Journal of Statistical Planning and Inference 139 (2009) 734 – 748

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

New approaches to model-free dimension reduction for bivariate regression

Xuerong Meggie Wen[a,*], R. Dennis Cook[b]

[a]*Department of Mathematics and Statistics, University of Missouri, Rolla, MI 65409, USA*
[b]*School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA*

ARTICLE INFO

Article history:
Received 7 December 2006
Received in revised form
19 July 2007
Accepted 31 January 2008
Available online 11 June 2008

Keywords:
Bivariate dimension reduction
Central subspaces
Intra-slice information
Testing predictor effects
Censoring regression

ABSTRACT

Dimension reduction with bivariate responses, especially a mix of a continuous and categorical responses, can be of special interest. One immediate application is to regressions with censoring. In this paper, we propose two novel methods to reduce the dimension of the covariates of a bivariate regression via a model-free approach. Both methods enjoy a simple asymptotic chi-squared distribution for testing the dimension of the regression, and also allow us to test the contributions of the covariates easily without pre-specifying a parametric model. The new methods outperform the current one both in simulations and in analysis of a real data. The well-known PBC data are used to illustrate the application of our method to censored regression.

1. Introduction

The overarching goal of sufficient dimension reduction (SDR) in a regression analysis with response Y and a vector of random predictors $\mathbf{X}=(X_1,\ldots,X_p)^{\mathrm{T}}\in\mathbb{R}^p$ is to find a parsimonious characterization of the conditional distribution of $Y|\mathbf{X}$ without requiring a parametric model. We pursue this goal by seeking *sufficient predictors*, $\boldsymbol{\eta}_j^{\mathrm{T}}\mathbf{X}$, $j=1,\ldots,d$, a minimal set of linear combinations of $\mathbf{X}$, such that $Y|(\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X},\ldots,\boldsymbol{\eta}_d^{\mathrm{T}}\mathbf{X})$ has the same distribution as $Y|\mathbf{X}$ for all values of $\mathbf{X}$. More formally, we search for subspaces $\mathscr{S}\subseteq\mathbb{R}^p$ such that

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

where $\perp\!\!\!\perp$ indicates independence and $P_{(.)}$ stands for a projection operator with respect to the standard inner product. Such an $\mathscr{S}$ is called a *dimension reduction subspace*. While a dimension reduction subspace always exists, it is not necessarily unique. Ideally we should search for the minimal dimension reduction subspace, the one with the smallest dimension. Under mild conditions (Cook, 1998) that almost always hold in practice, the minimal dimension reduction subspace is uniquely defined and coincides with the intersection of all dimension reduction subspaces. This intersection is called the *central subspace* (CS; Cook, 1994, 1998) of the regression and denoted as $\mathscr{S}_{Y|\mathbf{X}}$. The dimension of $\mathscr{S}_{Y|\mathbf{X}}$ is called the *structural dimension* of the regression; let $d=\dim(\mathscr{S}_{Y|\mathbf{X}})$.

In practice, d is often at most three and this allows a fully informative and direct visualization of the original regression through a plot of Y versus the estimated sufficient predictors. In this sense, SDR can facilitate data visualization and model building. And unlike other nonparametric approaches, SDR can often avoid the curse of dimensionality (Friedman, 1994). Many SDR methods enjoy $\sqrt{n}$ convergence rates since they exploit the global features of the dependence of Y on $\mathbf{X}$.

* Corresponding author.
E-mail address: wenx@umr.edu (X.M. Wen).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.01.017

In the past decade, many methods have been developed to estimate $\mathscr{S}_{Y|\mathbf{X}}$ when the response is univariate, but there is relatively little methodology available when the response is multivariate. In this paper, we focus on investigating estimation methods when $\mathbf{Y}$ is bivariate. We assume that the bivariate central subspace exists throughout this article.

There are many important regression settings where a bivariate response arises naturally, particularly regressions involving censored data. Li et al. (1999) discussed dimension reduction methods that allows for censoring. Cook (1995, 2003) presented a method called bivariate SIR that is applicable with bivariate responses. Li and Li (2004) applied bivariate SIR to study gene microarrays in the context of censored survival data.

For bivariate responses we consider three cases: both responses are categorical, one is categorical and the other is continuous, both are continuous. Slicing a continuous response (Li, 1991), which is a standard methodology in SDR, converts the second and third cases to the double categorical case where bivariate SIR seems to be the only methodological option that has been developed.

Cook and Ni (2006) recently pointed out that intra-slice information is lost when converting a continuous response to a categorical one, and they developed methodology for recovering intra-slice information in univariate regressions. In this paper we present two new dimension reduction methods for bivariate responses, both designed to recover intra-slice information when at least one response is continuous. One method, *bivariate estimation across responses* (BEAR), is mainly designed for bivariate regressions with one continuous response and one categorical response. The other method, *balanced bivariate estimation* (BBE), is for the cases with two continuous responses. Both methods allow us to test predictor effects easily, a simple chi-squared distribution can be used in inference methods for d and they both have optimal properties to be described later.

The rest of this paper is organized as follows. In Section 2 we first review the two major approaches to estimation in SDR. We then discuss dimension reduction for regressions with bivariate responses. A brief review of the existing method, bivariate SIR, is presented. In Section 3, we propose BEAR. Its asymptotics are also discussed. Section 4 is dedicated to the discussion of BBE and its asymptotic properties. Application of our methods to survival regression data is studied in Section 5. The performances of BEAR and BBE are compared to that of bivariate SIR via simulation studies in Section 6. An illustration is given via a real data analysis. Brief conclusions and a discussion on future research directions are given in Section 7. We assume the usual SDR linearity and coverage conditions throughout; see Cook and Ni (2005) for a discussion of these conditions in the context of this article.

2. Dimension reduction in regressions with bivariate responses

2.1. Two major approaches

Many SDR methods, such as sliced inverse regression (SIR; Li, 1991), sliced average variance estimation (SAVE; Cook and Weisberg, 1991; Cook, 2000), graphical regression (Cook, 1994, 1998), and parametric inverse regression (Bura and Cook, 2001) are all based on the following logic. Let $\mathbf{\Sigma} = \mathrm{Cov}(\mathbf{X})$. First, working with the standardized variable $\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \mathrm{E}(\mathbf{X}))$, find a symmetric population kernel matrix $\mathbf{M}$, which satisfies the property that $\mathrm{Span}(\mathbf{M}) \subseteq \mathscr{S}_{Y|\mathbf{Z}}$. This restriction to symmetry involves no loss of generality. Then, spectrally decompose $\widehat{\mathbf{M}}$, a consistent estimate of $\mathbf{M}$, and use the span of the eigenvectors corresponding to the d largest eigenvalues of $\widehat{\mathbf{M}}$ to estimate $\mathrm{Span}(\mathbf{M})$. The eigenvalues provide a test statistic for hypotheses on the structural dimension, and the eigenvectors can be linearly transformed back to the $\mathbf{X}$-scale. This is called the spectral decomposition approach since it is based on a spectral decomposition of the sample kernel matrix $\widehat{\mathbf{M}}$.

Cook and Ni (2005) introduced a novel approach based on minimizing the following quadratic discrepancy function:

$$F_d(\mathbf{B},\mathbf{C}) = (\mathrm{vec}(\hat{\boldsymbol{\zeta}}) - \mathrm{vec}(\mathbf{BC}))^{\mathrm{T}}\mathbf{V}_n(\mathrm{vec}(\hat{\boldsymbol{\zeta}}) - \mathrm{vec}(\mathbf{BC})), \tag{2.1}$$

where $\hat{\boldsymbol{\zeta}} \in \mathbb{R}^{p\times l}$ is the data matrix, $\mathbf{B} \in \mathbb{R}^{p\times d}$, $\mathbf{C} \in \mathbb{R}^{d\times l}$, $l > d$, and $\mathbf{V}_n \in \mathbb{R}^{pl\times pl} > 0$. Letting $(\hat{\mathbf{B}},\hat{\mathbf{C}}) = \arg\min F_d$, $\mathrm{Span}(\hat{\mathbf{B}})$ is a consistent estimator of $\mathscr{S}_{Y|\mathbf{X}}$ for any $\mathbf{V}_n > 0$ that converges to a positive definite matrix $\mathbf{V}$. This is called the minimum discrepancy approach (MDA). They also showed that many current methods including SIR belong to a suboptimal class of this family.

2.2. Bivariate SIR

In this section, we will study regressions with bivariate responses of the form (Y,U), where Y and U are scalars. Analogous to the definition of the central subspace, the bivariate central subspace $\mathscr{S}_{(Y,U)|\mathbf{X}}$ is defined as the intersection of all subspaces $\mathscr{S}$ satisfying the conditional independent condition $(Y,U) \perp\!\!\!\perp \mathbf{X}|P_{\mathscr{S}}\mathbf{X}$.

Although SIR was introduced in the context of regression problems with a univariate response variable, the same theory applies when the response variable is bivariate. Cook (1995) proposed a method called bivariate SIR based on double slicing the observations on (Y,U) to form a discrete bivariate response $(\tilde{Y},\tilde{U})$, where $\tilde{U}$ takes value in $\{1,\ldots,K\}$, and for each value of $\tilde{U}$, $\tilde{Y}$ takes value in $\{1,\ldots,h_{\tilde{U}}\}$. Hence, the discrete response $(\tilde{Y},\tilde{U})$ is constructed by first slicing on U and then slicing on Y within each value of $\tilde{U}$. Let $h = \sum_{u=1}^{K} h_u$.

Letting $J_u(U) = I\{\tilde{U}=u\}$, $f_u = \Pr(J_u = 1)$, $J_{uy}(Y,U) = I\{\tilde{U}=u, \text{ and } \tilde{Y}=y\}$, and $f_{uy} = \Pr(J_{uy}=1)$. Define $\boldsymbol{\xi}_{uy} = \mathbf{\Sigma}^{-1}\,\mathrm{Cov}(\mathbf{X}, J_{uy})$, $u = 1,\ldots,K$, and $y = 1,\ldots,h_u$. Starting with a random sample $(\mathbf{X}_i, Y_i, U_i)$, $i = 1,\ldots,n$, on $(\mathbf{X},Y,U)$, define $\hat{\boldsymbol{\xi}}_{uy}$ as $\mathbf{X}$ coefficients from the ordinary least-squares fit of $J_{uy}(Y_i,U_i)$ on $\mathbf{X}_i$, including an intercept.

Following the MDA, we now write a nonlinear objective function that reproduces the methodology of bivariate SIR:

$$F_d^{\text{bsir}}(\mathbf{B},\mathbf{C}) = \sum_{u=1}^{K}\sum_{y=1}^{h_u}(\hat{f}_{uy}\hat{\xi}_{uy} - \mathbf{B}\mathbf{C}_{uy})^{\mathrm{T}}\hat{f}_{uy}^{-1}\hat{\boldsymbol{\Sigma}}(\hat{f}_{uy}\hat{\xi}_{uy} - \mathbf{B}\mathbf{C}_{uy}),$$

where $\mathbf{B}\in\mathbb{R}^{p\times d}$, $\mathbf{C}_{uy}\in\mathbb{R}^d$, $\hat{f}_{uy}$ is the fraction of the sample points in slice (y,u), $u=1,\dots,K$, $y=1,\dots,h_u$, and $\hat{\boldsymbol{\Sigma}}$ is the sample version of $\boldsymbol{\Sigma}$. Here and in what follows we use $\hat{F}$ to denote the value of an objective function minimized over its arguments. For instance, $\hat{F}_d^{\text{bsir}} = F_d^{\text{bsir}}(\hat{\mathbf{B}},\hat{\mathbf{C}})$. Span$(\hat{\mathbf{B}})$ provides a consistent estimate of $\mathscr{S}_{(Y,U)|\mathbf{X}}$ and $n\hat{F}_m^{\text{bsir}}$ can be used to test hypothesis $d=m$ versus $d>m$.

3. Bivariate estimation across responses

Define

$$\boldsymbol{\eta}_w = \boldsymbol{\Sigma}^{-1}\operatorname{Cov}(\mathbf{X},J_w(U)),\quad w=1,\dots,K, \tag{3.1}$$

$$\boldsymbol{\beta}_{ws} = \boldsymbol{\Sigma}^{-1}\operatorname{Cov}(\mathbf{X},YJ_{ws}(Y,U)),\quad s=1,\dots,h_w, \tag{3.2}$$

where U is categorical. One special case is survival data where U takes two values and is constant within Y slices. Detailed discussion on this case is given in Section 5.

Most SDR methods require a *linearity condition* on the marginal distribution of $\mathbf{X}$. In the present context, the linearity condition says that $\mathrm{E}(\mathbf{X}|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X})$ must be a linear function of $\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}$, where the columns of $\boldsymbol{\rho}\in\mathbb{R}^{p\times d}$ form a basis for the bivariate central subspace. When $\mathbf{X}$ follows an elliptically contoured distribution like the multivariate normal, the linearity condition always holds (Eaton, 1986). Diaconis and Freedman (1984) showed that most lower dimension projections of high-dimensional data are close to normal. Hall and Li (1993) showed that as p increases with d fixed, the linearity condition holds to a good approximation in many problems. Shao et al. (2007) argued that for single-index regression model, the linearity condition seems to substitute for knowing the exact conditional distribution of $Y|\mathbf{X}$. In practice, we are free to use experimental design, and one-to-one predictor transformations to induce the linearity condition. Cook and Nachtsheim (1994) proposed a re-weighting method to force this condition when necessary without suffering complications when inferring about $Y|\mathbf{X}$. Since no model is assumed for $Y|\mathbf{X}$, these methods will not change the fundamental issues in the regression. Additional discussion of this condition, which is typically regarded as mild with quantitative predictors, was given by Cook and Ni (2005). Under the linearity condition it is easy to prove that

$$\boldsymbol{\eta}_w\in\mathscr{S}_{(Y,U)|\mathbf{X}}\quad\text{and}\quad\boldsymbol{\beta}_{ws}\in\mathscr{S}_{(Y,U)|\mathbf{X}}. \tag{3.3}$$

Taking a step further, we then assume the common *coverage condition* (for background, see Cook and Ni, 2005; Wen and Cook, 2007), that $\mathscr{S}_{(Y,U)|\mathbf{X}} = \text{Span}(\boldsymbol{\eta}_1,\dots,\boldsymbol{\eta}_K,\boldsymbol{\beta}_{11},\dots,\boldsymbol{\beta}_{Kh_K})$. Letting $\boldsymbol{\eta}=(\boldsymbol{\eta}_1,\dots,\boldsymbol{\eta}_K)$, and $\boldsymbol{\beta}=(\boldsymbol{\beta}_{11},\dots,\boldsymbol{\beta}_{Kh_K})$, we can always find $\boldsymbol{\gamma}\in\mathbb{R}^{d\times(h+K)}$ such that $(\boldsymbol{\eta},\boldsymbol{\beta})=\boldsymbol{\rho}\boldsymbol{\gamma}$.

Following Cook and Ni (2006), we divide $\boldsymbol{\beta}_{ws}$ into two parts:

$$\boldsymbol{\beta}_{ws} = f_{ws}\boldsymbol{\Sigma}^{-1}\operatorname{Cov}(\mathbf{X},Y|J_{ws}=1) + f_{ws}\mathrm{E}(Y|J_{ws}=1)\boldsymbol{\xi}_{ws}. \tag{3.4}$$

Eq. (3.4) shows us how the intra-slice information is being recovered for the continuous variable Y with BEAR, by using the intra-slice covariance between Y and $\mathbf{X}$. Let $\hat{\boldsymbol{\eta}}_w$ and $\hat{\boldsymbol{\beta}}_{ws}$ be the $\mathbf{X}$ coefficients from the ordinary least-squares fits of $J_w(U_i)$ and $Y_iJ_{ws}(Y_i,U_i)$ on $\mathbf{X}_i$, respectively, including an intercept. Letting $\hat{\boldsymbol{\eta}}=(\hat{\boldsymbol{\eta}}_1,\dots,\hat{\boldsymbol{\eta}}_K)$, and $\hat{\boldsymbol{\beta}}=(\hat{\boldsymbol{\beta}}_{11},\dots,\hat{\boldsymbol{\beta}}_{Kh_K})$, we then estimate $(\boldsymbol{\rho},\boldsymbol{\gamma})$ by minimizing the following quadratic discrepancy function:

$$(\operatorname{vec}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}}) - \operatorname{vec}(\mathbf{B}\mathbf{C}))^{\mathrm{T}}\mathbf{V}_n(\operatorname{vec}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}}) - \operatorname{vec}(\mathbf{B}\mathbf{C})), \tag{3.5}$$

where $\mathbf{B}\in\mathbb{R}^{p\times d}$, $\mathbf{C}\in\mathbb{R}^{d\times(h+K)}$, and $\mathbf{V}_n>0$ depends on the specific method.

Define $\boldsymbol{\theta}=(\theta_1,\dots,\theta_K)^{\mathrm{T}}$ and $\boldsymbol{\epsilon}=(\varepsilon_{11},\dots,\varepsilon_{ws},\dots,\varepsilon_{Kh_K})^{\mathrm{T}}$ where the elements θ_w and ε_{ws}, $w=1,\dots,K$, $s=1,\dots,h_w$, are the population residuals from the ordinary least squares fit of $UJ_w(U)$ and $YJ_{ws}(Y,U)$ on $\mathbf{X}$. The asymptotic distribution necessary to select $\mathbf{V}_n$ optimally is given in the following Lemma. The proof is omitted.

Lemma 1. *Assume that the data* $((Y_i,U_i),\mathbf{X}_i)$, $i=1,\dots,n$, *are a simple random sample of* $((Y,U),\mathbf{X})$ *with finite fourth moments. Then*

$$\sqrt{n}(\operatorname{vec}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}}) - \operatorname{vec}(\boldsymbol{\rho}\boldsymbol{\gamma}))\xrightarrow{\mathscr{D}}\text{Normal}(0,\boldsymbol{\Gamma}), \tag{3.6}$$

where $\boldsymbol{\Gamma}=\operatorname{Cov}(\operatorname{vec}(\boldsymbol{\Sigma}^{-1}\{\mathbf{X}-\mathrm{E}(\mathbf{X})\}(\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\epsilon}^{\mathrm{T}})))\in\mathbb{R}^{p(h+K)\times p(h+K)}$.

Letting $\hat{\boldsymbol{\Gamma}}$ be a consistent estimate of $\boldsymbol{\Gamma}$, our new method is obtained by minimizing

$$F_d^{\text{bear}}(\mathbf{B},\mathbf{C}) = (\text{vec}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC}))^{\mathrm{T}}\hat{\boldsymbol{\Gamma}}^{-1}(\text{vec}(\hat{\boldsymbol{\eta}},\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC})), \tag{3.7}$$

which is an application of (3.5) with $\mathbf{V}_n = \hat{\boldsymbol{\Gamma}}^{-1}$. The estimate of $\mathscr{S}_{(Y,U)|\mathbf{X}}$ constructed by minimizing (3.7) is called the BEAR estimator.

Since $\mathbf{X}$ and $(\boldsymbol{\theta},\boldsymbol{\epsilon})$ are uncorrelated, we can rewrite $\boldsymbol{\Gamma}$ as $\mathrm{E}[(\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\epsilon}^{\mathrm{T}})^{\mathrm{T}}(\boldsymbol{\theta}^{\mathrm{T}},\boldsymbol{\epsilon}^{\mathrm{T}})\otimes\boldsymbol{\Sigma}^{-1}\{\mathbf{X}-\mathrm{E}(\mathbf{X})\}\{\mathbf{X}-\mathrm{E}(\mathbf{X})\}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}]$. Assume that the data $((Y_i,U_i),\mathbf{X}_i)$, $i=1,\ldots,n$ are a simple random sample of $((Y,U),\mathbf{X})$, one choice of $\hat{\boldsymbol{\Gamma}}$ can be constructed straightforwardly by substituting sample versions for the population moments in $\boldsymbol{\Gamma}$:

$$\hat{\boldsymbol{\Gamma}} = \sum_{i=1}^{n}\frac{1}{n}\{(\hat{\boldsymbol{\theta}}^{\mathrm{T}},\hat{\boldsymbol{\epsilon}}^{\mathrm{T}})_i^{\mathrm{T}}(\hat{\boldsymbol{\theta}}^{\mathrm{T}},\hat{\boldsymbol{\epsilon}}^{\mathrm{T}})_i\otimes\hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{X}_i-\bar{\mathbf{X}}_\bullet)(\mathbf{X}_i-\bar{\mathbf{X}}_\bullet)^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-1}\},$$

where $\bar{\mathbf{X}}_\bullet$ is the sample average of X.

Let $\boldsymbol{\Delta}_{\text{bear}} \equiv (\boldsymbol{v}^{\mathrm{T}}\otimes I_p,\ I_{h+K}\otimes\boldsymbol{\rho})$, which is the Jacobian matrix

$$\boldsymbol{\Delta} = \left(\frac{\partial\,\text{vec}(\mathbf{BC})}{\partial\,\text{vec}(\mathbf{B})},\frac{\partial\,\text{vec}(\mathbf{BC})}{\partial\,\text{vec}(\mathbf{C})}\right)$$

evaluated at $(\mathbf{B}=\boldsymbol{\rho},\mathbf{C}=\boldsymbol{v})$. The associated asymptotic properties of BEAR are given in the following theorem. The proof is structurally similar to that of Theorem 2 in Cook and Ni (2005).

Theorem 1. *Assume that the data* $((Y_i,U_i),\mathbf{X}_i)$, $i=1,\ldots,n$, *are a simple random sample of* $((Y,U),\mathbf{X})$ *with finite fourth moments. Let* $\hat{F}_d^{\text{bear}}$ *be the minimum value of* (3.7), *and let*

$$(\hat{\boldsymbol{\rho}},\hat{\boldsymbol{v}}) = \arg\min_{\mathbf{B},\mathbf{C}} F_d^{\text{bear}}(\mathbf{B},\mathbf{C}).$$

Then,

1. *The estimate* $\text{vec}(\hat{\boldsymbol{\rho}}\hat{\boldsymbol{v}})$ *is asymptotically efficient, and*

$$\sqrt{n}(\text{vec}(\hat{\boldsymbol{\rho}}\hat{\boldsymbol{v}}) - \text{vec}(\boldsymbol{\rho}\boldsymbol{v})) \xrightarrow{\mathscr{D}} \text{Normal}(0,\boldsymbol{\Delta}_{\text{bear}}(\boldsymbol{\Delta}_{\text{bear}}^{\mathrm{T}}\boldsymbol{\Gamma}_{\text{bear}}^{-1}\boldsymbol{\Delta}_{\text{bear}})^{-}\boldsymbol{\Delta}_{\text{bear}}^{\mathrm{T}}).$$

2. $n\hat{F}_d^{\text{bear}}$ *has an asymptotic chi-squared distribution with degrees of freedom* $(p-d)(h+K-d)$.

Theorem 1 provides a basis for inference about $\mathscr{S}_{(Y,U)|\mathbf{X}}$. In particular, the second conclusion can be used to test the hypothesis $d=m$ versus $d>m$, rejecting if $n\hat{F}_m^{\text{bear}}$ exceeds a selected quantile of the chi-squared distribution with $(p-m)(h+K-m)$ degrees of freedom. A useful property of BEAR is that $n\hat{F}_d^{\text{bear}}$ follows an asymptotic chi-squared distribution, while the corresponding statistics for bivariate SIR is distributed as a linear combination of independent chi-squared random variables. Asymptotic efficiency means that the estimator has minimum asymptotic variance within family (3.5).

The alternating least-squares algorithm for *inverse regression estimation* (IRE: Cook and Ni, 2005; Ruhe and Wedin, 1980) can be adapted for the minimization of $\hat{F}_d^{\text{bear}}$. Since the form of the inner-product matrix in (3.7) depends on $\mathscr{S}_{(Y,U)|\mathbf{X}}$, we could adapt the iterative algorithm to reduce the variability of this inner-product matrix. Here is a sketch of this idea. First, obtain $\text{Span}(\hat{\boldsymbol{\rho}})$, an estimate of $\mathscr{S}_{(Y,U)|\mathbf{X}}$, via the alternating least-squares method. Second, update the inner-product matrix using $\text{Span}(\hat{\boldsymbol{\rho}})$. Then, re-run the alternating least-squares algorithm to update $\text{Span}(\hat{\boldsymbol{\rho}})$ applying this new inner-product matrix. Carroll and Ruppert (1988) recommended at least two cycles. We use a three-cycle iterative computation algorithm for BEAR.

4. Balanced bivariate IRE

In this section, we propose a method, called BBE, to deal with the bivariate regressions with two continuous responses. For $w=1,\ldots,K$, and $s=1,\ldots,h_w$, within each bivariate slice, we consider $(\boldsymbol{\alpha}_{ws},\boldsymbol{\beta}_{ws})$, where

$$\boldsymbol{\alpha}_{ws} = \boldsymbol{\Sigma}^{-1}\,\text{Cov}(\mathbf{X},UJ_{ws}). \tag{4.1}$$

Let $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_{11},\ldots,\boldsymbol{\alpha}_{Kh_K})$, and define $\upsilon=(\upsilon_{11},\ldots,\upsilon_{ws},\ldots,\upsilon_{Kh_K})^{\mathrm{T}}$, and $\boldsymbol{\epsilon}=(\boldsymbol{\varepsilon}_{11},\ldots,\boldsymbol{\varepsilon}_{ws},\ldots,\boldsymbol{\varepsilon}_{Kh_K})^{\mathrm{T}}$ where the elements υ_{ws} and $\boldsymbol{\varepsilon}_{ws}$, $w=1,\ldots,K$, $s=1,\ldots,h_w$, are the population residuals from the ordinary least squares fit of $UJ_{ws}(Y,U)$ and $YJ_{ws}(Y,U)$ on $\mathbf{X}$.

Lemma 2. *Assume that the data* $((Y_i,U_i),\mathbf{X}_i)$, $i=1,\ldots,n$, *are a simple random sample of* $((Y,U),\mathbf{X})$ *with finite fourth moments. Then*

$$\sqrt{n}(\text{vec}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \text{vec}(\boldsymbol{\rho}\boldsymbol{\gamma})) \xrightarrow{\mathscr{D}} \text{Normal}(0,\boldsymbol{\Gamma}_{\text{bbe}}), \tag{4.2}$$

where $\boldsymbol{\Gamma}_{\text{bbe}} = \text{Cov}(\text{vec}(\boldsymbol{\Sigma}^{-1}\{\mathbf{X} - \text{E}(\mathbf{X})\}(\boldsymbol{v}^{\text{T}}, \boldsymbol{\epsilon}^{\text{T}}))) \in \mathbb{R}^{2ph\times 2ph}$.

Similar to the construction of BEAR, based on Lemma 2, the BBE estimator of $\mathscr{S}_{(Y,U)|\mathbf{X}}$ is constructed by minimizing

$$F_d^{\text{bbe}}(\mathbf{B},\mathbf{C}) = (\text{vec}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC}))^{\text{T}}\hat{\boldsymbol{\Gamma}}_{\text{bbe}}^{-1}(\text{vec}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}) - \text{vec}(\mathbf{BC})), \tag{4.3}$$

where $\hat{\boldsymbol{\Gamma}}_{\text{bbe}}$ is a consistent estimate of $\boldsymbol{\Gamma}_{\text{bbe}}$; the following sample version is one possible choice:

$$\hat{\boldsymbol{\Gamma}}_{\text{bbe}} = \sum_{i=1}^{n}\frac{1}{n}\{(\hat{\boldsymbol{v}}^{\text{T}},\hat{\boldsymbol{\epsilon}}^{\text{T}})_i^{\text{T}}(\hat{\boldsymbol{v}}^{\text{T}},\hat{\boldsymbol{\epsilon}}^{\text{T}})_i \otimes \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{X}_i - \bar{\mathbf{X}}_\bullet)(\mathbf{X}_i - \bar{\mathbf{X}}_\bullet)^{\text{T}}\hat{\boldsymbol{\Sigma}}^{-1}\}.$$

Let $\boldsymbol{\Delta}_{\text{bbe}} \equiv (\boldsymbol{v}^{\text{T}} \otimes I_p,\ I_{2h} \otimes \boldsymbol{\rho})$, which is the Jacobian matrix

$$\boldsymbol{\Delta} = \left(\frac{\partial\,\text{vec}(\mathbf{BC})}{\partial\,\text{vec}(\mathbf{B})}, \frac{\partial\,\text{vec}(\mathbf{BC})}{\partial\,\text{vec}(\mathbf{C})}\right)$$

evaluated at $(\mathbf{B} = \boldsymbol{\rho}, \mathbf{C} = \boldsymbol{v})$. The following theorem provides a basis for BBEs inference about $\mathscr{S}_{(Y,U)|\mathbf{X}}$. BBE also enjoys a simple asymptotic chi-squared test for testing the dimension of $\mathscr{S}_{(Y,U)|\mathbf{X}}$. Similarly a three-cycle iterative computation algorithm adapted from the alternating least-squares algorithm are used for the computation of BBE.

Theorem 2. *Assume that the data* $((Y_i,U_i),\mathbf{X}_i)$, $i = 1,\ldots,n$, *are a simple random sample of* $((Y,U),\mathbf{X})$ *with finite fourth moments. Let* $\hat{F}_d^{\text{bbe}}$ *be the minimum value of* (4.3), *and let*

$$(\hat{\boldsymbol{\rho}},\hat{\boldsymbol{v}}) = \underset{\mathbf{B},\mathbf{C}}{\arg\min}\, F_d^{\text{bbe}}(\mathbf{B},\mathbf{C}).$$

Then,

1. *The estimate* $\text{vec}(\hat{\boldsymbol{\rho}}\hat{\boldsymbol{v}})$ *is asymptotically efficient, and*

$$\sqrt{n}(\text{vec}(\hat{\boldsymbol{\rho}}\hat{\boldsymbol{v}}) - \text{vec}(\boldsymbol{\rho}\boldsymbol{v})) \xrightarrow{\mathscr{D}} \text{Normal}(0, \boldsymbol{\Delta}_{\text{bbe}}(\boldsymbol{\Delta}_{\text{bbe}}^{\text{T}}\boldsymbol{\Gamma}_{\text{bbe}}^{-1}\boldsymbol{\Delta}_{\text{bbe}})^{-}\boldsymbol{\Delta}_{\text{bbe}}^{\text{T}}).$$

2. $n\hat{F}_d^{\text{bbe}}$ *has an asymptotic chi-squared distribution with degrees of freedom* $(p-d)(2h-d)$.

We are free to interchange the role of Y and U during the development of BBE. As we stated in Section 1, BBE is designed to handle the case when both Y and U are continuous, where we extract the intra-slice information from both responses. If Y and U are both constants in each bivariate slice, the estimates of $\boldsymbol{\alpha}_{ws}$ and $\boldsymbol{\beta}_{ws}$ will span the same subspace so we need only one of them. Hence, BBE is reduced to bivariate SIR, or bivariate IRE, depending on the inner-product.

4.1. Testing predictor effects

Both BEAR and BBE methods proposed by us allow us to test conditional independence hypotheses by using various quadratic inference functions. Without loss of generality, we limit our discussion to BBE. We consider testing hypotheses of the forms: $(Y,U) \perp\!\!\!\perp P_{\mathscr{H}}\mathbf{X} | (Q_{\mathscr{H}}\mathbf{X})$, where $\mathscr{H}$ is an r-dimensional user-selected subspace of the predictor space. This is equivalent to testing hypotheses $P_{\mathscr{H}}\mathscr{S}_{(Y,U)|\mathbf{X}} = \mathscr{O}_p$ under the usual linearity and coverage conditions (Cook and Ni, 2005).

Since the marginal predictor hypothesis $P_{\mathscr{H}}\mathscr{S}_{(Y,U)|\mathbf{X}} = \mathscr{O}_p$ does not require specification of d, equivalently, we test the hypothesis $\boldsymbol{H}^{\text{T}}(\boldsymbol{\alpha},\boldsymbol{\beta}) = 0$, where $\boldsymbol{H} \in \mathbb{R}^{p\times r}$ be an orthonormal basis for $\mathscr{H}$. It follows from Theorem 2 that a Wald test statistic of the form

$$T^{\text{bm}}(\mathscr{H}) = n\,\text{vec}(\boldsymbol{H}^{\text{T}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}))^{\text{T}}\{(I_{2h-1}\otimes\boldsymbol{H}^{\text{T}})\hat{\boldsymbol{\Gamma}}_{\text{bbe}}(I_{2h-1}\otimes\boldsymbol{H})\}^{-1}\,\text{vec}(\boldsymbol{H}^{\text{T}}(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})) \tag{4.4}$$

can be used to test a marginal predictor hypothesis. Following Theorem 2 and Slutsky's theorem, it can be shown that, under the hypothesis, (4.4) is distributed asymptotically as a chi-squared random variable with degrees of freedom of $r(2h-1)$.

Let $\boldsymbol{H}_0 \in \mathbb{R}^{p\times(p-r)}$ be an orthonormal basis for $\text{Span}(Q_{\mathscr{H}})$, then the joint dimension-predictor hypothesis $P_{\mathscr{H}}\mathscr{S}_{\boldsymbol{\alpha},\beta} = \mathscr{O}_p$ and $d = m$, is equivalent to $(\boldsymbol{\alpha},\boldsymbol{\beta}) = Q_{\mathscr{H}}(\boldsymbol{\alpha},\boldsymbol{\beta}) = Q_{\mathscr{H}}\boldsymbol{\rho}\boldsymbol{v} = \boldsymbol{H}_0\boldsymbol{\rho}_0\boldsymbol{v}$, where $\boldsymbol{\rho}_0$ contains the coordinates of $\boldsymbol{\rho}$ represented in terms of the basis $\boldsymbol{H}_0$. We then can fit under the joint hypothesis by minimizing the following constrained optimal discrepancy function:

$$F_{m,H}^{\text{bbe}}(\mathbf{B},\mathbf{C}) = (\text{vec}((\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})) - \text{vec}(\boldsymbol{H}_0\mathbf{BC}))^{\text{T}}\hat{\boldsymbol{\Gamma}}_{\text{bbe}}^{-1}(\text{vec}((\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}})) - \text{vec}(\boldsymbol{H}_0\mathbf{BC})) \tag{4.5}$$

over $\mathbf{B} \in \mathbb{R}^{(p-r)\times m}$ and $\mathbf{C} \in \mathbb{R}^{m\times(2h-1)}$. Values of $\mathbf{B}$ and $\mathbf{C}$ that minimize (4.5) provide estimates of $\boldsymbol{\rho}_0$ and $\boldsymbol{v}$. Following the result of Theorem 2, we know that, under the null hypothesis, the statistic $n\hat{F}^{\text{bbe}}_{d,H}$ has an asymptotic chi-squared distribution with degrees of freedom $(p-d)(2h-1-d)+dr$.

Finally, when d is specified, or when inference on d using marginal dimension tests results in a firm estimate, we might consider the conditional hypothesis $P_{\mathscr{H}}\mathscr{S}_{\boldsymbol{\alpha},\beta} = \mathscr{O}_p$ given d. The difference in minimum discrepancies

$$T^{\text{bc}}(\mathscr{H}|d) = n\hat{F}^{\text{bbe}}_{d,H} - n\hat{F}^{\text{bbe}}_{d} \tag{4.6}$$

is used to test a conditional predictor hypothesis. Under the null hypothesis, $T^{\text{bc}}(\mathscr{H}|d)$ has an asymptotic chi-squared distribution with degrees of freedom rd. It can be shown that the conditional predictor test statistic and the marginal dimension statistic are asymptotically independent.

5. Dimension reduction in regressions with censoring

SDR can be of practical interest to censored regression analysis. As pointed out by Zeng (2004), when there are many predictors, nonparametric approaches may be infeasible due to the "curse of dimensionality". Moreover, for semiparametric models, the parametric functions are likely to be misspecified. In contrast, the bivariate SDR methods we proposed can be carried out without pre-specifying any parametric model, and they can often avoid the curse of dimensionality. After reduction of $\mathbf{X}$ to the estimated sufficient predictors, many traditional methodologies of survival analysis can be applied.

The unique aspects of regressions with censoring make BEAR attractive as a dimension reduction method. Since the censoring status takes only two values, the double slicing required by BEAR may be feasible in many applications. Given the fact that we could use different symbols and colors to indicate the censoring status, a summary plot of Y versus estimated sufficient predictors will give us a direct and informative visualization of the survival data. This may provide insights into model building, and can identify influential subjects or outliers. In some cases, summary plots may provide a useful tool in the study of the censoring mechanism.

Common model-based methods of dimension reduction, such as subset selection on linear regression or the lasso, achieve dimension reduction by removing predictors, thereby reducing p to some smaller number. Many nonparametric methods achieve dimension reduction by replacing $\mathbf{X}$ with a few linear combinations, but do not allow possibility of variable selection. BEAR permits both reduction by variable selection and reduction via linear combinations, without specifying a parsimonious model for the regression. The importance of variable selection in survival analysis was pointed out by Raftery et al. (1995): "survival analysis is concerned with finding models to predict the survival of patients or to assess the efficacy of a clinical treatment. A key part of the model-building process is the selection of the predictor variables." In short, BEAR may be an effective *pre-modeling* tool, aiming to reduce complicated high-dimensional regressions to simpler ones with lower dimensions.

To adapt BEAR for censored survival data, we first introduce the following notation:

- T the true unobservable survival time,
- C the censoring time,
- δ the censoring indicator; $\delta = 1$ if $T \leqslant C$ and $\delta = 0$ otherwise,
- Y the observed time, $T\delta + C(1-\delta)$.

We assume that the observed data $(Y_i, \delta_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, are obtained from i.i.d. realizations of $(T, C, \mathbf{X})$. We also assume that $T \perp\!\!\!\perp C|\mathbf{X}$, the usual independence assumption (Ebrahimi et al., 2003; Tsiatis, 1975), to ensure the identifiability of T.

The goal of SDR for survival data is to infer about the central subspace $\mathscr{S}_{T|\mathbf{X}}$. However, since T is not fully observable, we can estimate only the central subspace $\mathscr{S}_{(Y,\delta)|\mathbf{X}}$ for the bivariate regression of the observable (Y, δ) on $\mathbf{X}$. The following proposition provides a connection between $\mathscr{S}_{(Y,\delta)|\mathbf{X}}$ and $\mathscr{S}_{T|\mathbf{X}}$. Its proof is given in an Appendix.

Proposition 1. *Let the columns of* $\mathbf{B}$ *be an orthonormal basis for* $\mathscr{S}_{T|\mathbf{X}}$.

1. *If* $T \perp\!\!\!\perp C|\mathbf{X}$, *then* $T \perp\!\!\!\perp C|\mathbf{B}^{\mathrm{T}}\mathbf{X}$ *and*

$$\mathscr{S}_{(Y,\delta)|\mathbf{X}} = \mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}. \tag{5.1}$$

2. *If* $T \perp\!\!\!\perp C|\mathbf{X}$ *and* $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{T|\mathbf{X}}$ *then* $\mathscr{S}_{T|\mathbf{X}} = \mathscr{S}_{(T,C)|\mathbf{X}}$.

The condition of $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{T|\mathbf{X}}$ is not as restrictive as it might appear. For example, the widely used Koziol and Green (1976) model assumes that the distribution function of C is a positive power of that of T. In this case, $\mathscr{S}_{C|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}}$.

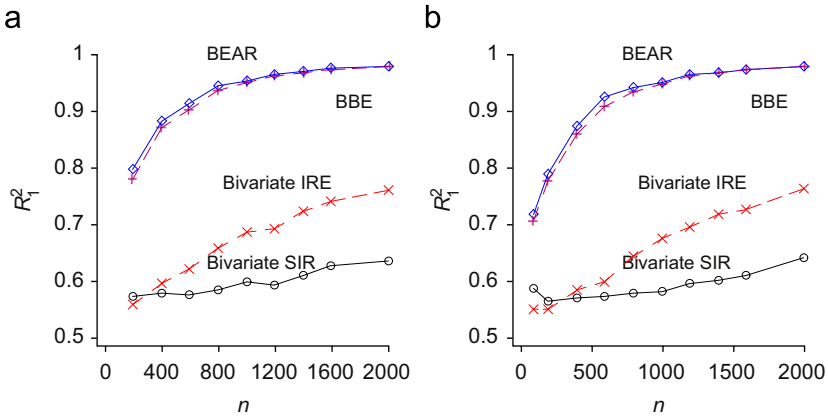

Fig. 1. Model A—estimation accuracy: R_1^2. Simulation results at various sample sizes with 1000 runs. (a) $U = Y_2$, (b) $U = Y_1$.

6. Simulation studies and data analysis

In this section, we report results from two simulation models to support our theoretical conclusions regarding BBE. To complement our simulation study of BBE, we also apply BEAR to the well-known PBC data (Fleming and Harrington, 1991).

6.1. Simulation studies

Model A: We first consider a three-dimensional model:

$$Y_1 = 1.5(5 + X_1)(2 + X_2 + X_3) + 0.5\varepsilon_2,$$
$$Y_2 = (1.5 + X_4)\varepsilon_1,$$

where ε_1 and ε_2 are independent standard normal random variates, $X_1 = W_1, X_2 = V_1 + W_2/2, X_3 = -V_1 + W_2/2, X_4 = V_2 + V_3$, and $X_5 = V_2 - V_3$. The V_i's and W_j's are independent with V_i's drawn from a $t_{(5)}$ distribution and the W_j's from gamma(0.2, 1) distribution. Versions of this model were also used by Li (1991), Velilla (1998) and others in simulation studies related to the performance of SIR. It can be shown that the linearity condition holds under this model. Nevertheless, this is a difficult test case for dimension reduction since the predictors are skewed or have heavy tails, and are prone to outliers.

The number of slices h is a tuning parameter much like the tuning parameter encountered in the smoothing literature (Li, 1987; Härdle et al., 1988). Experience indicates that good results are often obtained by choosing h to be somewhat larger than $d + 1$, trying a few different values of h as necessary. Since traditional asymptotic results in SDR are based on the number of observations per slice going to infinity, in practice this suggests relatively few slices. Choosing h very much larger than d should generally be avoided due to the conflicts between the requirements of asymptotic approximations and recovering intra-slice information. Based on our experience, at least 20 sample points within each slice are required to achieve satisfactory results.

We set $K = 3$ and $h_1 = h_2 = h_3 = 3$. Simulations are run with $Y_2 = U$ and $Y_1 = U$, respectively, to study the effects of asymmetric slicing. We calculated R_1^2, R_2^2 and R_3^2, the R^2 values from the regressions of X_1, $X_2 + X_3$ and X_4 on the $d = 3$ estimated sufficient predictors $\hat{\boldsymbol{\rho}}_1^{\mathrm{T}}\mathbf{X}$, $\hat{\boldsymbol{\rho}}_2^{\mathrm{T}}\mathbf{X}$, and $\hat{\boldsymbol{\rho}}_3^{\mathrm{T}}\mathbf{X}$ to measure estimation accuracy. Since the results for R_3^2 and R_2^2 are essentially the same as those for R_1^2, we will use just R_1^2 in the following discussion. As shown in Fig. 1, both BBE and BEAR clearly dominated over bivariate SIR and bivariate IRE. The performances of BBE and BEAR were almost identical, while bivariate IRE showed some advantages over bivariate SIR due to the MDA. It seems that incorporation of intra-slice information into estimation greatly improved the estimation accuracy. We did not detect significant differences due to different slicing techniques. We then compared the performances setting $U = Y_2$. When $n = 800$, the average of R_1^2 from 1000 replications was 0.940 from BBE, 0.946 from BEAR, 0.659 from bivariate IRE, and 0.586 from bivariate SIR. Also, the R_1^2's from BBE, BEAR and bivariate IRE exceeded the R_1^2's from bivariate SIR 95.3%, 96.3% and 70.9% of the time, respectively. While the R_1^2's from BBE and BEAR exceeded the R_1^2's from bivariate IRE 93.4% and 94.7% of the time, respectively.

As we mentioned previously, we use a series of dimension tests to estimate $d = \dim(\mathscr{S}_{(Y,U)|\mathbf{X}})$. Since we use the same level for all tests, the best we could expect is that the leading tests of $d = 0$, 1 and 2 have power 1, and that all of the estimation error arises from the level α of the test of $d = 3$. We first consider the actual levels of the test under the null $d = 3$. The percentage of correct dimension decisions in 1000 simulated data sets at various sample sizes using nominal level $\alpha = 0.05$ with two different slicing methods are shown in Fig. 2. We did not find any difference between the slicing orders, as shown in Fig. 2(a) and (b). The performances of BBE and BEAR are very similar for most of the time, with slightly better performance for BBE when sample size is small. Both of them did much better than bivariate SIR. Except for bivariate SIR, all other methods that respond to the change

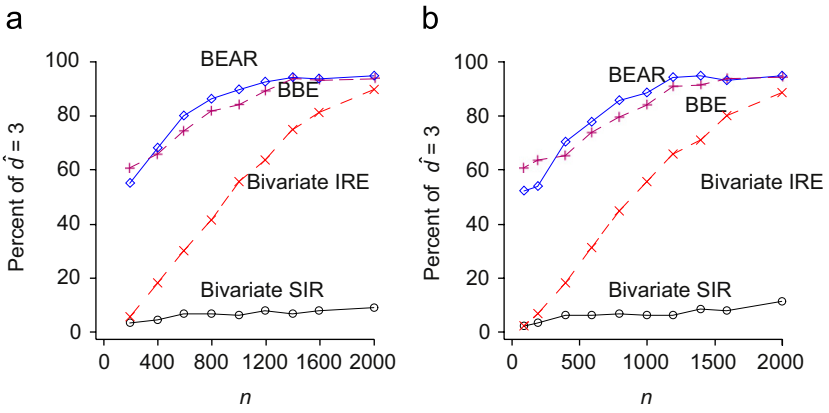

Fig. 2. Model A—percentage of correct estimates $\hat{d}=3$ using 5% tests. (a) $U=Y_2$, (b) $U=Y_1$.

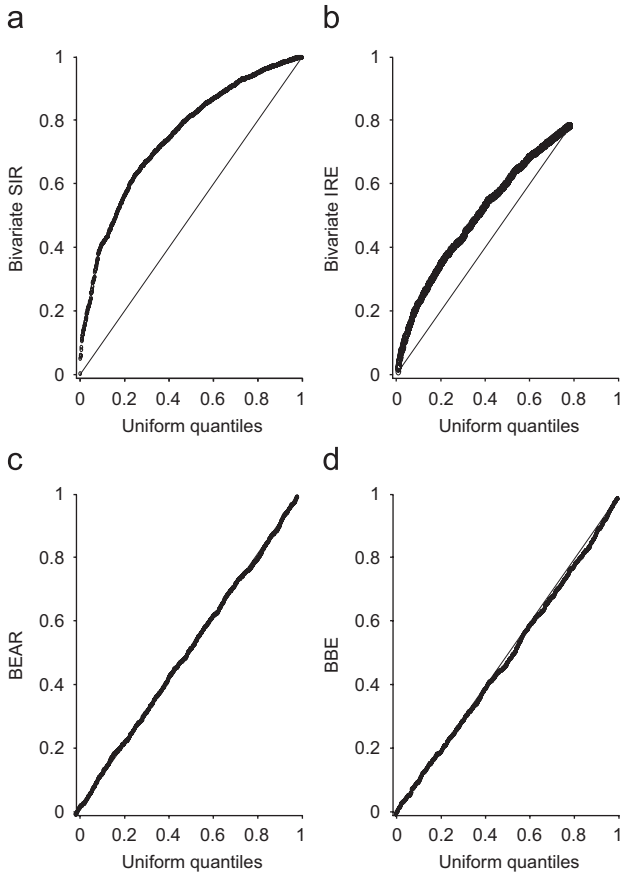

Fig. 3. Model A—uniform quantile plots of p-values for the hypothesis d=3 from 1000 replications with n=800. (a) Bivariate SIR, (b) bivariate IRE, (c) BEAR, (d) BBE.

of sample sizes perform closer to the best rate of correct decisions with increasing n. The increment of sample size seems to affect bivariate IRE the most: the correct dimension decisions it made changed dramatically from 5.8% for $n=200$ to 89.8% when $n=2000$.

Fig. 3 shows uniform quantile plots of p-values for testing $H_0: d=3$ in Model A with 1000 replications when $n=800$. The sampling distributions of both BBE and BEAR's test statistics is closer to the asymptotic approximations, which suggests a close agreement between the actual and nominal levels for them.

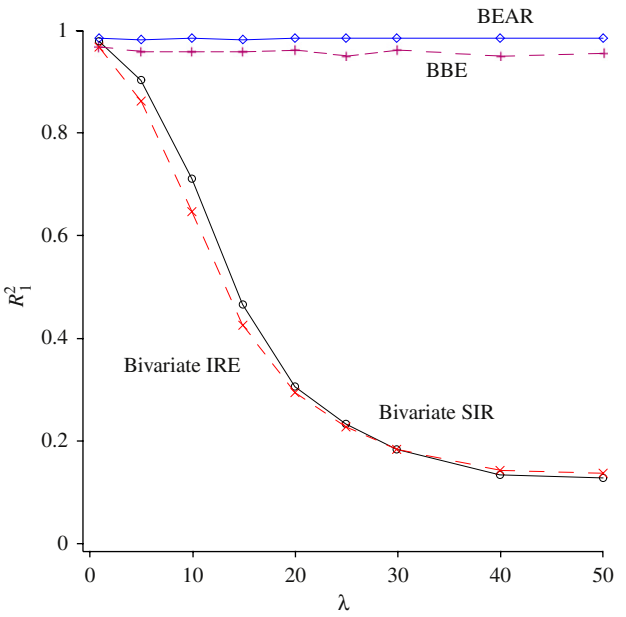

Fig. 4. Model B—estimation accuracy: R_1^2. Simulation results at various λ values with 1000 runs.

We also applied the marginal predictor tests using BBE to all the five predictors in turn at level of 5%. The predictor was selected if the testing p-value was less than 5%. The average number of predictors selected from 1000 simulation runs were $(4.3, 4.2, 4.1)$ for $n = (200, 400, 800)$, respectively, and the correct four predictors (X_1, X_2, X_3, X_4) were nearly always selected.

Model B: We now consider an inverse regression model:

$$Y_1 = \lambda(Z_1 + Z_2) + (Z_1 - Z_2),$$
$$Y_2 = \lambda(Z_1 + Z_2) - (Z_1 - Z_2),$$

$\mathbf{X} \in \mathbb{R}^{10}$ and $\mathbf{X} = \boldsymbol{\alpha}(Y_1 - Y_2) + \boldsymbol{\epsilon}$, where Z_1 and Z_2 are independent standard normal random variates, $\boldsymbol{\alpha} = (0, \ldots, 0, 1) \in \mathbb{R}^{10}$, $\boldsymbol{\epsilon} \perp\!\!\!\perp (Z_1, Z_2)^{\mathrm{T}}$ is a 10-dimensional standard normal vector, $\lambda \in \mathbb{R}^1$ is a constant. The same model was discussed by Yoo and Cook (2007). Letting $\mathbf{Y} = (Y_1, Y_2)$, we have $\mathscr{S}_{\mathbf{Y}|\mathbf{X}} = \mathscr{S}_{\boldsymbol{\alpha}}$.

The number of slices was 3×3 with $U = Y_1$ for each simulation run. We used R_1^2, the R^2 values from the regression of X_{10} on the first estimated sufficient predictor, to measure the estimation accuracy. The performances of all four estimators were compared at nine different values of λ with fixed sample size $n = 200$.

The curves shown in Fig. 4 are plots of average R_1^2 from 1000 replications versus each λ, for values of λ between 1 and 50. Both bivariate SIR and bivariate IRE are quite sensitive to the change of λ, their performances deteriorate fast as λ increases; while BBE and BEAR are robust to the increment. The average R_1^2 from bivariate SIR drops dramatically from 0.98 with $\lambda = 1$ to 0.13 with $\lambda = 50$, while the average values from BBE are always above 0.95 regardless the change of λ. BBE and BEAR show the clear advantage due to the incorporation of the intra-slice information.

Model C: In this model, we consider the case of one continuous response and one categorical response. In particular, we use the technique suggested by Bender et al. (2005) to generate a Cox proportional hazard model as follows:

$$Y = \min(T, C),$$
$$\delta = \text{Indicator}\{T \leqslant C\}$$

with $T = 5\log[1 - 200\log(e_1)\mathrm{e}^{1.5(5+X_1)(X_2+X_3+2)}]$ and $C = \mathrm{e}^{5+X_1+\log(-\log(e_2))}$; where e_1 and e_2 are random uniform variates on $[0, 1]$. For $i = 1, \ldots, 5$, X_i's are generated in the same way as that of Model A; and for $i = 6, \ldots, 10$, X_i's are standard normal variates. The censoring proportion is about 40%. For this model, $\mathscr{S}_{(Y,\delta)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} = \mathscr{S}_{(\beta_1,\beta_2)}$, with $\boldsymbol{\beta}_1^{\mathrm{T}}\mathbf{X} = X_1$ and $\boldsymbol{\beta}_2^{\mathrm{T}}\mathbf{X} = X_2 + X_3$.

The number of slices was 2×3 with $U = \delta$ for each simulation run. R_1^2 and R_2^2, the R^2 values from the regression of X_1 and $X_2 + X_3$ on the $d = 2$ estimated sufficient predictors, were used to measure estimation accuracy. The two methods (BBE and BEAR) which took the intra-slice information into estimation again showed similar performances: both of them were better than the other two methods with respect to estimation accuracy. For example, with $n = 800$, the average of R_2^2 values from 1000

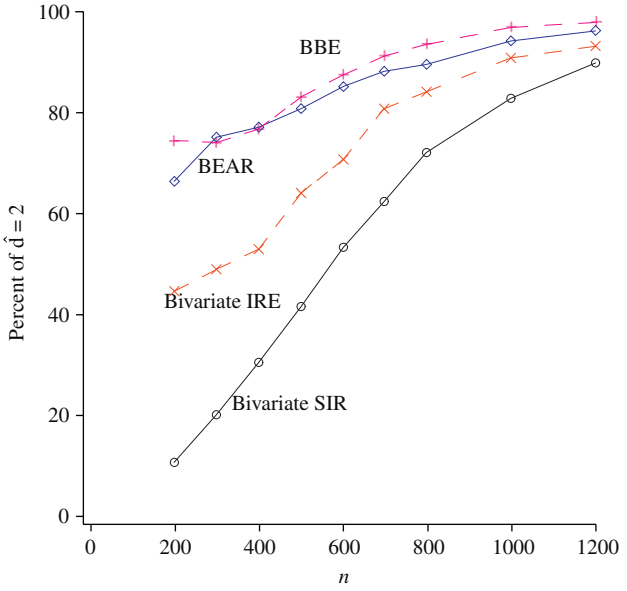

Fig. 5. Model C—percentage of correct estimates $\hat{d} = 2$ using 5% tests.

replications was 0.904 from BBE, 0.900 from BEAR, 0.794 from bivariate IRE and 0.791 from bivariate SIR. Also, the R_2^2's from BBE and BEAR exceeded the R_2^2's from bivariate IRE and bivariate SIR about 90.8% and 93.1% of the time, respectively.

Fig. 5 shows the percentage of correct dimension estimations for all four methods in 1000 simulation runs at various sample sizes using nominal levels $\alpha = 0.05$. BBE and BEAR showed clear advantages over the other two methods. Simulation results suggest that intra-slice information plays a more significant role than the use of MDA (Cook and Ni, 2005) under most of the cases.

6.2. The PBC data

In this section, we consider the well-known PBC (primary biliary cirrhosis) data set using BEAR. This data set contains survival time and other information on 312 PBC patients participating the randomized placebo controlled trial in the Mayo Clinic between 1974 and 1984. Fleming and Harrington (1991) provided a detailed description of this data set. They selected five predictors from the original 17 covariates. Li et al. (1999) studied this data set using bivariate SIR.

We will consider the following regression:

Y	The observed time, number of days between registration and the earlier of death or censoring.
δ	The censoring indicator; $\delta = 1$ if death occurred and $\delta = 0$ otherwise.
Age	Age in years.
Edema	Presence of edema. 0 = no edema and no diuretic therapy for edema; 0.5 = edema present for which no diuretic therapy was given, or edema resolved with diuretic therapy; 1 = edema despite diuretic therapy.
SerumB	Serum bilirubin in mg/dl.
Albumin	Albumin in gm/dl.
SGOT	SGOT in U/ml.
Triglycerides	Triglycerides in mg/dl.
Platelet	Platelet count.
Prothrombintime	Prothrombin time in seconds.

Until recently including categorical predictors like *Edema* in SDR methods was problematic because of the linearity condition. However, it follows from the recent work of Cook (2007, Section 3.3) that categorical predictors can be used for dimension reduction if one assumes an exponential family distribution for them in place of the linearity condition. Cook's results are based on assuming conditionally independent predictors, but the same conclusion can be shown if a quadratic exponential model (Prentice and Zhao, 1991) is used. Thus under a modest additional assumption, the inclusion of *Edema* is no longer problematic.

A major goal of survival analysis is to assess the efficacy of a clinical treatment. For the Mayo PBC data, previous studies have shown that there was no therapeutic differences between control and D-penicillamine-treated patients. Consequently, we ignored the drug factor throughout this illustration. Cases with missing values were ignored as well. The remaining 278 cases

Table 1
PBC data: p-values of dimension tests from BEAR and bivariate SIR.

	$d=0$	$d=1$	$d=2$	$d=3$
BEAR	0	0	0.184	0.504
Bivariate SIR	0	0	0.009	0.332

Table 2
PBC data: p-values from marginal predictor tests using BEAR.

Predictor	Step 1	Step 2	Step 3	Step 4
Age	0.007	0.007	0.008	0.040
Albumin	0.000	0.000	0.000	0.000
Platelet	0.253	0.248	Deleted	
Edema	0.002	0.002	0.003	0.003
Prothrombin time	0.000	0.000	0.000	0.000
Serumb	0.000	0.000	0.000	0.000
SGOT	0.172	0.153	0.104	Deleted
Triglycerides	0.271	Deleted		

Table 3
PBC data: the first two directions from BEAR.

First direction	(0.0144308, −0.121755, 0.912173, 0.387744, 0.0506238)
Second direction	(0.0146395, −0.960727, −0.0301523, −0.247165, 0.121609)

were analyzed using BEAR and bivariate SIR with $K=2$ and $h_1=h_2=4$. That is, we applied double slicing with four slices within the censored group and event group, respectively. The same outliers as Li et al. (1999) reported were found and removed in further analysis.

Table 1 provides the p-values of the asymptotic tests for $d=0$, 1, 2 and 3 using BEAR and bivariate SIR. BEAR inferred that the central subspace is two-dimensional, which agrees with the findings of Li et al. (1999). Judging from our simulation studies, we think that bivariate SIR over-estimated the dimension.

An important advantage of BEAR is that it allows us to test the predictor effects without assuming a model or specifying d. For this example, we used marginal predictor tests as the basis of a model-free backward elimination procedure (Li et al., 2005). As shown in Table 2, three predictors were screened using 5% tests, leaving five predictors for further analysis. This screening requires only independence condition $T \perp\!\!\!\perp C|\mathbf{X}$, because then $\mathscr{S}_{T|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$. If $\mathscr{S}_{C|\mathbf{X}} \nsubseteq \mathscr{S}_{T|\mathbf{X}}$ (cf. Proposition 1, part 2), then some extraneous predictors may remain.

We next re-estimated the dimension of the regression based on these five predictors. Using BEAR, we again inferred that two sufficient predictors are required. The coefficients associated with the first two BEAR predictors are reported in Table 3 are strongly correlated with the covariate suggested by Fleming and Harrington (1991):

$$FH = 0.03\text{Age} - 3.06\log(\text{Albumin}) + 0.78\text{Edema} + 3.11\log(\text{Prothrombin time}) + 0.88\log(\text{Serumb}). \tag{6.1}$$

The correlation between FH and its fitted values from the linear regression of FH on the two BEAR predictors is about 0.94.

Plots of the response Y versus the first and second BEAR predictors are shown in Figs. 6 and 7, along with separate smooths for the event and censored cases. Fig. 6 indicates that Y is independent of the first BEAR predictor within the censored cases, but is dependent on the predictor within the event cases. Fig. 6 suggests that Y is dependent on the second BEAR predictor within both the event and censored cases. These interpretations suggest that $\dim(\mathscr{S}_{C|\mathbf{X}})=1$, $\dim(\mathscr{S}_{T|\mathbf{X}})=2$ and $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{T|\mathbf{X}}$ (cf. Proposition 1). It also suggests that the independence condition hold for PBC data (cf. proof of Lemma 4).

A three-dimensional plot of Y versus both BEAR predictors (not shown) supports this interpretation. As a further diagnostic check of these conclusions, we applied univariate IRE within the event cases and found the p-values for the null hypotheses $d=m$, $m=0,1,2$, to be 0, 0 and 0.554, providing support for the inference $\dim(\mathscr{S}_{T|\mathbf{X}})=2$. The same procedure applied within the censored cases gave p-values of 0 and 0.230, sustaining the conclusion that $\dim(\mathscr{S}_{C|\mathbf{X}})=1$.

Turning to the modeling phase, Dickson et al. (1989) developed a prognosis model based on patient's age, total serum bilirubin and serum albumin concentrations, prothrombin time and severity of edema, the same five predictors that remained after applying backward elimination (cf. Table 1). Since these measurements can be obtained without requiring liver biopsy, the Mayo model—represented here by (6.1)—is perhaps the most widely used one among all available prognostic models. It has also been validated in several other survival studies (Prince and Jones, 2000). In order to compare our method with the Mayo model, we next fitted a Cox proportional model using the first two BEAR predictors. The resulting estimated risk function, which assigns a risk score to each individual patient, was

$$\hat{R}(\text{BEAR}_1, \text{BEAR}_2) = 1.157 \times \text{BEAR}_1 - 1.930 \times \text{BEAR}_2 - 0.303 \times \text{BEAR}_2^2,$$

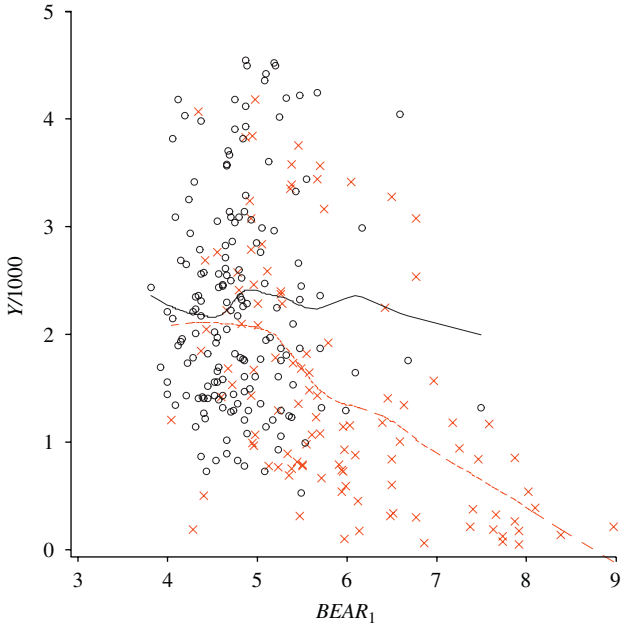

Fig. 6. PBC data: Y versus the first BEAR predictor. ×: event; ○: censored.

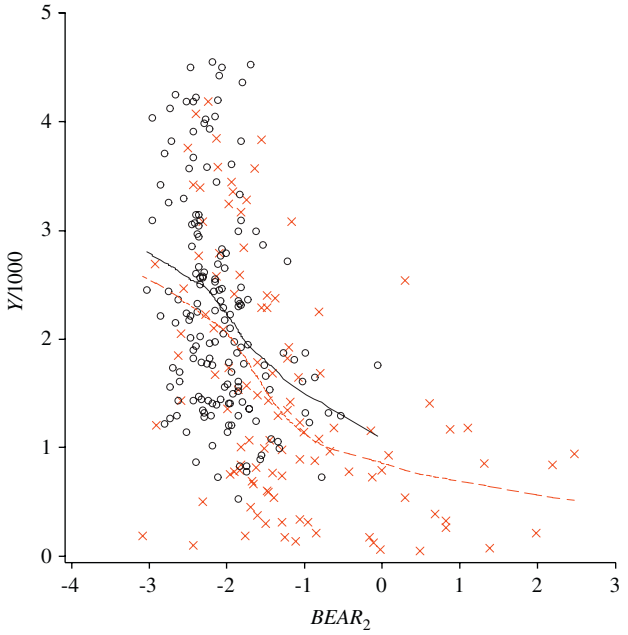

Fig. 7. PBC data: Y versus the second BEAR predictor. ×: event; ○: censored.

where our hazard model is $\lambda_i(t)=\lambda_0(t)\exp R(\mathrm{BEAR}_1, \mathrm{BEAR}_2)$. We also fitted a Cox proportional model using the first two directions identified by Li et al. (1999) resulting in the following estimated risk function (only the first direction is significant):

$$\hat{R}^{\mathrm{LI}} = 1.13 \times \mathrm{LI}_1,$$

where $\mathrm{LI}_1 = 0.02 \times \mathrm{Age} - 0.62 \times \mathrm{Albumin} + 0.9 \times \mathrm{Edema} + 0.38 \times \mathrm{Prothrombin} + 0.09 \times \mathrm{Serumb}$. We then cross-validate our model using an independent set of 140 Mayo clinic PBC patients consisting of 106 cases who were eligible for the trial but refused to

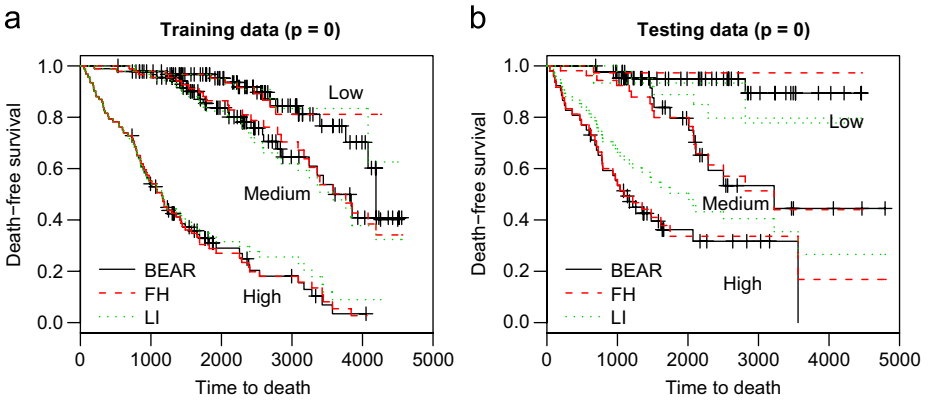

Fig. 8. PBC data: risk tertiles comparing survival curves from the BEAR-based model, the Mayo model and the model of Li et al. (1999): (a) 276 patients in the training set; (b) 140 patients in the testing set.

participate, plus 310−276 = 34 cases from the randomized participants with missing values on the "nonsignificant" predictors (Dickson et al., 1989).

Fig. 8a shows the Kaplan–Meier estimates of the survival curves for three groups of patients: the high-risk patients with risk scores above 7.88, the medium-risk patients with risk scores between 6.81 and 7.88, and the low-risk patients with risk scores less than 6.81. Those cut-off points are the $\frac{1}{3}$, $\frac{2}{3}$ percentiles of scores from the training data. The p-value from the log-rank test of difference among the three risk groups is highly significant. Shown in Fig. 8b is the survival curves for the 140 testing patients, where the same cut-off points as above are used to decide the risk groups among patients. The difference among the three groups is still highly significant. From both plots, we can see that our method performs well compared to the Mayo model, while the model of Li et al. (1999) does a much worse job on identifying the low-risk and medium-risk groups.

7. Discussion

In this paper, we have proposed two new dimension reduction methods for bivariate regressions. Both methods are designed to recover intra-slice information when at least one response is continuous. Simulations and real data examples showed that our proposed methods are superior to those based on SIR because they provide better estimates of the structural dimension of the regression, improves the estimation accuracy of the central subspace, and allows prior screening of predictors. The application to the PBC data showed the effectiveness of this model-free variable selection approach in practice. The model we developed based on BEAR is essentially equivalent to previous models, which may provide some assurance on the empirical performance of the proposed method.

Li et al. (2003) discussed how to reduce the dimension of Y in multivariate regression. They first applied multivariate (bivariate) SIR to reduce the predictors $\mathbf{X}$ to a lower-dimensional space, say, $b^{\mathrm{T}}\mathbf{X}$. After exchanging the roles of Y and $b^{\mathrm{T}}\mathbf{X}$, they then reduced the dimension of Y via multivariate SIR. This two-step process may be iterated. Their method should work much better with BBE (BEAR) as its core when Y is bivariate.

Independent (noninformative) censoring means that, conditional on covariates at each duration, the censored subjects are "representative" of those under observation at the same time. In other words, the death rates among censored subjects are the same as the death rates among uncensored subjects. Hence subjects may not be censored (withdrawn) because they have a higher or lower risk than the average, given the covariates. Independence condition is the standard assumption in survival analysis (Li et al., 1999). Both the Cox proportional model and the nonparametric Kaplan–Meier estimator require this assumption to obtain unbiased estimates of the true survival curve.

When the independence condition is violated, information about the censoring mechanism is needed to adjust the estimation bias due to censoring. Koziol–Green (1976) model commonly used to model the possible information contained in the informative censoring. Lee and Wolfe (1998) proposed a two-stage follow-up procedure to test the independent condition. Other approaches have been proposed to identify and account for dependent censoring (Zheng and Klein, 1995; Lin et al., 1996; Scharfstein and Robins, 2002). Research on developing dimension reduction methods with dependent censoring is ongoing.

Acknowledgments

We are grateful to the two referees and the editors for their constructive comments. R.D.C. was supported in part by National Science foundation Grants DMS-0405360 and DMS-0704098.

Appendix

Proof of Proposition 1. We first introduce the following lemmas.

Lemma 3. *$\mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}^{(C)}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}$, where $\mathscr{S}^{(C)}_{T|X}$ denote the intersection of all subspaces spanned by the columns of $\boldsymbol{\eta}$, which satisfies $T \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, C)$.*

Proof. Let $\boldsymbol{\rho}$ be an orthonormal basis for $\mathscr{S}_{(T,C)|\mathbf{X}}$, then $(T,C) \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}$. Immediately, $C \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}$, and $(T,C) \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}, C) \Longrightarrow T \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}, C)$. Hence, $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{(T,C)|\mathbf{X}}$ and $\mathscr{S}^{(C)}_{T|\mathbf{X}} \subseteq \mathscr{S}_{(T,C)|\mathbf{X}}$.

Also, let $\boldsymbol{\eta}$ be an orthonormal basis for $\mathscr{S}^{(C)}_{T|\mathbf{X}}$, $\boldsymbol{\zeta}$ be an orthonormal basis for $\mathscr{S}_{C|\mathbf{X}}$. By definition, $T \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, C)$, and $C \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X}$. We then have the following two conditions:

(a1) $T \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X}, C)$, (a2) $C \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X})$.

By Proposition 4.6 from Cook (1998), $(T,C) \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X})$. Hence, $\mathscr{S}_{(T,C)|\mathbf{X}} \subseteq \mathscr{S}^{(C)}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}$. □

Lemma 4. *Under the independent assumption $T \perp\!\!\!\perp C|\mathbf{X}$, $\mathscr{S}^{(C)}_{T|\mathbf{X}} \subseteq \mathscr{S}_{T|\mathbf{X}}$.*

Proof. Let $\mathbf{B}$ be an orthonormal basis for $\mathscr{S}_{T|\mathbf{X}}$. $T \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^{\mathrm{T}}\mathbf{X}$ and $T \perp\!\!\!\perp C|\mathbf{X} \Leftrightarrow T \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^{\mathrm{T}}\mathbf{X}$ and $T \perp\!\!\!\perp C|(\mathbf{B}^{\mathrm{T}}\mathbf{X}, \mathbf{X}) \Leftrightarrow T \perp\!\!\!\perp \mathbf{X}|(\mathbf{B}^{\mathrm{T}}\mathbf{X}, C)$ and $T \perp\!\!\!\perp C|\mathbf{B}^{\mathrm{T}}\mathbf{X}$. Therefore, $\mathscr{S}^{(C)}_{T|\mathbf{X}} \subseteq \mathscr{S}_{T|\mathbf{X}}$. □

Lemma 5. *Under the independent assumption $T \perp\!\!\!\perp C|\mathbf{X}$, $\mathscr{S}_{T|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$ and $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$.*

Proof. The proof follows Proposition 1 from Ebrahimi et al. (2003). Let $S_T(t) = \mathrm{pr}(T \geqslant t|\mathbf{X})$, $S_C(t) = \mathrm{pr}(C \geqslant t|\mathbf{X})$, $S_Y(t) = \mathrm{pr}(Y \geqslant t|\mathbf{X})$, and $S_\delta(t) = \mathrm{pr}(Y \geqslant t, \delta = 1|X)$. Following the proof of Ebrahimi et al. (2003), assuming that $T \perp\!\!\!\perp C|\mathbf{X}$, we have

$$S_T(t) = S_Y(t) \exp^{-\int_0^t \mathrm{d}S_\delta(u)/S_Y(u)}.$$

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

Since $\mathscr{S}_{T|X} \subseteq \mathscr{S}_{(T,C)|X}$ and $\mathscr{S}_{C|X} \subseteq \mathscr{S}_{(T,C)|X}$, consequently, $\mathscr{S}_{T|X} + \mathscr{S}_{C|X} \subseteq \mathscr{S}_{(T,C)|X}$. Also, by Lemmas 3 and 4, $\mathscr{S}_{(T,C)|X} = \mathscr{S}^{(C)}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{T|X} + \mathscr{S}_{C|X}$. Hence under the independent assumption, $\mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}$.

Also, note that (Y,δ) is a function of (T,C), therefore, $\mathscr{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathscr{S}_{(T,C)|\mathbf{X}}$. By Lemma 5, $\mathscr{S}_{T|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$ and $\mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$. Thus, $\mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$.

We have proved that

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

References

Bender, R., Augustin, T., Blettner, M., 2005. Generating survival times to simulate Cox proportional hazards models. Statist. Med. 24, 1713–1723.
Bura, E., Cook, R.D., 2001. Extending sliced inverse regression: the weighted chi-squared test. J. Amer. Statist. Assoc. 96, 996–1003.
Carroll, R., Ruppert, D., 1988. Transformation and Weighting in Regression. Chapman and Hall, London.
Cook, R.D., 1994. Using dimension-reduction subspaces to identify important inputs in models of physical systems. In: Proceedings of Section on Physical and Engineering Sciences.American Statistical Association, Alexandria, VA, pp. 18–25.
Cook, R.D., 1995. Graphics for studying net effects of regression predictors. Statist. Sinica 5, 689–708.
Cook, R.D., 1998. Regression Graphics. Wiley, New York.
Cook, R.D., 2000. SAVE: A method for dimension reduction and graphics in regression. Comm. Statist.: Theory and Methods 29, 161–175.
Cook, R.D., 2003. Dimension reduction and graphical exploration in regression including survival analysis. Statist. Med. 22, 1399–1413.
Cook, R.D., 2007. Fisher lecture: Dimension reduction in regression. Statist. Sci. 22, 1–26.
Cook, R.D., Nachtsheim, C.J., 1994. Reweighting to achieve elliptically contoured covariates in regression. J. Amer. Statist. Assoc. 89, 592–599.
Cook, R.D., Ni, L., 2005. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. J. Amer. Statist. Assoc. 100, 410–428.
Cook, R.D., Ni, L., 2006. Using intra slice covariances for improved estimation of the central subspace in regression. Biometrika 93, 65–74.
Cook, R.D., Weisberg, S., 1991. Discussion of sliced inverse regression for dimension reduction. J. Amer. Statist. Assoc. 86, 316–342.
Diaconis, P., Freedman, D., 1984. Asymptotics of graphical projection pursuit. Ann. Statist. 12, 793–815.
Dickson, E.R., Grambsch, P.M., Fleming, T.R., et al., 1989. Prognosis in primary biliary cirrhosis: model for decision making. Hepatology 10, 1–7.
Eaton, M.L., 1986. A characterization of spherical distributions. J. Multivariate Anal. 20, 272–276.
Ebrahimi, N., Molefe, D., Ying, Z., 2003. Identifiability and censored data. Biometrika 90, 724–727.
Fleming, T.R., Harrington, D.P., 1991. Counting Processes and Survival Analysis. Wiley, New York, NY.

Friedman, J.H., 1994. An overview of predictive learning and function, approximation from statistics to neural networks. In: Cherkassy, V., Friedman, J.H., Wechsler, H. (Eds.), NATO ASI Series F, vol. 136. Springer, New York.
Hall, P., Li, K.C., 1993. On almost linearity of low dimensional projections from high dimensional data. Ann. Statist. 21, 867–889.
Härdle, W., Hall, P., Marron, S., 1988. How far are automatically chosen regression smoothing parameters from their optimum. J. Amer. Statist. Assoc. 83, 86–101.
Koziol, J.A., Green, S.B., 1976. A Cramër–Von Mises statistic for randomly censored data. Biometrika 63, 465–474.
Lee, S.Y., Wolfe, R.A., 1998. A simple test for independent censoring under the proportional hazards model. Biometrics 54, 1176–1182.
Li, K.C., 1987. Asymptotic optimality for C_p, C_L cross-validation and generalized cross-validation: discrete index set. Ann. Statist. 15, 958–975.
Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342.
Li, K.C., Wang, J.L., Chen, C.H., 1999. Dimension reduction for censored regression data. Ann. Statist. 27, 1–23.
Li, K.C., Aragon, Y., Shedden, K., Agnan, C.T., 2003. Dimension reduction for multivariate response data. J. Amer. Statist. Assoc. 98, 99–109.
Li, L., Li, H., 2004. Dimension reduction methods for microarrays with applications to censored survival data. Bioinformatics 20, 3406–3412.
Li, L., Cook, R.D., Nachtsheim, C.J., 2005. Model-free variable selection. J. Roy. Statist. Soc. Ser. B 67, 285–300.
Lin, D.Y., Robins, J.M., Wei, L.J., 1996. Comparing two failure time distributions on the presence of dependent censoring. Biometrika 83, 381–393.
Prentice, R.L., Zhao, L.P., 1991. Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47, 825–838.
Prince, M.I., Jones, D.E.J., 2000. Primary biliary cirrhosis: new perspectives in diagnosis and treatment. Postgraduate Med. J. 76, 199–206.
Raftery, A., Madigan, D., Volinsky, C.T., 1995. Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance. Bayesian Statist., vol. 5. Oxford University Press, pp. 323–349.
Ruhe, A., Wedin, P.A., 1980. Algorithms for separable nonlinear least squares problems. SIAM Rev. 22, 318–337.
Scharfstein, D.O., Robins, J.M., 2002. Estimation of the failure time distribution in the presence of informative censoring. Biometrika 89, 617–634.
Shao, Y., Cook, R.D., Weisberg, S., 2007. The linearity condition and adaptive estimation in single-index regressions. Biometrika 94, 285–296.
Tsiatis, A., 1975. A nonidentifiability aspect of the problem of competing risks. Proc. Natl. Acad. Sci., USA 72, 20–22.
Velilla, S., 1998. Assessing the number of linear components in a general regression problem. J. Amer. Statist. Assoc. 93, 1088–1098.
Wen, X., Cook, R.D., 2007. Optimal sufficient dimension reduction in regressions with categorical predictors. J. Statist. Plann. Inference 137, 1961–1978.
Yoo, P., Cook, R.D., 2007. Optimal sufficient dimension reduction for the conditional mean in multivariate regression. Biometrika 94, 231–242.
Zeng, D., 2004. Estimating marginal survival function by adjusting for dependent censoring using many covariates. Ann. Statist. 32, 1533–1555.
Zheng, M., Klein, J.P., 1995. Estimates of marginal survival for dependent competing risks based on assumed copula. Biometrika 82, 127–138.