New Feature Screening Methods for Massive Interval-censored Failure Time Data

Huiqiong Li[a], Zhimiao Cao[a], Jianguo Sun[b] and Niansheng Tang[a]

[a] *Yunnan Key Laboratory of Statistical Modeling and Data Analysis,*

Yunnan University, Kunming,650091, China

[b] *Department of Statistics, University of Missouri,Columbia, MO, 65211, U.S.A.*

Supplementary Material

The supplementary material includes four sections. Some algorithms are presented in Section S1. Section S2 presents additional results from both simulation and application studies. Some preliminary lemmas are presented in Section S3, and technical proofs for all theorems appeared in the paper are presented in Section S4. The indices of assumptions, theorems in this supplement are the same as those in the main paper.

S1 Related Algorithms

This section introduces three algorithms discussed in the main paper: the Orthogonal Subsampling Algorithm, Simple Average Distance Correlation Screening based on OSS (SDC-OSS), and Jackknife Debiased Average Distance Correlation Screening based on OSS (JDC-OSS).

Algorithm 1 Orthogonal subsampling algorithm

Input: $\mathbf{X}, K$.

Output: $\widetilde{A}$.

1: Scale the values of each covariate to lie in $[-1, 1]$, and the new design matrix is denoted $\mathbf{Z}$. Set $i = 1$. Find the point $\mathbf{Z}_1^*$ in $\mathbf{Z}$ with the largest Euclidean norm, include it in $\mathbf{Z}_{\widetilde{A}\cdot}$, and remove it from $\mathbf{Z}$. Let $\mathscr{L}$ be an $(N-1)$-vector with each component corresponding to each remaining data point in $\mathbf{Z}$. Set all components of $\mathscr{L}$ to be 0.

2: **for all** $i = 1, 2, \ldots$ **do**

3: For each $\mathbf{z} \in \mathbf{Z}$, add $l(\mathbf{z}|\mathbf{Z}_{\widetilde{A}\cdot}^{i-1})$ to its corresponding component in $\mathscr{L}$, where

$$L(\mathbf{Z}_{\widetilde{A}\cdot}^{i}) = l(\mathbf{Z}_i^*|\mathbf{Z}_{\widetilde{A}\cdot}^{i-1}) + L(\mathbf{Z}_{\widetilde{A}\cdot}^{i-1}); \quad l(\mathbf{z}|\mathbf{Z}_{\widetilde{A}\cdot}^{i-1}) = \sum_{\mathbf{Z}_j^* \in \mathbf{Z}_{\widetilde{A}\cdot}^{i-1}} \left[p - \frac{\|\mathbf{z}\|^2}{2} - \frac{\|\mathbf{Z}_j^*\|^2}{2} + \xi(\mathbf{z}, \mathbf{Z}_j^*) \right]^2.$$

Find $\mathbf{z}$ with the smallest component in $\mathscr{L}$ and add it to $\mathbf{Z}_{\widetilde{A}\cdot}$.

4: **if** $N \geq K^2$ **then**

5: $k_i = N/i$

6: **else**

7: $k_i = N/i^{r-1}$, where $r = \log(N)/\log(K)$

8: **end if**

9: Keep k_i points in $\mathbf{Z}$ that correspond to the k_i smallest components in $\mathscr{L}$. Remove $\mathbf{z}$ which picked out in the last step and the other selected points from $\mathbf{Z}$, as well as their corresponding components from $\mathscr{L}$.

10: **if** $|\widetilde{A}| = K$ **then**

11: Stop iteration

12: **else**

13: $i = i + 1$

14: **end if**

15: **end for**

16: **return:** The index set of the subsample obtained by OSS is $\widetilde{A}$.

Algorithm 2 SDC-OSS screening

Input: $\mathbf{Y}, \mathbf{X}, n, B, d_0$.

Output: $\widehat{\mathcal{M}}$.

The nB subsample was extracted by the OSS algorithm and randomly divided into B segments. Each segment subsample was denoted as $\mathcal{D}_b, b = 1, \ldots, B$

for all $j = 1, 2, \ldots, p$ **do**

  for all $b = 1, \ldots, B$ **do**

    Construct the statistics $\widehat{dcov^2_{(b)}}(X_j, \mathbf{Y})$, $\widehat{dcov^2_{(b)}}(X_j, X_j)$ and $\widehat{dcov^2_{(b)}}(\mathbf{Y}, \mathbf{Y})$ based on the samples $\mathbf{X}_{\mathcal{D}_b}, \mathbf{Y}_{\mathcal{D}_b}$

    $\hat{\omega}_{(b),j} = \frac{\widehat{dcov^2_{(b)}}(X_j, \mathbf{Y})}{\sqrt{\widehat{dcov^2_{(b)}}(X_j, X_j)\widehat{dcov^2_{(b)}}(\mathbf{Y}, \mathbf{Y})}}$

  end for

  $\hat{\omega}_j^{SDC} = \frac{1}{B}\sum_{b=1}^{B} \hat{\omega}_{(b),j}$

end for

Arrange $\hat{\omega}_j^{SDC}, j = 1, \ldots, p$ in order from largest to smallest, denoting $\widehat{\mathcal{M}}^{SDC} = \{j : \hat{\omega}_j^{SDC} \geq \hat{\omega}_{(d_0)}^{SDC}\}$, where $\hat{\omega}_{(d_0)}^{SDC}$ is the d_0-th largest statistic.

return: $\widehat{\mathcal{M}}^{SDC}$ as the estimates of the set of significant covariates $\mathcal{M}$.

Algorithm 3 JDC-OSS screening

Input: $\mathbf{Y}, \mathbf{X}, n, B, d_0$.

Output: $\widehat{\mathcal{M}}$.

The nB subsample was extracted by the OSS algorithm and randomly divided into B segments. Each segment subsample was denoted as $\mathcal{D}_b, b = 1, \ldots, B$

for all $j = 1, 2, \ldots, p$ **do**

  for all $b = 1, \ldots, B$, $m = 1, \ldots, n$ **do**

    Construct $\widehat{dcov^2_{(b,-m)}}(X_j, \mathbf{Y})$, $\widehat{dcov^2_{(b,-m)}}(X_j, X_j)$ and $\widehat{dcov^2_{(b,-m)}}(\mathbf{Y}, \mathbf{Y})$ based on $\mathbf{X}_{(b,-m)} = (\mathbf{X}_i : i \in \mathcal{D}_b, j \neq m)^\top$, and $\mathbf{Y}_{(b,-m)} = (\mathbf{Y}_i : i \in \mathcal{D}_b, j \neq m)^\top$

    Calculate $\widehat{\triangle}_{(b)}(X_j, \mathbf{Y}), \widehat{\triangle}_{(b)}(X_j, X_j)$, and $\widehat{\triangle}_{(b)}(\mathbf{Y}, \mathbf{Y})$ based on Equation (3.4)

    $\widehat{dcov^2}^{JDC}(X_j, \mathbf{Y}) = \frac{1}{B}\sum_{b=1}^{B}\{\widehat{dcov^2}_{(b)}(X_j, \mathbf{Y}) - \widehat{\triangle}_{(b)}(X_j, \mathbf{Y})\}$

    $\widehat{dcov^2}^{JDC}(X_j, X_j) = \frac{1}{B}\sum_{b=1}^{B}\{\widehat{dcov^2}_{(b)}(X_j, X_j) - \widehat{\triangle}_{(b)}(X_j, X_j)\}$

    $\widehat{dcov^2}^{JDC}(\mathbf{Y}, \mathbf{Y}) = \frac{1}{B}\sum_{b=1}^{B}\{\widehat{dcov^2}_{(b)}(\mathbf{Y}, \mathbf{Y}) - \widehat{\triangle}_{(b)}(\mathbf{Y}, \mathbf{Y})\}$

  end for

  $\widehat{dcov^2}^{JDC}(X_j, \mathbf{Y}) = \frac{1}{B}\sum_{b=1}^{B}\{\widehat{dcov^2}_{(b)}(X_j, \mathbf{Y}) - \widehat{\triangle}_{(b)}(X_j, \mathbf{Y})\}$

end for

Arrange $\hat{\omega}_j^{JDC} = \frac{\widehat{dcov^2}^{JDC}(X_j, \mathbf{Y})}{\sqrt{\widehat{dcov^2}^{JDC}(X_j, X_j)\widehat{dcov^2}^{JDC}(\mathbf{Y}, \mathbf{Y})}}, j = 1, \ldots, p$ in order from largest to smallest, denoting $\widehat{\mathcal{M}}^{JDC} = \{j : \hat{\omega}_j^{JDC} \geq \hat{\omega}_{(d_0)}^{JDC}\}$, where $\hat{\omega}_{(d_0)}^{JDC}$ is the d_0-th largest statistic.

return: $\widehat{\mathcal{M}}^{JDC}$ as the estimates of the set of significant covariates $\mathcal{M}$.

S2 Supplementary Material for Simulation Studies and Application

In addition to the evaluation above, we also carried out the assessment of the two proposed procedures in terms of the False Discovery Rate (FDR) control. For this, we considered a new knockoff of covariate observations generated by constructing Gaussian distributions corresponding to the sample expectation and sample covariance of covariate data. The statistics $\omega_j^{JDC}(X_j, \mathbf{Y})$ (or $\omega_j^{SDC}(X_j, \mathbf{Y})$) and $\omega_j^{JDC}(\widetilde{X}_j, \mathbf{Y})$ (or $\omega_j^{SDC}(\widetilde{X}_j, \mathbf{Y})$) were then constructed under the proposed JDC-OSS (or SDC-OSS) method based on the orthogonal subsample of the true observed data $\mathbf{X}_{\widetilde{A}}$ of the covariate and the corresponding knockoff data $\widetilde{\mathbf{X}}$. Furthermore, define the statistics

$$W_j^{JDC} = \omega_j^{JDC}(X_j, \mathbf{Y}) - \omega_j^{JDC}(\widetilde{X}_j, \mathbf{Y}) \ , \ (W_j^{SDC} = \omega_j^{SDC}(X_j, \mathbf{Y}) - \omega_j^{SDC}(\widetilde{X}_j, \mathbf{Y})) \, ,$$

and the threshold T_α

$$T_\alpha = \min\left\{ t \in \mathcal{W} : \frac{1 + \#\{j : W_j \le t\}}{\#\{j : W_j \ge t\}} \le \alpha \right\}$$

for the given FDR control level α by following Liu et al.(2022), where $\mathcal{W} = \{|W_j| : 1 \le j \le p\}/\{0\}$.

Table 1: Simulation results under Model 2 (a).

						S				
n	$B(d_0)$	$Method$	time(sec)	$AUC(\%)$	PA	5%	25%	50%	75%	95%
Right-censored rate = 20%										
10	60(18)	JDC-OSS	78.54	99.893	0.97	4	4	4	4	13
		SDC-OSS	55.18	99.939	0.96	4	4	4	4	12
		JDC-RSS	26.50	99.757	0.915	4	4	4	7	26.05
		SDC-RSS	3.61	99.858	0.925	4	4	4	6	24
	100(20)	JDC-OSS	140.94	99.998	1	4	4	4	4	5
		SDC-OSS	109.32	99.999	1	4	4	4	4	4
		JDC-RSS	36.98	99.960	0.99	4	4	4	4	7
		SDC-RSS	5.67	99.991	0.995	4	4	4	4	5
20	25(17)	JDC-OSS	76.79	99.988	0.995	4	4	4	4	6
		SDC-OSS	49.56	99.983	0.985	4	4	4	4	6.05
		JDC-RSS	28.71	99.899	0.945	4	4	4	5	23.6
		SDC-RSS	2.15	99.909	0.955	4	4	4	5	14.1
	40(19)	JDC-OSS	119.44	99.999	1	4	4	4	4	4
		SDC-OSS	85.49	99.998	1	4	4	4	4	4
		JDC-RSS	36.19	99.996	1	4	4	4	4	4
		SDC-RSS	2.84	99.996	1	4	4	4	4	4.05
Right-censored rate = 40%										
10	60(18)	JDC-OSS	78.84	99.998	1	4	4	4	4	5
		SDC-OSS	54.77	99.997	1	4	4	4	4	5
		JDC-RSS	27.04	99.983	0.99	4	4	4	4	5
		SDC-RSS	3.60	99.977	0.995	4	4	4	4	5
	100(20)	JDC-OSS	149.90	100.000	1	4	4	4	4	4
		SDC-OSS	115.65	100.000	1	4	4	4	4	4
		JDC-RSS	40.10	100.000	1	4	4	4	4	4
		SDC-RSS	5.97	100.000	1	4	4	4	4	4
20	25(17)	JDC-OSS	75.43	99.999	1	4	4	4	4	4
		SDC-OSS	47.45	99.999	1	4	4	4	4	4
		JDC-RSS	29.47	99.999	1	4	4	4	4	4
		SDC-RSS	1.98	99.998	1	4	4	4	4	4
	40(19)	JDC-OSS	109.62	100.000	1	4	4	4	4	4
		SDC-OSS	78.97	100.000	1	4	4	4	4	4
		JDC-RSS	32.58	100.000	1	4	4	4	4	4
		SDC-RSS	2.92	100.000	1	4	4	4	4	4
Right-censored rate = 60%										
10	60(18)	JDC-OSS	75.95	99.996	1	4	4	4	4	5
		SDC-OSS	53.33	99.997	1	4	4	4	4	5
		JDC-RSS	25.83	99.988	0.99	4	4	4	4	6
		SDC-RSS	3.40	99.982	0.99	4	4	4	4	6.05
	100(20)	JDC-OSS	134.20	100.000	1	4	4	4	4	4
		SDC-OSS	104.63	100.000	1	4	4	4	4	4
		JDC-RSS	34.51	100.000	1	4	4	4	4	4
		SDC-RSS	5.56	100.000	1	4	4	4	4	4
20	25(17)	JDC-OSS	78.24	100.000	1	4	4	4	4	4
		SDC-OSS	48.88	99.998	1	4	4	4	4	4
		JDC-RSS	31.51	99.998	1	4	4	4	4	4
		SDC-RSS	2.04	99.997	1	4	4	4	4	4
	40(19)	JDC-OSS	110.64	100.000	1	4	4	4	4	4
		SDC-OSS	79.44	100.000	1	4	4	4	4	4
		JDC-RSS	33.16	100.000	1	4	4	4	4	4
		SDC-RSS	2.92	100.000	1	4	4	4	4	4

Table 2: Simulation results under Model 2 (b).

n	$B(d_0)$	*Method*	time(sec)	$AUC(\%)$	PA	S 5%	25%	50%	75%	95%
				Right-censored rate = 20%						
10	100(20)	JDC-OSS	240.15	99.917	0.845	10	10	10	14	63
		SDC-OSS	176.41	99.883	0.815	10	10	10	15	76.15
		JDC-RSS	48.69	99.768	0.695	10	10	13	24	86.6
		SDC-RSS	9.09	99.752	0.690	10	10	13	23.25	97.1
	150(21)	JDC-OSS	360.13	99.990	0.985	10	10	10	10	13
		SDC-OSS	296.07	99.986	0.975	10	10	10	10	14
		JDC-RSS	60.24	99.941	0.900	10	10	10	12	40.2
		SDC-RSS	11.78	99.958	0.935	10	10	10	12	33.35
15	40(18)	JDC-OSS	159.60	99.895	0.785	10	10	11	15.25	47.15
		SDC-OSS	96.60	99.876	0.795	10	10	11	16	67.15
		JDC-RSS	40.52	99.556	0.515	10	11	16.5	45.5	200.45
		SDC-RSS	5.46	99.581	0.540	10	11	16	44.25	175.7
	60(19)	JDC-OSS	219.11	99.975	0.920	10	10	10	11	26
		SDC-OSS	150.09	99.971	0.925	10	10	10	11	24.1
		JDC-RSS	50.01	99.862	0.805	10	10	10	14.25	102.75
		SDC-RSS	7.61	99.869	0.825	10	10	11	14	72.4
	100(21)	JDC-OSS	333.28	1.000	1.000	10	10	10	10	10
		SDC-OSS	271.05	1.000	1.000	10	10	10	10	10
		JDC-RSS	60.66	1.000	0.995	10	10	10	10	11.05
		SDC-RSS	8.21	1.000	0.980	10	10	10	10	11
				Right-censored rate = 40%						
10	100(20)	JDC-OSS	238.24	99.964	0.940	10	10	10	10.25	21
		SDC-OSS	174.54	99.953	0.935	10	10	10	11	21.05
		JDC-RSS	47.76	99.930	0.895	10	10	10	13.25	34.1
		SDC-RSS	9.05	99.903	0.855	10	10	10	13	43.05
	150(21)	JDC-OSS	370.59	99.999	1.000	10	10	10	10	10.05
		SDC-OSS	307.65	99.997	0.995	10	10	10	10	11
		JDC-RSS	62.32	99.988	0.980	10	10	10	10	13.05
		SDC-RSS	13.04	99.988	0.975	10	10	10	10	15
15	40(18)	JDC-OSS	160.02	99.963	0.905	10	10	10	11.25	25.05
		SDC-OSS	96.36	99.956	0.900	10	10	10	12	27.05
		JDC-RSS	39.99	99.885	0.790	10	10	11	16	71
		SDC-RSS	5.61	99.863	0.780	10	10	11	16	59.25
	60(19)	JDC-OSS	216.32	99.996	0.995	10	10	10	10	11
		SDC-OSS	148.99	99.989	0.995	10	10	10	10	11
		JDC-RSS	49.02	99.972	0.935	10	10	10	10	21.15
		SDC-RSS	7.08	99.977	0.945	10	10	10	10	22
	100(21)	JDC-OSS	325.63	1.000	1.000	10	10	10	10	10
		SDC-OSS	272.76	1.000	1.000	10	10	10	10	10
		JDC-RSS	56.68	1.000	0.990	10	10	10	10	10
		SDC-RSS	8.27	1.000	0.990	10	10	10	10	10
				Right-censored rate = 60%						
10	100(20)	JDC-OSS	236.64	99.970	0.950	10	10	10	11	19.25
		SDC-OSS	177.16	99.956	0.940	10	10	10	11	22.1
		JDC-RSS	47.70	99.899	0.860	10	10	10	13	52.2
		SDC-RSS	9.92	99.891	0.845	10	10	10	14	58.45
	150(21)	JDC-OSS	364.63	99.999	0.995	10	10	10	10	10
		SDC-OSS	305.40	99.999	0.995	10	10	10	10	10
		JDC-RSS	64.41	99.990	0.980	10	10	10	10	14
		SDC-RSS	13.01	99.990	0.985	10	10	10	10	12.05
15	40(18)	JDC-OSS	153.79	99.963	0.900	10	10	10	12	30.25
		SDC-OSS	94.08	99.946	0.895	10	10	10	11	37.05
		JDC-RSS	40.94	99.918	0.820	10	10	11	15	51.2
		SDC-RSS	6.75	99.910	0.780	10	10	11	17	55.5
	60(19)	JDC-OSS	216.31	99.997	0.995	10	10	10	10	12
		SDC-OSS	151.28	99.995	0.990	10	10	10	10	11
		JDC-RSS	50.36	99.961	0.945	10	10	10	11	29.1
		SDC-RSS	7.68	99.954	0.940	10	10	10	10.25	26.1
	100(21)	JDC-OSS	328.68	1.000	1.000	10	10	10	10	10
		SDC-OSS	271.97	1.000	1.000	10	10	10	10	10
		JDC-RSS	57.70	1.000	1.000	10	10	10	10	10
		SDC-RSS	7.91	1.000	1.000	10	10	10	10	10

Table 3: Simulation results under Model 2 (c).

						S				
n	$B(d_0)$	*Method*	time(sec)	$AUC(\%)$	PA	5%	25%	50%	75%	95%
				Right-censored rate = 20%						
10	100(20)	JDC-OSS	227.49	97.748	0.295	7	16	49	150.25	435.15
		SDC-OSS	165.37	97.439	0.240	7	21	56.5	155.25	500.65
		JDC-RSS	49.63	95.387	0.080	13.95	56	129.5	297	610.3
		SDC-RSS	7.87	95.108	0.060	18.95	58.75	131	315.25	677.4
	200(23)	JDC-OSS	500.17	99.246	0.575	6	7	17	55.25	205.4
		SDC-OSS	434.97	99.206	0.555	6	7	16	56.25	220.8
		JDC-RSS	79.51	98.209	0.350	6	16.75	39	115	425.25
		SDC-RSS	13.63	98.025	0.325	6	17	49.5	152	432.55
	300(24)	JDC-OSS	598.92	0.999	1.000	6	6	6.5	12.75	20.3
		SDC-OSS	508.54	0.998	0.800	6	6.25	8.5	22	53.7
		JDC-RSS	107.70	0.991	0.400	6	16.5	61.5	85	122.3
		SDC-RSS	15.18	0.989	0.300	7.35	19.75	62	104.5	147.2
20	80(22)	JDC-OSS	250.56	99.462	0.730	6	6	8.5	26	238.35
		SDC-OSS	205.91	99.485	0.715	6	6	9	27.25	165.65
		JDC-RSS	49.27	98.650	0.570	6	8	18	64.25	429.2
		SDC-RSS	5.15	98.663	0.575	6	8.75	17.5	58.75	389.7
	100(23)	JDC-OSS	329.16	99.692	0.855	6	6	7	13	103.45
		SDC-OSS	258.02	99.710	0.850	6	6	7	13	106
		JDC-RSS	79.10	99.164	0.620	6	7	13.5	42	255.75
		SDC-RSS	5.57	99.187	0.615	6	8	16	49.25	247.25
				Right-censored rate = 40%						
10	100(20)	JDC-OSS	131.53	97.981	0.285	7	17.75	43	142	466.45
		SDC-OSS	101.52	98.084	0.315	7	14.75	48.5	122	418.75
		JDC-RSS	34.93	96.028	0.075	16	39.75	120.5	258.75	565.55
		SDC-RSS	5.39	95.860	0.075	16.9	54.5	130	272.75	566.2
	200(23)	JDC-OSS	325.29	99.522	0.720	6	7	9	25.75	159.4
		SDC-OSS	263.24	99.512	0.685	6	7	10	27.25	159.35
		JDC-RSS	73.86	98.504	0.435	6	9.75	31	104.25	391.4
		SDC-RSS	9.88	98.442	0.450	6	11	30	104.75	374.65
	300(24)	JDC-OSS	593.41	99.936	0.900	6	6	6	8.25	23.95
		SDC-OSS	503.65	99.896	0.900	6	6	6	6.75	39.5
		JDC-RSS	106.40	99.190	0.600	6.9	9.25	17	51.5	198.25
		SDC-RSS	14.72	99.155	0.600	6.9	9.25	16.5	80.5	179.3
20	80(22)	JDC-OSS	384.31	99.768	0.870	6	6	6	9	98.1
		SDC-OSS	313.02	99.796	0.895	6	6	6	10.25	72.2
		JDC-RSS	65.52	99.457	0.745	6	6	9	23	199.3
		SDC-RSS	7.54	99.419	0.755	6	6	9	21.25	202.35
	100(23)	JDC-OSS	524.90	99.922	0.945	6	6	6	7	24.35
		SDC-OSS	441.82	99.918	0.945	6	6	6	7	26
		JDC-RSS	73.99	99.719	0.820	6	6	7	13.25	71.05
		SDC-RSS	9.09	99.715	0.835	6	6	7	14.25	87.8
				Right-censored rate = 60%						
10	100(20)	JDC-OSS	135.23	98.021	0.350	7	15.75	42	122.25	465.25
		SDC-OSS	105.50	97.972	0.320	7	15	39	134.75	449.35
		JDC-RSS	35.18	95.707	0.090	11	56.5	132	281.5	641.35
		SDC-RSS	5.31	95.349	0.080	13.9	60	136.5	314.25	669.25
	200(23)	JDC-OSS	499.71	99.493	0.685	6	6	10	37	140.15
		SDC-OSS	435.07	99.477	0.615	6	6	10	42.25	140.8
		JDC-RSS	79.22	98.619	0.465	6	11	26.5	115.25	313.05
		SDC-RSS	13.63	98.504	0.445	6	11.75	32.5	108	343.3
	300(24)	JDC-OSS	609.56	0.999	0.900	6	6	7	9.5	39.4
		SDC-OSS	517.65	0.998	0.800	6	6	7	16	48.9
		JDC-RSS	108.72	0.993	0.800	6.45	7.25	8	16	219.05
		SDC-RSS	15.16	0.995	0.900	6	6	7	14	141.8
20	80(22)	JDC-OSS	241.72	99.869	0.885	6	6	6	8	54.85
		SDC-OSS	183.21	99.890	0.885	6	6	6	8	50.05
		JDC-RSS	65.47	99.558	0.745	6	6	9	23	140.3
		SDC-RSS	4.81	99.557	0.765	6	6	9	20	143.05
	100(23)	JDC-OSS	370.51	99.958	0.970	6	6	6	7	21
		SDC-OSS	311.35	99.958	0.950	6	6	6	7	21.15
		JDC-RSS	65.45	99.682	0.830	6	6	7	14	114.2
		SDC-RSS	6.55	99.690	0.830	6	6	7	14	79.55

Table 4: Simulation results under Model 3.

						S				
n	$B(d_0)$	*Method*	time(sec)	$AUC(\%)$	PA	5%	25%	50%	75%	95%
Right-censored rate = 20%										
8	100(19)	JDC-OSS	97.56	97.648	0.055	19	60	116.5	207.75	503.6
		SDC-OSS	68.43	97.031	0.015	26.95	72.75	132	272.75	618.1
		JDC-RSS	34.68	96.526	0.015	46	95.5	175.5	318.25	633.55
		SDC-RSS	4.81	95.697	0.005	44.95	111.5	205	358	672.45
	150(20)	JDC-OSS	160.88	99.092	0.235	12	21	42.5	111	277.95
		SDC-OSS	121.28	98.865	0.19	12	25.75	53	118.25	299.5
		JDC-RSS	48.13	98.607	0.115	14.95	32	62.5	139.25	345.15
		SDC-RSS	7.25	98.157	0.09	17.95	41.75	84	171.25	482.65
	200(22)	JDC-OSS	236.07	99.529	0.54	10	13	20.5	48	208.55
		SDC-OSS	186.45	99.384	0.45	10	14	26.5	68.25	250.45
		JDC-RSS	61.07	99.197	0.33	11	18.75	36.5	90.25	260.55
		SDC-RSS	9.76	98.986	0.22	12	24.75	47	112.25	300.2
10	80(19)	JDC-OSS	97.77	98.483	0.105	16	33	74.5	152.75	359.95
		SDC-OSS	67.66	98.164	0.06	16.95	37.75	87.5	180.5	429.5
		JDC-RSS	34.97	97.554	0.06	19	50.75	109.5	262.5	560.6
		SDC-RSS	4.03	97.179	0.03	22	62	121	274.75	607
	100(20)	JDC-OSS	127.40	98.987	0.335	10	17	41.5	109.75	352.3
		SDC-OSS	92.22	98.926	0.255	11	20	42.5	117.25	360.1
		JDC-RSS	41.40	98.537	0.095	14.95	35	63.5	142	359.5
		SDC-RSS	4.95	98.158	0.085	17	36	78.5	181.5	440.75
Right-censored rate = 40%										
8	100(19)	JDC-OSS	97.40	98.514	0.145	13	33	62.5	149.25	410.15
		SDC-OSS	68.51	98.295	0.115	14.95	30.75	68	190.25	461.05
		JDC-RSS	34.25	97.917	0.065	18	44.75	93	208.25	470.5
		SDC-RSS	4.76	97.717	0.045	20	47	102	211.75	475.45
	150(20)	JDC-OSS	161.32	99.421	0.435	10	13.75	23	75	222.6
		SDC-OSS	121.58	99.336	0.425	10	14	27	81	226.3
		JDC-RSS	48.22	99.327	0.295	11	18	36	74	243.35
		SDC-RSS	7.27	99.256	0.265	12	19.75	39.5	81.25	214.2
	200(22)	JDC-OSS	235.17	99.685	0.65	10	11	15	32.25	134.6
		SDC-OSS	185.76	99.605	0.595	10	11	17	43	197.2
		JDC-RSS	60.89	99.598	0.565	10	12	19	43.25	184.1
		SDC-RSS	9.81	99.541	0.57	10	12	18.5	49	191.7
10	80(19)	JDC-OSS	97.22	99.116	0.33	11	17	35.5	77.5	325
		SDC-OSS	67.75	99.036	0.285	11	17.75	35.5	93	317.3
		JDC-RSS	34.14	98.796	0.16	12.95	22	45	129.5	371.45
		SDC-RSS	4.00	98.650	0.14	14	26	49.5	148.5	374.4
	100(20)	JDC-OSS	129.14	99.482	0.465	10	12	22	52.5	189.25
		SDC-OSS	93.28	99.463	0.505	10	12	20	55.25	245.55
		JDC-RSS	41.70	99.425	0.4	11	15.75	27.5	64	202
		SDC-RSS	5.05	99.331	0.345	11	17	31	72	224.85
Right-censored rate = 60%										
8	100(19)	JDC-OSS	97.77	98.459	0.125	15	32	68	153.75	467.25
		SDC-OSS	68.53	98.288	0.08	15.95	39	78	171	530
		JDC-RSS	34.69	97.807	0.065	19	45.5	89	208	525.65
		SDC-RSS	4.82	97.656	0.04	20.95	48.75	100.5	224.25	583.7
	150(20)	JDC-OSS	159.60	99.384	0.38	10	15	28	76.25	211.35
		SDC-OSS	120.64	99.292	0.35	10	15.75	33.5	85.25	239.55
		JDC-RSS	47.36	99.217	0.32	11	17	34	86.75	247.6
		SDC-RSS	7.18	99.034	0.27	11	19.75	41	88	323.1
	200(22)	JDC-OSS	237.29	99.676	0.655	10	11	14	33	168.2
		SDC-OSS	187.39	99.639	0.64	10	11	15	39.25	167.15
		JDC-RSS	61.55	99.593	0.61	10	12	17	40.25	158.45
		SDC-RSS	9.74	99.507	0.555	10	12	21	48.75	221.1
10	80(19)	JDC-OSS	97.51	99.248	0.325	11	17.75	35	78.25	286.15
		SDC-OSS	67.82	99.208	0.275	11	19	34.5	88	290.25
		JDC-RSS	34.44	98.777	0.21	13	23	52	129.5	380.05
		SDC-RSS	3.97	98.652	0.19	13	24	58	138.5	457.6
	100(20)	JDC-OSS	127.44	99.479	0.42	10	13	25.5	53	232.1
		SDC-OSS	92.30	99.491	0.445	10	14	26.5	54	218.9
		JDC-RSS	41.18	99.186	0.3	11	17	36.5	87.75	257.1
		SDC-RSS	4.92	99.109	0.31	10.95	18	35	97.5	231.55

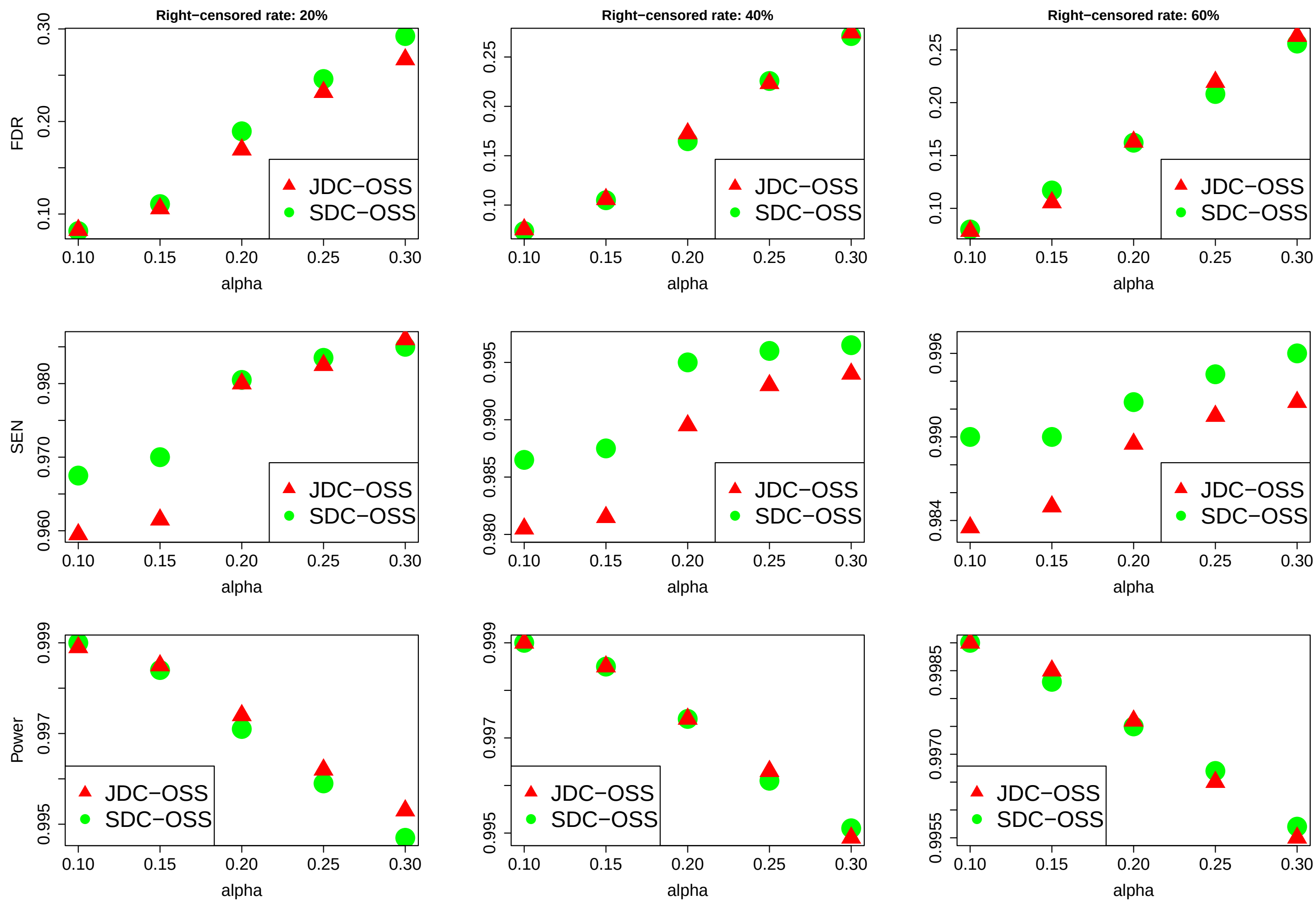

Figure 1: Empirical FDR, SEN and powers.

Figure 1 presents the FDR control for the two proposed methods with $n = 20$, $B = 100$, $p = 5000$ and $\alpha = 0.10, 0.15, 0.20, 0.25$ or 0.30 based on 200 replications. Here the true failure times were generated under **Model 1** with the true parameter being $\boldsymbol{\beta} = (\mathbf{2}_{10}, \mathbf{0}_{p-10})^\top$ and the censoring rate is 20%, 40% or 60%. That is, there exist ten important or relevant variables.

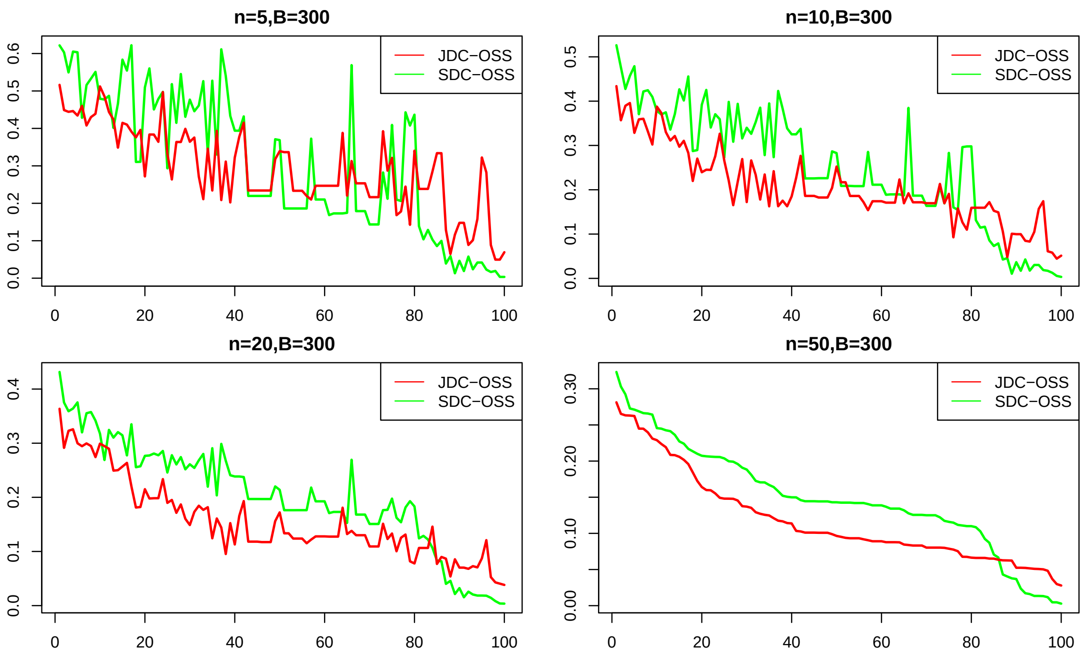

Figure 2: The comparison of JDC-OSS and SDC-OSS in different n with $B = 300$.

Table 5: The screening results for the SEER data with 50 covariates and their interaction terms.

Covariates	DC (B=100)	JDC-OSS (B=100,n=15)	SDC-OSS (B=100,n=15)	JDC-RSS (B=100,n=15)	SDC-RSS (B=100,n=15)
Year.of.diagnosis*CS.Schema.AJCC.6th.Edition	√	√	√	√	√
Year.of.diagnosis	√	√	√	√	√
Year.of.diagnosis*Site.recode.ICD.O.3.WHO.2008	√	√	√	√	√
Year.of.diagnosis*TNM.7.CS.v0204..Schema.recode	√	√			√
Year.of.diagnosis*Site.recode.ICD.O.3.WHO.2008.for.SIRs	√	√		√	√
Year.of.diagnosis*Histology.recode.Brain.groupings	√	√	√	√	
Year.of.diagnosis*SEER.Brain.and.CNS.Recode	√	√	√	√	√
Year.of.diagnosis*Behavior.code.ICD.O.3	√	√	√	√	√
Year.of.diagnosis*Behavior.recode.for.analysis	√	√	√	√	√
Year.of.diagnosis*AYA.site.recode.WHO.2008	√		√	√	√
Sex*Year.of.diagnosis	√		√		
Year.of.diagnosis*TNM.7.CS.v0204.Schema.thru.2017	√	√	√		√
Site.recode.rare.tumors*Breast.T		√	√		
Primary.Site*Histologic.Type.ICD.O.3		√	√		
Site.recode.rare.tumors*RX.Summ.Surg.Prim.Site		√	√	√	
Site.recode.rare.tumors*Breast.Stage		√	√		

In the figure, three metrics were calculated and they are the average of empirical FDR (FDR), the average of empirical sensitivity (SEN) and the average of empirical powers (Power). One can see from the figure that overall both methods gave good and consistent performances. Also the two procedures are similar in terms of FDR and Power but the SDC-OSS method seems to yield higher SEN values than the JDC-OSS method.

S3 Preliminary Lemmas

Before the proof of our main results, we introduce the following preliminary lemmas. The first two are extracted from Lemma 5.6.1.A and Theorem 5.6.1.A of Serfling (1980) for which the proof is omitted.

Lemma 1. *Let $\mu = E(Y)$. If $Pr(a \le Y \le b) = 1$, then*

$$E\left[\exp\{s(Y-\mu)\}\right] \le \exp(s^2(b-a)^2/8), \quad \textit{for any} \quad s > 0.$$

Lemma 2. *Let $h(x_1, \ldots, x_m)$ be a kernel of the U-statistic U_n, and $\theta = E(h(x_1, \ldots, x_m))$. If $a \le h(x_1, \ldots, x_m) \le b$, then for any $t > 0$ and $n \ge m$,*

$$Pr(|U_n - \theta| \ge t) \le 2\exp(-2\lfloor n/m \rfloor t^2/(b-a)^2),$$

where $\lfloor n/m \rfloor$ denotes the integer part of n/m.

Lemma 3. *Suppose $\widehat{\gamma}_1, \widehat{\gamma}_2$, and $\widehat{\gamma}_3$ are estimates of parameters γ_1, γ_2, and γ_3 based on a size-n sample, respectively. Assume $\gamma_2 > 0, \gamma_3 > 0$ and*

$M \geq 2\max\{\gamma_1, \gamma_2, \gamma_3, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3\}$. If

$$P(|\widehat{\gamma}_k - \gamma_k| > \varepsilon) \leq c_1 \exp(-c_2 n^{1-\kappa}\varepsilon^2) + n\exp(-c_3 n^{\kappa}), k = 1, 2, 3,$$

for some positive constants c_1, c_2, c_3. Then we have

$$Pr\left(\left|\frac{\widehat{\gamma}_1}{\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3}} - \frac{\gamma_1}{\gamma_2\gamma_3}\right| > \varepsilon\right) \leq 5c_1 \exp\{-c_2 n^{1-\kappa}\varepsilon^2\gamma_0\} + 5n\exp(-c_3 n^{\kappa}),$$

where $\gamma_0 = \min\{\gamma_2^2\gamma_3^2/4M^4, \gamma_2^3\gamma_3^3/4M^4\}$.

Proof of Lemma 3. Some $\gamma_1, \gamma_2, \gamma_3, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3$ are bounded by $M/2$. It is easy to verify that

$$Pr\big(|\widehat{\gamma}_2\widehat{\gamma}_3 - \gamma_2\gamma_3| > 2\varepsilon\big) = Pr\big(|\widehat{\gamma}_2(\widehat{\gamma}_3 - \gamma_3) + (\widehat{\gamma}_2 - \gamma_2)\gamma_3| > 2\varepsilon\big)$$
$$\leq Pr\big(|\widehat{\gamma}_3 - \gamma_3| > \frac{\varepsilon}{\widehat{\gamma}_2}\big) + Pr\big(|\widehat{\gamma}_2 - \gamma_2| > \frac{\varepsilon}{\gamma_3}\big)$$
$$\leq Pr\big(|\widehat{\gamma}_3 - \gamma_3| > \frac{2\varepsilon}{M}\big) + Pr\big(|\widehat{\gamma}_2 - \gamma_2| > \frac{2\varepsilon}{M}\big)$$
$$\leq 2c_1 \exp(-c_2 n^{1-\kappa} 4\varepsilon^2/M^2) + 2n\exp(-c_3 n^{\kappa}).$$

If event $\{|\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3} - \sqrt{\gamma_2\gamma_3}| > 2\varepsilon\}$ is true, then we have

$$2\varepsilon\sqrt{\gamma_2\gamma_3} < \sqrt{\gamma_2\gamma_3}|\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3} - \sqrt{\gamma_2\gamma_3}| < |\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3} + \sqrt{\gamma_2\gamma_3}|\cdot|\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3} - \sqrt{\gamma_2\gamma_3}| = |\widehat{\gamma}_2\widehat{\gamma}_3 - \gamma_2\gamma_3|.$$

Thus,

$$Pr\big(|\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3} - \sqrt{\gamma_2\gamma_3}| > 2\varepsilon\big) \tag{S3.1}$$
$$\leq Pr\big(|\widehat{\gamma}_2\widehat{\gamma}_3 - \gamma_2\gamma_3| > 2\varepsilon\sqrt{\gamma_2\gamma_3}\big)$$
$$\leq 2c_1 \exp(-c_2 n^{1-\kappa}\gamma_2\gamma_3 4\varepsilon^2/M^2) + 2n\exp(-c_3 n^{\kappa}).$$

Let $\gamma = \sqrt{\gamma_2\gamma_3}$ and $\widehat{\gamma} = \sqrt{\widehat{\gamma}_2\widehat{\gamma}_3}$. For any $0 < \varepsilon < 1$, we have

$$
\begin{aligned}
&Pr\Big(|\frac{1}{\widehat{\gamma}} - \frac{1}{\gamma}| > \varepsilon\Big) = Pr\big(|\widehat{\gamma} - \gamma| > |\widehat{\gamma}\gamma|\varepsilon\big) \\
\le& Pr\Big(|\widehat{\gamma} - \gamma| > |\widehat{\gamma}\gamma|\varepsilon, |\widehat{\gamma}| \ge \frac{\gamma}{2}\Big) + Pr\Big(|\widehat{\gamma}| < \frac{\gamma}{2}\Big) \\
\le& Pr\Big(|\widehat{\gamma} - \gamma| > \frac{\gamma^2}{2}\varepsilon\Big) + Pr\Big(|\widehat{\gamma} - \gamma| > \frac{\gamma}{2}\Big) \\
\le& Pr\Big(|\widehat{\gamma} - \gamma| > \frac{\gamma^2}{2}\varepsilon\Big) + Pr\Big(|\widehat{\gamma} - \gamma| > \frac{\gamma}{2}\varepsilon\Big) \\
\le& 2Pr\Big(|\widehat{\gamma} - \gamma| > \min\{\gamma^2, \gamma\}\frac{\varepsilon}{2}\Big).
\end{aligned}
$$

The second inequality sign in the above equation is due to $\{|\widehat{\gamma}| < \gamma/2\} = \{\widehat{\gamma} - \gamma < -\gamma/2\} = \{\gamma - \widehat{\gamma} > \gamma/2\}$ and $|\widehat{\gamma} - \gamma| > \gamma - |\widehat{\gamma}| > \gamma/2$. Therefore, we have $\{|\widehat{\gamma}| < \gamma/2\} \subseteq \{|\widehat{\gamma} - \gamma| > \gamma/2\}$. The third inequality is because $|\widehat{\gamma} - \gamma| > \gamma/2 > \frac{\gamma}{2}\varepsilon$. From (S3.1), we have

$$
\begin{aligned}
&Pr\Big(\Big|\frac{1}{\widehat{\gamma}} - \frac{1}{\gamma}\Big| > \varepsilon\Big) \\
\le& 4c_1 \exp\big(-c_2 n\gamma^2 \min\{\gamma^4, \gamma^2\}\frac{\varepsilon^2}{4M^2}\big) \\
\le& 4c_1 \exp\big(-c_2 n^{1-\kappa}\gamma^{'}\frac{\varepsilon^2}{4M^2}\big) + 4n\exp(-c_3 n^{\kappa}),
\end{aligned}
$$

where $\gamma^{'} = \min\{\gamma_2^3\gamma_3^3, \gamma_2^2\gamma_3^2\}$. As a result

$$
\begin{aligned}
&Pr\Big(\Big|\frac{\widehat{\gamma}_1}{\sqrt{\widehat{\gamma}_2\widehat{\gamma}_3}} - \frac{\gamma_1}{\sqrt{\gamma_2\gamma_3}}\Big| > \varepsilon\Big) \\
=& Pr\Big(\Big|\frac{\widehat{\gamma}_1}{\widehat{\gamma}} - \frac{\gamma_1}{\gamma}\Big| > \varepsilon\Big) = Pr\Big(\Big|\widehat{\gamma}_1\big(\frac{1}{\widehat{\gamma}} - \frac{1}{\gamma}\big) + (\widehat{\gamma}_1 - \gamma_1)\frac{1}{\gamma}\Big| > \varepsilon\Big) \\
\le& Pr\Big(|\frac{1}{\widehat{\gamma}} - \frac{1}{\gamma}| > \frac{\varepsilon}{2\widehat{\gamma}_1}\Big) + Pr\big(|\widehat{\gamma}_1 - \gamma_1| > \frac{\varepsilon\gamma}{2}\big)
\end{aligned}
$$

$$\leq Pr\Big(|\frac{1}{\widehat{\gamma}} - \frac{1}{\gamma}| > \frac{\varepsilon}{M}\Big) + Pr\big(|\widehat{\gamma}_1 - \gamma_1| > \frac{\varepsilon\gamma}{2}\big)$$

$$\leq 4c_1 \exp\big(-c_2 n^{1-\kappa}\gamma' \frac{\varepsilon^2}{4M^4}\big) + 4n\exp(-c_3 n^\kappa) + c_1 \exp\big(-c_2 n^{1-\kappa}\frac{\varepsilon^2\gamma^2}{4}\big) + n\exp(-c_3 n^\kappa)$$

$$\leq 5c_1 \exp\big(-c_2 n^{1-\kappa}\varepsilon^2\gamma_0\big) + 5n\exp(-c_3 n^\kappa),$$

where $\gamma_0 = \min\Big\{\frac{\gamma_2^3\gamma_3^3}{4M^4}, \frac{\gamma_2^2\gamma_3^2}{4M^4}, \frac{\gamma_2\gamma_3}{4}\Big\}$. According to $\gamma_2 \leq M/2$ and $\gamma_3 \leq M/2$, $\frac{\gamma_2^3\gamma_3^3}{4M^4} = \frac{\gamma_2\gamma_3}{4}\frac{\gamma_2^2\gamma_3^2}{M^4} < \frac{\gamma_2\gamma_3}{4}$ can be obtained. Thus $\gamma_0 = \min\Big\{\frac{\gamma_2^3\gamma_3^3}{4M^4}, \frac{\gamma_2^2\gamma_3^2}{4M^4}, \frac{\gamma_2\gamma_3}{4}\Big\} = \min\Big\{\frac{\gamma_2^3\gamma_3^3}{4M^4}, \frac{\gamma_2^2\gamma_3^2}{4M^4}\Big\}$.

S4 Technical Proofs

Proof of Theorem 1.

In the following, we will derive the order of the variance for $\widehat{\omega}_j^{SDC}$ and $\widehat{\omega}_j^{JDC}$, respectively.

Part I: Variance order derivation for $\widehat{\omega}_j^{SDC}$.

Define $\boldsymbol{D}_j = (S_{j1}, \ldots, S_{j8})^\top$, and $\widehat{\boldsymbol{D}}_{(b),j} = (\hat{S}_{(b),j1}, \ldots, \hat{S}_{(b),j8})^\top$ with $\hat{S}_{(b),jh}, h = 1, \ldots, 8$ are the corresponding estimates of $S_{(b),jh}, h = 1, \ldots, 8$ based on the data segment $\mathcal{D}_b$. Define function $g(\boldsymbol{x}) = \frac{x_1 + x_2 x_3 - 2x_4}{\sqrt{x_5 + x_2^2 - 2x_6}\sqrt{x_7 + x_3^2 - 2x_8}}$ with $\boldsymbol{x} = (x_1, \ldots, x_8)^\top$. In order to reduce the symbolic complexity, we omit the covariate index in the subsequent proof. So that $g(\widehat{\boldsymbol{D}}_{(b)}) = \hat{\omega}_{(b)}$, $g(\boldsymbol{D}) = \omega$ and $\hat{\omega}^{SDC} = B^{-1}\sum_{b=1}^B \hat{\omega}_{(b)} = B^{-1}\sum_{b=1}^B g(\widehat{\boldsymbol{D}}_{(b)})$.

Take Taylor expansion of $g(\widehat{\boldsymbol{D}}_{(b)})$ and since the derivative of $g(\cdot)$ is bounded, there is $C_1 > 0$ such that $\|g'(\boldsymbol{D})\|_{\max} \leq C_1$. We can get the following inequality

$$g(\widehat{\boldsymbol{D}}_{(b)}) = g\big(E\widehat{\boldsymbol{D}}_{(b)}\big) + g'\big(E\widehat{\boldsymbol{D}}_{(b)}\big)^{\top}\big(\widehat{\boldsymbol{D}}_{(b)} - E\widehat{\boldsymbol{D}}_{(b)}\big).$$

According to Condition (C2) and the proof of proposition 1 in Li et al. (2020), we can get the variance of $\hat{S}_{(b),1}, \ldots, \hat{S}_{(b),8}$ as follow

$$Var(\hat{S}_{(b),1}) = Var\Big(\binom{n}{2}^{-1}\sum_{i,k}^{n}\frac{1}{2!}\sum_{\Omega\{i,k\}}\|X_{ji}-X_{jk}\|_1\|\mathbf{Y}_i-\mathbf{Y}_k\|_2\Big) = O\Big(\frac{1}{n^2}\Big),$$

$$Var(\hat{S}_{(b),2}) = Var\Big(\binom{n}{2}^{-1}\sum_{i,k}^{n}\frac{1}{2!}\sum_{\Omega\{i,k\}}\|X_{ji}-X_{jk}\|_1\Big) = O\Big(\frac{1}{n^2}\Big),$$

$$Var(\hat{S}_{(b),3}) = Var\Big(\binom{n}{2}^{-1}\sum_{i,k}^{n}\frac{1}{2!}\sum_{\Omega\{i,k\}}\|\mathbf{Y}_i-\mathbf{Y}_k\|_2\Big) = O\Big(\frac{1}{n^2}\Big),$$

$$Var(\hat{S}_{(b),4}) = Var\Big(\binom{n}{3}^{-1}\sum_{i,k}^{n}\frac{1}{3!}\sum_{\Omega\{i,k,l\}}\|X_{ji}-X_{jl}\|_1\|\mathbf{Y}_k-\mathbf{Y}_l\|_2\Big) = O\Big(\frac{1}{n^3}\Big),$$

$$\ldots,$$

$$Var(\hat{S}_{(b),8}) = Var\Big(\binom{n}{3}^{-1}\sum_{i,k}^{n}\frac{1}{3!}\sum_{\Omega\{i,k,l\}}\|\mathbf{Y}_i-\mathbf{Y}_l\|_2\|\mathbf{Y}_k-\mathbf{Y}_l\|_2\Big) = O\Big(\frac{1}{n^3}\Big).$$

According to the relationship between correlation coefficient and covariance and the value range of correlation coefficient (that is, $0 \leq \frac{Cov(X,Y)^2}{Cov(X)Cov(Y)} \leq 1$), it can be obtained

$$\big(Cov(\hat{S}_{(b),1}, \hat{S}_{(b),2})\big)^2 \leq Var(\hat{S}_{(b),1})Var(\hat{S}_{(b),2}) = O\Big(\frac{1}{n^4}\Big),$$

and then $Cov(\hat{S}_{(b),1}, \hat{S}_{(b),2}) = O\Big(\frac{1}{n^2}\Big)$. Similarly, we can get $Cov(\hat{S}_{(b),1}, \hat{S}_{(b),3}) = O\Big(\frac{1}{n^2}\Big), Cov(\hat{S}_{(b),1}, \hat{S}_{(b),4}) = O\Big(\frac{1}{n^{5/2}}\Big), \ldots, Cov(\hat{S}_{(b),7}, \hat{S}_{(b),8}) = O\Big(\frac{1}{n^{5/2}}\Big)$.

Thus

$$
\begin{aligned}
&Var(g(\widehat{\boldsymbol{D}}_{(b)})) \\
\leq &\sum_{h=1}^{8} C_1^2 Var\Big(\hat{S}_{(b),h} - E(\hat{S}_{(b),h})\Big) + \sum_{h_1 \neq h_2}^{8} C_1^2 Cov\Big(\hat{S}_{(b),h_1} - E(\hat{S}_{(b),h_1}), \hat{S}_{(b),h_2} - E(\hat{S}_{(b),h_2})\Big) \\
= &\sum_{h=1}^{8} C_1^2 Var\Big(\hat{S}_{(b),h}\Big) + \sum_{h_1 \neq h_2}^{8} C_1^2 Cov\Big(\hat{S}_{(b),h_1}, \hat{S}_{(b),h_2}\Big) \\
= &O\Big(\frac{1}{n^2}\Big) + O\Big(\frac{1}{n^{5/2}}\Big) + O\Big(\frac{1}{n^3}\Big).
\end{aligned}
$$

Then, combined with condition $B = O(N^{\alpha})$ and $n = O(N^{\iota})$, we get

$$
\begin{aligned}
Var(\widehat{\omega}^{SDC}) = Var\Big(\frac{1}{B}\sum_{b=1}^{B} g(\widehat{\boldsymbol{D}}_{(b)})\Big) = \frac{1}{B} Var(g(\widehat{\boldsymbol{D}}_{(b)})) \\
= O\Big(\frac{1}{N^{\alpha+2\iota}}\Big) + O\Big(\frac{1}{N^{\alpha+5/2\iota}}\Big) + O\Big(\frac{1}{N^{\alpha+3\iota}}\Big),
\end{aligned}
$$

for $j \in 1, \ldots, p$. Therefore,

$$
\max_{j=1,\ldots,p} Var(\widehat{\omega}_j^{SDC}) = O\Big(\frac{1}{N^{\alpha+2\iota}}\Big) + O\Big(\frac{1}{N^{\alpha+5/2\iota}}\Big) + O\Big(\frac{1}{N^{\alpha+3\iota}}\Big).
$$

Part II: Variance order derivation for $\widehat{\omega}_j^{JDC}$.

Define $\boldsymbol{D}_j = (S_{j1}, S_{j2}, S_{j3}, S_{j4})^{\top}$, and $\widehat{\boldsymbol{D}}_{(b),j} = (\hat{S}_{(b),j1}, \hat{S}_{(b),j2}, \hat{S}_{(b),j3}, \hat{S}_{(b),j4})^{\top}$ with $\hat{S}_{(b),jh}, h = 1, \ldots, 4$ are the corresponding estimates of $S_{(b),jh}, h = 1, \ldots, 4$ based on the data segment $\mathcal{D}_b$. After removing the m-th sample in data segment $\mathcal{D}_b$, the estimation based on the remaining $(n-1)$ samples is denoted as $\widehat{\boldsymbol{D}}_{(b,-m)} = (\hat{S}_{(b,-m),j1}, \hat{S}_{(b,-m),j2}, \hat{S}_{(b,-m),j3}, \hat{S}_{(b,-m),j4})^{\top}$. Define function $g(\boldsymbol{x}) = x_1 + x_2 x_3 - 2x_4$ with $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^{\top}$. In order to reduce the symbolic complexity, we omit the covariate index in the subsequent proof. So that the distance covariance of the population is $dcov^2(X, \mathbf{Y}) = g(\boldsymbol{D})$, the distance covariance estimation for the data segment $\mathcal{D}_{(b,-m)}$ is $g(\widehat{\boldsymbol{D}}_{(b,-m)})$, and for the data segment $\mathcal{D}_{(b)}$ is $g(\widehat{\boldsymbol{D}}_{(b)})$.

There have

$$\begin{aligned}\widehat{dcov^2}_{(b)}^{JDC}(X, \mathbf{Y}) &= g(\widehat{\boldsymbol{D}}_{(b)}) - \frac{n-1}{n}\sum_{m=1}^{n}\left[g(\widehat{\boldsymbol{D}}_{(b,-m)}) - (n-1)g(\widehat{\boldsymbol{D}}_{(b)})\right] \\ &= ng(\widehat{\boldsymbol{D}}_{(b)}) - \frac{n-1}{n}\sum_{m=1}^{n} g(\widehat{\boldsymbol{D}}_{(b,-m)}). \qquad \text{(S4.1)}\end{aligned}$$

Similar to the above proof procedure of $Var(\widehat{\omega}^{SDC})$, we can also derive

$$Var(g(\widehat{\boldsymbol{D}}_{(b)})) = O\Big(\frac{1}{n^2}\Big) + O\Big(\frac{1}{n^{5/2}}\Big) + O\Big(\frac{1}{n^3}\Big), \qquad \text{(S4.2)}$$

$$Var(g(\widehat{\boldsymbol{D}}_{(b,-m)})) = O\Big(\frac{1}{(n-1)^2}\Big) + O\Big(\frac{1}{(n-1)^{5/2}}\Big) + O\Big(\frac{1}{(n-1)^3}\Big), \qquad \text{(S4.3)}$$

and

$$\begin{aligned}&Cov\Big(ng(\widehat{\boldsymbol{D}}_{(b)}), \frac{n-1}{n}\sum_{m=1}^{n} g(\widehat{\boldsymbol{D}}_{(b,-m)})\Big) \\ =&n\frac{n-1}{n}\sum_{m=1}^{n} Cov\Big(g(\widehat{\boldsymbol{D}}_{(b)}), g(\widehat{\boldsymbol{D}}_{(b,-m)})\Big) \\ =&n\frac{n-1}{n}\sum_{m=1}^{n}\Big(Var\Big(g(\widehat{\boldsymbol{D}}_{(b)})\Big)Var\Big(g(\widehat{\boldsymbol{D}}_{(b,-m)})\Big)\Big)^{1/2} \\ =&O(1) + O\Big(\frac{1}{n^{1/2}}\Big) + O\Big(\frac{1}{n}\Big). \qquad \text{(S4.4)}\end{aligned}$$

Combining equations S4.1,S4.2, S4.3, and S4.4, we can obtain

$$
\begin{aligned}
&Var(\widehat{dcov^2}_{(b)}^{JDC}(X,\mathbf{Y}))\\
=&n^2Var(g(\widehat{\boldsymbol{D}}_{(b)}))+(\frac{n-1}{n})^2\sum_{m=1}^{n}Var(g(\widehat{\boldsymbol{D}}_{(b,-m)}))\\
&+Cov\Big(ng(\widehat{\boldsymbol{D}}_{(b)}),\frac{n-1}{n}\sum_{m=1}^{n}g(\widehat{\boldsymbol{D}}_{(b,-m)})\Big)\\
=&O(1)+O\Big(\frac{1}{n^{1/2}}\Big)+O\Big(\frac{1}{n}\Big).
\end{aligned}
$$

By the definition of $\widehat{dcov^2}^{JDC}(X,\mathbf{Y})$,

$$
\begin{aligned}
&Var(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}))=Var\Big(\frac{1}{B}\sum_{b=1}^{B}\widehat{dcov^2}_{(b)}^{JDC}(X,\mathbf{Y})\Big)\\
=&O\Big(\frac{1}{B}\Big)+O\Big(\frac{1}{Bn^{1/2}}\Big)+O\Big(\frac{1}{Bn}\Big).
\end{aligned}
$$

And similarly, we can get $Var(\widehat{dcov^2}^{JDC}(X,X))=O\Big(\frac{1}{B}\Big)+O\Big(\frac{1}{Bn^{1/2}}\Big)+O\Big(\frac{1}{Bn}\Big)$ and $Var(\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))=O\Big(\frac{1}{B}\Big)+O\Big(\frac{1}{Bn^{1/2}}\Big)+O\Big(\frac{1}{Bn}\Big)$.

Given the definition of $\hat{\omega}^{JDC}$, we construct the function $f(x_1,x_2,x_3)=\frac{x_1}{\sqrt{x_2}\sqrt{x_3}}$ such that $\widehat{\omega}^{JDC}=f(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}),\widehat{dcov^2}^{JDC}(X,X),\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))$. It can be obtained that the elements of the first derivative of the function $f(x_1,x_2,x_3)$ are bounded. By combining condition $B=O(N^{\alpha})$ and $n=O(N^{\iota})$ with the Taylor expansion of

$$
f(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}),\widehat{dcov^2}^{JDC}(X,X),\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))
$$

at point $(dcov^{2^{JDC}}(X,\mathbf{Y}), dcov^{2^{JDC}}(X,X), dcov^{2^{JDC}}(\mathbf{Y},\mathbf{Y}))$, we obtain

$$
\begin{aligned}
&Var(\widehat{\omega}^{JDC})\\
=&O\Bigg(Var(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}))+Var(\widehat{dcov^2}^{JDC}(X,X))+Var(\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))\\
&+Cov\Big(\widehat{dcov^2}^{JDC}(X,\mathbf{Y})),\widehat{dcov^2}^{JDC}(X,X)\Big)+Cov\Big(\widehat{dcov^2}^{JDC}(X,\mathbf{Y})),\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y})\Big)\\
&+Cov\Big(\widehat{dcov^2}^{JDC}(X,X),\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y})\Big)\Bigg)\\
=&O\Bigg(Var(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}))+Var(\widehat{dcov^2}^{JDC}(X,X))+Var(\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))\\
&+\Big(Var(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}))Var(\widehat{dcov^2}^{JDC}(X,X))\Big)^{1/2}\\
&+\Big(Var(\widehat{dcov^2}^{JDC}(X,\mathbf{Y}))Var(\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))\Big)^{1/2}\\
&+\Big(Var(\widehat{dcov^2}^{JDC}(X,X))Var(\widehat{dcov^2}^{JDC}(\mathbf{Y},\mathbf{Y}))\Big)^{1/2}\Bigg)\\
=&O\Big(\frac{1}{N^{\alpha}}\Big)+O\Big(\frac{1}{N^{\alpha+1/2\iota}}\Big)+O\Big(\frac{1}{N^{\alpha+\iota}}\Big),
\end{aligned}
$$

for all $j \in 1,\dots,p$. Thus,

$$\max_{j=1,\dots,p} Var(\widehat{\omega}_j^{JDC}) = O\Big(\frac{1}{N^{\alpha}}\Big)+O\Big(\frac{1}{N^{\alpha+1/2\iota}}\Big)+O\Big(\frac{1}{N^{\alpha+\iota}}\Big).$$

This completes the proof of Theorem 1.

Proof of Theorem 2.

We will prove the sure screening properties of SDC-OSS and JDC-OSS, separately.

Part I: The proof of the sure screening property of SDC-OSS

Building upon the symbols and notation introduced in **Theorem 1,**

Part I, which are hereby adopted for use in this part, we proceed with the analysis of the Taylor expansion of the difference $\hat{\omega}^{SDC} - \omega$. Given that the derivative of the function $g(\cdot)$ is bounded, there exists a constant $C_1 > 0$ such that $\|g'(\boldsymbol{D})\|_{\max} \leq C_1$. Thus, we can derive the following inequality

$$|\hat{\omega}^{SDC} - \omega| = \Big|B^{-1}\sum_{b=1}^{B} g'(\boldsymbol{D})^{\top}\big(\widehat{\boldsymbol{D}}_{(b)} - \boldsymbol{D}\big)\Big| \leq \Big|B^{-1}\sum_{b=1}^{B}\sum_{h=1}^{8} C_1[\hat{S}_{(b),h} - S_h]\Big|.$$

Therefore, the event satisfies relation

$$\{|\hat{\omega}^{SDC} - \omega| > \epsilon\} \subseteq \cup_{h=1,\ldots,8}\Big\{C_1\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h} - S_h\big| \geq \epsilon/8\Big\},$$

and for probability we have

$$Pr(|\hat{\omega}^{SDC} - \omega| > \epsilon) \leq \sum_{h=1,\ldots,8} Pr\Big(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h} - S_h\Big| \geq \epsilon/(8C_1)\Big). \tag{S4.5}$$

Now we just need to focus on the relationship between each component estimate and its truth value. First deal with $B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),1}$. Let $h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) = \|X_i - X_k\|_1 \|\mathbf{Y}_i - \mathbf{Y}_k\|_2$ be the kernel of the U-statistic. We decompose the kernel function h_1 into two parts: $h_1 = h_1 I(h_1 > M) + h_1 I(h_1 \leq M)$ where M will be specified later. The U-statistic can now be written as follows,

$$\begin{aligned}\hat{S}_{(b),1} =& \frac{1}{n(n-1)}\sum_{i\neq k}^{n} h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) I(h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) \leq M)\\ &+ \frac{1}{n(n-1)}\sum_{i\neq k}^{n} h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) I(h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) > M)\end{aligned}$$

$$\triangleq \hat{S}_{(b),11} + \hat{S}_{(b),12}.$$

Accordingly, we decompose S_1 into two parts:

$$\begin{aligned} S_1 =& E[h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) I(h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) \leq M)] \\ &+ E[h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) I(h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) > M)] \\ \triangleq& S_{11} + S_{12}. \end{aligned}$$

Clearly, $\hat{S}_{(b),11}$ and $\hat{S}_{(b),12}$ are unbiased estimators of S_{11} and S_{12}, respectively.

We deal with the consistency $B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),11}$ of first. With the Markov's inequality, for any $t > 0$, we can obtain that

$$Pr\bigg(B^{-1} \sum_{b=1}^{B} \hat{S}_{(b),11} - S_{11} \geq \epsilon \bigg) \leq \exp(-t\epsilon) \exp(-tS_{11}) E\Big\{ \exp\Big(tB^{-1} \sum_{b=1}^{B} \hat{S}_{(b),11} \Big) \Big\}.$$

Serfling (1980, Section 5.1.6) showed that any U-statistic can be represented as an average of averages of independent and identically distributed (i.i.d) random variables. That is, $\hat{S}_{(b),11} = (n!)^{-1} \sum_{n!} \Omega_1(X_{(b),1}, \mathbf{Y}_{(b),1}; \ldots; X_{(b),n}, \mathbf{Y}_{(b),n})$, where $\sum_{n!}$ denotes the summation over all possible permutations of $(1, \ldots, n)$, and each $\Omega_1(X_{(b),1}, \mathbf{Y}_{(b),1}; \ldots; X_{(b),n}, \mathbf{Y}_{(b),n})$ is an average of $m = [n/2]$ i.i.d random variables (i.e.,$\Omega_1 = m^{-1} \sum_{r} h_1^{(r)} I(h_1^{(r)} \leq M)$). Since the exponential function is convex, it follows from Jensen's inequality that, for $0 < t \leq 2s_0$,

$$E\Big\{ \exp\Big(tB^{-1} \sum_{b=1}^{B} \hat{S}_{(b),11} \Big) \Big\}$$

$$
\begin{aligned}
&= E\Big\{ \exp\Big(tB^{-1}\sum_{b=1}^{B}(n!)^{-1}\sum_{n!}\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n})\Big)\Big\} \\
&\le (n!)^{-1}\sum_{n!}E\Big\{ \exp\Big(tB^{-1}\sum_{b=1}^{B}\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n})\Big)\Big\} \\
&= E^{mB}\Big\{ \exp\Big(t(mB)^{-1}h_1^{(r)}I(h_1^{(r)}\le M)\Big)\Big\},
\end{aligned}
$$

which together with Lemma 1 entails immediately that

$$
\begin{aligned}
&Pr\Big(B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),11}-S_{11}\ge\epsilon\Big) \\
&\le \exp(-t\epsilon)E^{mB}\Big\{ \exp\Big(t(mB)^{-1}[h_1^{(r)}I(h_1^{(r)}\le M)-S_{11}]\Big)\Big\} \\
&\le \exp(-t\epsilon+M^2t^2/(8mB)).
\end{aligned}
$$

By choosing $t=4\epsilon mB/M^2$, we have $Pr\big(B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),11}-S_{11}\ge\epsilon\big)\le\exp(-2\epsilon^2 mB/M^2)$. Therefore, by the symmetry of U-statistic, we can obtain easily that

$$
Pr\Big(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),11}-S_{11}\Big|\ge\epsilon\Big)\le 2\exp(-2\epsilon^2 mB/M^2). \tag{S4.6}
$$

Next we show the consistency of $B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),12}$. With Cauchy-Schwartz and Markov's inequality, for any $s'>0$, we have

$$
\begin{aligned}
S_{12}^2 &\le E\{h_1^2(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k)\}Pr(h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k)>M) \\
&\le E\{h_1^2(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k)\}E[\exp(s'h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k))]/\exp(s'M).
\end{aligned}
$$

Using the fact $(a^2+b^2)/2\ge(a+b)^2/4\ge|ab|$, we have

$$
h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k)=\{(X_i-X_k)^2(\mathbf{Y}_i-\mathbf{Y}_k)^\top(\mathbf{Y}_i-\mathbf{Y}_k)\}^{1/2}
$$

$$\leq 2\{(X_i^2 + X_k^2)(\|\mathbf{Y}_i\|_2^2) + \|\mathbf{Y}_k\|_2^2\}^{1/2}$$

$$\leq \{(X_i^2 + X_k^2 + \|\mathbf{Y}_i\|_2^2) + \|\mathbf{Y}_k\|_2^2)^2\}^{1/2}$$

$$= X_i^2 + X_k^2 + \|\mathbf{Y}_i\|_2^2 + \|\mathbf{Y}_k\|_2^2,$$

which yields that

$$E[\exp(s'h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k))] \leq E[\exp(s'(X_i^2 + X_k^2 + \|\mathbf{Y}_i\|_2^2 + \|\mathbf{Y}_k\|_2^2))]$$
$$\leq E\{\exp(2s'X_i^2)\}E\{\exp(2s'\|\mathbf{Y}_i\|_2^2)\}.$$

The last inequality follows from the Cauchy-Schwartz inequality. If we choose $M = C_2 N^{(\alpha+\iota)\gamma}$ for $0 < \gamma < 1/2 - \kappa$ and a nonnegative constant C_2, then $S_{12} \leq \epsilon/2$ when $N^{(\alpha+\iota)}$ is sufficiently large. Consequently,

$$Pr\left(\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12} - S_{12}\right| \geq \epsilon\right) \leq Pr\left(\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12}\right| \geq \epsilon/2\right).$$

It remains to bound the probability $Pr\left(\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12}\right| \geq \epsilon/2\right)$. We observe that the events satisfy

$$\left\{\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12}\right| \geq \epsilon/2\right\} \subseteq \{X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 > M/2, \text{ for some } 1 \leq i \leq n, 1 \leq b \leq B\}.$$

To see this, we assume that $X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 \leq M/2$ for all $1 \leq i \leq n, 1 \leq b \leq B$. This assumption will lead to a contradiction. To be precise, under this assumption, $h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) \leq M$. Consequently, $\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12}\right| = 0$, which is a contrary to the event $\left|B^{-1}\sum_{b=1}^{B} \hat{S}_{(b),12}\right| \geq \epsilon/2$. This proves that the above event inclusion relation is correct.

By invoking condition (C2) and Markov's inequality, there must exist a constant C_3 such that, for $s > 0$

$$
\begin{aligned}
& Pr\big(X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 > M/2\big) \\
\le\ & Pr\big(\|X_{(b),i}\|_1^2 \ge M/2\big) + Pr\big(\|\mathbf{Y}_{(b),i}\|_2^2 \ge M/2\big) \\
\le\ & \frac{E\{\exp(s\|X_{(b),i}\|_1^2)\}}{\exp(sM/4)} + \frac{E\{\exp(s\|\mathbf{Y}_{(b),i}\|_2^2)\}}{\exp(sM/4)} \\
\le\ & 2C_3 \exp(-sM/4).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
& Pr\Bigg(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),12} - S_{12}\Big| \ge \epsilon\Bigg) \qquad \text{(S4.7)} \\
\le\ & Pr\Bigg(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),12}\Big| \ge \epsilon/2\Bigg) \\
\le\ & \sum_{b=1}^{B}\sum_{i=1}^{n} Pr\big((X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2) > M/2\big) \\
\le\ & 2N^{(\alpha+\iota)} C_3 \exp(-sM/4).
\end{aligned}
$$

Recall that $M = C_2 N^{(\alpha+\iota)^\gamma}$. Combining the results (S4.6), and (S4.7), we have

$$
\begin{aligned}
& Pr\Bigg(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),1} - S_1\Big| \ge 4\epsilon\Bigg) \qquad \text{(S4.8)} \\
\le\ & Pr\Bigg(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),11} - S_{11}\Big| \ge \epsilon\Bigg) + Pr\Bigg(\Big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),12} - S_{12}\Big| \ge \epsilon\Bigg) \\
\le\ & 2\exp\big(-\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/C_2^2\big) + 2N^{(\alpha+\iota)} C_3 \exp\big(-sC_2 N^{(\alpha+\iota)\gamma}/4\big).
\end{aligned}
$$

Following arguments for proving (S4.8), we can show that

$$Pr\Big(\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h}-S_h\big|\geq 4\epsilon\Big) \tag{S4.9}$$

$$\leq 2\exp\big(-\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/C_2^2\big)+2N^{(\alpha+\iota)}C_3\exp\big(-sC_2N^{(\alpha+\iota)\gamma}/4\big), h=2,3,5,7.$$

In the sequel, we turn to $B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),4}$ where $\hat{S}_{(b),4}=\frac{1}{n(n-1)(n-2)}\sum_{i\neq k\neq l}^{n}\|X_{(b),i}-X_{(b),l}\|_1\cdot\|\mathbf{Y}_{(b),k}-\mathbf{Y}_{(b),l}\|_2$. Here, we define $h_2(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k;X_l,\mathbf{Y}_l)=\|X_i-X_l\|_1\|\mathbf{Y}_k-\mathbf{Y}_l\|_2$, which is the kernel of U-statistic $\hat{S}_4$. Following the arguments to deal with $B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),1}$, we decomposed h_2 into two parts: $h_2=h_2I(h_2>M)+h_2I(h_2\leq M)$. Accordingly

$$\hat{S}_{(b),4}=\frac{1}{n(n-1)(n-2)}\sum_{i\neq k\neq l}^{n}h_{2(b)}I(h_{2(b)}\leq M)+\frac{1}{n(n-1)(n-2)}\sum_{i\neq k\neq l}^{n}h_{2(b)}I(h_{2(b)}>M)$$

$$\triangleq\hat{S}_{(b),41}+\hat{S}_{(b),42},$$

where

$$h_{2(b)}=h_2(X_{(b),i},\mathbf{Y}_{(b),i};X_{(b),k},\mathbf{Y}_{(b),k};X_{(b),l},\mathbf{Y}_{(b),l}),$$

and

$$S_4=E\{h_2I(h_2\leq M)\}+E\{h_2I(h_2>M)\}\triangleq S_{41}+S_{42}.$$

Following similar arguments for proving (S4.6), we can show that

$$Pr\left(\left|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),41}-S_{41}\right|\geq\epsilon\right)\leq 2\exp(-2\epsilon^2 m'B/M^2), \tag{S4.10}$$

where $m'=[n/3]$ because $\hat{S}_{(b),41}$ is a third-order U-statistic.

Then we deal with $B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),42}$. We observe that

$$h_2(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k;X_l,\mathbf{Y}_l)=4(X_i^2+X_k^2+X_l^2+\|\mathbf{Y}_i\|_2^2+\|\mathbf{Y}_k\|_2^2+\|\mathbf{Y}_l\|_2^2)/6,$$

which will be smaller than M if $X_{(b),i}^2+\|\mathbf{Y}_{(b),i}\|_2^2\le M/2$ for all $1\le i\le n, 1\le b\le B$. Thus, for any $\epsilon>0$, the event satisfy

$$\left\{\left|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),42}\right|\ge\epsilon/2\right\}\subseteq\{X_{(b),i}^2+\|\mathbf{Y}_{(b),i}\|_2^2>M/2,\text{ for some }1\le i\le n,1\le b\le B\}.$$

By using the similar arguments to prove (S4.7), it follows that

$$Pr\left(\left|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),42}-S_{42}\right|\ge\epsilon\right)\le 2N^{(\alpha+\iota)}\exp(-sM/4). \tag{S4.11}$$

Then combine the results (S4.10) and (S4.11) with $M=C_2N^{(\alpha+\iota)\gamma}$, for some $0\le\gamma\le 1/2-\kappa$, we can obtain

$$\begin{aligned}&Pr\Big(\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),4}-S_4\big|\ge 4\epsilon\Big) \qquad\qquad (S4.12)\\ \le& 2\exp\big(-2\epsilon^2N^{(\alpha+\iota)(1-2\gamma)}/(3C_2^2)\big)+2N^{(\alpha+\iota)}C_3\exp\big(-sC_2N^{(\alpha+\iota)\gamma}/4\big).\end{aligned}$$

In addition, following arguments for proving (S4.12), we can show that

$$\begin{aligned}&Pr\Big(\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h}-S_h\big|\ge 4\epsilon\Big) \qquad\qquad (S4.13)\\ \le& 2\exp\big(-2\epsilon^2N^{(\alpha+\iota)(1-2\gamma)}/(3C_2^2)\big)+2N^{(\alpha+\iota)}C_3\exp\big(-sC_2N^{(\alpha+\iota)\gamma}/4\big),h=6,8.\end{aligned}$$

Combining (S4.5), (S4.8), (S4.9), (S4.12) and (S4.13), let $\epsilon=cN^{-(\alpha+\iota)\kappa}$, where $0\le\kappa+\gamma\le 1/2$, we thus have

$$Pr(|\hat{\omega}^{SDC}-\omega|>cN^{-(\alpha+\iota)\kappa})=O\Big(\exp\big(-c_1N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big)+N^{(\alpha+\iota)}\exp\big(-c_2N^{(\alpha+\iota)\gamma}\big)\Big),$$

for some positive constants c_1 and c_2. Therefore,

$$\begin{aligned}
&Pr(\max_{1\le j\le p}|\hat{\omega}_j^{SDC}-\omega_j|>cN^{-(\alpha+\iota)\kappa}) \qquad \text{(S4.14)}\\
=&O\Big(p\big[\exp\big(-c_1N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big)+N^{(\alpha+\iota)}\exp\big(-c_2N^{(\alpha+\iota)\gamma}\big)\big]\Big).
\end{aligned}$$

The first part of Theorem 2 for SDC-OSS is proven.

Now we deal with the second part of Theorem 2 for SDC-OSS. If $\mathcal{M}\nsubseteq\widehat{\mathcal{M}}$, then there must exist some $k\in\mathcal{M}$ such that $\hat{\omega}_j^{SDC}<cN^{-(\alpha+\iota)\kappa}$. It follows from condition (C3) that $|\hat{\omega}_j^{SDC}-\omega_j|>cN^{-(\alpha+\iota)\kappa}$ for some $k\in\mathcal{M}$, indicating that the events satisfy $\{\mathcal{M}\nsubseteq\widehat{\mathcal{M}}\}\subseteq\{|\hat{\omega}_j^{SDC}-\omega_j|>cN^{-(\alpha+\iota)\kappa}$, for some $k\in\mathcal{M}\}$, and hence $\mathscr{E}=\{\max_{k\in\mathcal{M}}|\hat{\omega}_k^{SDC}-\omega_k|\le cN^{-(\alpha+\iota)\kappa}\}\subseteq\{\mathcal{M}\subseteq\widehat{\mathcal{M}}\}$. Consequently,

$$\begin{aligned}
Pr\big(\mathcal{M}\subseteq\widehat{\mathcal{M}}\big)\ge Pr\big(\mathscr{E}\big)=1-Pr\big(\mathscr{E}^c\big)&=1-Pr\big(\min_{k\in\mathcal{M}}|\hat{\omega}_k^{SDC}-\omega_k|\ge cN^{-(\alpha+\iota)\kappa}\big)\\
&=1-|\mathcal{M}|Pr\big(|\hat{\omega}_k^{SDC}-\omega_k|\ge cN^{-(\alpha+\iota)\kappa}\big)\\
&\ge 1-O\Big(|\mathcal{M}|\big[\exp\big(-c_1N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big)+N^{(\alpha+\iota)}\exp\big(-c_2N^{(\alpha+\iota)\gamma}\big)\big]\Big),
\end{aligned}$$

where $|\mathcal{M}|$ is the cardinality of $\mathcal{M}$. This completes the proof of the second part of the method SDC-OSS.

Part II: The proof of the sure screening property of JDC-OSS

The symbols and notation introduced in **Theorem 1, Part II**, are hereby adopted for use in this Part. Then, the jackknife debaised estimation based on data segment $\mathcal{D}_b$ is expressed by the function

$$\widehat{dcov^2}_{(b)}^{JDC}(X,\mathbf{Y}) = g(\widehat{\boldsymbol{D}}_{(b)}) - \frac{n-1}{n}\sum_{m=1}^{n}\left[g(\widehat{\boldsymbol{D}}_{(b,-m)}) - (n-1)g(\widehat{\boldsymbol{D}}_{(b)})\right].$$

This leads to the jackknife debiased simple average distance covariance estimator defined as follows:

$$\widehat{dcov^2}^{JDC}(X,\mathbf{Y}) = \frac{n}{B}\sum_{b=1}^{B} g(\widehat{\boldsymbol{D}}_{(b)}) - \frac{n-1}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}\left[g(\widehat{\boldsymbol{D}}_{(b,-m)}) - g(\widehat{\boldsymbol{D}}_{(b)})\right].$$

For $|\widehat{dcov^2}^{JDC}(X,\mathbf{Y}) - dcov^2(X,\mathbf{Y})|$, we can organize it into the following form,

$$\begin{aligned}
&|\widehat{dcov^2}^{JDC}(X,\mathbf{Y}) - dcov^2(X,\mathbf{Y})| \\
=&\left|\frac{1}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}\left[n\left[g(\widehat{\boldsymbol{D}}_{(b)}) - g(\boldsymbol{D})\right] - (n-1)\left[g(\widehat{\boldsymbol{D}}_{(b,-m)}) - g(\boldsymbol{D})\right]\right]\right| \\
=&\left|\frac{1}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}\left[g'(\boldsymbol{\xi}_1)^\top\left[n\widehat{\boldsymbol{D}}_{(b)} - (n-1)\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\right] - (n-1)(g'(\boldsymbol{\xi}_2) - g'(\boldsymbol{\xi}_1))^\top\left[\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\right]\right]\right|, \\
\le&\left|\frac{1}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n} g'(\boldsymbol{\xi}_1)^\top\left[n\widehat{\boldsymbol{D}}_{(b)} - (n-1)\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\right]\right| + \left|\frac{(n-1)}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}\left[g'(\boldsymbol{\xi}_2) - g'(\boldsymbol{\xi}_1)\right]^\top\left[\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\right]\right|,
\end{aligned}$$

where $g'(\cdot)$ is the derivative of $g(\cdot)$, and $\boldsymbol{\xi}_1 = \boldsymbol{D} + \theta_1(\widehat{\boldsymbol{D}}_{(b)} - \boldsymbol{D}), \boldsymbol{\xi}_2 = \boldsymbol{D} + \theta_2(\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D})$ with $0 < \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 < 1$. There exist a positive constant C_2 such that $\|(n-1)(g'(\xi_2) - g'(\xi_1))\|_\infty = C_2$. As for $\frac{1}{n}\sum\limits_{m=1}^{n}\left[n\widehat{\boldsymbol{D}}_{(b)} - (n-1)\widehat{\boldsymbol{D}}_{(b,-m)}\right] = n\widehat{\boldsymbol{D}}_{(b)} - \frac{n-1}{n}\sum\limits_{m=1}^{n}\widehat{\boldsymbol{D}}_{(b,-m)}$, using $\sum\limits_{m=1}^{n}\sum\limits_{i\neq k\neq m}^{n} = (n-2)\sum\limits_{i\neq k}^{n}$ and $\sum\limits_{m=1}^{n}\sum\limits_{i\neq k\neq l\neq m}^{n} = (n-3)\sum\limits_{i\neq k\neq l}^{n}$, we have

$$\begin{aligned}
&n\hat{S}_{(b),1} - \frac{n-1}{n}\sum_{m=1}^{n}\hat{S}_{(b,-m),1} \\
=&\frac{1}{n-1}\sum_{i\neq k}^{n} h_{1(b)}(i,k) - \frac{1}{n(n-2)}\sum_{m=1}^{n}\sum_{i\neq k\neq m}^{n} h_{1(b)}(i,k)
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n-1}\sum_{i\neq k}^{n} h_{1(b)}(i,k)-\frac{1}{n}\sum_{i\neq k}^{n} h_{1(b)}(i,k)\\
=&\frac{1}{n(n-1)}\sum_{i\neq k}^{n} h_{1(b)}(i,k)=\hat{S}_{(b),1},
\end{aligned}
$$

and

$$
\begin{aligned}
&n\hat{S}_{(b),4}-\frac{n-1}{n}\sum_{m=1}^{n}\hat{S}_{(b,-m),4}\\
=&\frac{1}{(n-1)(n-2)}\sum_{i\neq k\neq l}^{n} h_{2(b)}(i,k,l)-\frac{1}{n(n-2)(n-3)}\sum_{m=1}^{n}\sum_{i\neq k\neq l\neq m}^{n} h_{2(b)}(i,k,l)\\
=&\frac{1}{(n-1)(n-2)}\sum_{i\neq k\neq l}^{n} h_{2(b)}(i,k,l)-\frac{1}{n(n-2)}\sum_{i\neq k\neq l}^{n} h_{2(b)}(i,k,l)\\
=&\frac{1}{n(n-1)(n-2)}\sum_{i\neq k\neq l}^{n} h_{2(b)}(i,k,l)=\hat{S}_{(b),4},
\end{aligned}
$$

where

$$h_{1(b)}(i,k)=h_1(X_{(b),i},\mathbf{Y}_{(b),i};X_{(b),k},\mathbf{Y}_{(b),k})$$

and

$$h_{2(b)}(i,k,l)=h_2(X_{(b),i},\mathbf{Y}_{(b),i};X_{(b),k},\mathbf{Y}_{(b),k};X_{(b),l},\mathbf{Y}_{(b),l}).$$

Similarly, we have $n\hat{S}_{(b),j}-\frac{n-1}{n}\sum_{m=1}^{n}\hat{S}_{(b,-m),j}=\hat{S}_{(b),j}, j=2,3,5,6,7,8$ and $\frac{1}{n}\sum_{m=1}^{n}\left[n\widehat{\boldsymbol{D}}_{(b)}-(n-1)\widehat{\boldsymbol{D}}_{(b,-m)}\right]=\widehat{\boldsymbol{D}}_{(b)}$. Therefore,

$$
\begin{aligned}
&|\widehat{dcov^2}^{JDC}(X,\mathbf{Y})-dcov^2(X,\mathbf{Y})|\\
\leq&\left|\frac{1}{B}\sum_{b=1}^{B} g'(\xi_1)^\top\left[\widehat{\boldsymbol{D}}_{(b)}-\boldsymbol{D}\right]\right|+\left|\frac{(n-1)}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}[g'(\xi_2)-g'(\xi_1)]^\top\left[\widehat{\boldsymbol{D}}_{(b,-m)}-\boldsymbol{D}\right]\right|.
\end{aligned}
$$

Thus, the event satisfies

$$\big\{|\widehat{dcov^2}^{JDC}(X,\mathbf{Y}) - dcov^2(X,\mathbf{Y})| > \epsilon\big\}$$
$$\subseteq\Bigg\{\bigg\{\bigg|\frac{1}{B}\sum_{b=1}^{B} g'(\xi_1)^\top\big[\widehat{\boldsymbol{D}}_{(b)} - \boldsymbol{D}\big]\bigg| > \frac{\epsilon}{2}\bigg\} \cup \bigg\{\bigg|\frac{(n-1)}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}[g'(\xi_2) - g'(\xi_1)]^\top\big[\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\big]\bigg| > \frac{\epsilon}{2}\bigg\}\Bigg\}.$$

The corresponding probability satisfies the following inequality:

$$\begin{aligned}
&Pr\big(|\widehat{dcov^2}^{JDC}(X,\mathbf{Y}) - dcov^2(X,\mathbf{Y})| > \epsilon\big) \qquad \text{(S4.15)}\\
\le &Pr\bigg\{\bigg|\frac{1}{B}\sum_{b=1}^{B} g'(\xi_1)^\top\big[\widehat{\boldsymbol{D}}_{(b)} - \boldsymbol{D}\big]\bigg| > \frac{\epsilon}{2}\bigg\}\\
&+ Pr\bigg\{\bigg|\frac{(n-1)}{nB}\sum_{b=1}^{B}\sum_{m=1}^{n}[g'(\xi_2) - g'(\xi_1)]^\top\big[\widehat{\boldsymbol{D}}_{(b,-m)} - \boldsymbol{D}\big]\bigg| > \frac{\epsilon}{2}\bigg\}\\
\le &\sum_{h=1,\dots,4} Pr\bigg(\bigg|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h} - S_h\bigg| \ge \frac{\epsilon}{4C_1}\bigg)\\
&+ \sum_{h=1,\dots,4} Pr\bigg(\bigg|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),h} - S_h\bigg| \ge \frac{\epsilon}{4C_2}\bigg)\\
\triangleq &\triangle_1 + \triangle_2 .
\end{aligned}$$

As for $\triangle_1$, the proof in **Part I** of this theorem gives us

$$\triangle_1 = O\Big(\exp\big(-c_1\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}\big) + N^{(\alpha+\iota)}\exp\big(-c_2 N^{(\alpha+\iota)\gamma}\big)\Big). \qquad \text{(S4.16)}$$

For $\triangle_2$, we have get

$$\triangle_2 = \sum_{h=1,\dots,4} Pr\bigg(\bigg|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),h} - S_h\bigg| \ge \frac{\epsilon}{4C_2}\bigg). \qquad \text{(S4.17)}$$

First of all, we will deal with $(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),1}$. Let $h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k) = \|X_i - X_k\|_1\|\mathbf{Y}_i - \mathbf{Y}_k\|_2$ be the kernel of the U-statistic. We decompose the kernel function h_1 into two parts: $h_1 = h_1 I(h_1 > M) + h_1 I(h_1 \leq M)$ where M will be specified later. The U-statistic can now be written as follows,

$$\begin{aligned}
&\hat{S}_{(b,-m),1} \\
=&\frac{1}{(n-1)(n-2)} \sum_{i \neq k \neq m}^{n} h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) I(h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) \leq M) \\
&+ \frac{1}{(n-1)(n-2)} \sum_{i \neq k \neq m}^{n} h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) I(h_1(X_{(b),i}, \mathbf{Y}_{(b),i}; X_{(b),k}, \mathbf{Y}_{(b),k}) > M) \\
\triangleq&\hat{S}_{(b,-m),11} + \hat{S}_{(b,-m),12}.
\end{aligned}$$

Accordingly, we decompose S_1 into two parts:

$$\begin{aligned}
S_1 =&E[h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) I(h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) \leq M)] \\
&+ E[h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) I(h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) > M)] \\
\triangleq&S_{11} + S_{12}.
\end{aligned}$$

Clearly, $\hat{S}_{(b),11}$ and $\hat{S}_{(b),12}$ are unbiased estimators of S_{11} and S_{12}, respectively.

We deal with the consistency $(nB)^{-1} \sum_{b=1}^{B} \sum_{m=1}^{n} \hat{S}_{(b,-m),11}$ of first. With the Markov's inequality, for any $t > 0$, we can obtain that

$$\begin{aligned}
&Pr\Big((nB)^{-1} \sum_{b=1}^{B} \sum_{m=1}^{n} \hat{S}_{(b,-m),11} - S_{11} \geq \epsilon\Big) \\
\leq&\exp(-t\epsilon)\exp(-tS_{11}) E\Big\{\exp\Big(t(nB)^{-1} \sum_{b=1}^{B} \sum_{m=1}^{n} \hat{S}_{(b,-m),11}\Big)\Big\}.
\end{aligned}$$

Serfling (1980, Section 5.1.6) showed that any U-statistic can be represented as an average of averages of independent and identically distributed (i.i.d) random variables. That is,

$$
\begin{aligned}
n^{-1}\sum_{m=1}^{n}\hat{S}_{(b,-m),11} &= \frac{1}{n}\sum_{m=1}^{n}\frac{1}{(n-1)(n-2)}\sum_{i\neq m}^{n}\sum_{k\neq m;k\neq m}^{n}\|X_i - X_k\|_1\|\mathbf{Y}_i - \mathbf{Y}_k\|_2 \\
&= \frac{1}{n(n-1)(n-2)}\sum_{i\neq k\neq m}^{n}\|X_i - X_k\|_1\|\mathbf{Y}_i - \mathbf{Y}_k\|_2 \\
&= (n!)^{-1}\sum_{n!}\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n}),
\end{aligned}
$$

where $\sum_{n!}$ denotes the summation over all possible permutations of $(1,\ldots,n)$, and each $\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n})$ is an average of $m' = [n/3]$ i.i.d random variables (i.e.,$\Omega_1 = (m')^{-1}\sum_r h_1^{(r)}I(h_1^{(r)} \leq M)$). Since the exponential function is convex, it follows from Jensen's inequality that, for $0 < t \leq 2s_0$,

$$
\begin{aligned}
&E\Big\{\exp\Big(t(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),11}\Big)\Big\} \\
&= E\Big\{\exp\Big(tB^{-1}\sum_{b=1}^{B}(n!)^{-1}\sum_{n!}\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n})\Big)\Big\} \\
&\leq (n!)^{-1}\sum_{n!}E\Big\{\exp\Big(tB^{-1}\sum_{b=1}^{B}\Omega_1(X_{(b),1},\mathbf{Y}_{(b),1};\ldots;X_{(b),n},\mathbf{Y}_{(b),n})\Big)\Big\} \\
&= E^{m'B}\Big\{\exp\Big(t(m'B)^{-1}h_1^{(r)}I(h_1^{(r)} \leq M)\Big)\Big\},
\end{aligned}
$$

which together with Lemma 1 entails immediately that

$$
Pr\Big((m'B)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),11} - S_{11} \geq \epsilon\Big)
$$

$$\leq \exp(-t\epsilon)E^{m'B}\Big\{\exp\Big(t(m'B)^{-1}[h_1^{(r)}I(h_1^{(r)} \leq M) - S_{11}]\Big)\Big\}$$

$$\leq \exp(-t\epsilon + M^2t^2/(8m'B)).$$

By choosing $t = 4\epsilon B/M^2$, we have

$$Pr\Big((nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),11} - S_{11} \geq \epsilon\Big) \leq \exp(-2\epsilon^2 mB/M^2).$$

Therefore, by the symmetry of U-statistic, we can obtain easily that

$$Pr\Bigg(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),11} - S_{11}\Big| \geq \epsilon\Bigg) \leq 2\exp(-2\epsilon^2 m'B/M^2) \tag{S4.18}$$

Next we show the consistency of $(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}$. With Cauchy-Schwartz and Markov's inequality, for any $s' > 0$, we have

$$S_{12}^2 \leq E\{h_1^2(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k)\}Pr(h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) > M)$$

$$\leq E\{h_1^2(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k)\}E[\exp(s'h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k))]/\exp(s'M).$$

Using the fact $(a^2 + b^2)/2 \geq (a + b)^2/4 \geq |ab|$, we have

$$\begin{aligned} h_1(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k) &= \{(X_i - X_k)^2(\mathbf{Y}_i - \mathbf{Y}_k)^\top(\mathbf{Y}_i - \mathbf{Y}_k)\}^{1/2} \\ &\leq 2\{(X_i^2 + X_k^2)(\|\mathbf{Y}_i\|_2^2) + \|\mathbf{Y}_k\|_2^2\}^{1/2} \\ &\leq \{(X_i^2 + X_k^2 + \|\mathbf{Y}_i\|_2^2) + \|\mathbf{Y}_k\|_2^2)^2\}^{1/2} \\ &= X_i^2 + X_k^2 + \|\mathbf{Y}_i\|_2^2 + \|\mathbf{Y}_k\|_2^2, \end{aligned}$$

which yields that

$$E[\exp(s'h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k))] \le E[\exp(s'(X_i^2+X_k^2+\|\mathbf{Y}_i\|_2^2+\|\mathbf{Y}_k\|_2^2))]$$
$$\le E\{\exp(2s'X_i^2)\}E\{\exp(2s'\|\mathbf{Y}_i\|_2^2)\}.$$

The last inequality follows from the Cauchy-Schwartz inequality. If we choose $M=C_3N^{(\alpha+\iota)\gamma}$ for $0<\gamma+\kappa<1/2$ and a nonnegative constant C_3, then $S_{12}\le\epsilon/2$ when $N^{(\alpha+\iota)}$ is sufficiently large. Consequently,

$$Pr\Big(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}-S_{12}\Big|\ge\epsilon\Big)\le Pr\Big(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big|\ge\epsilon/2\Big).$$

It remains to bound the probability $Pr\Big(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big|\ge\epsilon/2\Big)$. We observe that the events satisfy

$$\begin{aligned}
&\Big\{\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big|\ge\epsilon/2\Big\}\\
&=\Big\{\Big|B^{-1}\sum_{b=1}^{B}\frac{1}{n(n-1)(n-2)}\sum_{i\ne k\ne m}^{n}h_1(h_1>M)\Big|\ge\epsilon/2\Big\}\\
&\subseteq\{X_{(b),i}^2+\|\mathbf{Y}_{(b),i}\|_2^2>M/2,\ \text{for some } 1\le i\le n, 1\le b\le B\}.
\end{aligned}$$

To see this, we assume that $X_{(b),i}^2+\|\mathbf{Y}_{(b),i}\|_2^2\le M/2$ for all $1\le i\le n, 1\le b\le B$. This assumption will lead to a contradiction. To be precise, under this assumption, $h_1(X_i,\mathbf{Y}_i;X_k,\mathbf{Y}_k)\le M$. Consequently, $\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big|=0$, which is a contrary to the event $\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big|\ge\epsilon/2$. This proves that the above event inclusion relation is correct.

By invoking condition (C2) and Markov's inequality, there must exist a constant C_4 such that, for $s > 0$

$$
\begin{aligned}
& Pr\big(X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 > M/2\big) \\
\le\ & Pr\big(\|X_{(b),i}\|_1^2 \ge M/2\big) + Pr\big(\|\mathbf{Y}_{(b),i}\|_2^2 \ge M/2\big) \\
\le\ & \frac{E\{\exp(s\|X_{(b),i}\|_1^2)\}}{\exp(sM/4)} + \frac{E\{\exp(s\|\mathbf{Y}_{(b),i}\|_2^2)\}}{\exp(sM/4)} \\
\le\ & 2C_4 \exp(-sM/4).
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
& Pr\Bigg(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12} - S_{12}\Big| \ge \epsilon\Bigg) \\
\le\ & Pr\Bigg(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12}\Big| \ge \epsilon/2\Bigg) \\
\le\ & \sum_{b=1}^{B}\sum_{i=1}^{n} Pr\big((X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2) > M/2\big) \\
\le\ & 2N^{(\alpha+\iota)}C_4 \exp(-sM/4).
\end{aligned}
\tag{S4.19}
$$

Recall that $M = C_3 N^{(\alpha+\iota)\gamma}$. Combining the results (S4.18), and (S4.19), we have

$$
\begin{aligned}
& Pr\Bigg(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),1} - S_1\Big| \ge 4\epsilon\Bigg) \\
\le\ & Pr\Bigg(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),11} - S_{11}\Big| \ge \epsilon\Bigg)
\end{aligned}
\tag{S4.20}
$$

$$+ Pr\Big(\big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),12} - S_{12}\big| \geq \epsilon\Big)$$
$$\leq 2\exp\big(-\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/C_3^2\big) + 2N^{(\alpha+\iota)}C_4\exp\big(-sC_3N^{(\alpha+\iota)\gamma}/4\big).$$

Following arguments for proving (S4.20), we can show that

$$Pr\Big(\big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),h} - S_h\big| \geq 4\epsilon\Big) \tag{S4.21}$$
$$\leq 2\exp\big(-\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/C_3^2\big) + 2N^{(\alpha+\iota)}C_4\exp\big(-sC_3N^{(\alpha+\iota)\gamma}/4\big), h = 2,3,5,7.$$

In the sequel, we turn to $(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),4}$ where

$$n^{-1}\sum_{m=1}^{n}\hat{S}_{(b,-m),4} = \frac{1}{n}\sum_{m=1}^{n}\frac{1}{(n-1)(n-2)(n-3)}\sum_{i\neq k\neq l\neq m}^{n}\|X_{(b),i} - X_{(b),l}\|_1 \cdot \|\mathbf{Y}_{(b),k} - \mathbf{Y}_{(b),l}\|_2$$
$$= \frac{1}{n(n-1)(n-2)(n-3)}\sum_{i\neq k\neq l\neq m}^{n}\|X_{(b),i} - X_{(b),l}\|_1 \cdot \|\mathbf{Y}_{(b),k} - \mathbf{Y}_{(b),l}\|_2.$$

Here, we define $h_2(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k; X_l, \mathbf{Y}_l) = \|X_i - X_l\|_1\|\mathbf{Y}_k - \mathbf{Y}_l\|_2$, which is the kernel of U-statistic $\hat{S}_4$. Following the arguments to deal with $(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),1}$, we decomposed h_2 into two parts: $h_2 = h_2I(h_2 > M) + h_2I(h_2 \leq M)$. Accordingly

$$n^{-1}\sum_{m=1}^{n}\hat{S}_{(b,-m),4} = \frac{1}{n(n-1)(n-2)(n-3)}\sum_{i\neq k\neq l\neq m}^{n}h_{2(b,-m)}I(h_{2(b,-m)} \leq M)$$
$$+ \frac{1}{n(n-1)(n-2)(n-3)}\sum_{i\neq k\neq l\neq m}^{n}h_{2(b,-m)}I(h_{2(b,-m)} > M)$$
$$\triangleq n^{-1}\sum_{m=1}^{n}\hat{S}_{(b,-m),41} + n^{-1}\sum_{m=1}^{n}\hat{S}_{(b,-m),42},$$

where

$$h_{2(b,-m)} = h_2(X_{(b,-m),i}, \mathbf{Y}_{(b,-m),i}; X_{(b,-m),k}, \mathbf{Y}_{(b,-m),k}; X_{(b,-m),l}, \mathbf{Y}_{(b,-m),l}),$$

and

$$S_4 = E\{h_2 I(h_2 \le M)\} + E\{h_2 I(h_2 > M)\} \triangleq S_{41} + S_{42}.$$

Following similar arguments for proving (S4.18), we can show that

$$Pr\left(\left|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),41} - S_{41}\right| \ge \epsilon\right) \le 2\exp(-2\epsilon^2 m'' B/M^2), \tag{S4.22}$$

where $m'' = [n/4]$ because $\hat{S}_{(b,-m),41}$ is a fourth-order U-statistic.

Then we deal with $(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),42}$. We observe that

$$h_2(X_i, \mathbf{Y}_i; X_k, \mathbf{Y}_k; X_l, \mathbf{Y}_l) = 4(X_i^2 + X_k^2 + X_l^2 + \|\mathbf{Y}_i\|_2^2 + \|\mathbf{Y}_k\|_2^2 + \|\mathbf{Y}_l\|_2^2)/6,$$

which will be smaller than M if $X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 \le M/2$ for all $1 \le i \le n, 1 \le b \le B$. Thus, for any $\epsilon > 0$, the event satisfy

$$\begin{aligned}&\left\{\left|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),41}\right| \ge \epsilon/2\right\}\\ \subseteq&\{X_{(b),i}^2 + \|\mathbf{Y}_{(b),i}\|_2^2 > M/2, \text{ for some } 1 \le i \le n, 1 \le b \le B\}.\end{aligned}$$

By using the similar arguments to prove (S4.19), it follows that

$$Pr\left(\left|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),42} - S_{42}\right| \ge \epsilon\right) \le 2N^{(\alpha+\iota)}\exp(-sM/4). \tag{S4.23}$$

Then combine the results (S4.22) and (S4.23) with $M = C_3 N^{(\alpha+\iota)\gamma}$, for some $0 \leq \gamma \leq 1/2 - \kappa$, we can obtain

$$Pr\Big(\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),4} - S_4\big| \geq 4\epsilon\Big) \tag{S4.24}$$
$$\leq 2\exp\big(-2\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/(3C_3^2)\big) + 2N^{(\alpha+\iota)}C_4\exp\big(-sC_3N^{(\alpha+\iota)\gamma}/4\big).$$

In addition, following arguments for proving (S4.24), we can show that

$$Pr\Big(\big|B^{-1}\sum_{b=1}^{B}\hat{S}_{(b),h} - S_h\big| \geq 4\epsilon\Big) \tag{S4.25}$$
$$\leq 2\exp\big(-2\epsilon^2 N^{(\alpha+\iota)(1-2\gamma)}/(3C_3^2)\big) + 2N^{(\alpha+\iota)}C_4\exp\big(-sC_3N^{(\alpha+\iota)\gamma}/4\big), h = 6, 8.$$

Combining (S4.17), (S4.20), (S4.21), (S4.24) and (S4.25), with $0 \leq \kappa + \gamma \leq 1/2$, we thus have

$$\begin{aligned}\triangle_2 = &\sum_{h=1,\ldots,8} Pr\Big(\Big|(nB)^{-1}\sum_{b=1}^{B}\sum_{m=1}^{n}\hat{S}_{(b,-m),h} - S_h\Big| \geq \frac{\epsilon}{8C_2}\Big) \\ =&O\Big(\exp\big(-c_1\epsilon^2N^{(\alpha+\iota)(1-2\gamma)}\big) + N^{(\alpha+\iota)}\exp\big(-c_2N^{(\alpha+\iota)\gamma}\big)\Big),\end{aligned} \tag{S4.26}$$

for some positive constants c_1 and c_2. Combining (S4.15), (S4.16) and (S4.26), let $\epsilon = cN^{-(\alpha+\iota)\kappa}$, where $0 \leq \kappa + \gamma \leq 1/2$, we thus have

$$\begin{aligned}&Pr\big(|\widehat{dcov^2}^{JDC}(X, \mathbf{Y}) - dcov^2(X, \mathbf{Y})| > cN^{-(\alpha+\iota)\kappa}\big) \\ =&O\Big(\exp\big(-c_1N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big) + N^{(\alpha+\iota)}\exp\big(-c_2N^{(\alpha+\iota)\gamma}\big)\Big),\end{aligned}$$

for some positive constants c_1 and c_2. Using Lemma 3, we have

$$\gamma_0 = \min\Big\{\frac{dcov^4(X, X)dcov^4(\mathbf{Y}, \mathbf{Y})}{4M^4}, \frac{dcov^6(X, X)dcov^6(\mathbf{Y}, \mathbf{Y})}{4M^4}\Big\},$$

where

$$M \geq 2\max\{dcov(X,\mathbf{Y}), dcov(X,X), dcov(\mathbf{Y},\mathbf{Y}), \widehat{dcov}(X,\mathbf{Y}), \widehat{dcov}(X,X), \widehat{dcov}(\mathbf{Y},\mathbf{Y})\}.$$

Condition (C2) enables us to conclude that M and γ_0 are bounded constants. Thus

$$\begin{aligned}&Pr\big(|\hat{\omega}_j^{JDC} - \omega_j| > cN^{-(\alpha+\iota)\kappa}\big)\\ =&O\Big(\exp\big(-c_1 N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\gamma_0\big) + N^{(\alpha+\iota)}\exp\big(-c_2 N^{(\alpha+\iota)\gamma}\big)\Big),\end{aligned}$$

for some positive constants c_1 and c_2. Therefore,

$$\begin{aligned}&Pr\big(\max_{1\leq j\leq p}|\hat{\omega}_j^{JDC} - \omega_j| > cN^{-(\alpha+\iota)\kappa}\big)\\ =&O\Big(p\big[\exp\big(-c_1 N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big) + N^{(\alpha+\iota)}\exp\big(-c_2 N^{(\alpha+\iota)\gamma}\big)\big]\Big).\end{aligned} \tag{S4.27}$$

The first part of Theorem 2 for JDC-OSS is proven.

Now we deal with the second part of Theorem 2 for JDC-OSS. If $\mathcal{M} \nsubseteq \widehat{\mathcal{M}}$, then there must exist some $k \in \mathcal{M}$ such that $\hat{\omega}_j^{JDC} < cN^{-(\alpha+\iota)\kappa}$. It follows from condition (C3) that $|\hat{\omega}_j^{JDC} - \omega_j| > cN^{-(\alpha+\iota)\kappa}$ for some $k \in \mathcal{M}$, indicating that the events satisfy $\{\mathcal{M} \nsubseteq \widehat{\mathcal{M}}\} \subseteq \{|\hat{\omega}_j^{JDC} - \omega_j| > cN^{-(\alpha+\iota)\kappa}$, for some $k \in \mathcal{M}\}$, and hence $\mathscr{E} = \{\max_{k\in\mathcal{M}}|\hat{\omega}_k^{JDC} - \omega_k| \leq cN^{-(\alpha+\iota)\kappa}\} \subseteq \{\mathcal{M} \subseteq \widehat{\mathcal{M}}\}$. Consequently,

$$\begin{aligned}Pr\big(\mathcal{M} \subseteq \widehat{\mathcal{M}}\big) \geq Pr\big(\mathscr{E}\big) = 1 - Pr\big(\mathscr{E}^c\big) &= 1 - Pr\big(\min_{k\in\mathcal{M}}|\hat{\omega}_k^{JDC} - \omega_k| \geq cN^{-(\alpha+\iota)\kappa}\big)\\ &= 1 - |\mathcal{M}|Pr\big(|\hat{\omega}_k^{JDC} - \omega_k| \geq cN^{-(\alpha+\iota)\kappa}\big)\end{aligned}$$

$$\geq 1 - O\Big(|\mathcal{M}|\big[\exp\big(-c_1 N^{(\alpha+\iota)(1-2\gamma-2\kappa)}\big) + N^{(\alpha+\iota)}\exp\big(-c_2 N^{(\alpha+\iota)\gamma}\big)\big]\Big),$$

where $|\mathcal{M}|$ is the cardinality of $\mathcal{M}$. This completes the proof of the second part.

Proof of Theorem 3

Under Condition (C3), noting that $\min_{k\in\mathcal{M}}\omega_k \geq 2cN^{-(\alpha+\iota)\kappa}$ and combining (S4.14), we have

$$\begin{aligned}
Pr(\min_{k\in\mathcal{M}}\widehat{\omega}_k \leq \max_{k\notin\mathcal{M}}\widehat{\omega}_k) &= Pr(\max_{k\notin\mathcal{M}}\widehat{\omega}_k - \max_{k\notin\mathcal{M}}\omega_k - \min_{k\in\mathcal{M}}\widehat{\omega}_k + \min_{k\in\mathcal{M}}\omega_k \geq \min_{k\in\mathcal{M}}\omega_k)\\
&\leq Pr(\max_{k\notin\mathcal{M}}|\widehat{\omega}_k - \omega_k| \geq cN^{-(\alpha+\iota)\kappa}) + Pr(\max_{k\in\mathcal{M}}|\widehat{\omega}_k - \omega_k| \geq cN^{-(\alpha+\iota)\kappa})\\
&\leq 2Pr(\max_{1\leq k\leq p}|\widehat{\omega}_k - \omega_k| \geq cN^{-(\alpha+\iota)\kappa}),
\end{aligned}$$

where the first equation holds because the corresponding distance correlation measure for unimportant variables is 0. By combining Theorem 2, and plugging them into the equation above, we obtain

$$\begin{aligned}
&Pr\Big(\max_{j\notin\mathcal{M}}\hat{\omega}_j^{SDC} \leq \min_{j\in\mathcal{M}}\widehat{\omega}_j^{SDC}\Big) \geq 1 - O\left(p\left[\exp(-c_1 N^{(\alpha+\iota)(1-2\gamma-2\kappa)}) + N^{(\alpha+\iota)}\exp(-c_2 N^{(\alpha+\iota)\gamma})\right]\right),\\
&Pr\Big(\max_{j\notin\mathcal{M}}\hat{\omega}_j^{JDC} \leq \min_{j\in\mathcal{M}}\widehat{\omega}_j^{JDC}\Big) \geq 1 - O\left(p\left[\exp(-c_1 N^{(\alpha+\iota)(1-2\gamma-2\kappa)}) + N^{(\alpha+\iota)}\exp(-c_2 N^{(\alpha+\iota)\gamma})\right]\right).
\end{aligned}$$

This completes the proof of Theorem 3.

References

Serfling R. (1980). *Approximation theorems of mathematical statistics.* Vol. 162, John Wiley and Sons.

Liu, W., Ke, Y., Liu, J., and Li, R. (2022). Model-free feature screening and FDR control with knockoff features. *Journal of the American Statistical Association* **117(537)**, 428–443.

Li, X., Li, R., Xia, Z., and Xu, C. (2020). Distributed feature screening via componentwise debiasing. *Journal of Machine Learning Research* **21**, 1–32.