Robust Rank Canonical Correlation Analysis for Multivariate Survival Data

Di He[1] , Yong Zhou[2] and Hui Zou[3]

[1] *Nanjing University,* [2] *East China Normal University*

and [3] *University of Minnesota*

Abstract: Canonical correlation analysis (CCA) is widely applied in statistical analysis of multivariate data to find associations between two sets of multidimensional variables. However, we often cannot use CCA directly for survival data or their monotone transformations, owing to right-censoring in the data. In this paper, we propose a new robust rank CCA (RRCCA) method based on Kendall's τ correlation, and adjust it to deal with multivariate survival data, without requiring any model assumptions. Owing to the nature of rank correlation, the RRCCA is invariant against monotone transformations of the data. We establish the estimation consistency of the RRCCA approach under weak conditions. Simulation studies demonstrate the superior performance of the RRCCA in terms of estimation accuracy and empirical power. Lastly, we demonstrate the proposed method by applying it to Stanford heart transplant data.

Key words and phrases: Canonical correlation analysis; Right-censoring; Inverse probability of censoring weighting; Kendall's τ correlation.

1. Introduction

Canonical correlation analysis (CCA), introduced by Hotelling (1936), is a well-known statistical technique for finding associations between two sets of multidimensional variables. It searches for linear projections of each set of variables, such that the projected variables are maximally correlated. Extensions of the classical CCA have been proposed for particular kinds of practical data sets. For example, Akaho (2001) developed a kernel C-CA for discovering nonlinear correlations among variables, Va et al. (2007) proposed a generalization of the CCA that can handle several data sets. S-parse CCAs (Witten et al., 2009; Hardoon and Shawe-Taylor, 2011; Mai and Zhang, 2019; Chen et al., 2020) have been proposed for high-dimensional data sets, and supervised CCAs (Witten and Tibshirani, 2009; Golugula et al., 2011) are used when the two sets of variables are associated with the outcomes.

In medicine, demography, economics, and other fields, available data on the time to some event are not always exact and complete. These survival times, such as time to death, divorce, or the acceptance of a job offer, are subject to right-censoring. Here, it is important that we be able to measure the correlation between survival times, such as the times to blindness of the two eyes, and some works have been done for bivariate survival times

(Oakes, 1982; Newton and Rudel, 2007; Schemper et al., 2013; Kang et al., 2015). However, to the best of our knowledge, no works have focused on the canonical correlation between random vectors of multivariate survival data, even though such data arise naturally in many contexts. Typical examples include recurrent events data, for instance, repeated occurrences of ear infections for each individual, and clustered survival data, for instance, the possible failure of several dental fillings for an individual, or the lifetimes of related individuals in family groups (Aalen et al., 2008). Treating right-censored data as regular data in a CCA can lead to substantial bias and inaccuracy.

In this study, we address this gap in the literature by developing a new CCA method under multivariate survival data, where right-censoring can occur.First, we handle the censoring by using the inverse probability of censoring weighting technique. Next, we construct the proposed robust rank CCA (RRCCA) method based on Kendall's τ correlation, and adjust it to deal with multivariate survival data, without requiring any model assumptions. Owing to the nature of a rank correlation, the RRCCA is invariant against monotone transformations of any of the variables. This is a nice property, because survival times are often modeled using accelerated failure time models with a logarithmic transformation, or using transformation

models with an unknown increasing function. In addition, we establish the estimation consistency of the RRCCA approach under weak conditions.

The rest of the paper is organized as follows. Section 2 gives a brief review of the classical CCA. In Section 3, we present the RRCCA approach, and establish its estimation consistency is in Section 4. In Section 5, we use simulation studies to examine the estimation accuracy and empirical power of the proposed approach, and demonstrate it using Stanford heart transplant data in Section 6. Section 7 concludes the paper.

2. A Brief Review of CCA

For two random vectors $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \ldots, \widetilde{X}_p) \in \mathbb{R}^p$ and $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \ldots, \widetilde{Y}_q) \in \mathbb{R}^q$, denote the covariance matrix of $(\widetilde{\mathbf{X}}^{\mathrm{T}}, \widetilde{\mathbf{Y}}^{\mathrm{T}})^{\mathrm{T}}$ as

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_{XX}} & \mathbf{\Sigma_{XY}} \\ \mathbf{\Sigma_{YX}} & \mathbf{\Sigma_{YY}} \end{pmatrix}.$$

Note that we remove the tilde symbol from the subscript for simplicity of notation throughout the remainder of the paper. A CCA seeks vectors $\boldsymbol{a} \in \mathbb{R}^p$ and $\boldsymbol{b} \in \mathbb{R}^q$ that maximize $\rho^c = \mathrm{Cor}(\boldsymbol{a}^{\mathrm{T}}\widetilde{\mathbf{X}}, \boldsymbol{b}^{\mathrm{T}}\widetilde{\mathbf{Y}})$; that is

$$\rho^c = \max_{\boldsymbol{a},\boldsymbol{b}} \frac{\boldsymbol{a}^{\mathrm{T}}\mathbf{\Sigma_{XY}}\boldsymbol{b}}{\sqrt{\boldsymbol{a}^{\mathrm{T}}\mathbf{\Sigma_{XX}}\boldsymbol{a}}\sqrt{\boldsymbol{b}^{\mathrm{T}}\mathbf{\Sigma_{YY}}\boldsymbol{b}}}. \tag{2.1}$$

The correlation ρ^c is called the first canonical correlation, and the vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are the first pair of canonical vectors. If $\mathbf{\Sigma_{XX}}$ and $\mathbf{\Sigma_{YY}}$ are invertible,

ρ^c is equal to the square root of the largest eigenvalue of the matrices

$$\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}\Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}} \quad \text{and} \quad \Sigma_{\mathbf{YY}}^{-1}\Sigma_{\mathbf{YX}}\Sigma_{\mathbf{XX}}^{-1}\Sigma_{\mathbf{XY}}, \quad (2.2)$$

and $\boldsymbol{a}$ and $\boldsymbol{b}$ are the respective eigenvectors of (2.2) corresponding to the largest eigenvalue.

In practice, the sample covariance matrices of the observed data $(\widetilde{\mathbf{X}}^{(1)}, \widetilde{\mathbf{Y}}^{(1)})$, $\ldots, (\widetilde{\mathbf{X}}^{(n)}, \widetilde{\mathbf{Y}}^{(n)}) \in \mathbb{R}^p \times \mathbb{R}^q$ are computed, and the canonical vectors and correlations are obtained based on the eigenvectors and the eigenvalues of the sample covariance matrices

$$\widehat{\Sigma}_{\mathbf{XX}}^{-1}\widehat{\Sigma}_{\mathbf{XY}}\widehat{\Sigma}_{\mathbf{YY}}^{-1}\widehat{\Sigma}_{\mathbf{YX}} \quad \text{and} \quad \widehat{\Sigma}_{\mathbf{YY}}^{-1}\widehat{\Sigma}_{\mathbf{YX}}\widehat{\Sigma}_{\mathbf{XX}}^{-1}\widehat{\Sigma}_{\mathbf{XY}}. \quad (2.3)$$

3. Methodology

In practice, in addition to covariates, some or all of the components of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ may be survival times, or some monotone transformations of survival times, such as logarithmic survival times. Hence, in general, we assume each component varies from $-\infty$ to ∞. We show later that our proposed method is unaffected by monotone transformations. Because the survival times are subject to right-censoring, without loss of generality, we denote C_k as the censoring variable corresponding to $\widetilde{X}_k$, and D_l as the censoring variable of $\widetilde{Y}_l$, for $k = 1, \ldots, p$ and $l = 1, \ldots, q$. Let the censoring variable be positive infinity if there exists no censoring for some

$\widetilde{X}_k$ or $\widetilde{Y}_l$. We assume both C_k and D_l are independent of $\widetilde{X}_k$ and $\widetilde{Y}_l$.

For simplicity, we further assume that C_k and D_l are independent, as are

C_k and $C_{k'}$ and D_l and $D_{l'}$. Let $X_k^{(i)} = \min(\widetilde{X}_k^{(i)}, C_k^{(i)}), \delta_k^{(i)} = I(\widetilde{X}_k^{(i)} \leq$

$C_k^{(i)})$; $Y_l^{(i)} = \min(\widetilde{Y}_l^{(i)}, D_l^{(i)}), \phi_l^{(i)} = I(\widetilde{Y}_l^{(i)} \leq D_l^{(i)})$, for $i = 1, \ldots, n$. Then,

the independent and identically distributed (i.i.d.) sample we observe is

$(\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}, \boldsymbol{\delta}^{(i)}, \boldsymbol{\phi}^{(i)})$, where $\mathbf{X}^{(i)} = (X_1^{(i)}, \ldots, X_p^{(i)}), \mathbf{Y}^{(i)} = (Y_1^{(i)}, \ldots, Y_q^{(i)})$,

$\boldsymbol{\delta}^{(i)} = (\delta_1^{(i)}, \ldots, \delta_p^{(i)}), \boldsymbol{\phi}^{(i)} = (\phi_1^{(i)}, \ldots, \phi_q^{(i)})$.

3.1 Inverse probability of censoring weighting

Let $S_{C_k}(t) = \Pr(C_k > t)$ and $S_{D_l}(t) = \Pr(D_l > t)$ be the survival

functions of C_k and D_l, respectively. We obtain the following unbiased

variances and covariances in (2.2):

$$\mathrm{Cov}(\widetilde{X}_k, \widetilde{Y}_l) = E\{\frac{\delta_k}{S_{C_k}(X_k)} \frac{\phi_l}{S_{D_l}(Y_l)} X_k Y_l\} - E\{\frac{\delta_k}{S_{C_k}(X_k)} X_k\} E\{\frac{\phi_l}{S_{D_l}(Y_l)} Y_l\},$$

$$\mathrm{Cov}(\widetilde{X}_k, \widetilde{X}_{k'}) = \begin{cases} E\{\frac{\delta_k}{S_{C_k}(X_k)} \frac{\delta_{k'}}{S_{C_{k'}}(X_{k'})} X_k X_{k'}\} - E\{\frac{\delta_k}{S_{C_k}(X_k)} X_k\} E\{\frac{\delta_{k'}}{S_{C_{k'}}(X_{k'})} X_{k'}\}, & k \neq k'; \\ E\{\frac{\delta_k}{S_{C_k}(X_k)} X_k^2\} - [E\{\frac{\delta_k}{S_{C_k}(X_k)} X_k\}]^2, & k = k', \end{cases}$$

$$\mathrm{Cov}(\widetilde{Y}_l, \widetilde{Y}_{l'}) = \begin{cases} E\{\frac{\phi_l}{S_{D_l}(Y_l)} \frac{\phi_{l'}}{S_{D_{l'}}(Y_{l'})} Y_l Y_{l'}\} - E\{\frac{\phi_l}{S_{D_l}(Y_l)} Y_l\} E\{\frac{\phi_{l'}}{S_{D_{l'}}(Y_{l'})} Y_{l'}\}, & l \neq l'; \\ E\{\frac{\phi_l}{S_{D_l}(Y_l)} Y_l^2\} - [E\{\frac{\phi_l}{S_{D_l}(Y_l)} Y_l\}]^2, & l = l'. \end{cases}$$

Their sample versions can be estimated using corresponding moment esti-

mators. For example, $\mathrm{Cov}(\widetilde{X}_k, \widetilde{Y}_l)$ can be estimated by

$$\frac{1}{n} \sum_{i=1}^n \frac{\delta_k^{(i)}}{\widehat{S}_{C_k}(X_k^{(i)})} \frac{\phi_l^{(i)}}{\widehat{S}_{D_l}(Y_l^{(i)})} X_k^{(i)} Y_l^{(i)} - \left\{\frac{1}{n} \sum_{i=1}^n \frac{\delta_k^{(i)}}{\widehat{S}_{C_k}(X_k^{(i)})} X_k^{(i)}\right\} \left\{\frac{1}{n} \sum_{i=1}^n \frac{\phi_l^{(i)}}{\widehat{S}_{D_l}(Y_l^{(i)})} Y_l^{(i)}\right\},$$

where $\widehat{S}_{C_k}(t)$ and $\widehat{S}_{D_l}(t)$ are Kaplan–Meier estimators of S_{C_k} and S_{D_l}, respectively. Specifically, $\widehat{S}_{C_k}(t) = \prod_{s \leq t}(1 - \Delta N_{C_k}(s)/\overline{G_k}(s))$, where $N_{C_k}(t) = \sum_{i=1}^{n} N_{C_k}^{(i)}(t)$, with $N_{C_k}^{(i)}(t) = I(X_k^{(i)} \leq t, \delta_k^{(i)} = 0)$, and $\overline{G_k}(t) = \sum_{i=1}^{n} G_k^{(i)}(t)$, with $G_k^{(i)}(t) = I(X_k^{(i)} \geq t)$; $\widehat{S}_{D_l}(t)$ is defined similarly. Hence, a CCA can still use the sample estimators of the above adjusted variances and covariances in (2.3). For simplicity of presentation, we refer to the CCA with inverse probability of censoring weighting approach as CCA-IPCW.

Remark 1. The CCA-IPCW approach may not be robust for survival data, because the sample covariance matrices are used in solving the eigenproblem. This procedure is optimal for the classical CCA under a multivariate normal distribution, but is less efficient with heavier-tailed model distributions. As shown by (Romanazzi, 1992), the sample covariance matrices are highly sensitive to outliers, and a canonical analysis based on these matrices will yield unreliable results. The CCA-IPCW is also not invariant against a monotone transformation. We demonstrate these drawbacks in our simulation results in Section 5. Moreover, though some versions of the CCA are robust, for example, using the minimum covariance determinant estimator (Rousseeuw, 1985; Croux and Dehon, 2002) or the robust alternating regressions method (Filzmoser et al., 2000; Branco et al., 2005), they cannot be applied directly to right-censored survival data.

3.2 RRCCA

In survival analysis, a common approach to assess the effects of co-
variates on survival is to use a semiparametric regression model, such as
the transformation model. The transformation model, which includes the
proportional hazards model and the proportional odds model as special cas-
es, assumes that an unknown monotone transformation of the underlying
failure time is linearly related to the covariates with various error distri-
butions. When multivariate survival data are available, the analysis model
becomes more complicated, such as the multivariate frailty model (Aalen
et al., 2008). Therefore, we need a simpler statistical tool that is both
robust and invariant against monotone data transformations.

We propose a new RRCCA based on Kendall's τ correlation, without
needing any model assumptions. Because (2.1) is scale free, the problem
is equivalent if we replace the covariance matrices with their corresponding
correlation matrices. Hence, in this case, the solution (2.2) is no different.
We then replace the Pearson correlations of every pair of random variables
in the correlation matrices with Kendall's τ correlations, which are more
robust. When right-censoring occurs, it is necessary to adjust these rank
correlations to avoid severe bias. Specifically, using the inverse probabil-
ity of censoring weighting technique, we construct the following unbiased

Kendall's τ correlations for the survival data:

$$\widehat{\tau}_{X_k Y_l} = \frac{4}{n(n-1)} \sum_{i \neq j}^n \frac{\delta_k^{(j)}}{\widehat{S}_{C_k}^2(X_k^{(j)})} \frac{\phi_l^{(j)}}{\widehat{S}_{D_l}^2(Y_l^{(j)})} I(X_k^{(i)} > X_k^{(j)}) I(Y_l^{(i)} > Y_l^{(j)}) - 1,$$

$$\widehat{\tau}_{X_k X_{k'}} = \frac{4}{n(n-1)} \sum_{i \neq j}^n \frac{\delta_k^{(j)}}{\widehat{S}_{C_k}^2(X_k^{(j)})} \frac{\delta_{k'}^{(j)}}{\widehat{S}_{C_{k'}}^2(X_{k'}^{(j)})} I(X_k^{(i)} > X_k^{(j)}) I(X_{k'}^{(i)} > X_{k'}^{(j)}) - 1,$$

$$\widehat{\tau}_{Y_l Y_{l'}} = \frac{4}{n(n-1)} \sum_{i \neq j}^n \frac{\phi_l^{(j)}}{\widehat{S}_{D_l}^2(Y_l^{(j)})} \frac{\phi_{l'}^{(j)}}{\widehat{S}_{D_{l'}}^2(Y_l^{(j)})} I(Y_l^{(i)} > Y_l^{(j)}) I(Y_{l'}^{(i)} > Y_{l'}^{(j)}) - 1.$$

In this way, we can perform an RRCCA by solving the eigenvalues and eigenvectors of the following matrices:

$$\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1} \widehat{\mathbf{S}}_{\mathbf{XY}} \widehat{\mathbf{S}}_{\mathbf{YY}}^{-1} \widehat{\mathbf{S}}_{\mathbf{YX}}, \tag{3.1}$$

where

$$\widehat{\mathbf{S}}_{\mathbf{XX}} = (\widehat{\tau}_{X_k X_{k'}})_{p \times p}, \quad \widehat{\mathbf{S}}_{\mathbf{XY}} = (\widehat{\tau}_{X_k Y_l})_{p \times q}, \quad \text{and} \quad \widehat{\mathbf{S}}_{\mathbf{YY}} = (\widehat{\tau}_{Y_l Y_{l'}})_{q \times q} \tag{3.2}$$

are sample versions of Kendall's τ correlation matrices. We denote the maximum canonical correlation obtained this way as $\widehat{\tau}^c$. A similar inverse probability of censoring weighting technique for Kendall's τ correlation is used in Song et al. (2014), who consider the correlation measure between a right-censored response and a regular covariate, both univariate, in the variable screening problem. The RRCCA is invariant under any coordinatewise monotone data transformations, unlike the CCA-IPCW. We discuss the benefits of using a robust rank correlation in our numerical experiments.

4. Theory

In this section, we prove the estimation consistency of the RRCCA under general conditions. For two random variables $\widetilde{U}$ and $\widetilde{V}$ from a joint distribution, let $(\widetilde{U}_1, \widetilde{V}_1)$ and $(\widetilde{U}_2, \widetilde{V}_2)$ be two independent realizations without censoring. Then, the population Kendall's rank correlation is $\tau_{UV} = \text{Cov}\left\{\text{sgn}(\widetilde{U}_1 - \widetilde{U}_2), \text{sgn}(\widetilde{V}_1 - \widetilde{V}_2)\right\}$, and

$$\mathbf{S_{XX}} = (\tau_{X_k X_{k'}})_{p \times p}, \quad \mathbf{S_{XY}} = (\tau_{X_k Y_l})_{p \times q}, \quad \text{and} \quad \mathbf{S_{YY}} = (\tau_{Y_l Y_{l'}})_{q \times q}$$

are the population Kendall's rank correlation matrices that (3.2) estimates for. We denote the matrix spectral norm $\|\mathbf{A}\| \equiv \sup\{\|\mathbf{A}X\| : \|X\| = 1\} = \lambda_{\max}^{1/2}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$ for any matrix $\mathbf{A}$, which for symmetric matrices reduces to $\|\mathbf{A}\| = \max_i |\lambda_i(\mathbf{A})|$.

The following conditions are required:

(C1) There exist positive constants κ and ω, such that $\min\{\lambda_{\min}(\mathbf{S_{XX}}), \lambda_{\min}(\mathbf{S_{YY}})\} \geq \kappa$, $\lambda_{\max}^{1/2}(\mathbf{S_{XY}^{\mathrm{T}}}\mathbf{S_{XY}}) \leq 1/\omega$.

(C2) There exist positive constants u_k and v_l for every $k = 1, \ldots, p$ and $l = 1, \ldots, q$, such that $\Pr(C_k = u_k) > 0$, $\Pr(D_l = v_l) > 0$, and $\Pr(C_k > u_k) = \Pr(D_l > v_l) = 0$.

Condition (C1) ensures that the rank-based correlation matrices are well conditioned; similar conditions can be found in Bickel and Levina

(2008). Condition (C2) is common in the survival literature (Peng and Fine, 2009; Song et al., 2014) for asymptotic analysis, and is satisfied in many clinical settings. Note that both p and q can vary with the sample size n in the theorem.

Theorem 1. *Under Conditions (C1) and (C2), there exists a positive constant M, where, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, $p^2 \log np^2 = o(n)$, $q^2 \log nq^2 = o(n)$, and $pq \log npq = o(n)$ hold, then we have*

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\widehat{\mathbf{S}}_{\mathbf{YX}} - \mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{YX}}\| = o_p(1).$$

Theorem 1 confirms the consistency of the product of rank-based sample correlation matrices, and hence the canonical correlations and vectors obtained thereafter are also consistent. The technical proofs are relegated to the Appendix.

5. Simulation

5.1 Estimation accuracy

In this section, we use simulations to examine the estimation accuracy of CCA-IPCW and RRCCA. We consider two sample sizes, $n = 100$ and 200, and the number of replications is $M = 1000$. We set the same censoring rates for all components of the two random vectors in our simulation, and

10% and 30% censoring rates are considered.

The data-generating model is similar to those in Branco et al. (2005) and Taskinen et al. (2006), although we add a censoring mechanism. First, the following sampling distributions of $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ are considered: (i) a normal distribution: $N_{p+q}(\mathbf{0}, \boldsymbol{\Sigma})$; (ii) a multivariate t distribution with three degrees of freedom and scatter parameter $\boldsymbol{\Sigma}$; (iii) a normal distribution with contamination: $0.95 N_{p+q}(\mathbf{0}, \boldsymbol{\Sigma}) + 0.05 N_{p+q}(\mathbf{0}, 9\boldsymbol{\Sigma})$; and (iv) a lognormal distribution: $\log \widetilde{\mathbf{X}}$ and $\log \widetilde{\mathbf{Y}}$ are generated from $N_{p+q}(\mathbf{0}, \boldsymbol{\Sigma})$. We take the covariance matrices of $(\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}})$ $((\log \widetilde{\mathbf{X}}, \log \widetilde{\mathbf{Y}})$ for case (iv)) as

$$
\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I}_p & \mathbf{R} \\ \mathbf{R} & \mathbf{I}_p \end{pmatrix},
$$

where $\mathbf{R} = diag(\rho_1, \ldots, \rho_p)$. Our choices are (a) $\rho_1 = 0.8, \rho_2 = 0.2$, (b) $\rho_1 = 0.6, \rho_2 = 0.4$, and (c) $\rho_1 = 0.9, \rho_2 = 0.6, \rho_3 = 0.3$.

Next, we independently generate the censoring variables C_k and D_l as $C_k = \log G_k$ and $D_l = \log H_l$ for distributions (i)–(iii), and $C_k = G_k$ and $D_l = H_l$ for distribution (iv). Here, G_k and H_l are uniform random variables, defined on the intervals $(0, g_k)$ and $(0, h_l)$, respectively, where g_k and h_l are chosen to achieve the desired censoring ratio. Finally, $X_k^{(i)} = \min(\widetilde{X}_k^{(i)}, C_k^{(i)}), \delta_k^{(i)} = I(\widetilde{X}_k^{(i)} \leq C_k^{(i)})$ and $Y_l^{(i)} = \min(\widetilde{Y}_l^{(i)}, D_l^{(i)}), \phi_l^{(i)} = I(\widetilde{Y}_l^{(i)} \leq D_l^{(i)})$.

We need to find the population target of estimation in order to check the estimation accuracy.

The population canonical correlations: We use ρ_j^c and τ_j^c to denote the population canonical correlations for CCA-IPCW and RRCCA, respectively. For (i) the normal distribution or (iii) the contaminated normal case, by (2.2), ρ_j^c is just ρ_j. Furthermore, using $\tau = 2\pi^{-1}\arcsin\rho$, the well-known relations for the bivariate normal distribution with linear correlation ρ, we have that τ_j^c is $2\pi^{-1}\arcsin\rho_j$. Under (ii) the multivariate t distribution, ρ_j^c is still ρ_j, and τ_j^c remains $2\pi^{-1}\arcsin\rho_j$, because the relation between Kendall's τ and the linear correlation ρ holds more generally for all elliptical distributions with continuous marginals, including the bivariate Student t distribution (Lindskog et al., 2003). For (iv) the lognormal distribution, the invariance property against monotone transformations of rank correlations means that τ_j^c is still the same as that under the normal distribution. It can be shown that the variance matrix of $\widetilde{\mathbf{Z}} = (e^{\widetilde{\mathbf{X}}}, e^{\widetilde{\mathbf{Y}}})$ is $\mathrm{Var}(\widetilde{\mathbf{Z}}) = \{E(\widetilde{\mathbf{Z}})E(\widetilde{\mathbf{Z}})^{\mathrm{T}}\} \circ \{e^{\boldsymbol{\Sigma}} - \mathbf{1}_{(p+q)\times(p+q)}\}$, where $\circ$ represents element-wise multiplication. Hence, a simple calculation shows that ρ_j^c is equal to $(e^{\rho_j} - 1)/(e - 1)$.

The population canonical vectors: In most cases, the canonical vectors for an RRCCA differ from those for a CCA-IPCW, because they

have different population covariance matrices. However, in this example, it
is easy to see that these matrices are all diagonal, by the above argument,
and that the canonical vectors are the same unit vectors.

To assess the estimation accuracy, we adopt the following two criteria.

**Criterion 1. The mean squared error (MSE) when estimating the
direction of a canonical vector** (Branco et al., 2005). For the jth
canonical vector for $\mathbf{X}$, the MSE is measured by

$$MSE(\widehat{\boldsymbol{a}}_j) = \frac{1}{M} \sum_{m=1}^{M} \arccos \left\{ \frac{|\boldsymbol{a}_j^{\mathrm{T}} \widehat{\boldsymbol{a}}_j^{(m)}|}{\|\boldsymbol{a}_j\| \|\widehat{\boldsymbol{a}}_j^{(m)}\|} \right\},$$

where $\boldsymbol{a}_j$ is the jth population canonical vector for $\mathbf{X}$, and $\widehat{\boldsymbol{a}}_j^{(m)}$ is the es-
timate obtained from the mth generated sample. Using angles makes the
MSE invariant to whether we choose the standardized or unstandardized
canonical vectors. Because the results of $MSE(\widehat{\boldsymbol{b}}_j)$ for $\mathbf{Y}$ are similar in
our example, we omit them to conserve space. **Criterion 2. The mean
squared relative error (MSRE) when estimating the magnitude
of a canonical correlation** (Kudraszow and Maronna, 2011). For the
jth canonical correlation, the MSRE compares the jth canonical correla-
tion estimate with its corresponding theoretical one. Note that the two
approaches estimate different population quantities of the canonical corre-

lation. Therefore, we compute

$$MSRE(\widehat{\rho}_j^c) = \frac{1}{M} \sum_{m=1}^{M} \left\{ \frac{(\widehat{\rho}_j^c)^{(m)} - \rho_j^c}{\rho_j^c} \right\}^2,$$

$$MSRE(\widehat{\tau}_j^c) = \frac{1}{M} \sum_{m=1}^{M} \left\{ \frac{(\widehat{\tau}_j^c)^{(m)} - \tau_j^c}{\tau_j^c} \right\}^2,$$

where $(\widehat{\rho}_j^c)^{(m)}$ and $(\widehat{\tau}_j^c)^{(m)}$ are the corresponding estimates computed from

the mth replication.

The results of the simulation are presented in Tables 1 and 2, and show

that the RRCCA outperforms the CCA-IPCW in all cases. The advantage

of using the RRCCA is most evident under the lognormal distribution. The

RRCCA performs similarly under different distributions.

5.2 Empirical power of permutation test based on maximum CCA

A classical application of a CCA is to test the independence of two

sets of variables. If $\widetilde{\mathbf{X}}$ is independent of $\widetilde{\mathbf{Y}}$, then $\mathbf{\Sigma_{XY}}$ is $\mathbf{0}$, and all orders

of canonical correlations are consequently zero. For the multivariate nor-

mal distribution, the independence and the maximum canonical correlation

being zero are equivalent. For other distributions, the conclusion of depen-

dence can be drawn if the maximum canonical correlation is not zero, but

independence cannot be inferred, even if the latter is zero. Therefore, to

some extent, the maximum canonical correlation can be used to test the

independence of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$. We propose using the maximum RRCCA as the

Table 1: The mean squared error (MSE) of estimating the direction of the

jth canonical vector

n	censoring	distribution	methods	Case (a)		Case (b)		Case (c)		
				$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$	$j=3$
100	10%	normal	CCA-IPCW	0.12	0.14	0.37	0.39	0.22	0.34	0.31
			RRCCA	0.07	0.10	0.27	0.28	0.10	0.23	0.23
		$t(3)$	CCA-IPCW	0.16	0.21	0.46	0.49	0.29	0.52	0.47
			RRCCA	0.08	0.11	0.30	0.32	0.11	0.27	0.27
		contaminated normal	CCA-IPCW	0.13	0.16	0.40	0.42	0.23	0.40	0.36
			RRCCA	0.08	0.11	0.27	0.29	0.10	0.24	0.24
		lognormal	CCA-IPCW	0.36	0.36	0.62	0.60	0.65	0.85	0.70
			RRCCA	0.07	0.10	0.26	0.28	0.10	0.23	0.23
	30%	normal	CCA-IPCW	0.21	0.21	0.47	0.48	0.43	0.59	0.48
			RRCCA	0.11	0.13	0.34	0.36	0.18	0.33	0.31
		$t(3)$	CCA-IPCW	0.19	0.24	0.50	0.54	0.34	0.60	0.52
			RRCCA	0.12	0.15	0.37	0.39	0.17	0.36	0.35
		contaminated normal	CCA-IPCW	0.20	0.22	0.50	0.51	0.37	0.56	0.49
			RRCCA	0.11	0.14	0.35	0.36	0.17	0.34	0.32
		lognormal	CCA-IPCW	0.71	0.70	0.74	0.73	0.92	0.96	0.93
			RRCCA	0.11	0.14	0.35	0.36	0.17	0.34	0.32
200	10%	normal	CCA-IPCW	0.08	0.10	0.26	0.27	0.15	0.25	0.22
			RRCCA	0.05	0.07	0.18	0.19	0.06	0.16	0.16
		$t(3)$	CCA-IPCW	0.12	0.16	0.38	0.41	0.20	0.40	0.37
			RRCCA	0.06	0.08	0.20	0.22	0.07	0.18	0.18
		contaminated normal	CCA-IPCW	0.09	0.12	0.31	0.33	0.17	0.30	0.27
			RRCCA	0.05	0.07	0.19	0.21	0.07	0.16	0.16
		lognormal	CCA-IPCW	0.29	0.29	0.57	0.57	0.62	0.80	0.58
			RRCCA	0.05	0.07	0.17	0.19	0.07	0.16	0.16
	30%	normal	CCA-IPCW	0.15	0.16	0.40	0.41	0.32	0.44	0.36
			RRCCA	0.07	0.09	0.24	0.26	0.11	0.21	0.21
		$t(3)$	CCA-IPCW	0.15	0.21	0.44	0.49	0.26	0.48	0.45
			RRCCA	0.08	0.11	0.26	0.28	0.12	0.25	0.24
		contaminated normal	CCA-IPCW	0.15	0.16	0.42	0.44	0.29	0.44	0.38
			RRCCA	0.07	0.10	0.24	0.26	0.11	0.23	0.22
		lognormal	CCA-IPCW	0.67	0.64	0.76	0.76	0.88	0.97	0.91
			RRCCA	0.07	0.09	0.23	0.24	0.11	0.22	0.21

Table 2: The mean squared relative error (MSRE) of estimating the magnitude of the jth canonical correlation

n	censoring	distribution	methods	Case (a)		Case (b)		Case (c)		
				$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$	$j=3$
100	10%	normal	CCA-IPCW	0.01	0.30	0.03	0.08	0.01	0.03	0.15
			RRCCA	0.01	0.26	0.02	0.06	0.00	0.02	0.11
		$t(3)$	CCA-IPCW	0.02	0.54	0.04	0.14	0.01	0.04	0.23
			RRCCA	0.01	0.28	0.02	0.08	0.00	0.02	0.15
		contaminated normal	CCA-IPCW	0.01	0.35	0.03	0.10	0.01	0.03	0.17
			RRCCA	0.01	0.26	0.02	0.06	0.00	0.02	0.13
		lognormal	CCA-IPCW	0.12	1.27	0.32	0.40	0.06	0.19	0.44
			RRCCA	0.01	0.25	0.02	0.06	0.00	0.02	0.12
	30%	normal	CCA-IPCW	0.03	0.48	0.07	0.20	0.02	0.06	0.25
			RRCCA	0.02	0.34	0.03	0.12	0.01	0.03	0.18
		$t(3)$	CCA-IPCW	0.02	0.64	0.06	0.19	0.01	0.05	0.25
			RRCCA	0.02	0.35	0.03	0.15	0.01	0.04	0.23
		contaminated normal	CCA-IPCW	0.03	0.48	0.06	0.18	0.02	0.05	0.24
			RRCCA	0.02	0.34	0.03	0.12	0.01	0.03	0.20
		lognormal	CCA-IPCW	0.18	3.32	0.67	0.51	0.17	0.45	0.64
			RRCCA	0.02	0.33	0.03	0.12	0.01	0.03	0.20
200	10%	normal	CCA-IPCW	0.01	0.20	0.02	0.05	0.00	0.01	0.08
			RRCCA	0.00	0.14	0.01	0.03	0.00	0.01	0.06
		$t(3)$	CCA-IPCW	0.01	0.38	0.03	0.10	0.01	0.03	0.15
			RRCCA	0.00	0.17	0.01	0.04	0.00	0.01	0.07
		contaminated normal	CCA-IPCW	0.01	0.24	0.02	0.06	0.00	0.02	0.11
			RRCCA	0.00	0.15	0.01	0.04	0.00	0.01	0.07
		lognormal	CCA-IPCW	0.11	1.21	0.23	0.33	0.06	0.16	0.41
			RRCCA	0.00	0.14	0.01	0.03	0.00	0.01	0.06
	30%	normal	CCA-IPCW	0.02	0.35	0.05	0.14	0.02	0.04	0.20
			RRCCA	0.01	0.21	0.02	0.06	0.01	0.02	0.10
		$t(3)$	CCA-IPCW	0.02	0.49	0.04	0.12	0.01	0.03	0.18
			RRCCA	0.01	0.25	0.02	0.08	0.01	0.02	0.15
		contaminated normal	CCA-IPCW	0.02	0.34	0.04	0.12	0.01	0.04	0.18
			RRCCA	0.01	0.21	0.02	0.07	0.01	0.02	0.12
		lognormal	CCA-IPCW	0.17	2.07	0.50	0.43	0.13	0.32	0.50
			RRCCA	0.01	0.22	0.02	0.06	0.01	0.02	0.10

test statistic, and using the permutation test procedure to set the critical value. We consider a simulation design to check the type-I error rate and the power of the test with a significance level of 0.05, based on 2000 Monte Carlo replications.

We generate the data from the same model as in the previous subsection, except with a different $\mathbf{R} = diag(\rho_1, \rho_2)$. We use $\rho_1 = \rho_2 = 0, 0.1, 0.3$, and 0.5.

When $\rho_1 = \rho_2 = 0$, $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ are independent for distributions (i), (iii), and (iv), but not for the t distribution. Thus, we leave the t distribution out of this simulation study. The type-I error rates are given in Table 3. When ρ_1 and ρ_2 are not zero, $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{Y}}$ are correlated and dependent; the power is given in Table 4.

Table 3 shows that the CCA-IPCW has the correct type-I error under the normal distribution and the lognormal distribution with a low censoring rate, but that this error becomes inflated for the contaminated normal distribution, and is incorrect for the lognormal distribution with a higher censoring rate. Thus, its power becomes meaningless under the latter two distributions. In contrast, the RRCCA has the correct type-I error under all scenarios. Moreover, the RRCCA has better power than that of the CCA-IPCW in all cases. Not surprisingly, the power is monotone with

Table 3: Type-I error rate of the permutation test based on maximum CCA

n	censoring	method	normal	Distribution contaminated normal	lognormal
100	10%	CCA-IPCW	0.052	0.130	0.050
		RRCCA	0.051	0.062	0.054
	30%	CCA-IPCW	0.046	0.091	0.004
		RRCCA	0.044	0.069	0.050
200	10%	CCA-IPCW	0.051	0.151	0.048
		RRCCA	0.047	0.078	0.046
	30%	CCA-IPCW	0.044	0.098	0.007
		RRCCA	0.042	0.060	0.050

respect to ρ. When $\rho = 0.5$, the RRCCA has near perfect power, even with $n = 100$ and a 30% censoring rate. When $\rho = 0.1$, the power is much lower, indicating the difficulty of the testing problem, even under a sample size of $n = 200$. The more interesting case is $\rho = 0.3$, where the RRCCA still has relatively high power for $n = 100$ and a censoring rate of 10%.

6. Real-Data Analysis

Here, we demonstrate the proposed method by using it to analyze Stanford heart transplant data. The original data set can be found in Crowley and Hu (1977), and is reproduced by Kalbfleisch and Prentice (1980). We focus on the data of 69 patients who waited for a donor heart and received a transplantation. Denote the waiting time, in days, for these patients as Y_1, and their post-transplant survival as Y_2. The waiting times are all uncen-

Table 4: Power of the permutation test based on maximum CCA

n	censoring	ρ	method	Distribution		
				normal	contaminated normal	lognormal
100	10%	0.1	CCA-IPCW	0.100	0.225	0.062
			RRCCA	0.126	0.156	0.122
		0.3	CCA-IPCW	0.621	0.676	0.160
			RRCCA	0.826	0.838	0.814
		0.5	CCA-IPCW	0.984	0.980	0.360
			RRCCA	1.000	1.000	1.000
	30%	0.1	CCA-IPCW	0.068	0.133	0.006
			RRCCA	0.060	0.068	0.064
		0.3	CCA-IPCW	0.207	0.334	0.007
			RRCCA	0.470	0.475	0.462
		0.5	CCA-IPCW	0.564	0.659	0.007
			RRCCA	0.962	0.967	0.969
200	10%	0.1	CCA-IPCW	0.172	0.334	0.092
			RRCCA	0.248	0.261	0.225
		0.3	CCA-IPCW	0.923	0.929	0.237
			RRCCA	0.994	0.990	0.994
		0.5	CCA-IPCW	1.000	1.000	0.487
			RRCCA	1.000	1.000	1.000
	30%	0.1	CCA-IPCW	0.077	0.165	0.007
			RRCCA	0.097	0.114	0.101
		0.3	CCA-IPCW	0.340	0.528	0.013
			RRCCA	0.875	0.832	0.854
		0.5	CCA-IPCW	0.782	0.872	0.014
			RRCCA	1.000	1.000	1.000

sored among these patients, and 24 are censored for their post-transplant survival times. We include two covariates in the analysis: age at acceptance into the program (X_1), and the mismatch score (X_2). The latter is one of three measures of the degree to which a donor and a recipient are mismatched for tissue type, and is the only one found to be useful in previous analyses (Aitkin et al., 1983). Four of the transplanted patients have incomplete data on mismatch score, and so are not used in our analysis.

We first test the independence between (X_1, X_2) and (Y_1, Y_2). A permutation test of the CCA-IPCW with a significance level of 0.05 does not reject the hypothesis of independence. However, the corresponding RRCCA with the same significance level rejects the hypothesis of independence, which partially coincides with the dependence conclusion between Y_1 and Y_2 in Shih and Louis (1996). The histograms in Figure 1 show that the two survival times are heavy-tailed, and that the CCA-IPCW may not be robust. Even if we apply a log transformation to Y_1 and Y_2, the results of the permutation tests remain the same.

Next, we conduct a canonical analysis. The first and second canonical correlations for the RRCCA approach are 0.34 and 0.08, respectively. The coefficients of the canonical vectors, often used to interpret the canonical variates, are given in Table 5. It can be seen that Y_2 has a mild association
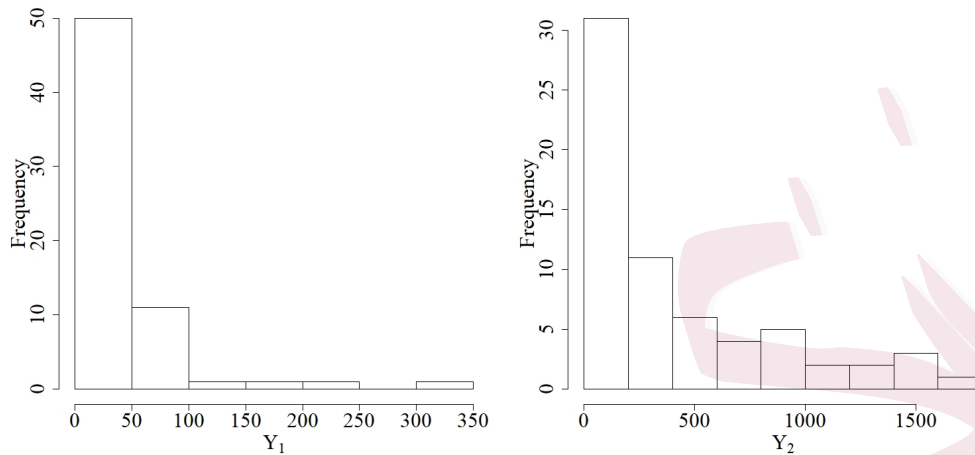
Figure 1: Histograms of Y_1 and Y_2 in the real-data example.

with X_1 and X_2, where $\widehat{\boldsymbol{a}}_1$ is determined mainly by X_1, and to a lesser extent, by X_2. Thus an older age and a greater mismatch score of tissue type may result in a shorter post-transplant survival.

Table 5: Canonical vectors of RRCCA in the real-data example

$\widehat{\boldsymbol{a}}_1$:	0.893	0.450	$\widehat{\boldsymbol{b}}_1$:	0.025	-1.000
$\widehat{\boldsymbol{a}}_2$:	-0.521	0.854	$\widehat{\boldsymbol{b}}_2$:	-0.984	0.178

7. Conclusion

We have used the inverse probability of censoring weighting and Kendall's rank correlation to construct an RRCCA for survival data with right-censoring. We have demonstrated the benefits of using the rank cor-

relation by means of simulations. We could also consider a version of the RRCCA approach based on Spearman's rank correlation, where we replace the unbiased Kendall's rank correlations in (3.1) with adjusted Spearman's rank correlations. Specifically, $\widehat{\tau}_{X_k Y_l}$ is replaced with

$$
\widehat{r}_{X_k Y_l} = \frac{12}{n(n^2 - 1)} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} \frac{\delta_k^{(i)}}{\widehat{S}_{C_k}^2(X_k^{(i)})} I(X_k^{(j)} > X_k^{(i)}) \right\} \left\{ \sum_{m=1}^{n} \frac{\phi_l^{(i)}}{\widehat{S}_{D_l}^2(Y_l^{(i)})} I(Y_l^{(m)} > Y_l^{(i)}) \right\} - \frac{3(n-1)}{n+1},
$$

and $\widehat{\tau}_{X_k X_k'}$ and $\widehat{\tau}_{Y_l Y_l'}$ are replaced similarly with $\widehat{r}_{X_k X_k'}$ and $\widehat{r}_{Y_l Y_l'}$, respectively. Using a numerical study, we also analyzed this Spearman's correlation version of the RRCCA, finding that it performs similarly to, but slightly worse than Kendall's version. For the full results, see the online Supplementary Material. Therefore, we have presented only the Kendall's τ version to conserve space.

Our theory assumes that the censoring variables are uncorrelated, and so it would be interesting to consider the case when they are correlated. Here, we would need to modify the probability of the censoring weight using a bivariate survival function. Estimating bivariate survival functions is well studied in the literature, yielding, for example, the bivariate Kaplan–Meier estimate (Dabrowska, 1988). The formulation of the corresponding RRCCA would be much more complicated, and thus we leave it for future

research.

Another possible avenue for future work is to extend the proposed method to include complex data types. The key idea is to construct an unbiased estimator of the rank correlation for the pairwise underlying population. For instance, we can cope with multivariate interval-censored data by using a modified Kendall's τ statistic (Kim et al., 2015) in our formulation. To handle multivariate missing data with some auxiliary information under a missing-at-random assumption, we can apply a similar inverse probability weighted complete-case estimator (Tsiatis, 2006) to the rank correlation. These are interesting research topics for future investigation.

Acknowledgments

Supplementary Material

The online Supplementary Material contains simulation results, theory,

and proofs for the Spearman rank CCA approach.

A. Appendix

To simplify the notations, we denote $\widehat{S}_{C_k}(X_k^{(i)})$ as $\widehat{S}_k^{(i)}$ and $\widehat{S}_{D_l}(Y_l^{(i)})$ as $\widehat{S}_l^{(i)}$, $i = 1, \ldots, n; k = 1, \ldots, p; l = 1, \ldots, q$. Let $\| \cdot \|_\infty$ be the maximum absolute value of all elements in a vector. Denote the i.i.d. sample as $\mathbf{W}^{(i)} = (\mathbf{X}^{(i)}, \mathbf{Y}^{(i)}, \boldsymbol{\delta}^{(i)}, \boldsymbol{\phi}^{(i)})$. For any matrix $\mathbf{A} \in \mathbb{R}^{p \times q}$, define Frobenius norm as $\|\mathbf{A}\|_{\mathrm{F}} = (\sum_{i=1}^p \sum_{j=1}^q |A_{ij}|^2)^{1/2} = \{\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{A})\}^{1/2}$, and maximum norm as $\|\mathbf{A}\|_{\max} = \max_{i,j} |A_{ij}|$. We have

$$\|\mathbf{A}\| = \lambda_{\max}^{1/2}(\mathbf{A}^{\mathrm{T}}\mathbf{A}) \le \{\sum_{i=1}^s \lambda_i(\mathbf{A}^{\mathrm{T}}\mathbf{A})\}^{1/2} = \|\mathbf{A}\|_{\mathrm{F}} \le (pq)^{1/2}\|\mathbf{A}\|_{\max}$$

$$(7.1)$$

where $s \le \min(p, q)$ is the rank of $\mathbf{A}$. We use the capital letter C and M to denote generic constants that could vary from line to line.

To prove Theorem 1, we need the following lemmas.

Lemma 1 (Bitouzé et al. (1999), Theorem 1). *Let $\{\widetilde{X}_k^{(i)}\}_{i=1}^n$ and $\{C_k^{(i)}\}_{i=1}^n$ be independent sequences of independently identically distributed nonnegative random variables with distribution functions F and G, respectively. Let $\widehat{F}_n$ be the Kaplan–Meier estimator of the distribution function F. There exists a positive constant M, such that for any positive constant λ,*

$$\Pr\left\{n^{1/2}\|(1 - G)(\widehat{F}_n - F)\|_\infty > \lambda\right\} \le 2.5 \exp(-2\lambda^2 + M\lambda).$$

Lemma 2 (Hoeffding (1963)). *Let $h = h(x_1, ..., x_m)$ be a symmetric kernel of the U-statistic, U, with $a \leq h(x_1, \ldots, x_m) \leq b$. For any $t > 0$ and $n \geq m$, we have*

$$\Pr\{|U - E(U)| \geq t\} \leq 2 \exp\left\{-\frac{2[n/m]t^2}{(b-a)^2}\right\}.$$

Lemma 3. *Under Condition (C2), there exists a positive constant M, for any $\epsilon > 0$, when $n > M(\frac{\epsilon}{1+\epsilon})^{-2}$, we have*

$$\Pr\left\{\max_i \left|(\frac{S_k^{(i)}}{\widehat{S}_k^{(i)}})^2(\frac{S_l^{(i)}}{\widehat{S}_l^{(i)}})^2 - 1\right| \geq \epsilon\right\} \leq Cn \exp\{-Cn(\frac{\epsilon}{\epsilon+1})^2\},$$

for every $k = 1, \ldots, p; l = 1, \ldots, q$.

Moreover, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have

$$\Pr\left\{\max_i \left|(\frac{S_k^{(i)}}{\widehat{S}_k^{(i)}})^2(\frac{S_l^{(i)}}{\widehat{S}_l^{(i)}})^2 - 1\right| \geq \epsilon\right\} \leq Cn \exp(-Cn\epsilon^2).$$

Proof of Lemma 3. Let $G_k(t) = \Pr(\widetilde{X}_k \leq t)$, $G_l(t) = \Pr(\widetilde{Y}_l \leq t)$, by Condition (C2), there exist $\xi_k > 0, \eta_l > 0$, such that $\xi_k \leq S_k^{(i)} \leq 1$, $\eta_l \leq S_l^{(i)} \leq 1$, $\xi_k \leq 1 - G_k(Y_k^{(i)}) \leq 1$, $\eta_l \leq 1 - G_l(Y_l^{(i)}) \leq 1$, and $0 \leq \widehat{S}_k^{(i)} \leq 1, 0 \leq \widehat{S}_l^{(i)} \leq 1$. Denote $\nu = \min_{k=1,\ldots,p;l=1,\ldots,q}\{\xi_k, \eta_l\}$.

If A and B are two positive constants, it can be shown that for $\epsilon > 0$,

$|A^{-1} - B^{-1}| \geq \epsilon B^{-1}$ implies $|A - B| \geq \epsilon B/(1 + \epsilon)$. Combine this fact

with Lemma 1, we have

$$\Pr\left\{\left|(\frac{S_k^{(i)}}{\widehat{S}_k^{(i)}})^2(\frac{S_l^{(i)}}{\widehat{S}_l^{(i)}})^2 - 1\right| \geq \epsilon\right\}$$

$$\leq\Pr\left\{|(\widehat{S}_k^{(i)})^2(\widehat{S}_l^{(i)})^2 - (S_k^{(i)})^2(S_l^{(i)})^2| > \frac{\epsilon}{1+\epsilon}(S_k^{(i)})^2(S_l^{(i)})^2\right\}$$

$$\leq\Pr\left\{(\widehat{S}_l^{(i)})^2|(\widehat{S}_k^{(i)})^2 - (S_k^{(i)})^2| + (S_k^{(i)})^2|(\widehat{S}_l^{(i)})^2 - (S_l^{(i)})^2| > \frac{\epsilon}{1+\epsilon}\nu^4\right\}$$

$$\leq\Pr\left\{|(\widehat{S}_k^{(i)})^2 - (S_k^{(i)})^2| > \frac{\nu^4\epsilon}{2(1+\epsilon)}\right\} + \Pr\left\{|(\widehat{S}_l^{(i)})^2 - (S_l^{(i)})^2| > \frac{\nu^4\epsilon}{2(1+\epsilon)}\right\}$$

$$\leq\Pr\left\{|\widehat{S}_k^{(i)} - S_k^{(i)}| > \frac{\nu^4\epsilon}{4(1+\epsilon)}\right\} + \Pr\left\{|\widehat{S}_l^{(i)} - S_l^{(i)}| > \frac{\nu^4\epsilon}{4(1+\epsilon)}\right\}$$

$$\leq\Pr\left\{n^{1/2}\|(1 - G_k)(\widehat{S}_{k_n} - S_k)\|_\infty > n^{1/2}\frac{\nu^5\epsilon}{4(1+\epsilon)}\right\} + \Pr\left\{n^{1/2}\|(1 - G_l)(\widehat{S}_{l_n} - S_l)\|_\infty > n^{1/2}\frac{\nu^5\epsilon}{4(1+\epsilon)}\right\}$$

$$\leq 5\exp\left[-2n\{\frac{\nu^5\epsilon}{4(1+\epsilon)}\}^2 + \tilde{M}n^{1/2}\frac{\nu^5\epsilon}{4(1+\epsilon)}\right],$$

for some constant $\tilde{M}$. When $\tilde{M}n^{1/2}\frac{\nu^5\epsilon}{4(1+\epsilon)} < n\{\frac{\nu^5\epsilon}{4(1+\epsilon)}\}^2$, that is, $n > 16\tilde{M}^2\nu^{-10}(\frac{\epsilon}{1+\epsilon})^{-2} \equiv M(\frac{\epsilon}{1+\epsilon})^{-2}$, we have

$$\Pr\left\{\max_i\left|(\frac{S_k^{(i)}}{\widehat{S}_k^{(i)}})^2(\frac{S_l^{(i)}}{\widehat{S}_l^{(i)}})^2 - 1\right| \geq \epsilon\right\} \leq n\Pr\left\{\left|(\frac{S_k^{(i)}}{\widehat{S}_k^{(i)}})^2(\frac{S_l^{(i)}}{\widehat{S}_l^{(i)}})^2 - 1\right| \geq \epsilon\right\} \leq Cn\exp\{-Cn(\frac{\epsilon}{\epsilon+1})^2\}.$$

If $0 < \epsilon < 1$, we have $\frac{\epsilon}{\epsilon+1} > \frac{\epsilon}{2}$, hence the second inequality holds.   □

Lemma 4. *Under Condition (C2), there exists a positive constant M, for*

any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have

$$\Pr(|\widehat{\tau}_{X_k Y_l} - \tau_{X_k Y_l}| \geq \epsilon) \leq Cn \exp(-Cn\epsilon^2),$$

$$\Pr(|\widehat{\tau}_{X_k X'_k} - \tau_{X_k X'_k}| \geq \epsilon) \leq Cn \exp(-Cn\epsilon^2),$$

$$\Pr(|\widehat{\tau}_{Y_l Y'_l} - \tau_{Y_l Y'_l}| \geq \epsilon) \leq Cn \exp(-Cn\epsilon^2).$$

Proof of Lemma 4. Rewrite $\widehat{\tau}_{X_k Y_l} = 4\binom{n}{2}^{-1} \sum_{i<j} h_1(\mathbf{W}^{(i)}, \mathbf{W}^{(j)}) - 1$, where

$$
\begin{aligned}
h_1(\mathbf{W}^{(i)}, \mathbf{W}^{(j)}) = \frac{1}{2} \Bigg\{ & \frac{\delta_k^{(i)}}{(\widehat{S}_k^{(i)})^2} \frac{\phi_l^{(i)}}{(\widehat{S}_l^{(i)})^2} I(X_k^{(j)} > X_k^{(i)}) I(Y_l^{(j)} > Y_l^{(i)}) \\
& + \frac{\delta_k^{(j)}}{(\widehat{S}_k^{(j)})^2} \frac{\phi_l^{(j)}}{(\widehat{S}_l^{(j)})^2} I(X_k^{(i)} > X_k^{(j)}) I(Y_l^{(i)} > Y_l^{(j)}) \Bigg\}
\end{aligned}
$$

is the symmetric kernel of $(\widehat{\tau}_{X_k Y_l} + 1)/4$. Hence, $(\widehat{\tau}_{X_k Y_l} + 1)/4$ is a U-statistic.

Let $U_n[f] \equiv \binom{n}{2}^{-1} \sum_{i<j} \{f(x^{(i)}, x^{(j)}) + f(x^{(j)}, x^{(i)})\}/2$ denote the empir-

ical function for this U-statistics. We have

$$
\frac{\widehat{\tau}_{X_k Y_l} - \tau_{X_k Y_l}}{4}
$$

$$
= U_n\Big[\Big\{\frac{1}{(\widehat{S}_k^{(i)})^2}\frac{1}{(\widehat{S}_l^{(i)})^2} - \frac{1}{(S_k^{(i)})^2}\frac{1}{(S_l^{(i)})^2}\Big\}\delta_k^{(i)}\phi_l^{(i)}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]
$$

$$
+ (U_n - E)\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]
$$

$$
\leq \max_i\Big|\frac{(S_k^{(i)})^2}{(\widehat{S}_k^{(i)})^2}\frac{(S_l^{(i)})^2}{(\widehat{S}_l^{(i)})^2} - 1\Big| \cdot \Big|U_n\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]\Big|
$$

$$
+ (U_n - E)\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]
$$

$$
\leq \Big\{\max_i\Big|\frac{(S_k^{(i)})^2}{(\widehat{S}_k^{(i)})^2}\frac{(S_l^{(i)})^2}{(\widehat{S}_l^{(i)})^2} - 1\Big| + 1\Big\} \cdot \Big|(U_n - E)\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]\Big|
$$

$$
+ \max_i\Big|\frac{(S_k^{(i)})^2}{(\widehat{S}_k^{(i)})^2}\frac{(S_l^{(i)})^2}{(\widehat{S}_l^{(i)})^2} - 1\Big| \cdot \Big|E\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]\Big|.
$$

Under Condition (C2) and the constant ν we denoted in the proof of Lemma 3, we have $0 \leq \frac{1}{2}\Big\{\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)}) + \frac{\delta_k^{(j)}}{(S_k^{(j)})^2}\frac{\phi_l^{(j)}}{(S_l^{(j)})^2}I(X_k^{(i)} > X_k^{(j)})I(Y_l^{(i)} > Y_l^{(j)})\Big\} \leq 1/\nu^4$. By Lemma 2, for any $\epsilon > 0$, we have

$$
\mathrm{Pr}\Big\{\Big|(U_n - E)\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]\Big| \geq \epsilon\Big\} \leq 2\exp(-Cn\epsilon^2).
$$

Because $\Big|E\Big[\frac{\delta_k^{(i)}}{(S_k^{(i)})^2}\frac{\phi_l^{(i)}}{(S_l^{(i)})^2}I(X_k^{(j)} > X_k^{(i)})I(Y_l^{(j)} > Y_l^{(i)})\Big]\Big| \leq 1$, therefore, by Lemma 3, there exists a positive constant M, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have

$$
\mathrm{Pr}(|\widehat{\tau}_{X_k Y_l} - \tau_{X_k Y_l}| \geq \epsilon) \leq Cn\exp(-Cn\epsilon^2).
$$

The other two inequalities can be shown in the same way.   □

Lemma 5. *Under Condition (C2), there exists a positive constant M, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have*

$$\Pr(\|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\| \geq \epsilon) \leq Cnpq \exp(-C\frac{n}{pq}\epsilon^2),$$

$$\Pr(\|\widehat{\mathbf{S}}_{\mathbf{XX}} - \mathbf{S}_{\mathbf{XX}}\| \geq \epsilon) \leq Cnp^2 \exp(-C\frac{n}{p^2}\epsilon^2),$$

$$\Pr(\|\widehat{\mathbf{S}}_{\mathbf{YY}} - \mathbf{S}_{\mathbf{YY}}\| \geq \epsilon) \leq Cnq^2 \exp(-C\frac{n}{q^2}\epsilon^2),$$

which means,

$$\|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\| = O_p(\sqrt{\frac{pq \log npq}{n}}),$$

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}} - \mathbf{S}_{\mathbf{XX}}\| = O_p(\sqrt{\frac{p^2 \log np^2}{n}}),$$

$$\|\widehat{\mathbf{S}}_{\mathbf{YY}} - \mathbf{S}_{\mathbf{YY}}\| = O_p(\sqrt{\frac{q^2 \log nq^2}{n}}).$$

Proof of Lemma 5. By Inequality 7.1, the subadditivity of probability and Lemma 4 , we have

$$\Pr(\|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\| \geq \epsilon) \leq \Pr\{\|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\|_{\max} \geq \epsilon(pq)^{-1/2}\}$$

$$\leq pq\Pr\{|\widehat{\tau}_{X_kY_l} - \tau_{X_kY_l}| \geq \epsilon(pq)^{-1/2}\}$$

$$\leq Cnpq \exp(-C\frac{n}{pq}\epsilon^2).$$

Let $\epsilon = O(\sqrt{\frac{pq \log npq}{n}})$, we have the desired conclusion. The other two inequalities can be shown in the same way.   □

Lemma 6. *Under Condition(C1) and (C2), there exists a positive constant*

M, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have

$$\|\widehat{\mathbf{S}}_{\mathbf{XY}}\| = O_p(\sqrt{\frac{pq \log npq}{n}}) + O_p(\frac{1}{\omega}).$$

If further $pq \log npq = o(n)$ holds, then

$$\|\widehat{\mathbf{S}}_{\mathbf{XY}}\| = O_p(1).$$

Proof of Lemma 6. Since

$$\|\widehat{\mathbf{S}}_{\mathbf{XY}}\| \leq \|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\| + \|\mathbf{S}_{\mathbf{XY}}\|,$$

it is a direct result from Lemma 5. □

Lemma 7. *Under Condition(C1) and (C2), there exists a positive constant*

M, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, we have

$$\Pr\left\{\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \geq \frac{1}{\kappa(1-\epsilon)}\right\} \leq Cnp^2 \exp(-C\frac{n}{p^2}\epsilon^2),$$

$$\Pr\left\{\|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\| \geq \frac{1}{\kappa(1-\epsilon)}\right\} \leq Cnq^2 \exp(-C\frac{n}{q^2}\epsilon^2).$$

If further $p^2 \log np^2 = o(n)$ and $q^2 \log nq^2 = o(n)$ hold, then

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| = O_p(1),$$

$$\|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\| = O_p(1).$$

Proof of Lemma 7. For any symmetric matrix $\mathbf{H}$, we have the fact that $\lambda_{\min}^{-1}(\mathbf{H}) = \lambda_{\max}(\mathbf{H}^{-1})$. If A and B are two positive constants, it is shown in the proof of Lemma 5 of Fan et al. (2011) that for $\epsilon \in (0, 1)$,

$$|A^{-1} - B^{-1}| \geq (\frac{1}{1-\epsilon} - 1)B^{-1} \quad \text{implies} \quad |A - B| \geq \epsilon B.$$

Combine these two facts, we have

$$
\begin{aligned}
\Pr\left\{ \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \geq \frac{1}{\kappa(1-\epsilon)} \right\} &\leq \Pr\left(\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \geq \frac{1}{1-\epsilon}\|\mathbf{S}_{\mathbf{XX}}^{-1}\| \right) \\
&\leq \Pr\left\{ \left| \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| - \|\mathbf{S}_{\mathbf{XX}}^{-1}\| \right| \geq (\frac{1}{1-\epsilon} - 1)\|\mathbf{S}_{\mathbf{XX}}^{-1}\| \right\} \\
&\leq \Pr\{ |\lambda_{\min}(\widehat{\mathbf{S}}_{\mathbf{XX}}) - \lambda_{\min}(\mathbf{S}_{\mathbf{XX}})| \geq \epsilon\lambda_{\min}(\mathbf{S}_{\mathbf{XX}}) \} \\
&\leq \Pr(\|\widehat{\mathbf{S}}_{\mathbf{XX}} - \mathbf{S}_{\mathbf{XX}}\| \geq \kappa\epsilon) \\
&\leq Cnp^2 \exp(-C\frac{n}{p^2}\epsilon^2).
\end{aligned}
$$

The fourth inequality follows from the fact that

$$|\lambda_{\min}(\mathbf{A}) - \lambda_{\min}(\mathbf{B})| \leq \max\{|\lambda_{\min}(\mathbf{A} - \mathbf{B})|, |\lambda_{\min}(\mathbf{B} - \mathbf{A})|\},$$

for any symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, which is also proved in the Lemma 5 of Fan et al. (2011). And the fifth inequality follows from Lemma 5.

For a fixed ϵ, if $\log np^2 - C\frac{n}{p^2}\epsilon^2 \to -\infty$, we have

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| = O_p(1).$$

The result for $\|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\|$ can be shown in the same way. $\qquad\square$

Lemma 8. *Under Condition(C1) and (C2), there exists a positive con-stant M, for any $0 < \epsilon < 1$, when $n > 4M\epsilon^{-2}$, if $p^2 \log np^2 = o(n)$ and $q^2 \log nq^2 = o(n)$ hold, we have*

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1} - \mathbf{S}_{\mathbf{XX}}^{-1}\| = o_p(1),$$

$$\|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1} - \mathbf{S}_{\mathbf{YY}}^{-1}\| = o_p(1).$$

Proof of Lemma 8. By the submultiplicativity of matrix norm, Lemma 5 and Lemma 7, we have

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1} - \mathbf{S}_{\mathbf{XX}}^{-1}\| \leq \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \cdot \|\mathbf{S}_{\mathbf{XX}} - \widehat{\mathbf{S}}_{\mathbf{XX}}\| \cdot \|\mathbf{S}_{\mathbf{XX}}^{-1}\| = O_p(1)o_p(1)O_p(1/\kappa) = o_p(1).$$

The second equation can be proved in the same way.   □

Now we are ready to prove Theorem 1.

Proof of Theorem 1. Combine triangle inequality with Lemma 5, Lemma

6, Lemma 7 and Lemma 8, we have

$$\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\widehat{\mathbf{S}}_{\mathbf{YX}} - \mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{YX}}\|$$

$$\leq \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1}\widehat{\mathbf{S}}_{\mathbf{YX}} - \widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{YX}}\|$$

$$+ \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{YX}} - \mathbf{S}_{\mathbf{XX}}^{-1}\mathbf{S}_{\mathbf{XY}}\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{YX}}\|$$

$$\leq \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\widehat{\mathbf{S}}_{\mathbf{XY}}\|\left(\|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1} - \mathbf{S}_{\mathbf{YY}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}^{\mathrm{T}}\| + \|\mathbf{S}_{\mathbf{YY}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}^{\mathrm{T}} - \mathbf{S}_{\mathbf{XY}}^{\mathrm{T}}\|\right)$$

$$+ \left(\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1} - \mathbf{S}_{\mathbf{XX}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}\| + \|\mathbf{S}_{\mathbf{XX}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}} - \mathbf{S}_{\mathbf{XY}}\|\right)\|\mathbf{S}_{\mathbf{YY}}^{-1}\mathbf{S}_{\mathbf{XY}}^{\mathrm{T}}\|$$

$$\leq \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}\|^2 \cdot \|\widehat{\mathbf{S}}_{\mathbf{YY}}^{-1} - \mathbf{S}_{\mathbf{YY}}^{-1}\| + \|\mathbf{S}_{\mathbf{YY}}^{-1}\| \cdot \|\mathbf{S}_{\mathbf{XY}}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1} - \mathbf{S}_{\mathbf{XX}}^{-1}\|$$

$$+ \left(\|\widehat{\mathbf{S}}_{\mathbf{XX}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}\| + \|\mathbf{S}_{\mathbf{XX}}^{-1}\| \cdot \|\mathbf{S}_{\mathbf{XY}}\|\right)\|\mathbf{S}_{\mathbf{YY}}^{-1}\| \cdot \|\widehat{\mathbf{S}}_{\mathbf{XY}}^{-1} - \mathbf{S}_{\mathbf{XY}}^{-1}\|$$

$$= O_p(1)\{O_p(1)\}^2 o_p(1) + O_p(\frac{1}{\kappa})O_p(\frac{1}{\omega})O_p(1)o_p(1)$$

$$+ \left\{O_p(1)O_p(1) + O_p(\frac{1}{\kappa})O_p(\frac{1}{\omega})\right\}O_p(\frac{1}{\kappa})o_p(1)$$

$$= o_p(1).$$

□

References

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and event history analysis: A process point of view*. New York: Springer.

Aitkin, M., Laird, N., and Francis, B. (1983). A reanalysis of the stanford heart transplant data. *Journal of the American Statistical Association*, 78(382):264–274.

Akaho, S. (2001). A kernel method for canonical correlation analysis. In *International Meeting of Psychometric Society*, Osaka.

Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199 – 227.

Bitouzé, D., Laurent, B., and Massart, P. (1999). A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 35(6):735–763.

Branco, J. A., Croux, C., Filzmoser, P., and Oliveira, M. R. (2005). Robust canonical correlations: A comparative study. *Computational Statistics*, 20(2):203–229.

Chen, S., Ma, S., Xue, L., and Zou, H. (2020). An alternating manifold proximal gradient method for sparse principal component analysis and sparse canonical correlation analysis. *INFORMS Journal on Optimization*, 2(3):192–208.

Croux, C. and Dehon, C. (2002). Analyse canonique basée sur des estimateurs robustes de la matrice de covariance. *Revue de Statistique Appliquée*, 50(2):5–26.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association*, 72(357):27–36.

Dabrowska, D. M. (1988). Kaplan-Meier estimate on the plane. *The Annals of Statistics*, 16(4):1475–1489.

Fan, J., Feng, Y., and Song, R. (2011). Nonparametric independence screening in sparse

ultra-high-dimensional additive models. *Journal of the American Statistical Association*, 106(494):544–557.

Filzmoser, P., Dehon, C., and Croux, C. (2000). Outlier resistant estimators for canonical correlation analysis. In Bethlehem, J. G. and van der Heijden, P. G. M., editors, *COMPSTAT*, pages 301–306, Heidelberg. Physica-Verlag HD.

Golugula, A., Lee, G., Master, S., Feldman, M., Tomaszewski, J., Speicher, D., and Madabhushi, A. (2011). Supervised regularized canonical correlation analysis: Integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC bioinformatics*, 12:483.

Hardoon, D. and Shawe-Taylor, J. (2011). Sparse canonical correlation analysis. *Machine Learning*, 83:331–353.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, 28(3-4):321–377.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data, Second Edition*. New York: Wiley.

Kang, L., Chen, W., Petrick, N. A., and Gallas, B. D. (2015). Comparing two correlated c indices with right-censored survival outcome: a one-shot nonparametric approach. *Statistics in Medicine*, 34(4):685–703.

Kim, Y., Lim, J., and Park, D. (2015). Testing independence of bivariate interval-censored data using modified kendall's tau statistic. *Biometrical Journal*, 57(6):1131–1145.

Kudraszow, N. L. and Maronna, R. A. (2011). Robust canonical correlation analysis: a predictive approach. *Working paper*.

Lindskog, F., McNeil, A., and Schmock, U. (2003). Kendall's tau for elliptical distributions. In Bol, G., Nakhaeizadeh, G., Rachev, S. T., Ridder, T., and Vollmer, K.-H., editors, *Credit Risk*, pages 149–156, Heidelberg. Physica-Verlag HD.

Mai, Q. and Zhang, X. (2019). An iterative penalized least squares approach to sparse canonical correlation analysis. *Biometrics*, 75(3):734–744.

Newton, E. and Rudel, R. (2007). Estimating correlation with multiply censored data arising from the adjustment of singly censored data. *Environmental Science and Technology*, 41(1):221–228.

Oakes, D. (1982). A concordance test for independence in the presence of censoring. *Biometrics*, 38(2):451–455.

Peng, L. and Fine, J. P. (2009). Competing risks quantile regression. *Journal of the American Statistical Association*, 104(488):1440–1453.

Romanazzi, M. (1992). Influence in canonical correlation analysis. *Psychometrika*, 57(2):237–259.

Rousseeuw, P. (1985). Multivariate estimation with high breakdown point. In *Mathematical*

Statistics and Applications, volume B, pages 283–297, Reidel, Dordrecht.

Schemper, M., Kaider, A., Wakounig, S., and Heinze, G. (2013). Estimating the correlation of bivariate failure times under censoring. *Statistics in Medicine*, 32(27):4781–4790.

Shih, J. H. and Louis, T. A. (1996). Tests of independence for bivariate survival data. *Biometrics*, 52(4):1440–1449.

Song, R., Lu, W., Ma, S., and Jeng, X. J. (2014). Censored rank independence screening for high-dimensional survival data. *Biometrika*, 101(4):799–814.

Taskinen, S., Croux, C., Kankainen, A., Ollila, E., and Oja, H. (2006). Influence functions and efficiencies of the canonical correlation and vector estimates based on scatter and shape matrices. *Journal of Multivariate Analysis*, 97(2):359 – 384.

Tsiatis, A. A. (2006). *Semiparametric theory and missing data*. New York: Springer.

Va, J., Santamara, I., and Prez, J. (2007). A learning algorithm for adaptive canonical correlation analysis of several data sets. *Neural Networks*, 20(1):139–152.

Witten, D., Tibshirani, R., and Hastie, T. (2009). A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics*, 10:515–34.

Witten, D. M. and Tibshirani, R. J. (2009). Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical Applications in Genetics and Molecular Biology*, 8(1):1–27.

School of Economics, Nanjing University, Nanjing, 210046, China

E-mail: (hedi@nju.edu.cn)

Key Laboratory of Advanced Theory and Application in Statistics and Data Science, MOE, and

Academy of Statistics and Interdisciplinary Sciences, East China Normal University, Shanghai

200062, China

E-mail: (yzhou@amss.ac.cn)

School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

E-mail: (zouxx019@umn.edu)