Supplementary Materials: Martingale difference correlation and its use in high dimensional variable screening

by Xiaofeng Shao and Jingsi Zhang

The supplementary material contains some additional simulation results in Section 8 and proofs of Theorems 3-6 in Section 9. For the sake of readership and completeness, we also provide a brief description of the model setting.

8 Additional Simulation Results

8.1 Example 1

We adopt the simple linear model from Fan and Lv (2008): $Y = 5X_1 + 5X_2 + 5X_3 + \epsilon$. The predictor vector ($X_1$,..., X_p) is drawn from a multivariate normal distribution $N(0, \Sigma)$ whose covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$ has entries $\sigma_{ii} = 1$, $i = 1, ..., p$, and $\sigma_{ij} = \rho$, $i \neq j$. The error term ϵ is independently generated from the standard normal distribution. We consider several different combinations of (p, n, ρ), i.e., $p = 100,\ 1000$, $n = 20,\ 50,\ 70$ and $\rho = 0,\ 0.1,\ 0.5,\ 0.9$.

Table 11: $\mathcal{P}_a$ for Example 1 with $d = 5, 10, 15$ and $p = 1000$

d	p	n	*Method*	*Results for the following values of* ρ:			
				$\rho = 0$	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
5	1000	20	*SIS*	0.032	0.042	0.016	0.004
			DC-SIS	0.022	0.020	0.006	0.006
			MDC-SIS	0.030	0.030	0.012	0.006
			SIRS	0.012	0.008	0.000	0.000
		50	*SIS*	0.780	0.680	0.338	0.214
			DC-SIS	0.636	0.548	0.208	0.132
			MDC-SIS	0.704	0.618	0.258	0.146
			SIRS	0.614	0.508	0.150	0.002
		70	*SIS*	0.938	0.926	0.666	0.498
			DC-SIS	0.900	0.870	0.470	0.318
			MDC-SIS	0.926	0.906	0.548	0.364
			SIRS	0.890	0.840	0.396	0.006
		100	*SIS*	0.998	0.996	0.904	0.834
			DC-SIS	0.998	0.988	0.820	0.650
			MDC-SIS	0.998	0.994	0.852	0.714
			SIRS	0.994	0.986	0.742	0.092
10	1000	20	*SIS*	0.108	0.124	0.058	0.022
			DC-SIS	0.078	0.084	0.038	0.012
			MDC-SIS	0.092	0.108	0.052	0.012
			SIRS	0.054	0.078	0.020	0.000
		50	*SIS*	0.906	0.860	0.584	0.442
			DC-SIS	0.840	0.794	0.438	0.300
			MDC-SIS	0.868	0.836	0.478	0.330
			SIRS	0.836	0.790	0.322	0.022
		70	*SIS*	0.978	0.976	0.848	0.722
			DC-SIS	0.958	0.954	0.684	0.562
			MDC-SIS	0.972	0.960	0.768	0.628
			SIRS	0.968	0.946	0.632	0.100
		100	*SIS*	0.998	0.998	0.980	0.944
			DC-SIS	1.000	0.998	0.930	0.856
			MDC-SIS	1.000	1.000	0.962	0.900
			SIRS	0.998	0.998	0.902	0.226
15	1000	20	*SIS*	0.184	0.204	0.116	0.052
			DC-SIS	0.116	0.140	0.064	0.030
			MDC-SIS	0.160	0.176	0.076	0.028
			SIRS	0.114	0.120	0.042	0.002
		50	*SIS*	0.936	0.904	0.672	0.558
			DC-SIS	0.884	0.858	0.536	0.412
			MDC-SIS	0.906	0.864	0.602	0.430
			SIRS	0.898	0.848	0.448	0.042
		70	*SIS*	0.990	0.986	0.902	0.824
			DC-SIS	0.972	0.964	0.808	0.690
			MDC-SIS	0.982	0.974	0.866	0.734
			SIRS	0.982	0.970	0.718	0.156
		100	*SIS*	1.000	0.998	0.990	0.968
			DC-SIS	1.000	1.000	0.964	0.912
			MDC-SIS	1.000	1.000	0.976	0.940
			SIRS	1.000	0.998	0.938	0.310

Table 12: $\mathcal{P}_a$ for Example 1 with $d = n$ under several SNRs (signal noise ratios)

SNR	*p*	*n*	*Method*	*Results for the following values of ρ:*			
				$\rho = 0$	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
SNR=75	1000	20	*SIS*	0.212	0.224	0.144	0.092
			DC-SIS	0.140	0.168	0.096	0.060
			MDC-SIS	0.174	0.212	0.116	0.062
			SIRS	0.112	0.132	0.068	0.008
		50	*SIS*	0.980	0.978	0.854	0.810
			DC-SIS	0.978	0.962	0.794	0.700
			MDC-SIS	0.978	0.972	0.822	0.718
			SIRS	0.966	0.954	0.728	0.226
		70	*SIS*	1	1	0.982	0.956
			DC-SIS	0.998	0.998	0.952	0.898
			MDC-SIS	1	0.996	0.964	0.920
			SIRS	0.998	0.998	0.926	0.468
SNR=27	1000	20	*SIS*	0.220	0.258	0.148	0.046
			DC-SIS	0.146	0.188	0.088	0.026
			MDC-SIS	0.186	0.236	0.112	0.030
			SIRS	0.140	0.154	0.048	0.002
		50	*SIS*	0.976	0.972	0.882	0.732
			DC-SIS	0.964	0.946	0.778	0.588
			MDC-SIS	0.970	0.958	0.818	0.614
			SIRS	0.958	0.950	0.728	0.176
		70	*SIS*	1.000	1.000	0.986	0.966
			DC-SIS	0.998	0.996	0.962	0.896
			MDC-SIS	0.998	0.998	0.976	0.912
			SIRS	1.000	0.998	0.934	0.440
SNR=1	1000	20	*SIS*	0.016	0.024	0.006	0.000
			DC-SIS	0.010	0.016	0.008	0.002
			MDC-SIS	0.010	0.014	0.004	0.002
			SIRS	0.008	0.022	0.000	0.000
		50	*SIS*	0.604	0.666	0.336	0.034
			DC-SIS	0.506	0.572	0.248	0.024
			MDC-SIS	0.572	0.632	0.286	0.018
			SIRS	0.526	0.570	0.240	0.010
		70	*SIS*	0.892	0.914	0.660	0.076
			DC-SIS	0.832	0.866	0.560	0.070
			MDC-SIS	0.864	0.896	0.580	0.066
			SIRS	0.836	0.866	0.550	0.060

Table 13: $\mathcal{P}_a$ for Example 1 with $p = 3000$ and $d = \lfloor n/log(n) \rfloor$

p	n	*Method*	*Results for the following values of ρ:*			
			$\rho = 0$	$\rho = 0.1$	$\rho = 0.5$	$\rho = 0.9$
3000	50	*SIS*	0.846	0.788	0.466	0.316
		DC-SIS	0.752	0.710	0.312	0.198
		MDC-SIS	0.804	0.752	0.360	0.232
		SIRS	0.748	0.672	0.222	0.010
	70	*SIS*	0.980	0.968	0.782	0.654
		DC-SIS	0.952	0.936	0.650	0.480
		MDC-SIS	0.962	0.944	0.712	0.540
		SIRS	0.950	0.924	0.548	0.048
	100	*SIS*	0.998	0.998	0.982	0.956
		DC-SIS	1.000	1.000	0.938	0.856
		MDC-SIS	1.000	1.000	0.958	0.900
		SIRS	0.998	0.998	0.900	0.232

8.2 Example 2

In this example, we consider two nonlinear additive models, which have been analyzed in Meier, Geer, and Bühlmann (2009) and Fan, Feng and Song (2011). Let $g_1(x) = x$, $g_2(x) = (2x-1)^2$, $g_3(x) = sin(2\pi x)/(2 - sin(2\pi x))$, and $g_4(x) = 0.1sin(2\pi x) + 0.2cos(2\pi x) + 0.3sin^2(2\pi x) + 0.4cos^3(2\pi x) + 0.5sin^3(2\pi x)$. The following cases are studied:

Case 2.a: $Y = 5g_1(X_1) + 3g_2(X_2) + 4g_3(X_3) + 6g_4(X_4) + \sqrt{1.74}\epsilon$, where the covariates X_j, $j = 1, ..., p$ are simulated according to iid Unif(0,1), and ϵ is independent from the covariates and follows the standard normal distribution.

Case 2.b: The covariates and the error term are simulated as in Case 2a, but the model structure is more involved with 8 additional predictor variables. $Y = g_1(X_1) + g_2(X_2) + g_3(X_3) + g_4(X_4) + 1.5g_1(X_5) + 1.5g_2(X_6) + 1.5g_3(X_7) + 1.5g_4(X_8) + 2g_1(X_9) + 2g_2(X_{10}) + 2g_3(X_{11}) + 2g_4(X_{12}) + \sqrt{0.5184}\epsilon$.

Table 14: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Example 2 with $p = 2000$ and $n = 200$.

	Method	5%	25%	50%	75%	95%
	SIS	100.70	453.25	910.00	1494.25	1892.05
	DC-SIS	21.95	107.00	250.50	465.25	1140.05
2.a	MDC-SIS	18.00	90.75	195.00	368.50	894.90
	NIS	54.85	383.50	1112.50	1626.25	1951.10
	SIRS	107.40	420.00	970.50	1492.75	1903.00
	SIS	914.50	1366.50	1669.00	1837.00	1973.05
	DC-SIS	370.85	708.75	1083.00	1435.50	1822.05
2.b	MDC-SIS	277.95	571.75	846.00	1262.00	1757.20
	NIS	1347.9	1664.0	1860.0	1949.0	1996.0
	SIRS	903.15	1439.50	1696.50	1867.00	1976.05

Table 15: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Example 2 with $d = \lfloor n/logn \rfloor$, $p = 2000$ and $n = 200$.

	Method	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	*ALL*
	SIS	0.998	0.014	0.996	0.998	0.022	0.026	0.026	0.030	0.014	0.020	0.012	0.024	0.014
	DC-SIS	0.990	0.080	0.998	1.000	0.018	0.018	0.022	0.022	0.016	0.022	0.018	0.018	0.080
2.a	MDC-SIS	0.998	0.108	1.000	1.000	0.016	0.018	0.026	0.034	0.014	0.016	0.010	0.022	0.108
	NIS	0.952	0.038	0.996	1.000	0.008	0.016	0.018	0.028	0.020	0.014	0.020	0.020	0.038
	SIRS	0.990	0.018	0.976	0.998	0.018	0.020	0.018	0.028	0.016	0.022	0.014	0.022	0.018
	SIS	0.322	0.028	0.400	0.220	0.682	0.026	0.814	0.498	0.942	0.014	0.976	0.8	0.000
	DC-SIS	0.294	0.036	0.486	0.334	0.600	0.080	0.886	0.844	0.896	0.184	0.990	1	0.000
2.b	MDC-SIS	0.328	0.030	0.530	0.376	0.650	0.084	0.908	0.862	0.932	0.216	0.992	1	0.000
	NIS	0.096	0.016	0.152	0.086	0.246	0.010	0.414	0.296	0.474	0.018	0.778	0.634	0.000
	SIRS	0.300	0.022	0.376	0.186	0.600	0.026	0.758	0.458	0.894	0.012	0.966	0.778	0.000

8.3 Example 4

This example consists of three cases:

Case 4.a $Y = X_1 + 0.8X_2 + 0.6X_3 + 0.4X_4 + 0.2X_5 + exp(X_{20} + X_{21} + X_{22}) \cdot \epsilon$.

Case 4.b $Y = X_1 + 0.8\sin(|X_2|) + 0.6\exp(|X_3|) + 0.4X_4 + 0.2X_5 + exp(X_{20} + X_{21} + X_{22}) \cdot \epsilon$.

Case 4.c $Y = X_1X_2 + 0.6X_3 + 0.4X_4 + 0.2X_5 + exp(X_{20} + X_{21} + X_{22}) \cdot \epsilon$.

In the above models, the error $\epsilon \sim N(0, 1)$ and is independent from the covariates. The predictor vector follows the multivariate normal distribution with the correlation structure described in Example 1 but with $\rho = 0.8$. All models in this example are heteroscedastic with the number of active variables being 5 at the median (i.e., $\tau = 0.5$) but 8 for other τs. Case 4.a is adapted from an example used in Zhu et al. (2011). Cases 4.b and 4.c are modified versions of Case 4.a by including nonlinear structure and interaction terms. We report $\mathcal{S}$, $\mathcal{P}_s$ and $\mathcal{P}_a$ with $d = \lfloor n/logn \rfloor$ for all three methods in Tables 6 and 7. Tables 16-21 below report the minimum model size and the proportions of P_s and P_a for Cases 4a, 4b and 4c with varying degree of signal-to-noise ratios.

Table 16: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Case 4.a with different SNR

Settings	τ	*Method*	5%	25%	50%	75%	95%
				Case 4.a	$(c = 0.5)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	5.00	5.00	8.00	34.00	337.55
		MDC-SISQ	5.00	5.75	9.00	34.00	317.05
		QaSIS	24.95	54.00	112.50	218.50	763.15
	0.75	SISQ	39.95	146.00	389.50	965.00	2574.35
		MDC-SISQ	39.95	160.50	412.50	1081.25	2586.15
		QaSIS	214.95	472.50	810.50	1368.75	2377.10
		DC-SIS	105.90	637.75	1224.50	1864.75	2494.30
		SIRS	10.00	16.00	29.00	65.25	372.95
$\frac{n=200}{p=3000}$	0.5	SISQ	5.00	5.00	5.00	5.00	14.05
		MDC-SISQ	5.00	5.00	5.00	5.00	14.00
		QaSIS	6.00	9.00	12.00	19.25	70.05
	0.75	SISQ	9.00	15.00	36.00	132.25	705.80
		MDC-SISQ	9.00	15.00	39.00	140.25	640.50
		QaSIS	48.00	123.75	236.00	472.75	1209.65
		DC-SIS	61.80	362.25	721.50	1257.25	2171.00
		SIRS	8.00	8.00	9.00	11.00	19.00
				Case 4.a	$(c = 2)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	5.00	5.00	5.00	7.00	34.05
		MDC-SISQ	5.00	5.00	5.00	7.00	41.00
		QaSIS	6.00	9.00	15.00	32.25	103.45
	0.75	SISQ	22.00	165.00	623.50	1542.50	2744.80
		MDC-SISQ	27.00	147.75	645.50	1548.25	2727.15
		QaSIS	37.00	110.50	245.00	615.00	2401.15
		DC-SIS	9.00	19.00	84.00	379.25	1474.35
		SIRS	21.00	64.00	142.00	275.00	732.20
$\frac{n=200}{p=3000}$	0.5	SISQ	5.00	5.00	5.00	5.00	5.00
		MDC-SISQ	5.00	5.00	5.00	5.00	5.00
		QaSIS	5.00	5.00	6.00	7.00	10.00
	0.75	SISQ	9.00	23.00	104.00	490.50	2059.30
		MDC-SISQ	9.00	20.75	97.50	480.75	2022.15
		QaSIS	9.00	14.00	22.00	49.25	977.15
		DC-SIS	8.00	10.00	16.00	75.00	839.95
		SIRS	10.00	14.00	22.00	38.00	110.00

Table 17: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Case 4.a with $d = \lfloor n/logn \rfloor$ and different SNR

Settings	τ	*Method*	X_1	X_2	X_3	X_4	X_5	X_{20}	X_{21}	X_{22}	*ALL*
					Case	*4.a*	$(c = 0.5)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.966	0.982	0.970	0.914	0.710	0.014	0.014	0.006	0.686
		MDC-SISQ	0.978	0.980	0.980	0.918	0.688	0.014	0.016	0.008	0.666
		QaSIS	0.372	0.514	0.408	0.230	0.068	0.656	0.742	0.674	0.036
	0.75	SISQ	0.602	0.656	0.584	0.476	0.314	0.506	0.560	0.488	0.014
		MDC-SISQ	0.588	0.646	0.580	0.450	0.294	0.530	0.590	0.506	0.018
		QaSIS	0.016	0.020	0.016	0.008	0.004	0.772	0.912	0.808	0.000
		DC-SIS	0.058	0.082	0.064	0.048	0.022	0.984	0.998	0.984	0.006
		SIRS	0.970	0.980	0.962	0.888	0.706	0.700	0.866	0.712	0.350
$\frac{n=200}{p=3000}$	0.5	SISQ	1.000	1.000	1.000	1.000	0.990	0.018	0.042	0.018	0.990
		MDC-SISQ	1.000	1.000	1.000	1.000	0.986	0.022	0.048	0.026	0.986
		QaSIS	1.000	1.000	1.000	0.990	0.896	0.762	0.826	0.772	0.894
	0.75	SISQ	0.912	0.944	0.928	0.842	0.706	0.876	0.910	0.866	0.504
		MDC-SISQ	0.910	0.942	0.920	0.840	0.676	0.892	0.928	0.866	0.482
		QaSIS	0.174	0.252	0.216	0.124	0.072	0.984	0.996	0.980	0.024
		DC-SIS	0.184	0.250	0.194	0.130	0.074	1.000	1.000	1.000	0.032
		SIRS	1.000	1.000	1.000	0.998	0.986	1.000	1.000	1.000	0.986
					Case	*4.a*	$(c = 2)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.998	1.000	1.000	0.996	0.930	0.006	0.014	0.008	0.930
		MDC-SISQ	0.998	1.000	1.000	0.996	0.924	0.004	0.008	0.008	0.924
		QaSIS	0.994	0.998	0.984	0.922	0.638	0.434	0.524	0.442	0.630
	0.75	SISQ	0.960	0.976	0.976	0.930	0.770	0.180	0.200	0.172	0.046
		MDC-SISQ	0.960	0.980	0.980	0.922	0.760	0.188	0.214	0.184	0.042
		QaSIS	0.298	0.372	0.306	0.202	0.092	0.690	0.808	0.662	0.012
		DC-SIS	0.640	0.682	0.656	0.532	0.344	0.952	0.990	0.946	0.290
		SIRS	1.000	1.000	1.000	0.996	0.964	0.146	0.226	0.130	0.052
$\frac{n=200}{p=3000}$	0.5	SISQ	1.000	1.000	1.000	1.000	1.000	0.032	0.024	0.018	1.000
		MDC-SISQ	1.000	1.000	1.000	1.000	1.000	0.028	0.028	0.020	1.000
		QaSIS	1.000	1.000	1.000	1.000	1.000	0.548	0.628	0.518	1.000
	0.75	SISQ	1.000	1.000	1.000	0.998	0.994	0.498	0.566	0.492	0.346
		MDC-SISQ	1.000	1.000	1.000	1.000	0.998	0.520	0.580	0.504	0.356
		QaSIS	0.984	0.992	0.980	0.948	0.828	0.908	0.974	0.926	0.670
		DC-SIS	0.862	0.890	0.876	0.818	0.692	1.000	1.000	1.000	0.690
		SIRS	1.000	1.000	1.000	1.000	1.000	0.830	0.900	0.840	0.742

Table 18: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Case 4.b with different SNR

Settings	τ	*Method*	5%	25%	50%	75%	95%
				Case 4.b	$(c = 0.5)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	9.00	73.75	323.50	998.00	2582.90
		MDC-SISQ	6.95	49.00	209.50	639.50	1948.60
		QaSIS	45.00	144.25	329.50	789.00	1910.30
	0.75	SISQ	58.00	284.75	903.00	1811.00	2854.25
		MDC-SISQ	54.75	261.50	718.00	1596.25	2701.85
		QaSIS	425.75	873.75	1430.00	2072.75	2752.45
		DC-SIS	547.35	1338.50	2007.00	2417.75	2793.10
		SIRS	15.00	80.75	334.50	900.00	2100.80
$\frac{n=200}{p=3000}$	0.5	SISQ	5.00	10.00	39.00	188.50	923.85
		MDC-SISQ	5.00	6.00	18.00	84.25	665.85
		QaSIS	8.00	17.00	40.00	99.00	726.50
	0.75	SISQ	12.00	50.50	179.50	628.50	2204.20
		MDC-SISQ	12.00	39.75	155.00	485.75	1623.20
		QaSIS	149.85	349.75	665.00	1190.50	2170.85
		DC-SIS	329.90	1077.50	1624.00	2161.50	2719.25
		SIRS	8.00	13.00	42.50	189.00	818.05
				Case 4.b	$(c = 2)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	5.00	17.00	88.00	415.25	1650.75
		MDC-SISQ	5.00	10.00	38.50	135.00	854.05
		QaSIS	9.00	24.75	60.50	185.00	754.10
	0.75	SISQ	42.95	231.25	602.50	1512.25	2753.65
		MDC-SISQ	40.00	188.75	513.50	1260.50	2594.35
		QaSIS	144.90	325.00	584.00	1088.25	2187.10
		DC-SIS	25.95	232.00	698.50	1361.50	2252.30
		SIRS	21.00	60.75	130.00	298.75	1377.30
$\frac{n=200}{p=3000}$	0.5	SISQ	5.00	5.00	7.00	22.00	218.75
		MDC-SISQ	5.00	5.00	5.00	8.25	72.35
		QaSIS	5.00	6.00	9.00	13.00	66.20
	0.75	SISQ	9.00	21.75	66.00	306.50	1515.05
		MDC-SISQ	9.00	17.00	52.00	230.50	1308.15
		QaSIS	20.00	45.00	87.00	199.00	766.50
		DC-SIS	12.00	57.00	248.50	748.00	1857.40
		SIRS	8.00	10.00	14.00	24.00	98.15

Table 19: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Case 4.b with $d = \lfloor n/logn \rfloor$ and different SNR

Settings	τ	*Method*	X_1	X_2	X_3	X_4	X_5	X_{20}	X_{21}	X_{22}	*ALL*
					Case	*4.b*	$(c = 0.5)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.696	0.428	0.260	0.348	0.250	0.016	0.028	0.024	0.116
		MDC-SISQ	0.750	0.560	0.514	0.410	0.266	0.026	0.028	0.032	0.156
		QaSIS	0.094	0.198	0.310	0.134	0.054	0.716	0.796	0.688	0.002
	0.75	SISQ	0.278	0.318	0.314	0.232	0.166	0.584	0.618	0.544	0.008
		MDC-SISQ	0.292	0.362	0.432	0.256	0.160	0.614	0.668	0.576	0.010
		QaSIS	0.006	0.004	0.006	0.002	0.002	0.776	0.910	0.800	0.000
		DC-SIS	0.016	0.014	0.010	0.008	0.002	0.986	0.998	0.984	0.002
		SIRS	0.622	0.396	0.280	0.328	0.226	0.906	0.980	0.902	0.088
$\frac{n=200}{p=3000}$	0.5	SISQ	0.974	0.856	0.648	0.722	0.638	0.044	0.040	0.026	0.488
		MDC-SISQ	0.984	0.966	0.958	0.844	0.662	0.038	0.052	0.034	0.624
		QaSIS	0.910	0.954	0.998	0.880	0.558	0.808	0.886	0.816	0.480
	0.75	SISQ	0.594	0.604	0.602	0.520	0.400	0.918	0.946	0.912	0.212
		MDC-SISQ	0.614	0.708	0.786	0.582	0.428	0.926	0.962	0.930	0.242
		QaSIS	0.036	0.048	0.044	0.036	0.026	0.986	0.998	0.988	0.000
		DC-SIS	0.030	0.042	0.046	0.028	0.026	1.000	1.000	1.000	0.002
		SIRS	0.964	0.828	0.664	0.690	0.598	1.000	1.000	1.000	0.474
					Case	*4.b*	$(c = 2)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.956	0.690	0.430	0.552	0.446	0.008	0.016	0.008	0.280
		MDC-SISQ	0.976	0.872	0.808	0.706	0.484	0.010	0.014	0.010	0.398
		QaSIS	0.714	0.808	0.948	0.676	0.296	0.504	0.616	0.514	0.230
	0.75	SISQ	0.714	0.662	0.632	0.554	0.424	0.250	0.312	0.236	0.014
		MDC-SISQ	0.740	0.800	0.880	0.668	0.442	0.282	0.314	0.254	0.022
		QaSIS	0.056	0.066	0.062	0.028	0.024	0.684	0.790	0.670	0.000
		DC-SIS	0.238	0.210	0.234	0.156	0.078	0.966	0.996	0.968	0.042
		SIRS	0.954	0.780	0.592	0.664	0.524	0.316	0.456	0.318	0.056
$\frac{n=200}{p=3000}$	0.5	SISQ	1.000	0.988	0.890	0.946	0.890	0.022	0.020	0.018	0.826
		MDC-SISQ	1.000	0.998	1.000	0.982	0.916	0.024	0.026	0.024	0.912
		QaSIS	1.000	1.000	1.000	1.000	0.928	0.626	0.716	0.640	0.928
	0.75	SISQ	0.976	0.960	0.956	0.926	0.848	0.626	0.698	0.622	0.382
		MDC-SISQ	0.990	0.998	0.996	0.984	0.894	0.652	0.728	0.646	0.440
		QaSIS	0.628	0.688	0.798	0.594	0.356	0.928	0.974	0.930	0.192
		DC-SIS	0.518	0.478	0.486	0.382	0.244	1.000	1.000	1.000	0.200
		SIRS	1.000	0.994	0.958	0.964	0.922	0.966	0.988	0.962	0.850

Table 20: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Case 4.c with different SNR

Settings	τ	*Method*	5%	25%	50%	75%	95%
				Case 4.c	$(c = 0.5)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	16.00	239.75	717.00	1772.75	2816.10
		MDC-SISQ	11.00	73.50	259.50	713.50	2032.05
		QaSIS	64.95	189.50	372.50	765.50	1842.30
	0.75	SISQ	95.95	479.50	1148.00	2249.50	2893.20
		MDC-SISQ	97.95	406.50	1024.00	1810.25	2783.20
		QaSIS	570.95	1013.75	1587.50	2253.75	2811.35
		DC-SIS	662.75	1505.00	2076.00	2520.75	2822.25
		SIRS	21.00	179.25	576.50	1357.25	2538.85
$\frac{n=200}{p=3000}$	0.5	SISQ	7.00	38.75	202.00	698.00	2365.75
		MDC-SISQ	5.00	9.00	30.00	100.00	574.65
		QaSIS	10.00	20.75	40.00	94.50	384.05
	0.75	SISQ	19.00	80.00	349.00	1241.00	2585.55
		MDC-SISQ	13.95	62.00	229.00	843.00	2104.15
		QaSIS	208.90	498.50	856.00	1472.75	2528.60
		DC-SIS	538.85	1275.50	1789.00	2289.00	2743.35
		SIRS	11.00	31.00	136.00	573.75	1771.85
				Case 4.c	$(c = 2)$		
$\frac{n=100}{p=3000}$	0.5	SISQ	8.00	52.50	249.50	884.25	2370.05
		MDC-SISQ	5.00	12.00	41.00	130.00	478.50
		QaSIS	9.00	26.00	52.00	127.25	455.40
	0.75	SISQ	47.95	249.50	666.00	1577.50	2723.15
		MDC-SISQ	44.95	173.00	483.50	1236.50	2459.35
		QaSIS	144.75	336.25	697.00	1202.75	2289.80
		DC-SIS	49.95	304.75	885.50	1512.00	2356.00
		SIRS	21.00	72.50	194.00	584.75	2198.95
$\frac{n=200}{p=3000}$	0.5	SISQ	6.00	16.00	63.00	312.25	1397.05
		MDC-SISQ	5.00	6.00	7.00	13.00	71.15
		QaSIS	5.00	7.00	8.00	12.00	26.00
	0.75	SISQ	10.95	28.75	94.00	361.00	1665.45
		MDC-SISQ	9.00	18.00	54.00	193.25	1380.90
		QaSIS	31.00	64.00	133.00	287.25	1084.70
		DC-SIS	18.00	130.75	431.50	953.50	1740.15
		SIRS	9.00	13.00	29.00	127.00	972.70

Table 21: The proportions of $\mathcal{P}_s$ and $\mathcal{P}_a$ for Case 4.c with $d = \lfloor n/logn \rfloor$ and different SNR

Settings	τ	*Method*	X_1	X_2	X_3	X_4	X_5	X_{20}	X_{21}	X_{22}	*ALL*
					Case	*4.c*	$(c = 0.5)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.086	0.206	0.554	0.602	0.474	0.024	0.022	0.020	0.062
		MDC-SISQ	0.166	0.358	0.634	0.636	0.478	0.026	0.026	0.022	0.094
		QaSIS	0.126	0.160	0.098	0.056	0.028	0.724	0.824	0.744	0.002
	0.75	SISQ	0.114	0.188	0.204	0.174	0.106	0.622	0.672	0.602	0.004
		MDC-SISQ	0.182	0.264	0.218	0.172	0.100	0.640	0.724	0.648	0.002
		QaSIS	0.008	0.002	0.006	0.002	0.004	0.784	0.916	0.812	0.000
		DC-SIS	0.006	0.008	0.006	0.010	0.000	0.986	0.998	0.984	0.000
		SIRS	0.094	0.218	0.494	0.536	0.394	0.944	0.986	0.950	0.052
$\frac{n=200}{p=3000}$	0.5	SISQ	0.262	0.528	0.920	0.956	0.860	0.042	0.038	0.034	0.248
		MDC-SISQ	0.624	0.836	0.976	0.972	0.872	0.036	0.052	0.036	0.562
		QaSIS	0.908	0.964	0.904	0.784	0.550	0.838	0.894	0.834	0.470
	0.75	SISQ	0.368	0.470	0.506	0.434	0.350	0.936	0.962	0.942	0.138
		MDC-SISQ	0.518	0.606	0.578	0.434	0.328	0.952	0.980	0.960	0.160
		QaSIS	0.020	0.018	0.034	0.030	0.020	0.988	0.998	0.990	0.000
		DC-SIS	0.010	0.026	0.042	0.026	0.028	1.000	1.000	1.000	0.000
		SIRS	0.300	0.542	0.874	0.894	0.794	1.000	1.000	1.000	0.274
					Case	*4.c*	$(c = 2)$				
$\frac{n=100}{p=3000}$	0.5	SISQ	0.174	0.418	0.850	0.886	0.794	0.010	0.012	0.008	0.152
		MDC-SISQ	0.470	0.734	0.930	0.932	0.816	0.008	0.014	0.014	0.374
		QaSIS	0.754	0.854	0.718	0.554	0.308	0.584	0.658	0.578	0.206
	0.75	SISQ	0.380	0.512	0.584	0.566	0.420	0.324	0.358	0.310	0.008
		MDC-SISQ	0.618	0.730	0.696	0.566	0.416	0.340	0.390	0.320	0.010
		QaSIS	0.028	0.040	0.030	0.018	0.028	0.696	0.834	0.704	0.000
		DC-SIS	0.070	0.128	0.186	0.180	0.106	0.976	0.996	0.972	0.028
		SIRS	0.274	0.510	0.878	0.896	0.852	0.402	0.518	0.374	0.052
$\frac{n=200}{p=3000}$	0.5	SISQ	0.412	0.806	0.998	0.998	0.996	0.018	0.020	0.020	0.406
		MDC-SISQ	0.912	0.994	1.000	1.000	0.996	0.020	0.022	0.024	0.906
		QaSIS	1.000	1.000	1.000	1.000	0.972	0.682	0.762	0.660	0.972
	0.75	SISQ	0.704	0.848	0.922	0.906	0.784	0.674	0.748	0.670	0.298
		MDC-SISQ	0.968	0.990	0.966	0.926	0.802	0.716	0.776	0.692	0.424
		QaSIS	0.522	0.580	0.482	0.402	0.262	0.944	0.978	0.946	0.104
		DC-SIS	0.170	0.274	0.384	0.358	0.226	1.000	1.000	1.000	0.100
		SIRS	0.580	0.884	1.000	1.000	0.994	0.994	0.998	0.990	0.564

8.4 Examples in He, Wang and Hong (2013)

Example HWH1: (additive model, n=400, p=1000). This example is adapted from Fan et al. (2011). Let $g_1(x) = x$, $g_2(x) = (2x-1)^2$, $g_3(x) = sin(2\pi x)/(2 - sin(2\pi x))$, and $g_4(x) = 0.1sin(2\pi x) + 0.2cos(2\pi x) + 0.3sin(2\pi x)^2 + 0.4cos(2\pi x)^3 + 0.5sin(2\pi x)^3$. The following cases are studied: Case 1a: $Y = 5g_1(X_1)+3g_2(X_2)+4g_3(X_3)+6g_4(X_4)+\sqrt{(}1.74)\epsilon$, where the vector of covariates X is generated from the multivariate normal distribution N(0,Σ) with $\sigma_{ij} = \rho^{|i-j|}$. In Case 1a, we consider $\rho = 0$; Case 1b: same as Case 1a except that $\rho = 0.8$; Case 1c: same as Case 1b except that $\epsilon \sim$ *Cauchy*.

Example HWH2: (index model, n=200,p=2000). This example is adapted from Zhu et al. (2011). The random data are generated from

$$Y = 2(X_1 + 0.8X_2 + 0.6X_3 + 0.4X_4 + 0.2X_5) + exp(X_{20} + X_{21} + X_{22}) \cdot \epsilon.$$

This model is heteroscedastic: the number of active variables is 5 at the median but 8 elsewhere.

Example HWH3: (a more complex structure, n=400,p=5000). Case 3a: $Y = 2(X_1^2 + X_2^2) + exp((X_1 + X_2 + X_{18} + X_{19} + \cdots + X_{30})/10) \cdot \epsilon$, where $\epsilon \sim N(0,1)$, and X follows the multivariate normal distribution with the correlation structure described in Case 1b. In this case, the number of active variables is 2 at the median but is 15 elsewhere. Case 3b: Same as Case 3a, but with $2(X_1^2 + X_2^2)$ replaced by $2((X_1 + 1)^2 + (X_2 + 2)^2)$.

Table 22: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Example HWH1

Case	τ	*Method*	5%	25%	50%	75%	95%
Case 1.a	0.5	SISQ	209.65	487.75	708.00	872.00	975.10
		MDC-SISQ	218.90	413.00	605.00	753.00	929.10
		QaSIS	210.85	516.50	689.00	858.00	968.05
	0.75	SISQ	260.80	502.75	715.00	885.50	975.00
		MDC-SISQ	208.00	461.00	698.50	863.25	973.05
		QaSIS	206.25	498.75	723.00	870.00	974.00
		DC-SIS	177.55	426.00	624.00	805.50	950.10
		SIRS	247.95	508.75	718.50	883.25	979.05
		NIS	224.90	491.25	697.00	856.25	972.00
Case 1.b	0.5	SISQ	5.95	28.00	116.50	382.25	865.15
		MDC-SISQ	4.00	4.00	4.00	4.00	9.00
		QaSIS	4.00	4.00	4.00	4.00	4.00
	0.75	SISQ	4.00	4.00	4.00	6.00	52.00
		MDC-SISQ	4.00	4.00	4.00	4.00	4.05
		QaSIS	4.00	4.00	4.00	4.00	5.00
		DC-SIS	4.00	4.00	4.00	4.00	4.00
		SIRS	4.00	4.00	6.00	11.00	63.10
		NIS	4.00	4.00	4.00	4.00	7.05
Case 1.c	0.5	SISQ	5.00	26.00	108.50	376.25	833.05
		MDC-SISQ	4.00	4.00	4.00	5.00	15.00
		QaSIS	4.00	4.00	4.00	4.00	4.00
	0.75	SISQ	4.00	4.00	5.00	9.25	107.00
		MDC-SISQ	4.00	4.00	4.00	4.00	5.00
		QaSIS	4.00	4.00	4.00	4.00	5.00
		DC-SIS	4.00	4.00	4.00	4.00	5.00
		SIRS	4.00	5.00	7.00	21.00	113.05
		NIS	4.00	7.00	22.00	227.75	954.10

Table 23: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Example HWH2

τ	*Method*	5%	25%	50%	75%	95%
	SISQ	5.00	5.00	5.00	5.00	5.00
0.5	MDC-SISQ	5.00	5.00	5.00	5.00	5.00
	QaSIS	5.00	5.00	6.00	7.00	9.00
	SISQ	10.00	24.00	94.50	322.25	1258.95
0.75	MDC-SISQ	10.00	21.75	85.00	314.50	1124.55
	QaSIS	9.00	13.00	19.00	36.25	561.15
	DC-SIS	8.00	10.00	14.00	49.25	440.80
	SIRS	10.00	13.00	18.50	30.00	65.05
	NIS	497.75	1287.50	1636.50	1847.00	1971.00

Table 24: The 5%, 25%, 50%, 75% and 95% quantiles of the minimum model size $\mathcal{S}$ for Example HWH3

	τ	*Method*	5%	25%	50%	75%	95%
Case 3.a	0.5	SISQ	70.00	633.75	1689.00	3456.25	4718.55
		MDC-SISQ	2.00	2.00	2.00	2.00	2.00
		QaSIS	2.00	2.00	2.00	4.00	7.00
	0.75	SISQ	180.00	1133.75	2688.00	3965.25	4811.40
		MDC-SISQ	17.00	38.00	117.00	410.25	1954.40
		QaSIS	34.00	72.00	130.50	241.00	1064.55
		DC-SIS	800.85	1797.75	2975.00	3964.75	4796.50
		SIRS	40.00	252.25	660.00	1296.25	2916.45
		NIS	1292.70	2841.50	3764.50	4374.25	4885.05
Case 3.b	0.5	SISQ	2.00	2.00	2.00	2.00	2.00
		MDC-SISQ	2.00	2.00	2.00	2.00	2.00
		QaSIS	2.00	2.00	2.00	2.00	5.00
	0.75	SISQ	29.95	174.75	669.00	1709.75	4003.05
		MDC-SISQ	37.00	180.75	666.50	1890.75	4239.65
		QaSIS	23.00	42.00	87.50	400.25	3242.00
		DC-SIS	768.85	1990.75	3065.00	4028.25	4820.00
		SIRS	19.95	24.00	33.00	49.00	115.05
		NIS	1259.85	2887.25	3709.50	4428.25	4884.10

9 Technical Appendix 2

Proof of Theorem 3: Write $MDD^2(V|U) = ||g_{V,U}(s) - g_V g_U(s)||^2$ and $MDD_n^2(V|U) = ||\xi_n(s)||^2$, where $\xi_n(s) = \frac{1}{n}\sum_{k=1}^n V_k e^{i<s,U_k>} - \frac{1}{n}\sum_{k=1}^n V_k \frac{1}{n}\sum_{k=1}^n e^{i<s,U_k>}$. After an elementary transformation, $\xi_n(s)$ can be expressed as

$$\xi_n(s) = \frac{1}{n}\sum_{k=1}^n \widetilde{U}_k \widetilde{V}_k - \frac{1}{n}\sum_{k=1}^n \widetilde{U}_k \frac{1}{n}\sum_{k=1}^n \widetilde{V}_k,$$

where $\widetilde{U}_k = \exp\{i < s, U_k >\} - E[\exp\{i < s, U >\}]$, $\widetilde{V}_k = V_k - E(V)$. Define the region $D(\delta) = \{s : \delta \leq |s|_q \leq 1/\delta\}$ for each $\delta > 0$, $MDD_{n,\delta}^2(V|U) = \int_{D(\delta)} |\xi_n(s)|^2 dw$, where $dw = w(s)ds$ and $w(s) = \frac{1}{c_q |s|_q^{1+q}}$. For any fixed $\delta > 0$, the weight function $w(s)$ is bounded on $D(\delta)$. Hence $MDD_{n,\delta}^2(V|U)$ is a combination of V-statistics with finite expectation. By the SLLN for V-statistics, it follows that almost surely

$$\lim_{n\to\infty} MDD_{n,\delta}^2(V|U) = MDD_{\cdot,\delta}^2(V|U) = \int_{D(\delta)} |g_{U,V}(s) - g_U(s) g_V|^2 dw.$$

Obviously $MDD_{\cdot,\delta}^2(V|U)$ converges to $MDD^2(V|U)$ when δ tends to zero.
Now, it remains to show that

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} |MDD_n^2(V|U) - MDD_{n,\delta}^2(V|U)| = 0.$$

For each $\delta > 0$,

$$|MDD_n^2(V|U) - MDD_{n,\delta}^2(V|U)| = \int_{|s|_q<\delta} |\xi_n(s)|^2 dw + \int_{|s|_q>\frac{1}{\delta}} |\xi_n(s)|^2 dw \qquad (8)$$

For $z = (z_1, z_2, ..., z_q) \in \mathbf{R}^q$, define the function $G(y) = \int_{|z|_q<y} \frac{1-\cos z_1}{|z|_q^{1+q}} dz$. Clearly $G(y)$ is bounded by c_q and $\lim_{y\to 0} G(y) = 0$. Applying the Cauchy-Schwarz inequality, we obtain

$$|\xi_n(s)|^2 \leq \frac{4}{n}\sum_{k=1}^n |\widetilde{U}_k|^2 \frac{1}{n}\sum_{k=1}^n |\widetilde{V}_k|^2. \qquad (9)$$

Hence the first summand in (8) satisfies that

$$\begin{aligned}\int_{|s|_q<\delta} |\xi_n(s)|^2 dw &\leq \frac{4}{n}\sum_{k=1}^n \int_{|s|_q<\delta} \frac{|\widetilde{U}_k|^2}{c_q |s|_q^{1+q}} ds \frac{1}{n}\sum_{k=1}^n |\widetilde{V}_k|^2 \\ &\leq \frac{4}{n}\cdot 2\sum_{k=1}^n E_U\{|U_k - U|_q G(|U_k - U|_q \delta)\} \cdot \frac{2}{n}\sum_{k=1}^n (|V_k|^2 + E(V^2)),\end{aligned}$$

where we used the inequalities $|a-b|^2 \leq 2(a^2+b^2)$ for $a, b \in \mathbf{R}$ and $(E(V))^2 \leq E(V^2)$ as well as the fact that $\int_{|s|_q<\delta} \frac{|\widetilde{U}_k|^2}{c_q|s|_q^{1+q}} ds \leq 2E_U\{|U_k-U|_q G(|U_k-U|_q\delta)\}$, as presented in Székely et al. (2007), page 2777. By the SLLN,

$$\limsup_{n\to\infty} \int_{|s|_q<\delta} |\xi_n(s)|^2 dw \leq 8E\{|U_1-U_2|_q G(|U_1-U_2|_q\delta)\} \cdot 4E(V^2) \quad a.s.$$

Therefore, by the Lebesgue dominated convergence theorem,

$$\limsup_{\delta\to 0} \limsup_{n\to\infty} \int_{|s|_q<\delta} |\xi_n(s)|^2 dw = 0 \quad a.s.$$

Now, consider the second summand in (8). Using the fact that $|\widetilde{U}_k|^2 \leq 4$ and the inequality (9) again, we can derive that

$$\begin{aligned} \int_{|s|_q>\frac{1}{\delta}} |\xi_n(s)|^2 dw &\leq \frac{16}{n}\sum_{k=1}^n \int_{|s|_q>\frac{1}{\delta}} \frac{1}{c_q|s|_q^{1+q}} ds \frac{1}{n}\sum_{k=1}^n |V_k - E(V)|^2 \\ &\leq 16h(\delta)\frac{2}{n}\sum_{k=1}^n \{V_k^2 + E(V^2)\}, \end{aligned}$$

where $h(\delta) = \int_{|s|_q>\frac{1}{\delta}} \frac{1}{c_q|s|_q^{1+q}} ds$ goes to zero as $\delta \to 0$; compare page 2778 of Székely et al. (2007). Thus, almost surely $\limsup_{\delta\to 0}\limsup_{n\to\infty} \int_{|s|_q>\frac{1}{\delta}} |\xi_n(s)|^2 dw = 0$, which implies that $MDD_n(V|U) \to_{a.s.} MDD(V|U)$. The consistency of $MDC_n(V|U)$ follows from the fact that $Var_n(V) \to Var(V)$ (SLLN) and $dVar_n(U) \to_{a.s.} dVar(U)$ (Theorem 2 in Székely et al. (2007)). The proof is then complete. ♢

Proof of Theorem 4: The argument is similar to that presented in the proofs of Theorem 5 and Corollary 2 of Székely et al. (2007). (a), Define the process $\Gamma_n(s) = \sqrt{n}\xi_n(s) = \sqrt{n}(g_{U,V}^n(s) - g_U^n(s)g_V^n)$. After some straightforward calculation, we can derive that

$$\begin{aligned} &E[\Gamma_n(s)] = 0 \\ &E[\Gamma_n(s)\overline{\Gamma_n(s_0)}] = (\frac{n-1}{n})^2 F(s-s_0) + \frac{n-1}{n} g_U(s-s_0)[\frac{1}{n}E(V^2) - (EV)^2] \\ &\quad + \frac{n-1}{n}[(EV)^2 + \frac{n-2}{n}E(V^2)]g_U(s)\overline{g_U(s_0)} - (\frac{n-1}{n})^2 F(s)\overline{g_U(s_0)} - (\frac{n-1}{n})^2 g_U(s)\overline{F(s_0)}. \end{aligned}$$

In particular,

$$\begin{aligned} E|\Gamma_n(s)|^2 &= \frac{n-1}{n}E(V^2)(1+\frac{n-2}{n}|g_U(s)|^2) - \frac{n-1}{n}(EV)^2(1-|g_U(s)|^2) \\ &\quad -(\frac{n-1}{n})^2[F(s)\overline{g_U(s)} + g_U(s)\overline{F(s)}]. \end{aligned}$$

In the sequel, we construct a sequence of random variables $\{Q_n(\delta)\}$, such that

$$\begin{aligned} (i) \quad & Q_n(\delta) \xrightarrow{D} Q(\delta) \text{ for each } \delta > 0,; \\ (ii) \quad & \limsup_{n\to\infty} E|Q_n(\delta) - ||\Gamma_n||^2| \to 0 \text{ as } \delta \to 0; \\ (iii) \quad & E|Q(\delta) - ||\Gamma||^2| \to 0 \text{ as } \delta \to 0. \end{aligned}$$

Then the weak convergence of $||\Gamma_n||^2$ to $||\Gamma||^2$ follows from Theorem 8.6.2 of Resnick (1999).

Following the construction in Sz$\acute{e}$kely et al. (2007), we define

$$Q_n(\delta) = \int_{D(\delta)} |\Gamma_n(s)|^2 dw \text{ and } Q(\delta) = \int_{D(\delta)} |\Gamma(s)|^2 dw.$$

Given $\epsilon = 1/p > 0, p \in \mathbb{N}$, choose a partition $\{D_k\}_{k=1}^N$ of $D(\delta)$ into $N = N(\epsilon)$ measurable sets with diameter at most ϵ. Then $Q_n(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma_n(s)|^2 dw$ and $Q(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma(s)|^2 dw$. Define $Q_n^p(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma_n(s_0(k))|^2 dw$ and $Q^p(\delta) = \sum_{k=1}^N \int_{D_k} |\Gamma(s_0(k))|^2 dw$, where $\{s_0(k)\}_{k=1}^N$ are a set of distinct points such that $s_0(k) \in D_k$. By multivariate CLT and continuous mapping theorem, $Q_n^p(\delta) \xrightarrow{D} Q^p(\delta)$, for any $p \in \mathbb{N}$. Then in view of Theorem 8.6.2 of Resnick (1999), (i) holds if we can show

$$\limsup_{p\to\infty} E|Q^p(\delta) - Q(\delta)| = 0 \tag{10}$$

and

$$\limsup_{p\to\infty} \limsup_{n\to\infty} E|Q_n^p(\delta) - Q_n(\delta)| = 0. \tag{11}$$

Let $\beta_n(\epsilon) = \sup_{s,s_0} E||\Gamma_n(s)|^2 - |\Gamma_n(s_0)|^2|$ and $\beta(\epsilon) = \sup_{s,s_0} E||\Gamma(s)|^2 - |\Gamma(s_0)|^2|$, where the supremum is taken over all s and s_0, under the restrictions: $\delta < |s|_q, |s_0|_q < 1/\delta$ and $|s - s_0|_q < \epsilon$. In view of the form of $Cov_\Gamma(s, s_0)$ (defined after Theorem 3) and by applying the Cauchy-Swartz inequality, we derive that

$$\begin{aligned} \beta(\epsilon) &= \sup_{s,s_0} E|(\Gamma(s) - \Gamma(s_0))\overline{\Gamma(s)} + \Gamma(s_0)(\overline{\Gamma(s)} - \overline{\Gamma(s_0)})| \\ &\leq \sup_{s,s_0} E^{1/2}|\Gamma(s) - \Gamma(s_0)|^2 (E^{1/2}|\Gamma(s)|^2 + E^{1/2}|\Gamma(s_0)|^2) \leq C \sup_{s,s_0} E^{1/2}|\Gamma(s) - \Gamma(s_0)|^2 \\ &\leq \sup_{s,s_0} C|Cov_\Gamma(s,s) - Cov_\Gamma(s,s_0) - Cov_\Gamma(s_0,s) + Cov_\Gamma(s_0,s_0)|^{1/2}. \end{aligned}$$

Since $g_U(s)$ and $F(s)$ are uniformly continuous in $s \in \mathbf{R}^q$, it can be easily shown that $\beta(\epsilon) \to 0$ as $\epsilon \to 0$. To show (10), we note that

$$\begin{aligned} E|Q^p(\delta) - Q(\delta)| &= E\left|\int_{D(\delta)} |\Gamma(s)|^2 dw - \sum_{k=1}^N \int_{D_k} |\Gamma(s_0(k))|^2 dw\right| \\ &= E\left|\sum_{k=1}^N \int_{D_k} (|\Gamma(s)|^2 - |\Gamma(s_0(k))|^2) dw\right| \\ &\leq \beta(1/p) \int_{D(\delta)} \frac{1}{c_q |s|_q^{1+q}} ds \to 0, \text{ as } p \to \infty \end{aligned}$$

Using exactly the same argument, we can show (11) and thus (i) holds.

On the other hand,

$$E|\int_{D(\delta)} |\Gamma_n(s)|^2 dw - \int_{\mathbf{R}^q} |\Gamma_n(s)|^2 dw| = \int_{|s|_q < \delta} E|\Gamma_n(s)|^2 dw + \int_{|s|_q > 1/\delta} E|\Gamma_n(s)|^2 dw$$

Following similar steps as in the proof of Theorem 3, we can derive that for any small ϵ, there exist δ_0, n_0, such that when $n \geq n_0$ and $\delta \leq \delta_0$, $\int_{|s|_q<\delta} E|\Gamma_n(s)|^2 dw < \epsilon$ and $\int_{|s|_q>1/\delta} E|\Gamma_n(s)|^2 dw < \epsilon$. Thus, we complete our proof for (ii). A similar argument also applies to $Q(\delta)$, so (iii) holds. Therefore $nMDD_n^2(V|U) = ||\Gamma_n||^2 \xrightarrow[n\to\infty]{D} ||\Gamma||^2$.

(b), According to the first assertion, under the assumption that $MDC(V|U) = 0$, $nMDD_n^2(V|U)$ converges in distribution to a quadratic form $||\Gamma||^2$. Note that

$$\begin{aligned} E||\Gamma||^2 &= \int_{\mathbf{R}^q} Cov_\Gamma(s,s) dw \\ &= \int_{\mathbf{R}^q} \{[E(V^2) - (EV)^2](1 - |g_U(s)|^2) + 2E(V^2)|g_U(s)|^2 - F(s)\overline{g_U(s)} - g_U(s)\overline{F(s)}\} dw \end{aligned}$$

Under the assumption that $E(V^2|U) = E(V^2)$, $F(s) = E(V^2) g_U(s)$, which implies that $E||\Gamma||^2 = E|U - U'|_q \cdot [E(V^2) - (EV)^2]$. By the SLLN for V-statistics, $S_n \xrightarrow[n\to\infty]{a.s.} E|U - U'|_q \cdot [E(V^2) - (EV)^2]$. Therefore

$$nMDD_n^2(V|U)/S_n \xrightarrow[n\to\infty]{D} Q,$$

where $E[Q] = 1$ and Q is a nonnegative quadratic form of centered Gaussian random variable following the argument in the proof of Corollary 2 of Székely et al. (2007).

(c), Suppose that $MDD(V|U) > 0$, then Theorem 3 implies that $MDD_n^2(V|U) \xrightarrow[n\to\infty]{a.s.} MDD^2(V|U) > 0$, and therefore $nMDD_n^2(V|U) \xrightarrow[n\to\infty]{P} \infty$. By the SLLN, S_n converges to a constant and therefore $nMDD_n^2(V|U)/S_n \xrightarrow[n\to\infty]{P} \infty$.

◇

Proof of Theorem 5: Our argument basically follows that in the proof of Theorem 1 of Li, Zhong and Zhu (2012) with a slight modification. For the sake of completeness, we present the details. In our proof, the positive constant C is generic and its value may vary from place to place.

We shall first show the uniform consistency of $\widehat{\omega}_j = (MDC_n^j)^2$ under the assumption (A1). Due to the similarity of its numerator and denominator, we only deal with its numerator, i.e., the uniform consistency of $(MDD_n^j)^2$. Let $S_1^j = E[YY'|X_j - X_j'|]$, $S_2^j = E[YY']E[|X_j - X_j'|]$ and $S_3^j = E[YY'|X_j - X_j''|]$, where (X_j', Y') and (X_j'', Y'') are iid copies of (X_j, Y). Correspondingly, denote their sample counterparts as

$$
\begin{aligned}
S_{1n}^j &= \frac{1}{n^2}\sum_{k,l=1}^n Y_k Y_l |X_{jk} - X_{jl}|, \\
S_{2n}^j &= \frac{1}{n^2}\sum_{k,l=1}^n Y_k Y_l \frac{1}{n^2}\sum_{k,l=1}^n |X_{jk} - X_{jl}|, \\
S_{3n}^j &= \frac{1}{n^3}\sum_{k,l,h=1}^n Y_k Y_h |X_{jk} - X_{jl}|.
\end{aligned}
$$

According to the proofs of Theorems 1 and 2, MDD^j and MDD_n^j can be expressed as $(MDD^j)^2 = -S_1^j - S_2^j + 2S_3^j$ and $(MDD_n^j)^2 = -S_{1n}^j - S_{2n}^j + 2S_{3n}^j$. We shall establish the consistency result for each part respectively.

Part I: Consistency of S_{1n}^j

Define a U-statistic $\widetilde{S}_{1n}^j = \{n(n-1)\}^{-1}\sum_{k\neq l} Y_k Y_l |X_{jk} - X_{jl}|$ with the kernel function $h_1(X_{jk}, Y_k; X_{jl}, Y_l) = Y_k Y_l |X_{jk} - X_{jl}|$. First, we shall show that the uniform consistency of S_{1n}^j can be derived by that of $\widetilde{S}_{1n}^j$. By using the Cauchy-Schwarz inequality, $S_1^j = E[YY'|X_j - X_j'|] \leq \{E[(YY')^2]\cdot E[|X_j - X_j'|^2]\}^{1/2} \leq \{(E(Y^4))^{1/2}\cdot (E[(Y')^4])^{1/2}\cdot 4E(X_j^2)\}^{1/2} = 2(E(X_j^2)\cdot E(Y^4))^{1/2}$. Under the assumption (A1), $\sup_p \max_{1\leq j\leq p} S_1^j < \infty$, i.e., $\{S_1^j\}_{j=1}^p$ are uniformly bounded. Thus, for any $\epsilon > 0$, there exists a sufficiently large n, s.t. $S_1^j/n \leq \epsilon$ for any $j = 1, \cdots, p$ (in the case $\epsilon = cn^{-\kappa}$ as will be specified later, this still holds). Then, $P(|S_{1n}^j - S_1^j| \geq 2\epsilon) = P(|\frac{n-1}{n}(\widetilde{S}_{1n}^j - S_1^j) - \frac{1}{n}S_1^j| \geq 2\epsilon) \leq P(|\widetilde{S}_{1n}^j - S_1^j| + |S_1^j/n| \geq 2\epsilon) \leq P(|\widetilde{S}_{1n}^j - S_1^j| \geq \epsilon)$.

Next, we shall establish the uniform consistency of $\widetilde{S}_{1n}^{j}$ based on the theories of U-statistics. Write $\widetilde{S}_{1n}^{j}$ as $\widetilde{S}_{1n}^{j}=\{n(n-1)\}^{-1}\sum_{k\neq l}h_1 I\{|h_1|\leq M\}+\{n(n-1)\}^{-1}\sum_{k\neq l}h_1 I\{|h_1|>M\}=\widetilde{S}_{1n,1}^{j}+\widetilde{S}_{1n,2}^{j}$. Correspondingly, its population counterpart can also be decomposed as $S_1^j=E[h_1 I\{|h_1|\leq M\}]+E[h_1 I\{|h_1|>M\}]=S_{1,1}^j+S_{1,2}^j$. Note that $\widetilde{S}_{1n,1}^{j}$ and $\widetilde{S}_{1n,2}^{j}$ are unbiased estimators of $S_{1,1}^j$ and $S_{1,2}^j$ respectively.

To show the consistency of $\widetilde{S}_{1n,1}^{j}$, we note that all U-statistics can be expressed as an average of averages of iid random variables, see Serfling (1980, Section 5.1.6). Denote $m=\lfloor n/2\rfloor$ and define

$$\Omega(X_{j1},Y_1;\cdots;X_{jn},Y_n)=\frac{1}{m}\sum_{r=0}^{m-1}h_1^{(r)}I\{|h_1^{(r)}|\leq M\},\text{ where }h_1^{(r)}=h_1(X_{j\ 1+2r},Y_{1+2r};X_{j\ 2+2r},Y_{2+2r}).$$

Then we have $\widetilde{S}_{1n,1}^{j}=(n!)^{-1}\sum_{n!}\Omega(X_{ji_1},Y_{i_1};\cdots;X_{ji_n},Y_{i_n})$, where $\sum_{n!}$ denote summation over all $n!$ permutations $(i_1,\cdots,i_n)$ of $(1,\cdots,n)$. By Jensen's inequality, for $t>0$,

$$\begin{aligned}E[\exp(t\widetilde{S}_{1n,1}^{j})] &= E[\exp\{t(n!)^{-1}\sum_{n!}\Omega(X_{ji_1},Y_{i_1};\cdots;X_{ji_n},Y_{i_n})\}]\\ &\leq (n!)^{-1}\sum_{n!}E[\exp(t\sum_{r=0}^{m-1}h_1^{(r)}I\{|h_1^{(r)}|\leq M\}/m)]\\ &= E^m[\exp(th_1^{(r)}I\{|h_1^{(r)}|\leq M\}/m)],\end{aligned}$$

which entails that

$$\begin{aligned}P(\widetilde{S}_{1n,1}^{j}-S_{1,1}^{j}\geq\epsilon) &\leq \exp(-t\epsilon)\exp(-tS_{1,1}^{j})E[\exp(t\widetilde{S}_{1n,1}^{j})]\\ &\leq \exp(-t\epsilon)\cdot E^m\{\exp[t(h_1^{(r)}I\{|h_1^{(r)}|\leq M\}-S_{1,1}^{j})/m]\}\\ &\leq \exp(-t\epsilon)\cdot\exp\{t^2M^2/(2m)\}\end{aligned}$$

where we have applied Markov's inequality and Hoeffding's inequality (see Lemma 1 of Li, Zhong and Zhu (2012)) in the first and third inequality above, respectively. Choosing $t=\epsilon m/M^2$ and utilizing the symmetry of U-statistic, we can obtain that $P(|\widetilde{S}_{1n,1}^{j}-S_{1,1}^{j}|\geq\epsilon)\leq 2\exp\{-\epsilon^2 m/(2M^2)\}$.

Next, we turn to the other part $\widetilde{S}_{1n,2}^{j}$. With the Cauchy-Schwarz inequality and Markov's inequality, $(S_{1,2}^j)^2=(E[h_1 I\{|h_1|>M\}])^2\leq E[h_1^2]\cdot P\{|h_1|>M\}\leq E[h_1^2]E[|h_1|^q]\cdot M^{-q}$ for any $q\in\mathbb{N}$. By applying the inequality $|ab|\leq(a^2+b^2)/2$, $a,b\in\mathbf{R}$ twice, we get $|h_1(X_{jk},Y_k;X_{jl},Y_l)|\leq Y_k^2Y_l^2/2+\frac{1}{2}|X_{jk}-X_{jl}|^2\leq Y_k^4+Y_l^4+X_{jk}^2+X_{jl}^2$, which yields

$E[|h_1|^q] \leq (2^q-1)^2 E[Y_k^{4q} + Y_l^{4q} + X_{jk}^{2q} + X_{jl}^{2q})] < \infty$ by the C_r inequality and assumption (A1).

Thus, if we choose $M = n^\gamma$ for $0 < \gamma < 1/2 - \kappa$, then $S_{1,2}^j \leq \epsilon/2$ for sufficiently large n (in the case $\epsilon = cn^{-\kappa}$ as will be specified later, q can be any integer greater than $2\kappa/\gamma$). Hence, $P(|\widetilde{S}_{1n,2}^j - S_{1,2}^j| \geq \epsilon) \leq P(|\widetilde{S}_{1n,2}^j| \geq \epsilon/2)$. Since the event $\{|\widetilde{S}_{1n,2}^j| \geq \epsilon/2\}$ implies the event $\{Y_k^4 + X_{jk}^2 \geq M/2 \text{ for some } 1 \leq k \leq n\}$, we have that

$$\begin{aligned} P\{|\widetilde{S}_{1n,2}^j| \geq \epsilon/2\} &\leq P(\cup_{k=1}^n \{Y_k^4 + X_{jk}^2 \geq M/2\}) \\ &\leq \sum_{k=1}^n P(\{Y_k^4 + X_{jk}^2 \geq M/2\}) = nP(\{Y_k^4 + X_{jk}^2 \geq M/2\}), \end{aligned}$$

where we have applied Bonferroni's inequality in the second inequality above. Invoking assumption (A1) and Markov's inequality, there must exist a constant C, s.t. $P(Y_k^4 + X_{jk}^2 \geq M/2) \leq P(Y_k^2 \geq \sqrt{M}/2) + P(X_{jk}^2 \geq M/4) \leq C\exp(-s\sqrt{M}/2)$ for any j, k, and $s \in (0, 2s_0]$. Consequently, for sufficiently large n, $\max_{1\leq j\leq p} P(|\widetilde{S}_{1n,2}^j - S_{1,2}^j| \geq \epsilon) \leq \max_{1\leq j\leq p} P(|\widetilde{S}_{1n,2}^j| \geq \epsilon/2) \leq \max_{1\leq j\leq p} nP(Y_k^4 + X_{jk}^2 \geq M/2) \leq Cn\exp(-s\sqrt{M}/2)$. In combination with the convergence result of $\widetilde{S}_{1n,1}^j$, we get that for large enough n,

$$\begin{aligned} P(|S_{1n}^j - S_1^j| \geq 4\epsilon) &\leq P(|\widetilde{S}_{1n}^j - S_1^j| \geq 2\epsilon) \\ &\leq P(|\widetilde{S}_{1n,1}^j - S_{1,1}^j| \geq \epsilon) + P(|\widetilde{S}_{1n,2}^j - S_{1,2}^j| \geq \epsilon) \\ &\leq 2\exp(-\epsilon^2 n^{1-2\gamma}/4) + Cn\exp(-sn^{\gamma/2}/2). \end{aligned}$$

Part II: Consistency of S_{2n}^j

Denote S_{2n}^j as $S_{2n}^j = S_{2n,1}^j \cdot S_{2n,2}^j$, where $S_{2n,1}^j = n^{-2}\sum_{k,l=1}^n |X_{jk} - X_{jl}|$, and $S_{2n,2}^j = n^{-2}\sum_{k,l=1}^n Y_k Y_l$. Similarly, write its population counterpart as $S_2^j = S_{2,1}^j \cdot S_{2,2}^j$, where $S_{2,1}^j = E|X_j - X_j'|$ and $S_{2,2}^j = E(YY')$. Following the similar arguments in Part I, we can show that

$$\begin{aligned} &P(|S_{2n,1}^j - S_{2,1}^j| \geq 4\epsilon) \leq 2\exp(-\epsilon^2 n^{1-2\gamma}/4) + Cn\exp(-sn^{2\gamma}/4) \\ &P(|S_{2n,2}^j - S_{2,2}^j| \geq 4\epsilon) \leq 2\exp(-\epsilon^2 n^{1-2\gamma}/4) + Cn\exp(-sn^{\gamma}). \end{aligned}$$

Assumption (A1) ensures that $S_{2,1}^j = E|X_j - X_j'| \leq (E|X_j - X_j'|^2)^{1/2} \leq [4E(|X_j|^2)]^{1/2}$ and $S_{2,2}^j = E(YY') \leq \frac{1}{2}E(Y^2 + Y'^2) = E(Y^2)$ are both uniformly bounded. Let C be a sufficiently large constant, which satisfies

$$C > \max\left(\{S_{2,1}^j\}_{j=1}^p, \{S_{2,2}^j\}_{j=1}^p, \{E[\exp(sX_j^2)]\}_{j=1}^p, E[\exp(sY^2)], 1\right) \quad \text{for } s \in (0, 2s_0].$$

Note that $S_{2n}^j - S_2^j = S_{2n,1}^j \cdot S_{2n,2}^j - S_{2,1}^j \cdot S_{2,2}^j = (S_{2n,1}^j - S_{2,1}^j)(S_{2n,2}^j - S_{2,2}^j) + S_{2,1}^j(S_{2n,2}^j - S_{2,2}^j) + S_{2,2}^j(S_{2n,1}^j - S_{2,1}^j)$. Therefore, by utilizing above inequalities repeatedly, we can show that

$$\begin{aligned} P(|(S_{2n,1}^j - S_{2,1}^j)(S_{2n,2}^j - S_{2,2}^j)| \geq \epsilon) &\leq P(|S_{2n,1}^j - S_{2,1}^j| \geq \sqrt{\epsilon}) + P(|S_{2n,2}^j - S_{2,2}^j| \geq \sqrt{\epsilon}) \\ &\leq 4\exp(-\epsilon n^{1-2\gamma}/64) + 2Cn\exp(-sn^\gamma), \end{aligned}$$

$$P(|S_{2,1}^j(S_{2n,2}^j - S_{2,2}^j)| \geq \epsilon) \leq P(|S_{2n,2}^j - S_{2,2}^j| \geq \epsilon/C) \leq 2\exp(-\epsilon^2 n^{1-2\gamma}/(64C^2)) + Cn\exp(-sn^\gamma),$$

and

$$P(|S_{2,2}^j(S_{2n,1}^j - S_{2,1}^j)| \geq \epsilon) \leq P(|S_{2n,1}^j - S_{2,1}^j| \geq \epsilon/C) \leq 2\exp(-\epsilon^2 n^{1-2\gamma}/(64C^2)) + Cn\exp(-sn^{2\gamma}/4).$$

It follows from Bonferroni's inequality that,

$$\begin{aligned} P(|S_{2n}^j - S_2^j| \geq 3\epsilon) \leq & P(|(S_{2n,1}^j - S_{2,1}^j)(S_{2n,2}^j - S_{2,2}^j)| \geq \epsilon) + P(|S_{2,1}^j(S_{2n,2}^j - S_{2,2}^j)| \geq \epsilon) \\ & + P(|S_{2,2}^j(S_{2n,1}^j - S_{2,1}^j)| \geq \epsilon) \\ \leq & 8\exp(-\epsilon^2 n^{1-2\gamma}/(64C^2)) + 4Cn\exp(-sn^\gamma). \end{aligned}$$

Part III: Consistency of S_{3n}^j

Define the corresponding U-statistic:

$$\begin{aligned} \widetilde{S}_{3n}^j &= \{n(n-1)(n-2)\}^{-1} \sum_{k<l<h} \Big[Y_k Y_l |X_{jk} - X_{jh}| + Y_k Y_h |X_{jk} - X_{jl}| + Y_l Y_k |X_{jl} - X_{jh}| \\ &\qquad + Y_l Y_h |X_{jl} - X_{jk}| + Y_h Y_k |X_{jh} - X_{jl}| + Y_h Y_l |X_{jh} - X_{jk}| \Big] \\ &= 6\{n(n-1)(n-2)\}^{-1} \sum_{k<l<h} h_3(X_{jk}, Y_k; X_{jl}, Y_l; X_{jh}, Y_h), \end{aligned}$$

where $h_3(X_{jk}, Y_k; X_{jl}, Y_l; X_{jh}, Y_h)$ is the kernel function. Following the same argument to deal with $\widetilde{S}_{1n}^j$, we write $\widetilde{S}_{3n}^j$ as $\widetilde{S}_{3n}^j = 6\{n(n-1)(n-2)\}^{-1} \sum_{k<l<h} h_3 I(|h_3| \leq M) + 6\{n(n-1)(n-2)\}^{-1} \sum_{k<l<h} h_3 I(|h_3| > M) = \widetilde{S}_{3n,1}^j + \widetilde{S}_{3n,2}^j$ and its population counterpart as $S_3^j = E[h_3 I\{|h_3| \leq M\}] + E[h_3 I\{|h_3| > M\}] = S_{3,1}^j + S_{3,2}^j$.

By using the same argument for $\widetilde{S}_{1n,1}^j$, we can show that

$$P(|\widetilde{S}_{3n,1}^j - S_{3,1}^j| \geq \epsilon) \leq 2\exp\{-\epsilon^2 m'/(2M^2)\},$$

where $m' = \lfloor n/3 \rfloor$ due to the fact $\widetilde{S}_{3n}^j$ is a third-order U-statistic. Now, it remains to establish the uniform convergency of the other part $\widetilde{S}_{3n,2}^j$. Note that $|h_3(X_{jk}, Y_k; X_{jl}, Y_l; X_{jh}, Y_h)| \leq$

$[Y_k^4+Y_l^4+Y_h^4+X_{jk}^2+X_{jh}^2+X_{jl}^2]$, so the event $\{|\widetilde{S}_{3n,2}^j| \geq \epsilon/2\}$ implies the event $\{Y_k^4+X_{jk}^2 > M/3, \ \ \text{for some } 1 \leq k \leq n\}$. Therefore, following a similar argument as presented in Part I, we have

$$\begin{aligned} P(|\widetilde{S}_{3n,2}^j - S_{3,2}^j| \geq \epsilon) &\leq P(|\widetilde{S}_{3n,2}^j| \geq \epsilon/2) \\ &\leq P(\cup_{k=1}^n [Y_k^4 + X_{jk}^2 \geq M/3]) \leq Cn\exp(-s\sqrt{M}/\sqrt{6}) \end{aligned}$$

for any j, k and $s \in (0, 2s_0]$. Combining the two convergence results for $\widetilde{S}_{3n,1}^j$ and $\widetilde{S}_{3n,2}^j$ with $M = n^\gamma$ for some $0 < \gamma < 1/2 - \kappa$, it follows that

$$P(|\widetilde{S}_{3n}^j - S_3^j| \geq 2\epsilon) \leq 2\exp(-\epsilon^2 n^{1-2\gamma}/6) + Cn\exp(-sn^{\gamma/2}/\sqrt{6}).$$

Note that $S_{3n}^j - S_3^j = \frac{(n-1)(n-2)}{n^2}(\widetilde{S}_{3n}^j - S_3^j) - \frac{3n-2}{n^2}S_3^j + \frac{n-1}{n^2}(\widetilde{S}_{1n}^j - S_1^j) + \frac{n-1}{n^2}S_1^j$. Following the similar argument in dealing with S_1^j, we can show S_3^j is also uniformly bounded in j. Therefore, with a sufficiently large n, $|(3n-2)S_3^j/n^2|$ and $|(n-1)S_1^j/n^2|$ are both smaller than ϵ (in the case $\epsilon = cn^{-\kappa}$, this also holds). Then,

$$\begin{aligned} P(|S_{3n}^j - S_3^j| \geq 4\epsilon) &\leq P(|\widetilde{S}_{3n}^j - S_3^j| \geq \epsilon) + P(|\widetilde{S}_{1n}^j - S_1^j| \geq \epsilon) \\ &\leq 4\exp(-\epsilon^2 n^{1-2\gamma}/24) + 2Cn\exp(-sn^{\gamma/2}/\sqrt{6}). \end{aligned}$$

This, together with the consistency in Part I and Part II, yields that

$$\begin{aligned} &P\{|(2S_{3n}^j - S_{1n}^j - S_{2n}^j) - (2S_3^j - S_1^j - S_2^j)| \geq \epsilon\} \\ \leq &P(|S_{3n}^j - S_3^j| \geq \frac{\epsilon}{4}) + P(|S_{2n}^j - S_2^j| \geq \frac{\epsilon}{4}) + P(|S_{1n}^j - S_1^j| \geq \frac{\epsilon}{4}) \\ = &O\{\exp(-c_1\epsilon^2 n^{1-2\gamma}) + n\exp(-c_2 n^{\gamma/2})\} \end{aligned}$$

for some positive constants c_1 and c_2 and the bound is uniform with respect to $j = 1, \cdots, p$. Analyzing the denominator of $\widehat{\omega}_j$ would generate the same form of convergence rate, so we omit the details here. Let $\epsilon = cn^{-\kappa}$, where κ satisfies $0 < \kappa + \gamma < 1/2$, we then have

$$\begin{aligned} &P\{\max_{1\leq j\leq p} |\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa}\} \leq p \max_{1\leq j\leq p} P\{|\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa}\} \\ \leq &O(p[\exp\{-c_1 n^{1-2(\kappa+\gamma)}\} + n\exp(-c_2 n^{\gamma/2})]) \end{aligned}$$

which finishes our proof for the first part of theorem.

If $\mathcal{D}_E \not\subseteq \widehat{\mathcal{D}}_E$, then there exist some $j \in \mathcal{D}_E$, such that $\widehat{\omega}_j < cn^{-\kappa}$. According to assumption (A2), this particular j would make $|\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa}$, which implies that

$$A = \{\mathcal{D}_E \not\subseteq \widehat{\mathcal{D}}_E\} \subseteq \{|\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa}, \ for\ some\ j \in \mathcal{D}_E\} = B$$

and hence $B^c \subseteq A^c$.

Therefore,

$$P(A^c) \geq P(B^c) = 1 - P(B) = 1 - P(|\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa},\ for\ some\ j \in \mathcal{D}_E)$$
$$\geq 1 - s_n \max_{j \in \mathcal{D}_E} P(|\widehat{\omega}_j - \omega_j| \geq cn^{-\kappa})$$
$$\geq 1 - O(s_n[\exp\{-c_1 n^{1-2(\kappa+\gamma)}\} + n\exp(-c_2 n^{\gamma/2})]),$$

where the first inequality above is due to Bonferroni's inequality. The proof is thus complete. ♢

Proof of Theorem 6: We shall show the uniform consistency of $\widehat{\omega}_j(\widehat{W}) := MDC_n^j(\widehat{W})^2$ under the assumptions (B1) and (B2). Due to the similarity of its numerator and denominator, we only deal with the numerator part, i.e., the consistency of $MDD_n^j(\widehat{W})^2$. First we demonstrate the consistency of $MDD_n^j(W)^2$, and then study the difference between $MDD_n^j(W)^2$ and $MDD_n^j(\widehat{W})^2$. Since W and W_j are uniformly bounded, we can adopt the argument in the proof of Theorem 1 in Li, Zhong and Zhu (2012) (also see the proof of Theorem 5 for a slightly modified argument, where the bound can be slightly improved under the assumption that the response variable is bounded) and get that for any $\gamma \in (0, 1/2 - \kappa)$, there exist positive constants c_1 and c_2 such that for a sufficiently small ϵ (say $\epsilon = cn^{-\kappa}$ as will be specified later),

$$P(|MDD_n^j(W)^2 - MDD^j(W)^2| \geq \epsilon) \leq C[\exp\{-c_1\epsilon^2 n^{1-2\gamma}\} + n\exp(-c_2 n^{\gamma})]. \tag{12}$$

Next we analyze the difference between $MDD_n^j(W)^2$ and $MDD_n^j(\widehat{W})^2$. Denote $\widehat{T}_{1n}^j = n^{-2}\sum_{k,l=1}^n \widehat{W}_k\widehat{W}_l|X_{jk} - X_{jl}|$, $\widehat{T}_{2n}^j = n^{-2}\sum_{k,l=1}^n \widehat{W}_k\widehat{W}_l\frac{1}{n^2}\sum_{k,l=1}^n |X_{jk} - X_{jl}|$, and $\widehat{T}_{3n}^j = n^{-3}\sum_{k,l,h=1}^n \widehat{W}_k\widehat{W}_h|X_{jk} - X_{jl}|$. Similarly T_{1n}^j, T_{2n}^j and T_{3n}^j are defined with $\{\widehat{W}_k\}_{k=1}^n$ replaced by $\{W_k\}_{k=1}^n$. Let $C_0 = \tau + 1$. By using the triangle inequality and the boundness of W_k and $\widehat{W}_k$, we can derive that

$$
\begin{aligned}
&|MDD_n^j(\widehat{W})^2 - MDD_n^j(W)^2| \le |\widehat{T}_{1n}^j - T_{1n}^j| + |\widehat{T}_{2n}^j - T_{2n}^j| + 2|\widehat{T}_{3n}^j - T_{3n}^j| \\
=&\Big|\frac{1}{n^2}\sum_{k,l=1}^n [\widehat{W}_k\widehat{W}_l - W_kW_l]|X_{jk}-X_{jl}|\Big| + \Big|\frac{1}{n^2}\sum_{k,l=1}^n [\widehat{W}_k\widehat{W}_l - W_kW_l]\Big|\frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}| \\
&+ 2\Big|\frac{1}{n^3}\sum_{k,l,h=1}^n [\widehat{W}_k\widehat{W}_h - W_kW_h]|X_{jk}-X_{jl}|\Big| \\
\le&\frac{1}{n^2}\sum_{k,l=1}^n [|\widehat{W}_k(\widehat{W}_l - W_l)| + |W_l(\widehat{W}_k - W_k)|]|X_{jk}-X_{jl}| \\
&+ \frac{1}{n^2}\sum_{k,l=1}^n [|\widehat{W}_k(\widehat{W}_l - W_l)| + |W_l(\widehat{W}_k - W_k)|]\frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}| \\
&+ \frac{2}{n^3}\sum_{k,l,h=1}^n [|\widehat{W}_k(\widehat{W}_h - W_h)| + |W_h(\widehat{W}_k - W_k)|]|X_{jk}-X_{jl}| \\
\le&\frac{4C_0}{n^2}\sum_{k,l=1}^n |\widehat{W}_l - W_l||X_{jk}-X_{jl}| + \frac{4C_0}{n}\sum_{k=1}^n |\widehat{W}_k - W_k|\frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}| =: \Delta_1 + \Delta_2.
\end{aligned}
$$

We first treat Δ_1. By the Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
(\Delta_1)^2 \le& 16C_0^2\frac{1}{n}\sum_{l=1}^n |\widehat{W}_l - W_l|^2 \frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}|^2 \\
&= 16C_0^2\Big[\frac{1}{n}\sum_{l=1}^n |\widehat{W}_l - W_l|^2 \frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}|^2 - \frac{1}{n}\sum_{l=1}^n |\widehat{W}_l - W_l|^2 E|X_{j1}-X_{j2}|^2\Big] \\
&\quad + 16C_0^2\frac{1}{n}\sum_{l=1}^n |\widehat{W}_l - W_l|^2 E|X_{j1}-X_{j2}|^2 =: D_1 + D_2
\end{aligned}
$$

Noting $n^{-1}\sum_{l=1}^n |\widehat{W}_l - W_l|^2 \le 4C_0^2$, then we have

$$
\begin{aligned}
P(|D_1| \ge \epsilon^2/2) &\le P\Big(\Big|\frac{1}{n^2}\sum_{k,l=1}^n |X_{jk}-X_{jl}|^2 - E|X_{j1}-X_{j2}|^2\Big| \ge \epsilon^2/(128C_0^4)\Big) \\
&\le C(\exp(-c_1\epsilon^4 n^{1-2\gamma}) + n\exp(-c_2 n^{2\gamma}))
\end{aligned}
$$

for some positive constants c_1 and c_2, based on equation (B.7) in Li, Zhong and Zhu (2012).

Under the assumption (B2), there exists a positive constant $C_2 < \infty$, such that $E|X_{j1} - X_{j2}|^2 \leq 4E|X_j|^2 < C_2$. Then by Proposition 2,

$$P(|D_2| \geq \epsilon^2/2) \leq P(\frac{1}{n}\sum_{l=1}^{n} |\widehat{W}_l - W_l|^2 \geq \epsilon^2/(32C_0^2C_2)) \leq C\exp(-nc_3\epsilon^4)$$

for small enough ϵ and some $c_3 > 0$. Combining the probability bounds we derived for D_1 and D_2,

$$\begin{aligned} P(|\Delta_1| \geq \epsilon) &\leq C\big(\exp(-c_1\epsilon^4 n^{1-2\gamma}) + n\exp(-c_2 n^{2\gamma}) + \exp(-c_3 n\epsilon^4)\big) \\ &\leq C\big(\exp(-c_1\epsilon^4 n^{1-2\gamma}) + n\exp(-c_2 n^{2\gamma})), \end{aligned}$$

where the third term on the right hand side can be absorbed into the first term. In a similar fashion, we can derive that $P(|\Delta_2| \geq \epsilon) \leq C\big(\exp(-c_1\epsilon^2 n^{1-2\gamma}) + n\exp(-c_2 n^{2\gamma})\big)$ for some positive constants c_1, c_2.

Consequently, in view of (12), we have that

$$\begin{aligned} &P(|MDD_n^j(\widehat{W})^2 - MDD^j(W)^2| \geq 3\epsilon) \leq P(|MDD_n^j(W)^2 - MDD^j(W)^2| \geq \epsilon) \\ &+P(|\Delta_1| \geq \epsilon) + P(|\Delta_2| \geq \epsilon) \leq C\big(\exp(-c_1\epsilon^4 n^{1-2\gamma}) + n\exp(-c_2 n^{\gamma})\big) \end{aligned}$$

for a sufficiently small ϵ and some positive constants c_1, c_2.

The analysis of the denominator of $MDC_n^j(\widehat{W})^2$ would generate a similar form of convergence rate. Therefore, if we set $\epsilon = cn^{-\kappa}$, where κ satisfies $0 < 2\kappa + \gamma < 1/2$, we would have

$$\begin{aligned} P\{\max_{1\leq j\leq p} |\widehat{\omega}_j(\widehat{W}) - \omega_j(W)| \geq cn^{-\kappa}\} &\leq p\max_{1\leq j\leq p} P\{|\widehat{\omega}_j(\widehat{W}) - \omega_j(W)| \geq cn^{-\kappa}\} \\ &\leq C(p[\exp(-c_1 n^{1-2(\gamma+2\kappa)}) + n\exp(-c_2 n^{\gamma})]) \end{aligned}$$

which proves the first assertion. The second assertion follows from the same argument used in proving the second statement in Theorem 5. The proof is complete.

$\diamondsuit$

References

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, 544-557.

Fan, J., and Lv, J. (2008), "Sure Independence Screening for Ultra-High Dimensional Feature Space," *Journal of the Royal Statistical Society, Series B*, 70, 849-911. (with discussions and rejoinder).

He, X., Wang, L., and Hong, H. G. (2013), "Quantile-Adaptive Model-Free Variable Screening for High-Dimensional Heterogeneous Data," *Annals of Statistics*, 41, 342-369.

Li, R., Zhong, W., and Zhu, L. (2012), "Feature Screening via Distance Correlation Learning," *Journal of the American Statistical Association*, 107, 1129-1139.

Meier, L., van de Geer, S., and Bühlmann, P. (2009), "High-Dimensional Additive Modeling," *Annals of Statistics*, 37, 3779-3821.

Resnick, S. I. (1999) A Probability Path. Birkhäuser, Boston.

Serfling, R. J. (1980) Approximation Theorems of Mathematical Statistics, New York: John Wiley & Sons Inc.

Székely, G. J., Rizzo, M. L., and Bakirov, N. K. (2007), "Measuring and Testing Dependence by Correlation of Distances," *Annals of Statistics*, 35, 2769-2794.

Székely, G. J., and Rizzo, M. L. (2009), "Brownian Distance Covariance," *The Annals of Applied Statistics*, 3, 1236-1265.

Zhu, L., Li, L., Li, R., and Zhu, L. (2011), "Model-Free Feature Screening for Ultrahigh-Dimensional Data," *Journal of the American Statistical Association*, 106, 1464-1474.