions of the estimated predictors and an unknown error, while our approach is

independent of this model assumption. The two lifetime SIR directions obtained in Li et al. (1999) are (0.02, 0.90, 0.09, -0.62, -0.00, 0.38) and (0.03, -2.3, 0.20, -0.28, -0.00, -0.68). The basis estimates and bootstrap standard error of the corresponding covariates are given in Table 2. We can see that basis estimates from both approaches have higher coefficients for edema, albumin and prothrombin time but edema contributes less to the linear combination using our proposed procedure.

Now we redo the analysis with all 17 predictors. 276 cases remained after discarding observations with missing data. We perform a similar procedure to the one described above. Using the marginal predictor test, we identify the covariates age, serum bilirubin, albumin, prothrombin time, sex and spiders to be important. The dimension test of $d = d_0$ indicate that the central subspace dimension is two yet again. The basis estimates and bootstrap standard error of corresponding covariates when $d = 2$ are given below in Table 3. From the table, we find that some covariates such as edema have high coefficients even if they are not identified as significant using a marginal test. This is probably because they contribute very little marginally but have higher impact when entering jointly. Fitting a cox proportional hazards regression model, we also list the estimates in Table 3. Our basis estimates reflect low effects from age, serum bilirubin, platelet, copper, alkaline phosphatase, SGOT, triglycerides and serum cholesterol, which is consistent with Cox regression estimates.

6 Future research and additions

We have shown the asymptotic normality of the discrepancy function. Future theoretical derivation of the variance of this limiting distribution of the discrepancy function can potentially improve efficiency in estimation. Namely, we can set V_n in the discrepancy function equal to a consistent estimate of the inverse of the basis estimate's asymptotic variance $\Gamma_{\widehat{\zeta}}$. In the context of dimension determination and variable selection, approaches have been developed based on the bootstrap procedure. However, we can also potentially develop methods for central subspace dimension determination and variable selection using the theoretical variance $\Gamma_{\widehat{\zeta}}$, which could reduce the computational burden significantly. In addition, we are interested in developing a conditional predictor test of

$$P_{\mathcal{H}} S_{\tilde{Y}|Z} = O_p \quad \text{given} \quad d \quad \text{versus} \quad P_{\mathcal{H}} S_{\tilde{Y}|Z} \neq O_p \quad \text{given} \quad d.$$

Conditional predictor hypotheses should have greater power than the marginal tests if we know the true dimension of the central subspace. Conditional on the dimension of the central subspace being d, we can obtain the basis estimate $B_{p\times d}$. To determine if a predictor is significant, we can utilize the difference in minimum discrepancies to carry out conditional testing, i.e.,

$$T(\mathcal{H}|d) = n\widehat{F}_{d,\mathbf{H}} - n\widehat{F}_d,$$

which has a well-defined distribution. Currently, we have difficulties implementing the conditional test since the F value obtained under our setting does not follow a chi-square distribution, but is a mixture of chi-square distributions. This problem should be solved if we can obtain the true limiting variance $\Gamma_{\widehat{\zeta}}$ of the discrepancy function.

The goal of this paper is to augment current methodology for variable selection and for selecting significant predictors. This work should prove to be a useful tool that will aid in

analysis of survival data. An R (http://www.r-project.org) package is being developed for practical implementation of the entire proposed methodology.

Acknowledgments

Yingqi Zhao and Michael Kosorok were supported in part by NCI grant CA075142. The authors thank the referees for their very helpful comments that led to a significantly improved paper.

Appendix

A.1. Proof of Theorem 1

To prove consistency of the estimator for $G_Z(t)$, we first show that the weighted version of the Nelson-Aalen estimator is consistent. Since the weighted Kaplan-Meier can be re-expressed as a continuous functional of the Nelson-Aalen estimator, consistency of the Nelson-Aalen estimator will suffice. Therefore, we will show that $\widehat{\Lambda}_z^*(t)$, the weighted version of the Nelson-Aalen estimator of the cumulative hazard is consistent for $\Lambda(t)$.

Since the class $\{z \mapsto az^2+bz+c : a, b, c \in \mathbb{R}\}$ is a vector space of dimension 3, the class $\{z \mapsto \|z-u\|^2/h^2 : u \in \mathbb{R}, h>0\}$ is a VC class by Lemma 9.6 of Kosorok (2008). Since the square root function is monotone, the class $\{z \mapsto \|z-u\|/h : u \in \mathbb{R}, h>0\}$ is also VC by Lemma 9.9(viii) of Kosorok (2008). Therefore, we have $K(\|z-u\|/h)$ is Donsker and bounded since K is monotone and bounded. Thus both $K(\|z-z_i\|/h)\,N_i^c(t)$ and $K(\|z-z_i\|/h)\,Y_i^c(t)$ are Donsker since products of bounded Donsker classes are Donsker. Hence

$$(\mathbb{P}_n - P)\,K\left(\frac{\|z-Z\|}{h}\right)N^c(s) = O_p\left(n^{-1/2}\right),$$
$$(\mathbb{P}_n - P)\,K\left(\frac{\|z-Z\|}{h}\right)Y^c(s) = O_p\left(n^{-1/2}\right).$$

Let

$$(\mathbb{P}_n - P)\,h^{-d}K\left(\frac{\|z-Z\|}{h}\right)N^c(s) + Ph^{-d}K\left(\frac{\|z-Z\|}{h}\right)N^c(s) - E\left[N^c(s)\,|Z=z\right]f_Z(z) = I + II. \tag{15}$$

Now we evaluate each of I and II separately. I can be re-expressed as $I = O_p(n^{-1/2})h^{-d} = O_p(n^{-1/2}h^{-d})$. Note that if $\|z-u\| \geq h$, then $K(\|z-u\|/h) = 0$. We then have

$$\begin{aligned} Ph^{-d}K\left(\frac{\|z-Z\|}{h}\right)N^c(s) &= \int_{\|z-u\|\leq h} h^{-d}K\left(\frac{\|z-u\|}{h}\right)N^c(s)\,\mathrm{d}F(u) \\ &= \int_{\|z-u\|\leq h} h^{-d}K\left(\frac{\|z-u\|}{h}\right)E\left(N^c(s)\,|z=u\right)\mathrm{d}F(u) \\ &= \int_{\|z-u\|\leq h} h^{-d}K\left(\frac{\|z-u\|}{h}\right)P_u(C\leq t, T>C)\,f(u)\,\mathrm{d}u. \end{aligned} \tag{16}$$

Also,

$$\begin{aligned} E\left(N^c(s)\,|Z=z\right)f(z) &= P_z(C\leq t, T>C)\,f(z) \\ &= \int_{\|z-u\|\leq h} h^{-d}K\left(\frac{\|z-u\|}{h}\right)P_z(C\leq t, T>C)\,f(z)\,du. \end{aligned}$$

In order to obtain the rate of II, we proceed as follows:

$$\begin{aligned} II &\doteq \int_{\|z-u\|\le h} h^{-d} K\left(\tfrac{\|z-u\|}{h}\right) P_u(C\le t, T>C)\left(f(z)+\dot{f}(z)\|u-z\|\right) du \\ &- \int_{\|z-u\|\le h} h^{-d} K\left(\tfrac{\|z-u\|}{h}\right) P_z(C\le t, T>C) f(z)\, du \\ &= \int_{\|z-u\|\le h} h^{-d} K\left(\tfrac{\|z-u\|}{h}\right) (P_u(C\le t, T>C) - P_z(C\le t, T>C)) f(z)\, du \\ &+ \int_{\|z-u\|\le h} h^{-d} K\left(\tfrac{\|z-u\|}{h}\right) \|u-z\| \dot{f}(z)\, du \\ &\le \int_{\|z-u\|\le h} h^{-d} K\left(\tfrac{\|z-u\|}{h}\right) c\|z-u\|^{\gamma} du = O_p(h^{\gamma}). \end{aligned}$$

This implies that $I + II = O_p(n^{-1/2}h^{-d} + h^{\gamma})$. Using similar arguments as the ones used before, we can conclude that,

$$\mathbb{P}_n h^{-1} K\left(\frac{\|z-Z\|}{h}\right) Y^c(s) - E(Y^c(s)|Z=z) f(z) = O_p\left(n^{-1/2}h^{-d}+h^{\gamma}\right).$$

Therefore, by the Hadamard differentiability of the map $(A, B) \mapsto \int_0^t dA(s)/B(s)$,

$$\begin{aligned} \widehat{\Lambda}_z^*(t) - \Lambda_z(t) &= \int_0^t \frac{\mathbb{P}_n h^{-d} K(\|z-Z\|h)\quad \mathrm{d}N^c(s) - \mathrm{d}E[N^c(s)|Z=z] f_Z(z) + \mathrm{d}E[N^c(s)|Z=z] f_Z(z)}{\mathbb{P}_n h^{-d} K(\|z-Z\|/h) Y^c(s) - E[Y^c(s)|Z=z] f_Z(z) + E[Y^c(s)|Z=z] f_Z(z)} - \Lambda_z(t) \\ &= O_p\left(n^{-1/2}h^{-d}+h^{\gamma}\right). \end{aligned}$$

Hence, the estimator of the cumulative hazard is consistent. By applying the product integral to the Nelson-Aalen estimator, we obtain the Kaplan-Meier estimator. Since the product integral is Hadamard differentiable, the desired uniform consistency of the Kaplan-Meier follows (van der Vaart (1998) Theorem 20.8 and Lemma 20.14), i.e.

$$\|\widehat{L}_Z(s) - L_Z(s)\|_{\infty} = O_p\left(n^{-1/2}h^{-d}+h^{\gamma}\right).$$

By letting $n^{-1/2}dh^{-d-1} + rh^{\gamma-1} \sim 0$, we obtain the optional $h \sim n^{-1/2(d+\gamma)}$, with optimal convergence rate $n^{-\gamma/2(d+\gamma)}$.

A.2. Proof of Lemma 1

To show that the estimator of the survival function of T is consistent, we first prove consistency of the weighted Nelson-Aalen estimator. Let $\widehat{\Lambda}_T(t)$ be the estimator of the true cumulative hazard $\Lambda_T(t)$. Consider,

$$\widehat{\Lambda}_T(t) - \Lambda_T(t) = \int_0^t \frac{\sum \frac{dN_i(s)}{\widehat{L}_Z(s-)}}{\sum \frac{Y_i(s)}{\widehat{L}_Z(s-)}} - \Lambda_T(t). \tag{17}$$

Therefore we have,

$$\begin{aligned}\widehat{\Lambda}_T(t) - \Lambda_T(t) &= \int_0^t \frac{\sum \frac{dN_i(s)}{\widehat{L}_Z(-)} - \sum \frac{dN_i(s)}{\widehat{L}_Z(s-)} + \sum \frac{dN_i(s)}{L_Z(s-)}}{\sum \frac{Y_i(s)}{\widehat{L}_Z(s-)} - \sum \frac{Y_i(s)}{\widehat{L}_Z(s-)} + \sum \frac{Y_i(s)}{L_Z(s-)}} - \Lambda_T(t) \\ &= \int_0^t \frac{-\sum \frac{dN_i(s)\left(\widehat{L}_Z(s-) - L_Z(s-)\right)}{\widehat{L}_Z(s-)L_Z(s-)} + \sum \frac{dN_i(s)}{L_Z(s-)}}{-\sum \frac{Y_i(s)\left(\widehat{L}_Z(s-) - L_Z(s-)\right)}{\widehat{L}_Z(s-)L_Z(s-)} + \sum \frac{Y_i(s)}{L_Z(s-)}} - \Lambda_T(t).\end{aligned} \tag{18}$$

Since we have already proved the consistency of the weighted Kaplan-Meier estimator for C, the above form reduces to

$$\begin{aligned}\widehat{\Lambda}_T(t) - \Lambda_T(t) &= \int_0^t \frac{\sum \frac{dN_i(s)}{L_Z(-)}}{\sum \frac{Y_i(s)}{L_Z(-)}} - \Lambda_T(t) + O_p^{[0,\tau]}\left(n^{-\gamma/2(d+\gamma)}\right) \\ &= \int_0^t \frac{d\bar{N}}{\bar{Y}} - \frac{dN_0}{Y_0} + O_p^{[0,\tau]}\left(n^{-\gamma/2(d+\gamma)}\right),\end{aligned} \tag{19}$$

where $O_p^{[0,\tau]}$ is a quantity bounded in probability uniformly over $t \in [0, \tau]$, and where $\bar{N}(t) = \mathbb{P}_n\left[I(T \le t, T \le C)/L_z(t-)\right]$ and $\bar{Y}(t) = \mathbb{P}_n\left[I(X \ge t)/L_z(t-)\right]$ are respectively the weighted number of events and number at risk. Therefore, the above expression reduces to

$$\int_0^t \frac{d\bar{N} - dN_0}{\bar{Y}} - \frac{dN_0\left(\bar{Y} - Y_0\right)}{\bar{Y}\,Y_0} + O_p^{[0,\tau]}\left(n^{-\gamma/2(d+\gamma)}\right), \tag{20}$$

where, $\bar{N}(t) - N_0(t)$ can be written as

$$(\mathbb{P}_n - P)\left[\int_0^t \frac{I(C \ge s)\,dG(s)}{L_Z(s-)}\right]. \tag{21}$$

Note that $I(C \ge T)$ and $I(T \le t)$ belong to Donsker classes. $L_Z(s)$ is a Lipschitz continuous function and therefore bounded. We can thereby argue that $\bar{N} - N_0$ can be represented as $\phi(\bar{N}, L_Z)$, where $\phi(H, L_Z) = \int_0^t dH/L_Z$. Since the standard Nelson-Aalen estimator for censored data is $\sqrt{n}$ consistent, and ϕ is Hadamard-differentiable, we can apply the functional delta method to this functional, and thus obtain $\sqrt{n}$ consistency for $\bar{N}$. In an identical fashion we can argue that $\bar{Y} - Y_0$ is also $\sqrt{n}$ consistent. Hence, the weighted estimator of the cumulative hazard based on known $L_Z(t-)$ is $\sqrt{n}$ consistent. We obtain the Kaplan-Meier by applying the product integral to the Nelson-Aalen estimator. Since the product integral is again Hadamard differentiable (van der Vaart (1998)), the weighted Kaplan-Meier estimator is $n^{-\gamma/2(d+\gamma)}$ consistent for the true survival function of T. Hence, $\hat{f}_{u_j} = \hat{S}(t_{j+1}) - \hat{S}(t_j)$, $j = 1, \ldots, h$, is also $O_p(n^{-\gamma/2(d+\gamma)})$ consistent for f_{u_j}. Therefore, we have $\widehat{P}\left(Y \in u_y\right) - P\left(Y \in u_y\right) = O_p^{[0,\tau]}\left(n^{-\gamma/2(d+\gamma)}\right)$.

A.3. Proof of Lemma 2

Consider,

$$\mathbb{P}_n\left[\frac{Z\delta I\left(Y\in u_y\right)}{\widehat{L}_Z(Y-)\widehat{P}\left(Y\in u_y\right)}\right]=\mathbb{P}_n\left[\frac{Z\delta I\left(Y\in u_y\right)}{\widehat{L}_Z(Y-)P\left(Y\in u_y\right)}\times\frac{P\left(Y\in u_y\right)}{\widehat{P}\left(Y\in u_y\right)}\right]. \tag{22}$$

Now, we have ,

$$\begin{aligned}\frac{P(Y\in u_y)}{\widehat{P}(Y\in u_y)} &= \frac{P(Y\in u_y)}{P(Y\in u_y)+O_p\left(n^{-\gamma/2(d+\gamma)}\right)}\\ &= \left[1+\frac{O_p\left(n^{-\gamma/2(d+\gamma)}\right)}{P(Y\in u_y)}\right]^{-1}\\ &= 1+O_p\left(n^{-\gamma/2(d+\gamma)}\right).\end{aligned}$$

Therefore the equation (22) reduces to

$$\mathbb{P}_n\left[\frac{Z\delta I\left(Y\in u_y\right)}{\widehat{L}_Z(Y-)P\left(Y\in u_y\right)}\right]\left(1+O_p\left(n^{-\gamma/2(d+\gamma)}\right)\right). \tag{23}$$

Since $\mathcal{A}$ is a VC class, $\left[\frac{Z\delta I\left(Y\in u_y\right)}{\widehat{L}_Z(Y-)P\left(Y\in u_y\right)}\right]$ is eventually contained in VC class. Hence, the above form reduces to

$$P\left[\frac{Z\delta I\left(Y\in u_y\right)}{L_Z(Y-)P\left(Y\in u_y\right)}\right]+O_p\left(n^{-\gamma/2(d+\gamma)}\right). \tag{24}$$

Now, consider,

$$\begin{aligned}P\left[\frac{Z\delta I(Y\in u_y)}{L_Z(Y-)P(Y\in u_y)}\right] &= P\left[\frac{Z}{P(Y\in u_y)}E\left[\frac{\delta I(Y\in u_y)}{L_Z(Y-)}\middle|Z\right]\right]\\ &= P\left[\frac{Z}{L_Z(Y-)P(Y\in_y)}E\left[\delta I\left(Y\in u_y\right)|Z\right]\right]\\ &= E\left[\frac{\delta Z}{L_Z(Y-)}\middle|I\left(Y\in u_y\right)\right].\end{aligned} \tag{25}$$

We have $\delta = I(C \geq T)$, and hence, $E[\delta Z/L_Z(Y-)|Y\in u_y]$ can be re-expressed as:

$$E\left[\frac{ZP\left(Y\in u_y|Z\right)}{P\left(Y\in u_y\right)}\right]=P\left(Z|Y\in u_y\right). \tag{26}$$

So, we can conclude that,

$$\mathbb{P}_n\left[\frac{Z\delta I\left(Y\in u_y\right)}{\widehat{L}_Z(Y-)\widehat{P}\left(Y\in u_y\right)}\right]=P\left[Z|Y\in u_y\right]+O_p\left(n^{-\gamma/2(d+\gamma)}\right). \tag{27}$$

Using similar but simpler arguments we can say the same when $u_y = (\tau, \infty)$. Hence we can conclude that $\bar{Z}_{y.}$ is consistent for $E[Z|Y]$.

A.4. Proof of Theorem 2

Let μ be the expected value of $\bar{Z}_{..}$ and μ_y be the expected value of $\bar{Z}_{y.}$. Let $\tilde{Z}$ be the standardized value of Z and ϵ^* the residual from the weighted regression of J_y on $\tilde{Z}$.

Consider,

$$\begin{aligned}
\widehat{f_{u_y}}\widehat{\xi_{u_y}} - f_{u_y}\xi_{u_y} &= \widehat{f_{u_y}}\widehat{\Sigma}^{-1}\left(\bar{Z}_{y.} - \bar{Z}_{..}\right) - f_{u_y}\Sigma^{-1}\left(\mu_y - \mu\right) \\
&= \widehat{f_{u_y}}\widehat{\Sigma}^{-1}\left(\bar{Z}_{y.} - \bar{Z}_{..}\right) - \widehat{f_{u_y}}\Sigma^{-1}\left(\mu_y - \mu\right) + \widehat{f_{u_y}}\Sigma^{-1}\left(\mu_y - \mu\right) - f_{u_y}\Sigma^{-1}\left(\mu_y - \mu\right) \\
&= \widehat{f_{u_y}}\left[\widehat{\Sigma}^{-1}\left(\bar{Z}_{y.} - \bar{Z}_{..}\right) - \Sigma^{-1}\left(\mu_y - \mu\right)\right] + \Sigma^{-1}\left(\widehat{f_{u_y}} - f_{u_y}\right)\left(\mu_y - \mu\right) \\
&= \left(\widehat{f_{u_y}} - f_{u_y}\right)\left[\widehat{\Sigma}^{-1}\left(\bar{Z}_{y.} - \bar{Z}_{..}\right) - \Sigma^{-1}\left(\mu_y - \mu\right)\right] + f_{u_y}\left[\widehat{\Sigma}^{-1}\left(\bar{Z}_{y.} - \bar{Z}_{..}\right) - \Sigma^{-1}\left(\mu_y - \mu\right)\right] + \Sigma^{-1}\left(\widehat{f_{u_y}} - f_{u_y}\right)\left(\mu_y - \mu\right) \\
&= O_p\left(n^{-\gamma/2(d+\gamma)}\right)
\end{aligned} \tag{28}$$

Therefore, using arguments similar to those in Hall (1991), we can claim that the limiting distribution of $vec\left(\widehat{\xi}D_{\widehat{f}}\right)$ is asymptotically normal with rate $n^{-}\gamma/2(d+\gamma)$. Hence, we can further claim that the limiting distribution of $\hat{F}_d$, the discrepancy function, is a mixture of chi-squared distributions with the same rate.

A.5. Proof of Theorem 3

To prove this theorem, we make use of Proposition 3.1 and 4.1 in Shapiro (1986). Shapiro's results are applicable for fixed V , and thus we need to modify for when V is random. We use Cook and Ni's results for random V to show that the results hold. Lemma A.3 in Cook and Ni (2005) permits connecting minimum discrepancy functions with fixed inner products to those with random inner products. We can then claim that the basis estimate is consistent for the true value, and, provided we use a consistent estimate for V , the asymptotic properties of the discrepancy function are preserved. The desired results now follow since the minimization of F_d always provides a consistent estimate of vec(βv) for any sequence $V_n > 0$ that converges to $V > 0$.

References

Cook RD. Graphics for regressions with a binary response. J. of American Statistical Association. 1996; 91:983–992.

Cook, RD. Regression Graphics. Wiley; New York: 1998.

Cook RD. Testing predictor contributions in sufficient dimension reduction. Annals of Statistics. 2004; 32:1062–1092.

Cook RD, Ni L. Sufficient dimension reduction via inverse regression: A minimum discrepancy approach. J. of American Statistical Association. 2005; 100:410–428.

Cox DR. Regression Models and Life-Tables (with discussion). J. of Royal Statistical Society. 1972; 34:187–202.

Dabrowska DM. Uniform consistency of the kernel conditional Kaplan-Meier estimate. Annals of Statistics. 1989; 17:1157–1167.

Diaconis P, Freedman D. Asymptotics of graphical regression pursuit. Annals of Statistics. 1984; 12:793–815.

Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. Annals of Statistics. 2002; 30:74–99.

Fleming, TR.; Harrington, DP. Counting processes and Survival Analysis. Wiley; New York: 1991.

Hall P. On converging rates of suprema. Probability Theory and Related Fields. 1991; 89:447–455.

Hall P, Li KC. On almost linearity of low dimensional projections from high dimensional data. Annals of Statistics. 1993; 21:867–889.

Keles S, van der Laan M, Dudoit S. Asymptotic optimal model selection method with right censored outcomes. Bernoulli. 2004; 6:1011–1037.

Kosorok, MR. Introduction to empirical processes and semiparametric inference. Springer-Verlag; New York: 2008.

Kosorok MR, Song R. Inference under right censoring for transformation models with a change-point based on a covariate threshold. Annals of Statistics. 2007; 35:957–989.

Li KC. Sliced inverse regression for dimension reduction. Annals of Statistics. 1991; 86:316–342.

Li KC, Wang J-L, Chen C-H. Dimension reduction for censored regression data. Annals of Statistics. 1999; 27:1–23.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large B-cell lymphoma. New England J. of Medicine. 2002:1937–1947.

Rotnitzky A, Robbins J. Inverse probability weighted estimation in survival analysis. IPW–Survival Encyclopedia. 2003

Shapiro A. Asymptotic theory of overparametrized structural models. J. of American Statistical Association. 1986; 81:142–149.

Tibshirani R. The lasso method for variable selection in the Cox model. Statistics in Medicine. 1997; 16:385–395. [PubMed: 9044528]

van der Vaart, A. Asymptotic Statistics. Cambridge University Press; New York: 1998.

Zhang HH, Lu W. Adaptive Lasso for Cox's proportional hazards model. Biometrika. 2007; 94:1–13.

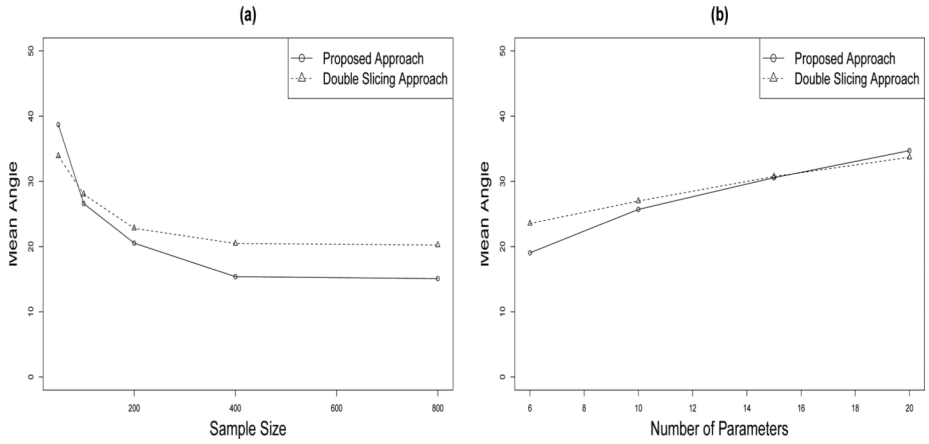

Figure 1.

Mean Angles between S_ξ and both the SIR estimate and proposed procedures under 100 simulation runs of Model 1: (a) Different sample sizes and (b) Different numbers of parameters.

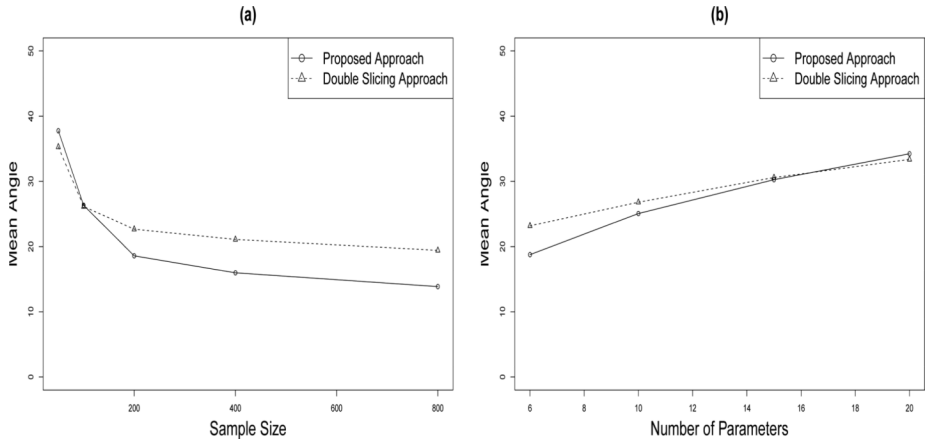

Figure 2.

Mean Angles between S_ξ and both the SIR estimate and proposed procedures under 100 simulation runs when z_1 is Rademacher distributed: (a) Different sample sizes and (b) Different numbers of parameters.

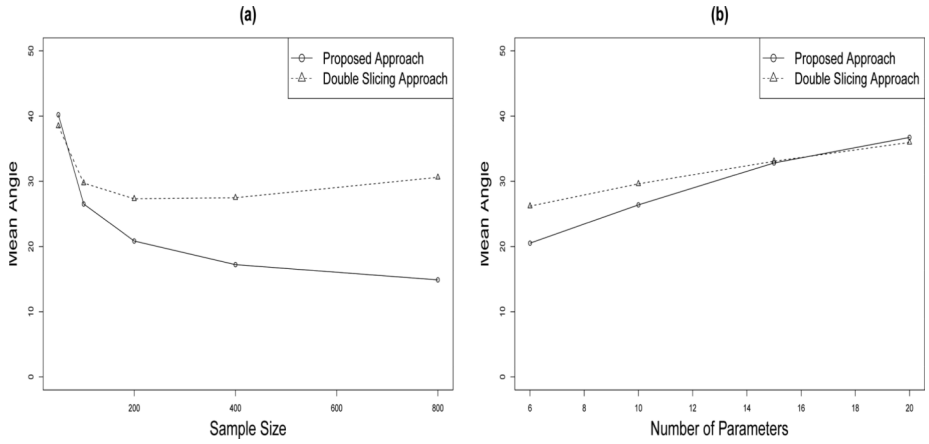

Figure 3.

Mean Angles between S_ξ and both the SIR estimate and proposed procedures under 100 simulation runs when z_2 is Rademacher distributed: (a) Different sample sizes and (b) Different numbers of parameters.

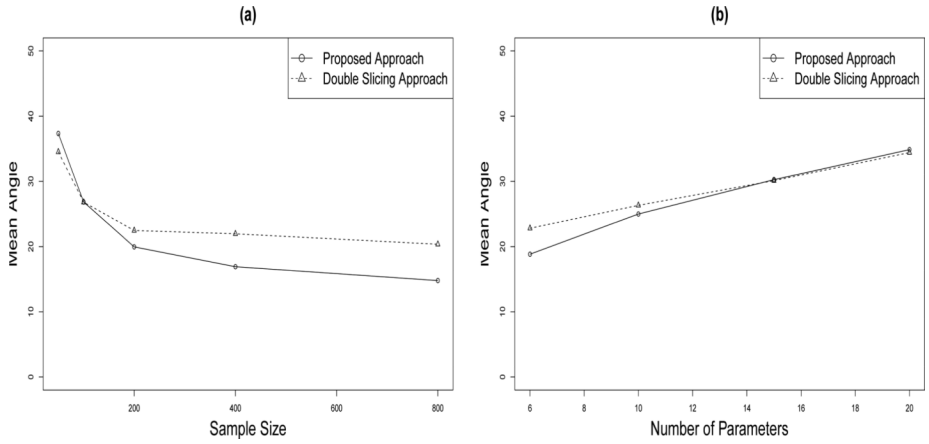

Figure 4.

Mean Angles between S_ξ and both the SIR estimate and proposed procedures under 100 simulation runs when z_4 is Rademacher distributed: (a) Different sample sizes and (b) Different numbers of parameters.

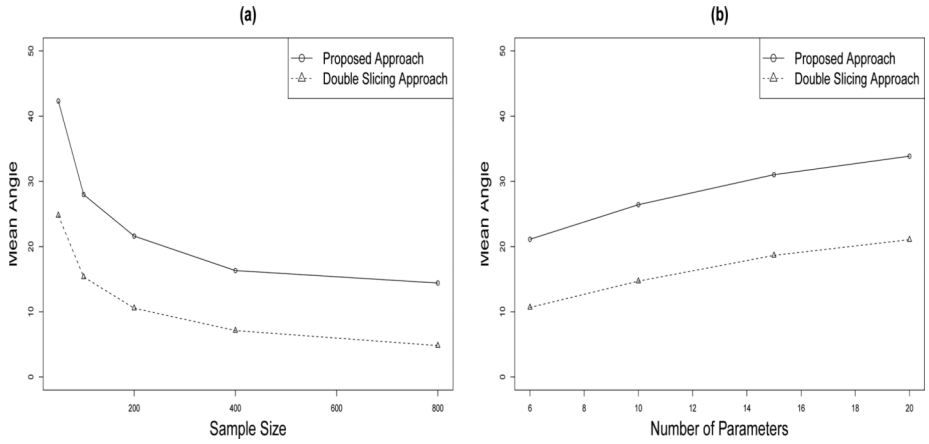

Figure 5.
Mean Angles between S_ξ and both the SIR estimate and proposed procedures under 100 simulation runs of Model 2 for different sample sizes: (a) Different sample sizes and (b) Different numbers of parameters.

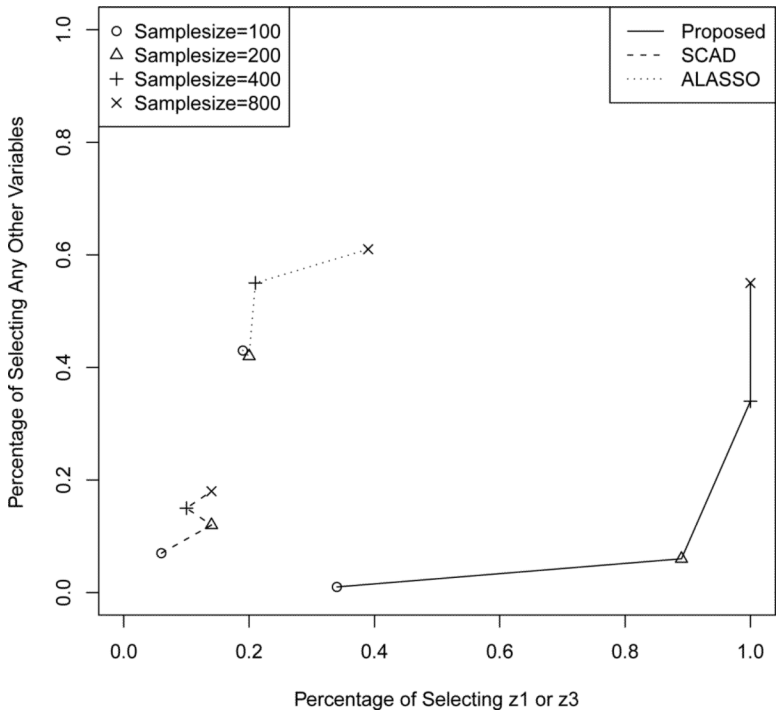

Figure 6.
Percentage of selecting significant/non-significant variables for different sample sizes using SCAD, ALASSO and the proposed method

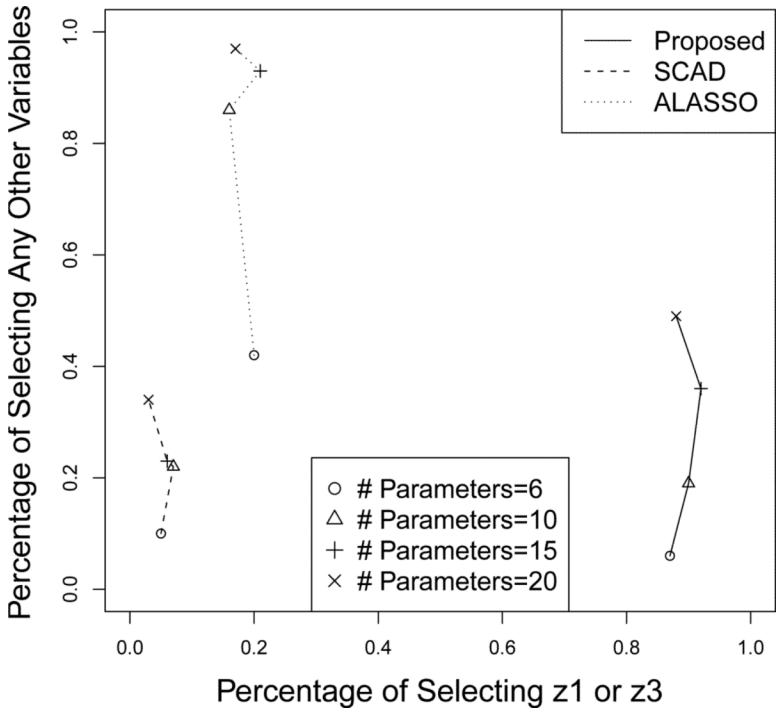

Figure 7.
Percentage of selecting significant/non-significant variables for different numbers of parameters using SCAD, ALASSO and the proposed method

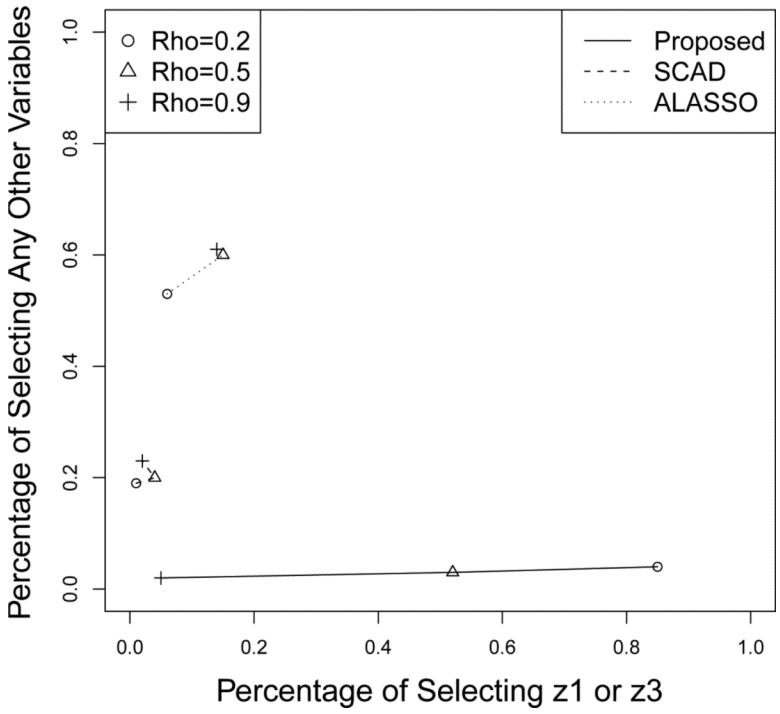

Figure 8.
Percentage of selecting significant/non-significant variables for different correlations between predictors using SCAD, ALASSO and the proposed method

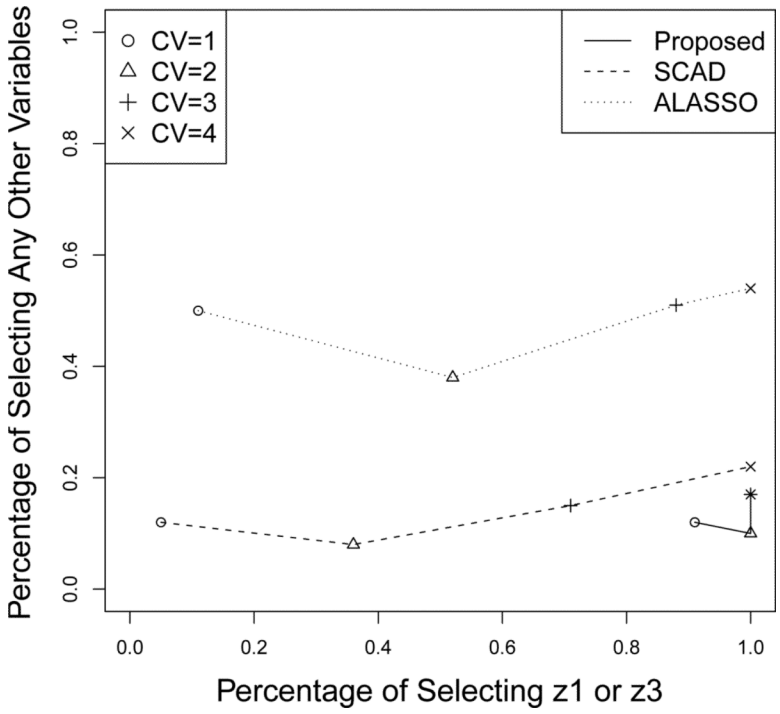

Figure 9.
Percentage of selecting significant/non-significant variables for different CVs using SCAD, ALASSO and the proposed method

Table 1

Estimates of the basis for $d = 2$ for the DLBCL data. Bootstrap standard errors are given in parentheses.

Basis estimate	Covariate
0.020(0.537)	ABC
0.029(0.640)	GCB
-0.251(0.161)	B-cell sig.
-0.212(0.152)	Lymph sig.
0.201(0.267)	Prolif. sig.
0.267(0.216)	BMP6
-0.266(0.187)	MHC sig.
-0.842(0.248)	Out.pred.score

Table 2

Estimates of the basis for $d = 2$ for the PBC data with 6 original covariates. Bootstrap standard errors are given in parentheses.

Basis estimate 1	Basis estimate 2	Covariate
0.012(0.018)	-0.004(0.023)	Age
-0.807(0.300)	-0.076(0.363)	Edema
0.034(0.047)	-0.000(0.056)	Serum bilirubin
0.486(0.511)	0.473(0.646)	Albumin
0.003(0.065)	0.000(0.061)	Platelet
-0.332(0.532)	0.878(0.587)	Prothrombin time

Table 3

Estimates of the basis for d =2 for the PBC data with 17 original covariates, with bootstrap standard errors given in parentheses, along with estimates (standard errors) using Cox regression.

Basis estimate 1	Basis estimate 2	Cox Model Estimate	Covariate
-0.015(0.066)	-0.027(0.015)	0.025(0.010)	Age
-0.025(0.112)	0.266(0.338)	0.711(0.460)	Edema
0.085(0.010)	0.173(0.029)	0.072(0.166)	Serum bilirubin
-0.313(0.230)	0.799(0.674)	2.651(1.030)	Albumin
0.033(0.009)	0.049(0.017)	0.002(0.001)	Platelet
-0.880(0.269)	-0.306(0.268)	0.743(1.270)	Prothrombin time
0.137(0.048)	-0.115(0.199)	0.268(0.203)	Treatment
-0.031(0.080)	0.121(0.291)	0.987(0.431)	Sex
0.039(0.010)	0.059(0.020)	0.001(0.001)	Copper
0.044(0.011)	0.064(0.021)	0.000(0.000)	Alkaline phosphatase
0.044(0.011)	0.064(0.021)	-0.001(0.002)	SGOT
0.043(0.011)	0.062(0.022)	-0.002(0.002)	Triglycerides
0.048(0.012)	-0.072(0.023)	-0.001(0.001)	Serum cholesterol
0.150(0.033)	-0.115(0.122)	0.137(0.136)	Histologic stage
0.139(0.124)	0.145(0.367)	0.610(0.469)	Ascites
0.003(0.051)	0.256(0.194)	-0.207(0.226)	Hepatomegaly
-0.216(0.057)	-0.124(0.214)	0.158(0.236)	Spiders