Conditional Quantile Correlation Learning for Ultrahigh Dimensional Varying Coefficient Models and Its Application in Survival Analysis

Xiaochao Xia[1], Jialiang Li[2,3,4], Bo Fu[5,6]

[1]*Huazhong Agricultural University*

[2]*National University of Singapore,* [3]*Duke-NUS Graduate Medical School,* [4]*Singapore Eye Research Institute*

[5]*Fudan University,* [6]*University College London*

Supplementary Material

In this supplementary file, we will present some additional simulation examples for illustration and give the detailed proofs to Theorems 2.1-2.3. The proofs of Theorems 3.1-3.3 follow similar lines to Theorems 2.1-2.3.

S1 Additional Simulations and Results

In addition to Examples 1 in the main paper, we provide some additional examples below.

Example 2. We generate response data as

$$H(Y) = 2T \cdot X_1 + (2T-1)^2 \cdot X_2 + \sin(2\pi T) \cdot X_3 + \varepsilon,$$

where we consider three monotone transform functions as: (T1) $H(y) = \log(0.5(e^{2y} - 1))$; (T2) $H(y) = \log(2y)$; and (T3) $H(y) = (2y)^{1/3}$. The rest of the model setup are the same as Example 1. Here we consider two types for model error distributions:

- Case (2a): the same as Case (1a), i.e., $\varepsilon \sim N(0, 1)$;
- Case (2b): the sam as Case (1b), i.e., $\varepsilon \sim C(0, 1)$.

The corresponding results are reported in Table 1. From the tabled results we can draw a similar conclusion to that for Example 1. Our method performs more satisfactorily than the NIS. This example illustrates the invariance of the CQCSIS under monotone transformation on the response variable.

				$\varrho = 0$			$\varrho = 0.4$			$\varrho = 0.8$		
$H(y)$	Case	Method(τ)	s_n	MMS	RSD	PS	MMS	RSD	PS	MMS	RSD	PS
(T1)	(2a)	CQCSIS(0.50)	3	9	17	0.850	3	0	1.000	3	0	1.000
		CQCSIS(0.75)	3	21	42	0.750	3	0	0.995	3	0	1.000
		NIS	3	5	12	0.900	3	0	1.000	3	0	1.000
	(2b)	CQCSIS(0.50)	3	36	115	0.580	3	1	0.985	3	0	1.000
		CQCSIS(0.75)	3	132	272	0.375	7	12	0.890	3	1	1.000
		NIS	3	587	374	0.190	551	378	0.160	286	426	0.305
(T2)	(2a)	CQCSIS(0.50)	3	7	20	0.830	3	0	0.995	3	0	1.000
		CQCSIS(0.75)	3	15	63	0.710	3	0	0.995	3	0	1.000
		NIS	3	270	296	0.170	154	221	0.290	9	36	0.805
	(2b)	CQCSIS(0.50)	3	63	132	0.515	3	1	1.000	3	0	1.000
		CQCSIS(0.75)	3	158	271	0.290	8	20	0.880	3	1	1.000
		NIS	3	643	631	0.270	659	613	0.305	622	642	0.295
(T3)	(2a)	CQCSIS(0.50)	3	8	28	0.845	3	0	1.000	3	0	1.000
		CQCSIS(0.75)	3	21	55	0.720	3	0	1.000	3	0	1.000
		NIS	3	68	114	0.480	15	43	0.770	3	0	0.995
	(2b)	CQCSIS(0.50)	3	46	117	0.605	3	1	0.995	3	0	1.000
		CQCSIS(0.75)	3	142	249	0.370	9	20	0.860	3	1	0.995
		NIS	3	820	170	0.000	791	204	0.000	746	281	0.010

Table 1: Results of the median of minimum model size (MMS), its robust standard deviation (RSD) and the proportion of truly active covariates selected (PS) with a pre-specified threshold size $d_n = \lfloor n/\log n \rfloor$ for Example 2.

Example 3. This example is modified from Example 3 in Fan et al. (2011). Specifically, let $\{W_1, \ldots, W_p\}$ be independent and standard normal random variables, and $\{U_1, U_2\}$ be independent random variables with unit uniform distribution. We generate response variable from the following model:

$$Y = 2X_1 + 3TX_2 + (T+1)^2X_3 + \frac{4\sin(2\pi T)}{2-\sin(2\pi T)}X_4 + \exp(T)X_5 + 2\sqrt{T}X_6 + \varepsilon,$$

where $X_j = (W_j + t_1U_1)/(1+t_1), j = 1, \ldots, p$, and $T = (U_2 + t_2U_1)/(1+t_2)$, implying that $\mathrm{Corr}(X_j, X_k) = t_1^2/(12+t_1^2)$ for $j \neq k$ and $\mathrm{Corr}(X_j, T) = t_1t_2/((12+t_1^2)(1+t_2^2))^{1/2}$ independent of j. Regarding the distribution of error ε, we consider four scenarios:

- Case (3a): the same as Case (1a), i.e., $\varepsilon \sim N(0,1)$;
- Case (3b): the same as Case (1b), i.e., $\varepsilon \sim C(0,1)$;
- Case (3c): the error follows the scaled Cauchy distribution, i.e., $\varepsilon = 0.5(\frac{\exp(T)}{1+\exp(T)}X_7 + (2T-1)^2X_8 + \cos(2\pi T)X_9 + \sqrt{T+1}X_{10})C(0,1)$;
- Case (3d): the error follows a scaled Chi-squared distribution, that is $\varepsilon = 0.5(\frac{\exp(T)}{1+\exp(T)}X_7 + (2T-1)^2X_8 + \cos(2\pi T)X_9 + \sqrt{T+1}X_{10}) \cdot (\varepsilon - Q_{\varepsilon,\tau})$ with $\varepsilon \sim \chi^2(1)$.

The results are summarized in Table 2. Eyeballing the table, we observe that CQCSIS and NIS perform similarly as in previous examples.

Example 4. We consider survival time data in a setting where the proportional hazard assumption is not satisfied. Specifically, we generate the response variable $Y = 2X_1 + 3TX_2 + (T+1)^2X_3 + \frac{4\sin(2\pi T)}{2-\sin(2\pi T)}X_4 + \varepsilon$, where X_j's and T are generated in the

			$(t_1,t_2)=(0,0)$			$(t_1,t_2)=(1,0)$			$(t_1,t_2)=(1,1)$		
Case	Method(τ)	s_n	MMS	RSD	PS	MMS	RSD	PS	MMS	RSD	PS
(3a)	CQCSIS(0.50)	6	7	4	0.940	8	6	0.950	26	32	0.795
	CQCSIS(0.75)	6	11	13	0.895	13	19	0.905	38	41	0.700
	NIS	6	6	0	1.000	6	1	1.000	6	0	1.000
(3b)	CQCSIS(0.50)	6	14	19	0.850	26	40	0.750	64	85	0.505
	CQCSIS(0.75)	6	33	53	0.685	73	111	0.485	152	176	0.255
	NIS	6	655	367	0.050	770	236	0.005	790	228	0.000
(3c)	CQCSIS(0.50)	6	10	12	0.905	9	8	0.925	24	25	0.810
	CQCSIS(0.75)	10	823	156	0.000	813	215	0.000	843	160	0.000
	NIS	10	832	164	0.000	837	148	0.000	844	147	0.000
(3d)	CQCSIS(0.50)	10	828	183	0.000	867	129	0.000	855	152	0.000
	CQCSIS(0.75)	6	11	18	0.850	10	11	0.905	26	42	0.720
	NIS	10	815	177	0.000	871	137	0.000	883	147	0.000

Table 2: Results of the median of minimum model size (MMS), its robust standard deviation (RSD) and the proportion of truly active covariates selected (PS) with a pre-specified threshold size $d_n = \lfloor n/\log n \rfloor$ for Example 3.

same way as Example 3 with $(t_1, t_2) = (0, 0)$. Let $\widetilde{Y} = \min(Y, Z)$, where censoring time Z is simulated from a mixture of normal distributions: (S1) $0.2N(-5, 4) + 0.1N(5, 1) + 0.7N(55, 1)$ and (S2) $0.4N(-5, 4) + 0.1N(5, 1) + 0.5N(55, 1)$. We consider two scenarios for the distribution of error ε:

- Case (4a): the same as Case (3a), i.e., $\varepsilon \sim N(0, 1)$;
- Case (4b): the same as Case (3b), i.e., $\varepsilon \sim C(0, 1)$.

Note that in scenarios (S1) and (S2), the censoring rates are roughly 20% and 35%, respectively.

Example 5. We consider an alternative setting where the censoring depends on covariate. The model setup is the same as that in Example 4 only except that the censoring time Z is generated from the covariate dependent mixture distribution: (S3) $0.4N(-5, 4) + 0.1N(5, 1) + 0.5N(55, 1)$ if $X_1 > 0$ and $0.5N(-5, 4) + 0.2N(5, 1) +$

$0.3N(55,1)$ if $X_1 < 0$. The censoring rates are about 40%. Error distributions the same as cases (4a) and (4b).

In both examples, we apply the proposed CQCSIS method at the median and the first quartile, respectively. Here, we compare our censored version of the CQCSIS (denoted by the CQCSIScens) in Section 4 with four existing approaches: the CQCSISnaive, under which we treat the censored observations as complete observations and then apply the CQCSIS, the quantile-adaptive screening (QaSIS) by He et al. (2013) which treats varying coefficients as constants, the conditional quantile sure independent screening (CQSIS) by Wu and Yin (2015) as well as the censored rank sure independent screening (CRSIS) by Song et al. (2014a). In earlier publications all these approaches were shown to be quite robust when the proportional hazards assumption is violated. Other screening methods in the survival literature are known to be less effective and thus are not chosen for comparison in this paper.

Table 3 presents the simulation results. Eyeballing the table, we can see that CQCSIScens performs much better than all other procedures. In this case, QaSIS and CQSIS are unsatisfactory because they fail to acknowledge the varying coefficients. In addition, the results of Example 5 displayed in Table 3 indicate that our proposed CQCSIScens also performs fairly well for various model errors even though the censoring mechanism is covariate-dependent.

Finally, in order to evaluate the performance of proposed two-stage variable selec-

	Censoring	Method(τ)	s_n	$\varepsilon \sim N(0,1)$ MMS	RSD	PS	$\varepsilon \sim C(0,1)$ MMS	RSD	PS
Example 4	(S1) 20%	CQCSIScens(0.50)	4	4	0	1.000	4	1	0.980
		CQCSIScens(0.25)	4	4	1	1.000	6	4	0.940
		CQCSISnaive(0.50)	4	5	4	0.980	10	14	0.900
		CQCSISnaive(0.25)	4	27	46	0.725	65	123	0.510
		QaSIS(0.50)	4	23	65	0.715	77	143	0.460
		QaSIS(0.25)	4	33	74	0.655	140	199	0.330
		CQSIS(0.50)	4	27	88	0.655	75	158	0.480
		CQSIS(0.25)	4	51	100	0.560	120	210	0.405
		CRSIS	4	633	254	0.000	778	190	0.005
	(S2) 35%	CQCSIScens(0.50)	4	4	1	0.995	5	2	0.975
		CQCSIScens(0.25)	4	4	1	0.995	6	10	0.950
		CQCSISnaive(0.50)	4	61	82	0.520	126	177	0.300
		CQCSISnaive(0.25)	4	165	195	0.195	210	259	0.135
		QaSIS(0.50)	4	56	111	0.540	127	180	0.315
		QaSIS(0.25)	4	77	151	0.475	180	232	0.225
		CQSIS(0.50)	4	45	101	0.590	71	216	0.495
		CQSIS(0.25)	4	49	148	0.560	115	240	0.370
		CRSIS	4	653	276	0.010	777	170	0.000
Example 5	(S3) 40%	CQCSIScens(0.50)	4	4	0	1.000	4	1	0.990
		CQCSIScens(0.25)	4	4	0	1.000	4	3	0.945
		CQCSISnaive(0.50)	4	88	127	0.410	170	201	0.230
		CQCSISnaive(0.25)	4	182	177	0.275	240	308	0.165
		QaSIS(0.50)	4	61	141	0.525	115	181	0.315
		QaSIS(0.25)	4	70	112	0.480	150	193	0.225
		CQSIS(0.50)	4	41	115	0.570	86	164	0.420
		CQSIS(0.25)	4	52	93	0.580	128	227	0.350
		CRSIS	4	652	254	0.000	754	237	0.000

Table 3: Results of the median of minimum model size (MMS), its robust standard deviation (RSD) and the proportion of truly active covariates selected (PS) with a pre-specified threshold size $d_n = \lfloor n/\log n \rfloor$ for Examples 4 and 5.

tion procedures, we consider one more example.

Example 6 We generate the response variable from the following model

$$Y = 2TX_1 + 5(2T-1)^2 X_2 + 3\sin(2\pi T)X_3 + \varepsilon,$$

where ε is considered from following two cases:

- Case (6a) (Normal): $\varepsilon \sim 0.8N(0,1)$;
- Case (6b) (Cauchy): $\varepsilon \sim \frac{2}{15}C(0,1)$.

The remaining of the model setup is the same as Example 1.

For Examples 6, we first conduct a variable screening procedure using the proposed CQCSIS method and then we carry out SCAD-penalized variable selection procedure. For this example, we compare the group SCAD-penalized quantile regression method with $\tau = 0.5$ stated in Section 4, abbreviated as RQSCAD(0.5), with the group SCAD-penalized mean regression method of Fan et al. (2011), abbreviated as LSSCAD.

We consider two cases for screening tuning parameter: $d_n = 10$ and 20. For each situation, we report: (C) the average of true positives (the average number of true nonzero coefficients that are correctly estimated to be nonzero), (I) the average of false positives (the average number of true zero coefficients that are incorrectly estimated to be nonzero), (CF) the proportion of correctly fitted models, (UF) the proportion of under-fitted models, (OF) the proportion of over-fitted models, and (MADE) the median of absolute deviation error (ADE) over N simulations, where ADE is defined as

$$\mathrm{ADE}(\widehat{\boldsymbol{\beta}}) = \frac{1}{n_{\mathrm{grid}}} \sum_{j=1}^{p} \sum_{l=1}^{n_{\mathrm{grid}}} \left| \widehat{\beta}_j(t_l) - \beta_j(t_l) \right|,$$

where the t_l's are the equally spaced grid points on the support of T with $n_{\mathrm{grid}} = 100$.

Tables 4 and 5 summarize the simulated results. Our method performs competitively with the LSSCAD method, and behaves much better for (6b) Cauchy error. This is anticipated since the LSSCAD is based on mean regression model and sensitive to heavy-tailed errors. We note that a larger screening threshold parameter gives higher proportion of correctly fitted models and larger number of correctly identified nonzero coefficients with a slight loss in estimation accuracy for non-vanishing varying coefficients. Obviously, the increase of sample size significantly improves the consistency of the variable selection and the accuracy of estimated functional coefficients.

Example 7 The data generating process of this example is the same as that in *Example 1*. The only difference is that we here consider four distinct combinations of dimensionality p and sample size n to see the performance affected by choosing different (n, p). The screening results are summarized in Table 6. From Table 6, we can see that in normal error setting, the NIS performs better as expected, however, in Cauchy error setting the proposed CQCSIS is more desirable than NIS. This finding is the same as before. In addition, an increase of p generally makes all the screening methods harder.

Case	Method (τ)	No. of estimated nonzeros C	I	Proportion of fitted models CF	UF	OF	MADE
			$(n, d_n, \varrho) = (200, 10, 0.4)$				
(6a)	RQSCAD(0.5)	2.86(0.00)	0.03(0.00)	0.91(0.00)	0.09(0.00)	0.01(0.00)	1.86(0.16)
	LSSCAD	2.86(0.00)	0.10(0.00)	0.89(0.00)	0.09(0.00)	0.02(0.00)	1.31(0.57)
(6b)	RQSCAD(0.5)	2.85(0.00)	0.04(0.00)	0.89(0.00)	0.10(0.00)	0.01(0.00)	0.89(0.30)
	LSSCAD	2.57(0.00)	0.60(0.75)	0.51(0.75)	0.25(0.00)	0.25(0.00)	19.68(41.84)
			$(n, d_n, \varrho) = (200, 10, 0.8)$				
(6a)	RQSCAD(0.5)	2.93(0.00)	0.00(0.00)	0.96(0.00)	0.04(0.00)	0.00(0.00)	1.13(0.21)
	LSSCAD	2.95(0.00)	0.08(0.00)	0.93(0.00)	0.03(0.00)	0.05(0.00)	1.52(0.36)
(6b)	RQSCAD(0.5)	2.88(0.00)	0.01(0.00)	0.92(0.00)	0.08(0.00)	0.01(0.00)	0.91(0.09)
	LSSCAD	2.07(1.49)	0.52(0.75)	0.30(0.75)	0.55(0.75)	0.16(0.00)	29.87(9.55)
			$(n, d_n, \varrho) = (200, 20, 0.4)$				
(6a)	RQSCAD(0.5)	2.98(0.00)	0.01(0.00)	0.98(0.00)	0.01(0.00)	0.01(0.00)	1.18(0.11)
	LSSCAD	2.99(0.00)	0.07(0.00)	0.96(0.00)	0.01(0.00)	0.04(0.00)	1.37(0.35)
(6b)	RQSCAD(0.5)	2.88(0.00)	0.01(0.00)	0.93(0.00)	0.07(0.00)	0.01(0.00)	1.14(0.25)
	LSSCAD	2.58(0.00)	1.44(1.49)	0.39(0.75)	0.25(0.00)	0.37(0.75)	79.33(43.20)
			$(n, d_n, \varrho) = (200, 20, 0.8)$				
(6a)	RQSCAD(0.5)	2.82(0.00)	0.10(0.00)	0.80(0.00)	0.19(0.00)	0.02(0.00)	1.91(4.07)
	LSSCAD	3.00(0.00)	0.13(0.00)	0.90(0.00)	0.01(0.00)	0.10(0.00)	1.39(0.27)
(6b)	RQSCAD(0.5)	2.76(0.00)	0.01(0.00)	0.84(0.00)	0.16(0.00)	0.01(0.00)	0.93(0.22)
	LSSCAD	1.98(1.49)	1.46(1.49)	0.17(0.00)	0.66(0.75)	0.17(0.00)	174.87(225.46)

Table 4: Simulated results for two-stage approaches for Example 6, where $n = 200$ and robust standard deviation is given in parenthesis

Case	Method (τ)	No. of estimated nonzeros C	I	Proportion of fitted models CF	UF	OF	MADE
			$(n, d_n, \varrho) = (400, 10, 0.4)$				
(6a)	RQSCAD(0.5)	3.00(0.00)	0.01(0.00)	1.00(0.00)	0.00(0.00)	0.01(0.00)	0.90(0.06)
	LSSCAD	3.00(0.00)	0.02(0.00)	0.99(0.00)	0.00(0.00)	0.02(0.00)	0.95(0.08)
(6b)	RQSCAD(0.5)	2.96(0.00)	0.06(0.00)	0.97(0.00)	0.02(0.00)	0.02(0.00)	1.02(0.17)
	LSSCAD	2.58(0.00)	0.51(0.75)	0.56(0.75)	0.21(0.00)	0.24(0.00)	9.67(8.96)
			$(n, d_n, \varrho) = (400, 10, 0.8)$				
(6a)	RQSCAD(0.5)	3.00(0.00)	0.01(0.00)	1.00(0.00)	0.00(0.00)	0.01(0.00)	1.01(0.11)
	LSSCAD	3.00(0.00)	0.01(0.00)	0.99(0.00)	0.00(0.00)	0.01(0.00)	1.03(0.10)
(6b)	RQSCAD(0.5)	2.96(0.00)	0.03(0.00)	0.98(0.00)	0.02(0.00)	0.01(0.00)	1.34(0.05)
	LSSCAD	2.05(1.49)	0.37(0.75)	0.35(0.75)	0.54(0.75)	0.12(0.00)	27.88(21.37)
			$(n, d_n, \varrho) = (400, 20, 0.4)$				
(6a)	RQSCAD(0.5)	3.00(0.00)	0.00(0.00)	1.00(0.00)	0.00(0.00)	0.00(0.00)	0.89(0.05)
	LSSCAD	3.00(0.00)	0.06(0.00)	0.96(0.00)	0.00(0.00)	0.04(0.00)	1.15(0.14)
(6b)	RQSCAD(0.5)	2.98(0.00)	0.01(0.00)	0.99(0.00)	0.01(0.00)	0.01(0.00)	0.78(0.03)
	LSSCAD	2.63(0.00)	1.12(1.49)	0.48(0.75)	0.19(0.00)	0.34(0.75)	50.15(12.19)
			$(n, d_n, \varrho) = (400, 20, 0.8)$				
(6a)	RQSCAD(0.5)	3.00(0.00)	0.00(0.00)	1.00(0.00)	0.00(0.00)	0.00(0.00)	0.99(0.12)
	LSSCAD	3.00(0.00)	0.06(0.00)	0.96(0.00)	0.00(0.00)	0.05(0.00)	1.20(0.20)
(6b)	RQSCAD(0.5)	2.94(0.00)	0.01(0.00)	0.96(0.00)	0.04(0.00)	0.01(0.00)	0.81(0.05)
	LSSCAD	2.08(0.93)	1.22(1.49)	0.21(0.00)	0.59(0.75)	0.21(0.00)	44.70(46.51)

Table 5: Simulated results for two-stage approaches for Example 6, where $n = 400$ and robust standard deviation is given in parenthesis

Case		MMS	RSD	PS	MMS	RSD	PS
		$(n, p) = (100, 1000)$			$(n, p) = (100, 2000)$		
(1a)	CQCIS(0.5)	11.5	8.2	0.800	19	11.2	0.595
	CQCIS(0.75)	13	9.9	0.735	20.5	12.9	0.520
	NIS	3	0.0	1.000	3	0.0	1.000
(1b)	CQCIS(0.5)	11	7.5	0.815	21	11.2	0.530
	CQCIS(0.75)	14	9.7	0.735	26.5	24.3	0.395
	NIS	3	31.5	0.730	3	82.6	0.665
(1c)	CQCIS(0.5)	66.5	121.3	0.145	169	383.6	0.020
	CQCIS(0.75)	14	10.1	0.735	27	23.5	0.385
	NIS	8	21.1	0.715	13	36.8	0.590
(1d)	CQCIS(0.5)	13	7.8	0.790	20	14.4	0.570
	CQCIS(0.75)	68.5	136.4	0.180	154.5	251.3	0.060
	NIS	12	84.5	0.575	15.5	339.0	0.535
		$(n, p) = (200, 1000)$			$(n, p) = (200, 5000)$		
(1a)	CQCIS(0.5)	4	1.5	0.990	8	4.5	0.965
	CQCIS(0.75)	4	1.5	1.000	8	4.5	0.980
	NIS	3	0.0	1.000	3	0.0	1.000
(1b)	CQCIS(0.5)	4	1.5	0.995	8	4.7	0.985
	CQCIS(0.75)	4	1.5	0.990	8.5	5.2	0.950
	NIS	3	41.4	0.735	3	125.0	0.660
(1c)	CQCIS(0.5)	13	14.4	0.810	40.5	78.4	0.480
	CQCIS(0.75)	4	1.5	1.000	7.5	4.5	0.980
	NIS	5	0.0	0.980	5	0.7	0.925
(1d)	CQCIS(0.5)	4	1.5	0.995	8	6.0	0.970
	CQCIS(0.75)	11	9.7	0.875	33	48.5	0.570
	NIS	5	45.1	0.710	6	134.7	0.675

Table 6: Simulation results for the Example 1 with $\rho = 0.8$ and different combinations of (n, p).

Example 8 As pointed out by one referee, how CQCSIS can be affected if τ varies. To answer this question, in this example we revisit *Example 3* with a range of τ and $(t_1, t_2) = (0, 0)$. Specifically, we take nine different values of τ $(= 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)$ to see the performance of CQCSIS at various quantile levels. In addition, we also consider a composite CQCSIS, denoted as CCQCSIS below, that uses the average of CQC at a grid of these given quantile levels as the screening utility as in Ma and Zhang (2016), which is defined by

$$u_j = E\{E[\rho^2_{\tau,j}(T)]\} = \int_0^1 E[\rho^2_{\tau,j}(T)]\mathrm{d}\tau.$$

An empirical utility is defined by $\widehat{u}_j = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{n}\sum_{i=1}^{n}\widehat{\rho}^2_{\tau_k,j}(T_i)$ with $\tau_k = \frac{k-0.5}{N}$ for a large N, for example, we take $N = 100$ in the simulation. Table 7 reports the results, which indicate that CQCSIS performs satisfactorily across quantile levels in the middle range of the overall interval, while a bit poor at quite low and high quantile levels. Also we can see that the performance of CCQCSIS is quite appealing. Theoretically speaking, the screening property of CCQCSIS can be similarly proved without much difficulty.

Example 9 This example is designed as same as *Example 4* with censoring rate of 20%. We here consider four cases of $(n, p) = (100, 1000)$, $(100, 2000)$, $(200, 2000)$, $(200, 5000)$. The results are summarized in Tables 8 and 9. We can make a similar observation as in *Example 7*

	Normal Error			Cauchy Error		
	MMS	RSD	PS	MMS	RSD	PS
CQCSIS(0.1)	37.0	119.2	0.59	195.0	267.7	0.22
CQCSIS(0.2)	7.0	7.6	0.91	15.0	30.8	0.81
CQCSIS(0.3)	5.0	1.5	1.00	6.5	6.7	0.94
CQCSIS(0.4)	5.0	0.7	1.00	6.0	3.7	0.99
CQCSIS(0.5)	5.0	1.5	0.99	6.0	3.0	0.97
CQCSIS(0.6)	5.0	1.5	0.99	7.0	5.2	0.94
CQCSIS(0.7)	5.0	1.5	1.00	6.0	6.9	0.92
CQCSIS(0.8)	6.0	2.4	0.98	14.0	23.9	0.85
CQCSIS(0.9)	12.0	20.7	0.84	89.5	166.6	0.41
CCQCSIS	5.0	1.5	0.97	6.0	2.2	0.93
NIS	5.0	0.7	1.00	432.5	382.1	0.14

Table 7: Simulated results for Example 8.

	MMS	RSD	PS	MMS	RSD	PS
S1 censoring with Normal Error case						
	$(n, p) = (100, 1000)$			$(n, p) = (100, 2000)$		
CQCISCens(0.5)	142	157	0.015	310	373	0.000
CQCISCens(0.25)	238	246	0.000	427	420	0.000
CQCISNaive(0.5)	277	200	0.005	629	533	0.000
CQCISNaive(0.25)	523	298	0.000	984	546	0.000
QaSISCens(0.5)	277	252	0.005	546	566	0.000
QaSISCens(0.25)	319	296	0.000	635	641	0.000
CQSISCens(0.5)	450	375	0.020	731	735	0.005
CQSISCens(0.25)	415	347	0.010	837	621	0.000
CRSIS	711	240	0.000	1407	503	0.000
	$(n, p) = (200, 2000)$			$(n, p) = (200, 5000)$		
CQCISCens(0.5)	24	32	0.645	56	73	0.360
CQCISCens(0.25)	40	51	0.480	82	157	0.255
CQCISNaive(0.5)	103	151	0.200	223	347	0.070
CQCISNaive(0.25)	459	500	0.025	836	1163	0.010
QaSISCens(0.5)	288	508	0.135	638	1084	0.015
QaSISCens(0.25)	375	552	0.065	699	1003	0.010
CQSISCens(0.5)	254	423	0.190	566	1188	0.060
CQSISCens(0.25)	318	520	0.135	607	1165	0.060
CRSIS	1246	457	0.000	3389	1116	0.000

Table 8: Simulation results for the Example 9 with different combinations of (n, p).

	MMS	RSD	PS	MMS	RSD	PS
S1 censoring with Cauchy Error case						
	$(n,p) = (100, 1000)$			$(n,p) = (100, 2000)$		
CQCISCens(0.5)	207	205	0.005	517	471	0.000
CQCISCens(0.25)	379	272	0.000	804	594	0.000
CQCISNaive(0.5)	431	289	0.000	778	611	0.000
CQCISNaive(0.25)	583	271	0.000	1137	613	0.000
QaSISCens(0.5)	421	262	0.000	766	605	0.000
QaSISCens(0.25)	412	340	0.000	933	520	0.000
CQSISCens(0.5)	498	384	0.000	1039	703	0.000
CQSISCens(0.25)	555	350	0.005	1134	641	0.000
CRSIS	780	190	0.000	1581	385	0.000
	$(n,p) = (200, 2000)$			$(n,p) = (200, 5000)$		
CQCISCens(0.5)	64	118	0.370	147	265.1	0.145
CQCISCens(0.25)	161	174	0.105	352	474.3	0.035
CQCISNaive(0.5)	259	299	0.065	587	813.8	0.010
CQCISNaive(0.25)	643	397	0.010	1579	1465.3	0.000
QaSISCens(0.5)	433	467	0.010	940	1145.9	0.000
QaSISCens(0.25)	525	574	0.005	1335	1549.8	0.000
CQSISCens(0.5)	500	569	0.045	1297.5	1776.9	0.020
CQSISCens(0.25)	646	646	0.010	1738	1743.5	0.000
CRSIS	1537	373	0.000	3900	1047.4	0.000

Table 9: Simulation results for the Example 9 with different combinations of (n, p).

Appendix: proofs

Before proving the main results, we introduce some notation first. For $i = 1, \ldots, n$ and $j = 1, \ldots, p$, we denote $f_{1,ij} = I(Y_i - Q_{\tau,Y} > 0)X_{ij}$, $f_{2,ij} = I(Y_i - Q_{\tau,Y} > 0)$, $f_{3,ij} = X_{ij}^2$, $f_{4,ij} = X_{ij}$, $f_{1w,ij} = w_i(F)I(\widetilde{Y}_i < Q_{\tau,Y})X_{ij}$, $f_{2w,ij} = w_i(F)I(\widetilde{Y}_i < Q_{\tau,Y})$ and $f_{3w,ij} = w_i^2(F)I(\widetilde{Y}_i < Q_{\tau,Y})$.

Appendix A: Some Properties of B-spline

According to the properties of normalized B-splines, we have that for each $j = 1, \cdots, p$ and $k = 1, \ldots, L_n$, (i) $B_k(t) \geq 0$ and $\sum_{k=1}^{L_n} B_k(t) = 1$ for $t \in \mathcal{T}$; (ii) there exist positive constants C_1, C_2 such that for any $\mathbf{a} \in \mathrm{R}^{L_n}$,

$$\frac{C_1}{L_n}\mathbf{a}'\mathbf{a} \leq \int \mathbf{a}'\mathbf{B}(t)(\mathbf{B}(t))'\mathbf{a}\mathrm{d}t \leq \frac{C_2}{L_n}\mathbf{a}'\mathbf{a}.$$

Under property (ii), we have that: (iii) there exist positive constants C_3 and C_4 such that for $k = 1, \ldots, L_n$,

$$C_3 L_n^{-1} \leq \mathbb{E}\{(B_k(T))^2\} \leq C_4 L_n^{-1},$$

where $C_3 = C_1 M_1$ and $C_4 = C_2 M_2$; and (iv)

$$C_3 L_n^{-1} \leq \lambda_{\min}\big(\mathbb{E}\{\mathbf{B}(T)(\mathbf{B}(T))'\}\big) \leq \lambda_{\max}\big(\mathbb{E}\{\mathbf{B}(T)(\mathbf{B}(T))'\}\big) \leq C_4 L_n^{-1},$$

Appendix B: Some Technical Lemmas

Lemma 1. *Suppose that random variable X has a conditional exponential tail $P(|X| > x|T) \leq K_1 \exp(-K_2^{-1}x)$ for positive constants K_1, K_2, uniformly on the support of T. Then, for all $r \geq 2$, $\mathbb{E}(|X|^r|T) \leq K_1 K_2^r r!$.*

Proof of Lemma 1 By the condition on X and a change of variable, we have

$$\begin{aligned}\mathbb{E}\{|X|^r|T\} &= \int_0^\infty P(|X| > x^{1/r})\mathrm{d}x \leq K_1 \int_0^\infty \exp(-K_2^{-1}x^{1/r})\mathrm{d}x \\ &= rK_1K_2^r \int_0^\infty \exp(-t)t^{r-1}\mathrm{d}t = rK_1K_2^r\Gamma(r) = K_1K_2^r r!.\end{aligned}$$

Hence, the lemma follows. □

According to this lemma, it is easily seen that $\mathbb{E}\{|X|^2|T\}$ is finite.

Lemma 2. *(Bernstein's inequality, Lemma 2.2.11, van der Vaart and Wellner, 1996) For independent random variables* $Y_1, \ldots, Y_n$ *with mean zero and* $\mathbb{E}\{|Y_i|^r\} \leq r!K^{r-2}v_i/2$ *for every* $r \geq 2$, $i = 1, \ldots, n$ *and some constants* K, v_i*. Then, for* $x > 0$*, we have*

$$P(|Y_1 + \cdots + Y_n| > x) \leq 2\exp(-x^2/(2(v + Kx))),$$

for $v \geq \sum_{i=1}^n v_i$*.*

Lemma 3. *(Bernstein's inequality, Lemma 2.2.9, van der Vaart and Wellner, 1996) For independent random variables* $Y_1, \ldots, Y_n$ *with mean zero and bounded range* $[-M, M]$*, then*

$$P(|Y_1 + \cdots + Y_n| > x) \leq 2\exp(-x^2/(2(v + Mx/3))),$$

for $v \geq \mathrm{Var}(Y_i + \cdots + Y_n)$*.*

Lemma 4. *(Hoeffding's inequality, Lemma 14.11, Bühlmann and van de Geer, 2011) For independent random variables* $Y_1, \ldots, Y_n$ *with mean zero such that* $P(Y_i \in [a_i, b_i]) =$

1 *for some* a_i *and* b_i *for all* $i = 1, \ldots, n$. *Then, for* $x > 0$, *we have*

$$P(|Y_1 + \cdots + Y_n| > x) \leq 2\exp\bigg(- \frac{x^2}{\sum_{i=1}^n (b_i - a_i)^2} \bigg).$$

Lemma 5. *Suppose that conditions (C1)-(C3) are satisfied. For any* $\delta > 0$, *there exist some positive constants* $c_1, \ldots, c_8$ *and* $\tilde{c}_1, \ldots, \tilde{c}_6$ *such that for* $j = 1, \ldots, p$, *and* $k = 1, \ldots, L_n$,

$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{1,ij} - \mathbb{E}\{B_k(T) m_{1j}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{c_1 L_n^{-1} n + c_2 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{2,ij} - \mathbb{E}\{B_k(T) m_{2j}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{c_3 L_n^{-1} n + c_4 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{3,ij} - \mathbb{E}\{B_k(T) m_{3j}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{c_5 L_n^{-1} n + c_6 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{4,ij} - \mathbb{E}\{B_k(T) m_{4j}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{c_7 L_n^{-1} n + c_8 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{1w,ij} - \mathbb{E}\{B_k(T) m_{1j,w}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{\tilde{c}_1 L_n^{-1} n + \tilde{c}_2 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{2w,ij} - \mathbb{E}\{B_k(T) m_{2j,w}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{\tilde{c}_3 L_n^{-1} n + \tilde{c}_4 \delta} \bigg),$$
$$P\bigg(\bigg| \frac{1}{n} \sum_{i=1}^n B_k(T_i) f_{3w,ij} - \mathbb{E}\{B_k(T) m_{3j,w}(T)\} \bigg| \geq \frac{\delta}{n} \bigg) \leq 2\exp\bigg(- \frac{\delta^2}{\tilde{c}_5 L_n^{-1} n + \tilde{c}_6 \delta} \bigg).$$

Proof of Lemma 5 We first prove the inequality on $f_{1,ij}$. Denote $W_{ijk} = B_k(T_i) f_{1,ij} -$

$\mathbb{E}\{B_k(T)m_{1j}(T)\}$. Then for every $r > 2$,

$$\begin{aligned} \mathbb{E}\{|W_{ijk}|^r\} &\leq 2^r\mathbb{E}\{|B_k(T_i)I(Y_i - Q_{\tau,Y} > 0)X_{ij}|^r\} \\ &\leq 2^r\mathbb{E}\{B_k^r(T_i)\mathbb{E}[|X_{ij}|^r|T]\} \leq 2^r\mathbb{E}\{B_k^2(T_i)K_1K_2^r r!\} \\ &\leq 8C_4L_n^{-1}K_1K_2^2(2K_2)^{r-2}r!/2, \end{aligned}$$

where the first inequality is due to the C_r inequality, the third inequality follows by Lemma 1, and the last line uses the properties of B-spline basis. Thus, an application of Lemma 2 yields that

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} W_{ijk}\right| \geq \frac{\delta}{n}\right) \leq 2\exp\left(-\frac{\delta^2}{2(8C_4K_1K_2^2L_n^{-1}n + 2K_2\delta)}\right).$$

Hence, the first result is proved by letting $c_1 = 16C_4K_1K_2^2$ and $c_2 = 4K_2$.

For the second inequality, write $U_{ijk} = B_k(T_i)f_{2,ij} - \mathbb{E}\{B_k(T)m_{2j}(T)\}$. Clearly, it follows from the boundedness of the indicator function that $\max_{i,j,k}|U_{ijk}| \leq 2$ and $\mathrm{Var}(U_{ijk}) \leq \mathbb{E}(|B_k(T_i)f_{2,ij}|^2) \leq C_4L_n^{-1}$ uniformly in i, j, k. So applying Lemma 3 yields

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} U_{ijk}\right| \geq \frac{\delta}{n}\right) \leq 2\exp\left(-\frac{\delta^2}{c_3L_n^{-1}n + c_4\delta}\right),$$

where $c_3 = 2C_4$ and $c_4 = 4/3$.

Similarly, noting that $0 \leq w(F) \leq 1$, we can prove the remaining inequalities of the lemma. Thus, the lemma follows. $\square$

Lemma 6. *Suppose that conditions (C1)-(C3) are satisfied. For any* $\delta > 0$*, there exist some positive constants* c_9, c_{10} *such that*

$$P\left(\left\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{B}(T_i)(\mathbf{B}(T_i))' - \mathbb{E}\{\mathbf{B}(T)(\mathbf{B}(T))'\}\right\| \geq \frac{L_n\delta}{n}\right) \leq 6L_n^2 \exp\left(-\frac{\delta^2}{c_9 L_n^{-1} n + c_{10}\delta}\right),$$

and for any positive constant h_1*, there exists some positive constant* h_2 *such that*

$$P\left(\left\|\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{B}(T_i)(\mathbf{B}(T_i))'\right)^{-1}\right\| \geq (1+h_1)\big\|(\mathbb{E}\{\mathbf{B}(T)(\mathbf{B}(T))'\})^{-1}\big\|\right) \leq 6L_n^2 \exp\big(-h_2 L_n^{-3} n\big),$$

where $\|\cdot\|$ *is the operator norm, that is,* $\|\mathbf{A}\| = \sqrt{\lambda_{\max}(\mathbf{A}'\mathbf{A})}$ *for any matrix* $\mathbf{A}$*.*

The above lemma is adapted from Lemma 7 of Fan, Ma and Dai (2014).

Lemma 7. *Suppose that* $g_1(t), g_2(t)$ *are two functions of* t *such that for* $N_1 > 0, N_2 > 0$ *such that* $\sup_{t\in\mathcal{T}} |g_1(t)| \leq N_1, \sup_{t\in\mathcal{T}} |g_2(t)| \leq N_2$*. For a given* $t \in \mathcal{T}$*,* $\widehat{G}_1(t)$ *and* $\widehat{G}_2(t)$ *are estimators of* $g_1(t)$ *and* $g_2(t)$ *based on a sample with size* n*. Let* $p_{n1}(x)$ *and* $p_{n2}(x)$ *be two functions of* x*, both of which consist of some exponential forms that might depend on some constants and both are less than 1. If the following two conditions hold for any* $\epsilon \in (0,1)$*:*

$$\sup_{t\in\mathcal{T}} P\big(|\widehat{G}_1(t) - g_1(t)| \geq \epsilon\big) \leq p_{n1}(\epsilon),$$
$$\sup_{t\in\mathcal{T}} P\big(|\widehat{G}_2(t) - g_2(t)| \geq \epsilon\big) \leq p_{n2}(\epsilon).$$

Then we have

$$\sup_{t\in\mathcal{T}} P\big(|\widehat{G}_1(t)\widehat{G}_2(t) - g_1(t)g_2(t)| \geq \epsilon\big) \leq p_{n1}(\epsilon) + p_{n1}\Big(\frac{\epsilon}{2N_2}\Big) + p_{n2}\Big(\frac{\epsilon}{2(N_1+1)}\Big),$$
$$\sup_{t\in\mathcal{T}} P\big(|(\widehat{G}_1(t))^2 - (g_1(t))^2| \geq \epsilon\big) \leq p_{n1}(\epsilon) + p_{n1}\Big(\frac{\epsilon}{2(N_1+1)}\Big) + p_{n1}\Big(\frac{\epsilon}{2N_1}\Big),$$
$$\sup_{t\in\mathcal{T}} P\big(|\{\widehat{G}_1(t) - \widehat{G}_2(t)\} - \{g_1(t) - g_2(t)\}| \geq \epsilon\big) \leq p_{n1}(\epsilon/2) + p_{n2}(\epsilon/2).$$

Furthermore, suppose that $g_2(u)$ *is uniformly bounded away from zero, in other words, there exists* $N_3 > 0$ *such that* $\inf_{t\in\mathcal{T}} |g_2(t)| > N_3$*, then*

$$\sup_{t\in\mathcal{T}} P\left(\left|\frac{\widehat{G}_1(t)}{\widehat{G}_2(t)} - \frac{g_1(t)}{g_2(t)}\right| \geq \epsilon\right) \leq p_{n1}\Big(\frac{N_4}{2}\epsilon\Big) + 2p_{n2}(\epsilon) + p_{n2}\Big(\frac{N_3N_4}{2N_1}\epsilon\Big),$$

where $N_4 = N_3 - \min(N_3/2, 1)$*. In addition, if we further assume that* $g_2(t) > 0$*, then*

$$\sup_{t\in\mathcal{T}} P\left(\left|\sqrt{\widehat{G}_2(t)} - \sqrt{g_2(t)}\right| \geq \epsilon\right) \leq p_{n2}(\epsilon) + p_{n2}\big((\sqrt{N_4} + \sqrt{N_3})\epsilon\big).$$

Proof of Lemma 7 We use the technique in the proof of Lemma 4 in Liu, Li and Wu (2014) to prove the above results. We only outline the details. Note that, using the inequality that $|x - y| \geq |x| - |y|$, we have

$$\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)\big| \geq N_1 + \epsilon\big) \leq \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t) - g_1(t)\big| \geq \epsilon\big) \leq p_{n1}(\epsilon).$$

It follows that

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)\widehat{G}_2(t)-g_1(t)g_2(t)\big|\ge\epsilon\big)\\
&\le \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)\big|\big|\widehat{G}_2(t)-g_2(t)\big|\ge\epsilon/2\big)+\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)-g_1(t)\big|\big|g_2(t)\big|\ge\epsilon/2\big)\\
&\le \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)\big|\ge N_1+\epsilon\big)+\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)-g_2(t)\big|\ge\epsilon/(2(N_1+1))\big)\\
&\qquad+\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t)-g_1(t)\big|\ge\epsilon/(2N_2)\big)\\
&\le p_{n1}(\epsilon)+p_{n2}\Big(\frac{\epsilon}{2(N_1+1)}\Big)+p_{n1}\Big(\frac{\epsilon}{2N_2}\Big).
\end{aligned}$$

Similarly, we can obtain the result on the square term. Because $P(|\xi-\zeta|>x)\le P(|\xi|>x/2)+P(|\zeta|>x/2)$ for any random variables ξ and ζ, so the result for the subtraction holds.

Next, we consider the term on the division. Since

$$\begin{aligned}
\left|\frac{\widehat{G}_1(t)}{\widehat{G}_2(t)}-\frac{g_1(t)}{g_2(t)}\right| &= \frac{\big|\{\widehat{G}_{1j}(t)-g_1(t)\}g_2(t)-g_1(t)\{\widehat{G}_2(t)-g_2(t)\}\big|}{\big|\widehat{G}_2(t)g_2(t)\big|}\\
&\le \frac{\big|\{\widehat{G}_{1j}(t)-g_1(t)\}g_2(t)\big|}{\big|\widehat{G}_2(t)g_2(t)\big|}+\frac{\big|g_1(t)\{\widehat{G}_2(t)-g_2(t)\}\big|}{\big|\widehat{G}_2(t)g_2(t)\big|},
\end{aligned}$$

we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} P\left(\left|\frac{\widehat{G}_1(t)}{\widehat{G}_2(t)}-\frac{g_1(t)}{g_2(t)}\right|\ge\epsilon\right)\\
&\le \sup_{t\in\mathcal{T}} P\left(\frac{\big|\widehat{G}_{1j}(t)-g_1(t)\big|}{\big|\widehat{G}_2(t)\big|}\ge\epsilon/2\right)+\sup_{t\in\mathcal{T}} P\left(\frac{\big|g_1(t)\{\widehat{G}_2(t)-g_2(t)\}\big|}{\big|\widehat{G}_2(t)g_2(t)\big|}\ge\epsilon/2\right). \qquad (S0.1)
\end{aligned}$$

Notice that for $t \in \mathcal{T}$,

$$\big\{\big|\widehat{G}_2(t) - g_2(t)\big| \geq \epsilon\big\} \supseteq \big\{\big|\widehat{G}_2(t)\big| \leq \big|g_2(t)\big| - \epsilon\big\} \supseteq \big\{\big|\widehat{G}_2(t)\big| \leq N_3 - \epsilon\big\},$$

which implies that

$$\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_3 - \epsilon\big) \leq \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t) - g_2(t)\big| \geq \epsilon\big) \leq p_{n2}(\epsilon).$$

Hence, letting $N_4 = N_3 - \min(N_2/2, 1)$, in other words, taking $\epsilon = \min(N_2/2, 1)$ on the left hand side of the last inequality, we have $\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_4\big) \leq p_{n2}(\epsilon)$. Thus, it follows that the first term on the right hand side of inequality (S0.1) is bounded by

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t) - g_1(t)\big| \geq \epsilon/2\big|\widehat{G}_2(t)\big|, \big|\widehat{G}_2(t)\big| > N_4\big) + \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_4\big)\\
&\leq \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_1(t) - g_1(t)\big| \geq N_4\epsilon/2\big) + \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_4\big) \leq p_{n1}\Big(\frac{N_4}{2}\epsilon\Big) + p_{n2}(\epsilon),
\end{aligned}$$

and the first term on the right hand side of inequality (S0.1)

$$\begin{aligned}
&\leq \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t) - g_2(t)\ \big| \geq N_3\epsilon/(2N_1)\big|\widehat{G}_2(t)\big|, \big|\widehat{G}_2(t)\big| > N_4\big) + \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_4\big)\\
&\leq \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t) - g_2(t)\big| \geq N_3N_4\epsilon/(2N_1)\big) + \sup_{t\in\mathcal{T}} P\big(\big|\widehat{G}_2(t)\big| \leq N_4\big)\\
&\leq p_{n2}\Big(\frac{N_3N_4}{2N_1}\epsilon\Big) + p_{n2}(\epsilon).
\end{aligned}$$

Therefore, this, together with inequality (S0.1), concludes the result.

Last, for the root term, it can be similarly derived by observing the fact that $\sqrt{A} - \sqrt{a} = (A-a)/(\sqrt{A}+\sqrt{a})$. Thus, the lemma follows. $\square$

In what follows, we are going to derive some exponential bounds for $\widehat{m}_{kj}(t), k = 1,2,3,4$.

Lemma 8. *Under conditions of Theorem 2.1, for any $\epsilon > 0$ such that $L_n^{-1/2+d}\epsilon \to \infty$ and $\epsilon L_n^{-3/2} n^{\iota} \to \infty$ with any $0 < \iota < 1/2$ as $n \to \infty$, then we have*

$$\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{1j}(t) - m_{1j}(t)\big| > \epsilon\Big) \le 18L_n^2 \exp(-h_2 L_n^{-3} n) + (4L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{15}L_n^{-1}n + c_{16}\delta}\right)$$
$$\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{2j}(t) - m_{2j}(t)\big| > \epsilon\Big) \le 18L_n^2 \exp(-h_2 L_n^{-3} n) + (4L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{17}L_n^{-1}n + c_{18}\delta}\right)$$
$$\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{3j}(t) - m_{3j}(t)\big| > \epsilon\Big) \le 12L_n^2 \exp(-h_2 L_n^{-3} n) + (2L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{19}L_n^{-1}n + c_{20}\delta}\right)$$
$$\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{4j}(t) - m_{4j}(t)\big| > \epsilon\Big) \le 12L_n^2 \exp(-h_2 L_n^{-3} n) + (2L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{21}L_n^{-1}n + c_{22}\delta}\right)$$

where $\delta \asymp n\epsilon L_n^{-5/2}$, and $c_{15}, \ldots c_{22}$ are some positive constants.

Proof of Lemma 8 We only give the proof of the first result because the rest can be proved in a similar manner. Note that $\widehat{m}_{1j}(t) = (\mathbf{B}(t))'(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{f}_{1j}$. By the approximation theory of B-spline, we know that there exists a $\boldsymbol{\gamma}_j^* \in \mathrm{R}^{L_n}$ such that $m_{1j}(t) = (\mathbf{B}(t))'\boldsymbol{\gamma}_j^* + \eta_j(t)$, where $\eta_j(t)$ is the approximation error satisfying

$$\sup_{1\le j\le p}\sup_{t\in\mathcal{T}} |\eta_j(t)| \le C_5 L_n^{-d}$$

for some positive constant C_5.

For the sake of simplicity, let $\widehat{f}_{1i}(Y, X_j)$ be the ith component of $\mathbf{f}_{1j}$, i.e. $\widehat{f}_{1i}(Y, X_j) = I(Y_i - \widehat{Q}_{\tau,Y} > 0)X_{ij}$ and denote $\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i)(\mathbf{B}(T_i))'$, $\mathbf{E}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i)[\widehat{f}_{1i}(Y, X_j) - f_{1,ij}]$, $\mathbf{D}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i) f_{1,ij}$, $\mathbf{A} = \mathbb{E}\big\{\mathbf{B}(T)(\mathbf{B}(T))'\big\}$ and $\mathbf{D} = \mathbb{E}\big\{\mathbf{B}(T)m_{1j}(T)\big\}$. Define

$$\begin{aligned}
\boldsymbol{\gamma}_j^M &= \arg\min_{\boldsymbol{\gamma}_j} \mathbb{E}\big\{[m_{1j}(T) - (\mathbf{B}(T))'\boldsymbol{\gamma}_j]^2\big\}, \\
\widehat{\boldsymbol{\gamma}}_j &= \arg\min_{\boldsymbol{\gamma}_j} \frac{1}{n}\sum_{i=1}^n \big[\widehat{f}_{1i}(Y, X_j) - (\mathbf{B}(T_i))'\boldsymbol{\gamma}_j\big]^2.
\end{aligned}$$

Then, $\boldsymbol{\gamma}_j^M = \mathbf{A}^{-1}\mathbf{D}$ and $\widehat{\boldsymbol{\gamma}}_j = \mathbf{A}_n^{-1}(\mathbf{E}_n + \mathbf{D}_n)$. Thus, we have

$$\begin{aligned}
\sup_{t\in\mathcal{T}} \big|\widehat{m}_{1j}(t) - m_{1j}(t)\big| &\le \sup_{t\in\mathcal{T}} \big|(\mathbf{B}(t))'(\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j^*)\big| + \sup_{t\in\mathcal{T}} \big|\eta_j(t)\big| \\
&\le \big\|\widehat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j^*\big\| + \sup_{t\in\mathcal{T}} \big|\eta_j(t)\big| \\
&\le \big\|\mathbf{A}_n^{-1}\mathbf{E}_n\big\| + \big\|\mathbf{A}_n^{-1}\mathbf{D}_n - \mathbf{A}^{-1}\mathbf{D}\big\| \\
&\quad + \big\|\boldsymbol{\gamma}_j^M - \boldsymbol{\gamma}_j^*\big\| + \sup_{t\in\mathcal{T}} \big|\eta_j(t)\big| \\
&\hat{=} I_{n1} + I_{n2} + I_{n3} + I_{n4} \quad \text{(say)},
\end{aligned}$$

where the third line uses the following fact that

$$\sup_{t\in\mathcal{T}} \big\|\mathbf{B}(t)\big\| = \sup_{t\in\mathcal{T}} \sqrt{\sum_{k=1}^{L_n} (B_k(t))^2} \le \sup_{t\in\mathcal{T}} \sqrt{\sum_{k=1}^{L_n} |B_k(t)| \|B_k\|_\infty} \le \sup_{t\in\mathcal{T}} \sqrt{\sum_{k=1}^{L_n} B_k(t)} \le 1.$$

On the one hand, we can derive that

$$
\begin{aligned}
I_{n3} &= \big\|\boldsymbol{\gamma}_j^M - \boldsymbol{\gamma}_j^*\big\| = \Big\|\big(\mathbb{E}\{\mathbf{B}(T)(\mathbf{B}(T))'\}\big)^{-1}\mathbb{E}\{\mathbf{B}(T)\eta_j(T)\}\Big\| \\
&\le C_3^{-1}L_n\big\|\mathbb{E}\{\mathbf{B}(T)\big\|\max_j\sup_{t\in\mathcal{T}}\big|\eta_j(T)\}\big| \\
&\le C_3^{-1}L_nC_5L_n^{-d}\big\|\mathbb{E}\{\mathbf{B}(T)\}\big\| \le C_6L_n^{1/2-d},
\end{aligned}
$$

where the last inequality is because $\mathbb{E}\{B_k(T)\} \le C_0L_n^{-1}$ for some constant C_0, and $C_6 = C_3^{-1}C_5C_0$. On the other hand, it is easily shown that $I_{n4} \le C_5L_n^{-d}$. Therefore, by setting $\epsilon/2 > C_6L_n^{1/2-d} + C_5L_n^{-d}$, it follows that

$$
\begin{aligned}
\sup_{t\in\mathcal{T}}P\big(\big|\widehat{m}_{1j}(t) - m_{1j}(t)\big| > \epsilon\big) &\le P\bigg(\sup_{t\in\mathcal{T}}\big|\widehat{m}_{1j}(t) - m_{1j}(t)\big| > \epsilon\bigg) \\
&\le P\big(\big|I_{n1} + I_{n2} + I_{n3} + I_{n4}\big| > \epsilon\big) \\
&\le P\big(\big|I_{n1}\big| + \big|I_{n2}\big| > \epsilon/2\big). \qquad \text{(S0.2)}
\end{aligned}
$$

Next, we first consider the term for I_{n2} in inequality (S0.2). Note that

$$
\begin{aligned}
I_{n2} &= \big\|\mathbf{A}_n^{-1}(\mathbf{D}_n - \mathbf{D}) + \mathbf{A}_n^{-1}(\mathbf{A} - \mathbf{A}_n)\mathbf{A}^{-1}\mathbf{D}\big\| \\
&\le \big\|\mathbf{A}_n^{-1}(\mathbf{D}_n - \mathbf{D})\big\| + \big\|\mathbf{A}_n^{-1}(\mathbf{A} - \mathbf{A}_n)\mathbf{A}^{-1}\mathbf{D}\big\| \\
&\hat{=} I_{n2}^{(1)} + I_{n2}^{(2)} \quad \text{(say)}. \qquad \text{(S0.3)}
\end{aligned}
$$

By Lemma 1, we have

$$
\begin{aligned}
\big|\mathbb{E}\{B_k(T)m_{1j}(T)\}\big| &\leq \big|\mathbb{E}\{B_k(T)[\mathbb{E}(|f_{1j}(Y,X_j)|^2|T)]^{1/2}\}\big| \\
&\leq \big|\mathbb{E}\{B_k(T)[\mathbb{E}(X_j^2|T)]^{1/2}\}\big| \\
&\leq (2K_1K_2^2)^{1/2}\mathbb{E}\{B_k(T)\} \leq C_7L_n^{-1},
\end{aligned}
$$

where $C_7 = (2K_1K_2^2)^{1/2}C_0$. Thus, it follows that

$$
\|\mathbf{D}\| = \left(\sum_{k=1}^{L_n}\big[\mathbb{E}\{B_k(T)m_{1j}(T)\}\big]^2\right)^{1/2} \leq C_7L_n^{-1/2}. \tag{S0.4}
$$

By Lemma 5, we have, for $\delta > 0$,

$$
\begin{aligned}
P\bigg(\big\|\mathbf{D}_n - \mathbf{D}\big\| \geq \frac{L_n^{1/2}\delta}{n}\bigg) &\leq \sum_{k=1}^{L_n} P\bigg(\bigg|\frac{1}{n}\sum_{i=1}^{n} B_k(T_i)f_{1,ij} - \mathbb{E}\{B_k(T)m_{1j}(T)\}\bigg| \geq \frac{\delta}{n}\bigg) \\
&\leq 2L_n\exp\left(-\frac{\delta^2}{c_1L_n^{-1}n + c_2\delta}\right). \qquad (S0.5)
\end{aligned}
$$

Furthermore, by Lemma 6, we have

$$
P\big(\big\|\mathbf{A}_n^{-1}\big\| \geq (1+h_1)C_3^{-1}L_n\big) \leq 6L_n^2\exp(-h_2L_n^{-3}n). \tag{S0.6}
$$

Combining (S0.3)-(S0.6), we have

$$
P\bigg(\big|I_{n2}^{(1)}\big| \geq \frac{C_8L_n^{3/2}\delta}{n}\bigg) \leq 2L_n\exp\left(-\frac{\delta^2}{c_1L_n^{-1}n + c_2\delta}\right) + 6L_n^2\exp(-h_2L_n^{-3}n). \tag{S0.7}
$$

where $C_8 = (1+h_1)C_3^{-1}$. As a result, by Lemma 6, we conclude that

$$\begin{aligned} P\Big(\big|I_{n2}^{(2)}\big| \geq \frac{C_9 L_n^{5/2}\delta}{n}\Big) &\leq P\Big(\big\|\mathbf{A}_n^{-1}(\mathbf{A}-\mathbf{A}_n)\mathbf{A}^{-1}\mathbf{D}\big\| \geq \frac{C_9 L_n^{5/2}\delta}{n}\Big) \\ &\leq P\Big(\big\|\mathbf{A}_n^{-1}(\mathbf{A}-\mathbf{A}_n)\big\| \geq \frac{(1+h_1)C_3^{-1}L_n^2\delta}{n}\Big) \\ &\leq P\big(\big\|\mathbf{A}_n^{-1}\big\| \geq (1+h_1)C_3^{-1}L_n\big) + P\Big(\big\|\mathbf{A}_n-\mathbf{A}\big\| \geq \frac{L_n\delta}{n}\Big) \\ &\leq 6L_n^2\exp(-h_2L_n^{-3}n) + 6L_n^2\exp\Big(-\frac{\delta^2}{c_9L_n^{-1}n + c_{10}\delta}\Big), \quad \text{(S0.8)} \end{aligned}$$

where $C_9 = C_3^{-2}C_7(1+h_1)$. According to (S0.7) and (S0.8), we obtain, for any $\delta > 0$,

$$\begin{aligned} &P\Big(\big|I_{n2}\big| \geq \frac{(C_8L_n^{3/2} + C_9L_n^{5/2})\delta}{n}\Big) \\ &\leq P\Big(\big|I_{n2}^{(1)}\big| \geq \frac{C_8L_n^{3/2}\delta}{n}\Big) + P\Big(\big|I_{n2}^{(2)}\big| \geq \frac{C_9L_n^{5/2}\delta}{n}\Big) \\ &\leq 12L_n^2\exp(-h_2L_n^{-3}n) + (2L_n + 6L_n^2)\exp\Big(-\frac{\delta^2}{c_{11}L_n^{-1}n + c_{12}\delta}\Big), \qquad \text{(S0.9)} \end{aligned}$$

where $c_{11} = \max(c_1, c_9)$ and $c_{12} = \max(c_2, c_{10})$.

Last, we consider the term for I_{n1} in inequality (S0.2). Write $a_{njk} = \frac{1}{n}\sum_{i=1}^n B_k(T_i)[\widehat{f}_{1i}(Y, X_j) - f_{1,ij}]$. Note that $|Q_{\tau,Y} - \widehat{Q}_{\tau,Y}| = O(n^{-1/2}(\log n)^{1/2})$ almost surely (Serfling, 1980, Section 2.5.1; He et al., 2013, Lemma 8.4) and the fact that for any y and any $\epsilon > 0$, $\sup_{|y_1-y|<\epsilon}\big|I(Y < y_1) - I(Y < y)\big| \leq I(y-\epsilon < Y < y+\epsilon)$. On the event $\{|\widehat{Q}_{\tau,Y} - Q_{\tau,Y}| \leq M_3\eta n^{-\iota}\}$ with any $\eta > 0$, $0 < \iota < 1/2$ and some constant $M_3 > 0$, we have

$$
\begin{aligned}
\left|a_{njk}\right| &\leq \frac{1}{n}\sum_{i=1}^{n} B_k(T_i)\left|\widehat{f}_{1i}(Y, X_j) - f_{1,ij}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n} B_k(T_i)|X_{ij}| I(Q_{\tau,Y} - M_3\eta n^{-\iota} < Y_i < Q_{\tau,Y} + M_3\eta n^{-\iota}) \\
&\hat{=} \frac{1}{n}\sum_{i=1}^{n} \xi_{ijk} \text{ (say)}.
\end{aligned}
$$

Write $q_\tau(Y_i, Q_{\tau,Y}) = I(Q_{\tau,Y} - M_3\eta n^{-\iota} < Y_i < Q_{\tau,Y} + M_3\eta n^{-\iota})$. Then, by Taylor's expansion and condition (C4), we have $\mathbb{E}\{q_\tau(Y_i, Q_{\tau,Y})|X_{ij}, T_i\} = \int_{Q_{\tau,Y}-M_3\eta n^{-\iota}}^{Q_{\tau,Y}+M_3\eta n^{-\iota}} f_{Y_i|(X_{ij},T_i)}(s)\mathrm{d}s = 2M_3\eta f_{Y_i|(X_{ij},T_i)}(Q_{\tau,Y})n^{-\iota}\{1 + o(1)\}$, which further implies that for sufficiently large n,

$$
\begin{aligned}
\mu_1 &\widehat{=} \mathbb{E}\{B_k(T_i)|X_{ij}|\mathbb{E}[q_\tau(Y_i, Q_{\tau,Y})|X_{ij}, T_i]\} \\
&\leq 2M_3\eta n^{-\iota}\mathbb{E}\{B_k(T_i)|X_{ij}| f_{Y_i|(X_{ij},T_i)}(Q_{\tau,Y})\}\{1 + o(1)\} \\
&\leq 4M_3\eta M_{f1} n^{-\iota}\mathbb{E}\{B_k(T_i)|X_{ij}|\} \\
&\leq 4M_3\eta M_f n^{-\iota} K_1 K_2 C_0 L_n^{-1} = M_4 n^{-\iota} L_n^{-1},
\end{aligned}
$$

where condition (C4) implies that there exists a constant M_{f1} such that $\sup_{|y-Q_{\tau,Y}|<\varepsilon} f_{Y|(X_j,T)}(y) \leq M_{f1} < \infty$ for any $\varepsilon > 0$, and $M_4 = 4M_3 M_{f1} K_1 K_2 C_0 \eta$. For every $r > 2$, by the C_r

inequality and the boundedness of an indicator function, we have

$$\begin{aligned}\mathbb{E}\{|\xi_{ijk}-\mu_1|^r\} &\le 2^r\mathbb{E}\{B_k^r(T_i)|X_{ij}|^r q_\tau^2(Y_i,Q_{\tau,Y})\}\\ &\le 2^r\mathbb{E}\{B_k^r(T_i)|X_{ij}|^r\}\le 2^r\mathbb{E}\{B_k^2(T_i)K_1K_2^r r!\}\\ &\le 2^rK_1K_2^r r!C_4L_n^{-1}=8K_1K_2^2C_4L_n^{-1}(2K_2)^{r-2}r!/2.\end{aligned}$$

By choosing suitable constants such that $\delta>2M_4n^{1-\iota}L_n^{-1}$, we have $\delta/n-\mu_1>\delta/(2n)$ for some constant M_{f2}. It follows from Lemma 2 that

$$\begin{aligned}&\max_{1\le j\le p}\max_{1\le k\le L_n}P\Big(\big|a_{njk}\big|>\frac{\delta}{n}\Big)\\ &\le \max_{1\le j\le p}\max_{1\le k\le L_n}P\Big(\frac{1}{n}\sum_{i=1}^n\xi_{ijk}-\mu_1>\frac{\delta}{2n}\Big)\\ &\le \max_{1\le j\le p}\max_{1\le k\le L_n}P\Big(\Big|\frac{1}{n}\sum_{i=1}^n\xi_{ijk}-\mu_1\Big|>\frac{\delta}{2n}\Big)\\ &\le 2\exp\Big(-\frac{\delta^2}{c_{13}L_n^{-1}n+c_{14}\delta}\Big),\end{aligned}$$

where $c_{13}=16K_1K_2^2C_4$ and $c_{14}=8K_2$. With the above results, we have

$$\begin{aligned}&P\Big(\big|I_{n1}\big|>\frac{(1+h_1)C_3^{-1}L_n^{3/2}\delta}{n}\Big)\\ &\le P\big(\big\|A_n^{-1}\big\|>(1+h_1)C_3^{-1}L_n\big)+\sum_{k=1}^{L_n}P\big(\big|a_{njk}\big|>\delta/n\big)\\ &\le 6L_n^2\exp(-h_2L_n^{-3}n)+2L_n\exp\Big(-\frac{\delta^2}{c_{13}L_n^{-1}n+c_{14}\delta}\Big).\end{aligned}\tag{S0.10}$$

Hence, combining (S0.2), (S0.9), (S0.10) and the condition of lemma and taking $\frac{(2C_8L_n^{3/2}+C_9L_n^{5/2})\delta}{n} = \epsilon/2$, we can obtain

$$\begin{aligned}&\sup_{t\in\mathcal{T}} P\left(\left|\widehat{m}_{1j}(t) - m_{1j}(t)\right| > \epsilon\right)\\&\le 18L_n^2\exp(-h_2L_n^{-3}n) + (4L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{15}L_n^{-1}n + c_{16}\delta}\right),\end{aligned}$$

where $c_{15} = \max(c_{11}, c_{13})$, $c_{16} = \max(c_{12}, c_{14})$, and $\delta = n\epsilon/(4C_8L_n^{3/2} + 2C_9L_n^{5/2}) \asymp n\epsilon L_n^{-5/2}$.

For the second result of the lemma, we can complete the proof in a similar manner, where the proof of (S0.10) is distinct and has to modify as follows. Write $b_{njk} = \frac{1}{n}\sum_{i=1}^n B_k(T_i)[\widehat{f}_{2i}(Y, X_j) - f_{2,ij}]$. On the event $\{|\widehat{Q}_{\tau,Y} - Q_{\tau,Y}| \le M_3\eta n^{-\iota}\}$, we have

$$\begin{aligned}\left|b_{njk}\right| &\le \frac{1}{n}\sum_{i=1}^n B_k(T_i)I(Q_{\tau,Y} - M_3\eta n^{-\iota} < Y_i < Q_{\tau,Y} + M_3\eta n^{-\iota})\\&\hat{=} \frac{1}{n}\sum_{i=1}^n \zeta_{ijk} \text{ (say)}.\end{aligned}$$

Also, condition (C4) implies that $\sup_{|y-Q_{\tau,Y}|<\varepsilon} f_{Y|T}(y) \le M_{f2} < \infty$. Similarly, we derive that $\mu_2 \hat{=} \mathbb{E}\{\zeta_{ijk}\} = \mathbb{E}\{B_k(T_i)\mathbb{E}[q_\tau(Y_i, Q_{\tau,Y})|T_i]\} \le 4\eta M_3M_{f2}C_0n^{-\iota}L_n^{-1}$ and

$$\begin{aligned}\text{Var}(\zeta_{ijk} - \mu_2) &\le \mathbb{E}(|\zeta_{ijk}|^2) = \mathbb{E}\{B_k^2(T_i)q_\tau(Y_i, Q_{\tau,Y})\}\\&\le 4M_3\eta n^{-\iota}M_{f2}\mathbb{E}\{B_k^2(T_i)\} \le M_5n^{-\iota}L_n^{-1},\end{aligned}$$

where $M_5 = 4\eta M_3 M_{f2} C_4$. In addition, observe that $\max_{i,j,k} |\zeta_{ijk} - \mu_2| \le 2$. With an application of Lemma 3 and choosing $\delta > 8\eta M_3 M_{f2} C_0 n^{1-\iota} L_n^{-1}$, we get

$$\begin{aligned}
&\max_{1\le j\le p} \max_{1\le k\le L_n} P\Big(\big|b_{njk}\big| > \frac{\delta}{n}\Big) \\
&\le \max_{1\le j\le p} \max_{1\le k\le L_n} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} \zeta_{ijk} - \mu_2\Big| > \frac{\delta}{2n}\Big) \\
&\le 2\exp\Big(-\frac{\delta^2/4}{2(M_5 n^{1-\iota} L_n^{-1} + 2\delta/3)}\Big) \\
&\le 2\exp\Big(-\frac{\delta^2}{8M_5 L_n^{-1} n + 16\delta/3}\Big).
\end{aligned}$$

Thus, we can obtain the second result. As in derivation of the first result, the rest of results of the lemma can be proved. □

The lemmas given below are useful for establishing screening properties for censoring data.

Lemma 9. *Let $\mathcal{F}$ be a class of distribution functions whose support is the same as that of F, and let $\mathcal{Y}$ be the support of Y. For any $\varepsilon > 0$, define $\mathcal{W}(\varepsilon) = \{F^* \in \mathcal{F} : \|F^* - F\|_\infty \equiv \sup_{y\in\mathcal{Y}} |F^*(y) - F(y)| \le \varepsilon, |Q_{\tau,Y^*} - Q_{\tau,Y}| \le \varepsilon\}$, where Y^* is generated from the distribution F^*. Then, we have (i)*

$$\begin{aligned}
&\sup_{\mathcal{W}(\varepsilon)} \Big| w(F^*) I\{\widetilde{Y} \le Q_{\tau,Y^*}\} - w(F) I\{\widetilde{Y} \le Q_{\tau,Y}\}\Big| \\
&\le \frac{\varepsilon}{1-\tau} + I\{Q_{\tau,Y} - \varepsilon < Y \le Q_{\tau,Y} + \varepsilon\} \\
&\qquad + 3I\{Q_{\tau,Y} - \varepsilon < Z \le Q_{\tau,Y} + \varepsilon\} + I\{F^{-1}(\tau - \varepsilon) \le Z < F^{-1}(\tau + \varepsilon)\},
\end{aligned}$$

and (ii)

$$\begin{aligned}
&\sup_{\mathcal{W}(\varepsilon)} \Big| w^2(F^*)I\{\widetilde{Y} \le Q_{\tau,Y^*}\} - w^2(F)I\{\widetilde{Y} \le Q_{\tau,Y}\} \Big| \\
&\le \frac{4\varepsilon}{1-\tau} + I\{Q_{\tau,Y} - \varepsilon < Y \le Q_{\tau,Y} + \varepsilon\} \\
&\qquad +3I\{Q_{\tau,Y} - \varepsilon < Z \le Q_{\tau,Y} + \varepsilon\} + 2I\{F^{-1}(\tau-\varepsilon) \le Z < F^{-1}(\tau+\varepsilon)\}.
\end{aligned}$$

Proof of Lemma 9 Since the part (i) is adapted immediately from Wu and Yin (2015), it remains to show part (ii). By using similar arguments in (A.2) of Wang and Wang (2009), we can derive that

$$\begin{aligned}
&w^2(F)I\{\widetilde{Y} \le Q_{\tau,Y}\} \\
&= I\{Y \le Z, Z \le Q_{\tau,Y}\} + I\{Y \le Q_{\tau,Y}, Q_{\tau,Y} < Z\} + I\{Y > Z, Z < Q_{\tau,Y}\} \\
&\qquad \times \Big[1 + \Big(\frac{1-\tau}{1-F(Z)}\Big)^2 I\{F(Z) < \tau\} - \frac{2(1-\tau)}{1-F(Z)} I\{F(Z) < \tau\}\Big].
\end{aligned}$$

Thus, it follows that

$$w^2(F^*)I\{\widetilde{Y} \le Q_{\tau,Y^*}\} - w^2(F)I\{\widetilde{Y} \le Q_{\tau,Y}\} \hat{=} \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3,$$

where

$$\mathcal{I}_1 = I\{Y \le Z, Z \le Q_{\tau,Y^*}\} - I\{Y \le Z, Z \le Q_{\tau,Y}\},$$

$$\mathcal{I}_2 = I\{Y \le Q_{\tau,Y^*}, Q_{\tau,Y^*} < Z\} - I\{Y \le Q_{\tau,Y}, Q_{\tau,Y} < Z\},$$

$$\begin{aligned}\mathcal{I}_3 = {} & I\{Y > Z, Z < Q_{\tau,Y^*}\}\Big[1 + \Big(\frac{1-\tau}{1-F^*(Z)}\Big)^2 I\{F^*(Z) < \tau\} - \frac{2(1-\tau)}{1-F^*(Z)} I\{F^*(Z) < \tau\}\Big] \\ & - I\{Y > Z, Z < Q_{\tau,Y}\}\Big[1 + \Big(\frac{1-\tau}{1-F(Z)}\Big)^2 I\{F(Z) < \tau\} - \frac{2(1-\tau)}{1-F(Z)} I\{F(Z) < \tau\}\Big].\end{aligned}$$

Then, we have

$$\begin{aligned}\sup_{\mathcal{W}(\varepsilon)} |\mathcal{I}_1| &\le \sup_{\mathcal{W}(\varepsilon)} |I\{Z \le Q_{\tau,Y^*}\} - I\{Z \le Q_{\tau,Y}\}| \le I\{Q_{\tau,Y} - \varepsilon < Z \le Q_{\tau,Y} + \varepsilon\}, \\ \sup_{\mathcal{W}(\varepsilon)} |\mathcal{I}_2| &\le \sup_{\mathcal{W}(\varepsilon)} |I\{Y \le Q_{\tau,Y^*}\} - I\{Y \le Q_{\tau,Y}\}| + |I\{Q_{\tau,Y^*} < Z\} - I\{Q_{\tau,Y} < Z\}| \\ &\le I\{Q_{\tau,Y} - \varepsilon < Y \le Q_{\tau,Y} + \varepsilon\} + I\{Q_{\tau,Y} - \varepsilon < Z \le Q_{\tau,Y} + \varepsilon\}, \\ \sup_{\mathcal{W}(\varepsilon)} |\mathcal{I}_3| &\le \sup_{\mathcal{W}(\varepsilon)} |I\{Y > Z, Z < Q_{\tau,Y^*}\} - I\{Y > Z, Z < Q_{\tau,Y}\}| \\ &\quad + \sup_{\mathcal{W}(\varepsilon)} \Big|\Big(\frac{1-\tau}{1-F^*(Z)}\Big)^2 I\{F^*(Z) < \tau\} - \Big(\frac{1-\tau}{1-F(Z)}\Big)^2 I\{F(Z) < \tau\}\Big| \\ &\quad + \sup_{\mathcal{W}(\varepsilon)} \Big|\frac{2(1-\tau)}{1-F^*(Z)} I\{F^*(Z) < \tau\} - \frac{2(1-\tau)}{1-F(Z)} I\{F(Z) < \tau\}\Big| \\ &\hat{=} I_{31} + I_{32} + I_{33} \text{ (say)},\end{aligned}$$

where we further have

$$I_{31} \le I\{Q_{\tau,Y} - \varepsilon < Z \le Q_{\tau,Y} + \varepsilon\},$$

and

$$
\begin{aligned}
I_{32} \quad \leq \quad & \sup_{\mathcal{W}(\varepsilon)} I\{F^*(Z) < \tau, F(Z) < \tau\}\Big|\Big(\frac{1-\tau}{1-F^*(Z)}\Big)^2 - \Big(\frac{1-\tau}{1-F(Z)}\Big)^2\Big| \\
& + \sup_{\mathcal{W}(\varepsilon)} I\{F^*(Z) < \tau < F(Z)\}\Big|\frac{1-\tau}{1-F^*(Z)}\Big|^2 \\
& + \sup_{\mathcal{W}(\varepsilon)} I\{F(Z) < \tau < F^*(Z)\}\Big|\frac{1-\tau}{1-F(Z)}\Big|^2 \\
\leq \quad & \frac{2}{1-\tau}\sup_{\mathcal{W}(\varepsilon)} |F^*(Z) - F(Z)| + \sup_{\mathcal{W}(\varepsilon)} I\{F^*(Z) < \tau < F(Z)\} \\
& + \sup_{\mathcal{W}(\varepsilon)} I\{F(Z) < \tau < F^*(Z)\} \\
\leq \quad & \frac{2\varepsilon}{1-\tau} + I\{F^{-1}(\tau-\varepsilon) < Z \leq F^{-1}(\tau+\varepsilon)\},
\end{aligned}
$$

and, similarly,

$$
I_{33} \leq \frac{2\varepsilon}{1-\tau} + I\{F^{-1}(\tau-\varepsilon) < Z \leq F^{-1}(\tau+\varepsilon)\}.
$$

Therefore, gathering the above terms yields the desired result. □

Lemma 10. *Under conditions of Theorem 3.1, for any* $\epsilon > 0$ *such that* $L_n^{-1/2+d}\epsilon \to \infty$ *and* $\epsilon L_n^{-3/2} n^{\iota} \to \infty$ *with any* $0 < \iota < 1/2$ *as* $n \to \infty$*, then we have*

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{1j,w}(t) - m_{1j,w}(t)\big| > \epsilon\Big) \leq 18L_n^2\exp(-h_2L_n^{-3}n) + (10L_n + 6L_n^2)\exp\Big(-\frac{\delta^2}{\tilde{c}_7L_n^{-1}n + \tilde{c}_8\delta}\Big) \\
&\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{2j,w}(t) - m_{2j,w}(t)\big| > \epsilon\Big) \leq 18L_n^2\exp(-h_2L_n^{-3}n) + (10L_n + 6L_n^2)\exp\Big(-\frac{\delta^2}{\tilde{c}_9L_n^{-1}n + \tilde{c}_{10}\delta}\Big) \\
&\sup_{t\in\mathcal{T}} P\Big(\big|\widehat{m}_{3j,w}(t) - m_{3j,w}(t)\big| > \epsilon\Big) \leq 18L_n^2\exp(-h_2L_n^{-3}n) + (10L_n + 6L_n^2)\exp\Big(-\frac{\delta^2}{\tilde{c}_{11}L_n^{-1}n + \tilde{c}_{12}\delta}\Big)
\end{aligned}
$$

where $\delta \asymp n\epsilon L_n^{-5/2}$, and $\tilde{c}_7, \ldots, \tilde{c}_{12}$ are some positive constants.

Proof of Lemma 10 By the theory of spline approximation, there exists a vector $\boldsymbol{\gamma}_{j,w}^*$ such that $m_{1j,w}(t) = (\mathbf{B}(t))'\boldsymbol{\gamma}_{j,w}^* + \eta_{j,w}(t)$, where $\eta_{j,w}(t)$ is the approximation error. Write $\mathbf{A}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i)(\mathbf{B}(T_i))'$, $\mathbf{E}_{n,w} = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i)X_{ij}[w_i(\widehat{F})I(\widetilde{Y}_i < \widehat{Q}_{\tau,Y}) - w_i(F)I(\widetilde{Y}_i < Q_{\tau,Y})]$, $\mathbf{D}_{n,w} = \frac{1}{n}\sum_{i=1}^n \mathbf{B}(T_i)X_{ij}w_i(F)I(\widetilde{Y}_i < Q_{\tau,Y})$, $\mathbf{A} = \mathbb{E}\big\{\mathbf{B}(T)(\mathbf{B}(T))'\big\}$, $\mathbf{D}_w = \mathbb{E}\big\{\mathbf{B}(T)m_{1j,w}(T)\big\}$. Then, following the proof of Lemma 8, we have

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} \big|\widehat{m}_{1j,w}(t) - m_{1j,w}(t)\big| \\
&\leq \big\|\mathbf{A}_n^{-1}\mathbf{E}_{n,w}\big\| + \big\|\mathbf{A}_n^{-1}\mathbf{D}_{n,w} - \mathbf{A}^{-1}\mathbf{D}_w\big\| + \big\|\boldsymbol{\gamma}_{j,w}^M - \boldsymbol{\gamma}_{j,w}^*\big\| + \sup_{t\in\mathcal{T}}\big|\eta_{j,w}(t)\big| \\
&\hat{=} II_{n1} + II_{n2} + II_{n3} + II_{n4} \quad \text{(say)},
\end{aligned} \tag{S0.11}$$

where $\boldsymbol{\gamma}_{j,w}^M = \mathbf{A}^{-1}\mathbf{D}_w$. Following the derivation of upper bounds of I_{n3} and I_{n4} in Lemma 8, we get that $II_{n3} \leq C_6 L_n^{1/2-d}$ and $II_{n4} \leq C_5 L_n^{-d}$.

Using similar arguments to those in the derivation of exponential tail probability of I_{n2} in Lemma 8 and invoking Lemma 5, we obtain that

$$\begin{aligned}
&P\Bigg(\big|II_{n2}\big| \geq \frac{(C_8 L_n^{3/2} + C_9 L_n^{5/2})\delta}{n}\Bigg) \\
&\leq 12L_n^2 \exp(-h_2 L_n^{-3} n) + (2L_n + 6L_n^2)\exp\left(-\frac{\delta^2}{c_{17}L_n^{-1}n + c_{18}\delta}\right)
\end{aligned} \tag{S0.12}$$

for some positive constants c_{17} and c_{18}. It remains to derive an upper bound of the exponential tail probability of II_{n1}.

Under conditions (D2) and (D3), $\|\widehat{F} - F\|_\infty = O(n^{-1/2}(\log n)^{1/2})$ and $|\widehat{Q}_{\tau,Y} - Q_{\tau,Y}| = O(n^{-1/2}(\log n)^{1/2})$ almost surely (see He, Wang and Hong, 2013, Lemma 8.4). Under condition (D1), there exists a constant c_f such that $1/f(y) \le c_f$ uniformly in a neighbourhood of $Q_{\tau,Y}$. For any $\eta > 0$, $0 < \iota < 1/2$ and some constants $\tilde{a}_1, \tilde{a}_2, \tilde{a}_3, \tilde{a}_4$, we write $\phi_{ijk}^{(1)} = \eta n^{-\iota}\tilde{a}_1^{-1}B_k(T_i)|X_{ij}|/(1-\tau)$, $\phi_{ijk}^{(2)} = B_k(T_i)|X_{ij}|I\{Q_{\tau,Y} - \eta n^{-\iota}\tilde{a}_2^{-1} < Y_i \le Q_{\tau,Y} + \eta n^{-\iota}\tilde{a}_2^{-1}\}$, $\phi_{ijk}^{(3)} = 3B_k(T_i)|X_{ij}|I\{Q_{\tau,Y} - \eta n^{-\iota}\tilde{a}_3^{-1} < Z_i \le Q_{\tau,Y} + \eta n^{-\iota}\tilde{a}_3^{-1}\}$, $\phi_{ijk}^{(4)} = B_k(T_i)|X_{ij}|I\{F^{-1}(\tau - \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1}) \le Z < F^{-1}(\tau + \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1})\}$, and denote $\mu_{kj}^{(l)} = \mathbb{E}\{\phi_{ijk}^{(l)}\}$, $l = 1, 2, 3$. Note that there exist positive constants M_{f1}, M_{h1} such that $f_{Y|(X_j,T)}(y) \le M_{f1} < \infty$ and $h_{Z|(X_j,T)}(y) \le M_{h1} < \infty$ uniformly in a neighbourhood of $Q_{\tau,Y}$, due to conditions (D1) and (D2). Combining these with Lemma 1, we obtain that $\max_{j,k}\mu_{jk}^{(1)} \le K_1K_2C_0\eta\tilde{a}_1^{-1}n^{-\iota}L_n^{-1}/(1-\tau)$, $\max_{j,k}\mu_{jk}^{(2)} \le 4K_1K_2C_0\eta\tilde{a}_2^{-1}M_{f1}n^{-\iota}L_n^{-1}$, $\max_{j,k}\mu_{jk}^{(3)} \le 12K_1K_2C_0\eta\tilde{a}_3^{-1}M_{h1}n^{-\iota}L_n^{-1}$ and for every $r > 2$,

$$\mathbb{E}\big\{\big|\phi_{ijk}^{(1)} - \mu_{jk}^{(1)}\big|^r\big\} \le 2^r\mathbb{E}\big\{B_k^r(T_i)|X_{ij}|^r\big\} \le (2K_2)^{r-2}r!(8K_1K_2^2C_4L_n^{-1})/2,$$
$$\mathbb{E}\big\{\big|\phi_{ijk}^{(2)} - \mu_{jk}^{(2)}\big|^r\big\} \le 2^r\mathbb{E}\big\{B_k^r(T_i)|X_{ij}|^r\big\} \le (2K_2)^{r-2}r!(8K_1K_2^2C_4L_n^{-1})/2,$$
$$\mathbb{E}\big\{\big|\phi_{ijk}^{(3)} - \mu_{jk}^{(3)}\big|^r\big\} \le 2^r\mathbb{E}\big\{3^rB_k^r(T_i)|X_{ij}|^r\big\} \le (6K_2)^{r-2}r!(72K_1K_2^2C_4L_n^{-1})/2.$$

Therefore, taking $\delta > M_6n^{1-\iota}L_n^{-1}$ with $M_6 = 2K_1K_2C_0\eta\max(\tilde{a}_1^{-1}/(1-\tau), 4\tilde{a}_2^{-1}M_{f1}, 12\tilde{a}_3^{-1}M_{h1}, 4\tilde{a}_4^{-1}M_{h1})$, it follows from Lemma 2 that

$$\begin{aligned} &\max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(1)} \ge \frac{\delta}{n}\Big) \\ &\le \max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(1)} - \mu_{jk}^{(1)}\Big| \ge \frac{\delta}{2n}\Big) \\ &\le 2\exp\Big(-\frac{\delta^2}{64K_1K_2^2C_4L_n^{-1}n + 8K_2\delta}\Big), \end{aligned} \tag{S0.13}$$

and, similarly,

$$\max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(2)} \ge \frac{\delta}{n}\Big) \le 2\exp\Big(-\frac{\delta^2}{64K_1K_2^2C_4L_n^{-1}n + 8K_2\delta}\Big), \tag{S0.14}$$

and

$$\max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(3)} \ge \frac{\delta}{n}\Big) \le 2\exp\Big(-\frac{\delta^2}{576K_1K_2^2C_4L_n^{-1}n + 24K_2\delta}\Big). \tag{S0.15}$$

An application of Taylor's expansion of $F^{-1}(\tau - \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1})$ and $F^{-1}(\tau + \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1})$ around τ gives $I\{F^{-1}(\tau - \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1}) \le Z < F^{-1}(\tau + \eta n^{-\iota}c_f^{-1}\tilde{a}_4^{-1})\} \le I\{Q_{\tau,Y} - \eta n^{-\iota}\tilde{a}_4^{-1} \le Z < Q_{\tau,Y} + \eta n^{-\iota}\tilde{a}_4^{-1})\}$. Using this and following the derivation of (S0.15), we get

$$\max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(4)} \ge \frac{\delta}{n}\Big) \le 2\exp\Big(-\frac{\delta^2}{64K_1K_2^2C_4L_n^{-1}n + 8K_2\delta}\Big). \tag{S0.16}$$

Let $d_{ijk} = \frac{1}{n}\sum_{i=1}^{n} B_k(T_i)X_{ij}[w_i(\widehat{F})I(\widetilde{Y}_i < \widehat{Q}_{\tau,Y}) - w_i(F)I(\widetilde{Y}_i < Q_{\tau,Y})]$. Taking $\tilde{a}_5^{-1} =$

$\min(\tilde{a}_1^{-1}, \tilde{a}_2^{-1}, \tilde{a}_3^{-1}, c_f^{-1}\tilde{a}_4^{-1}, \tilde{a}_4^{-1})$, applying part (i) of Lemma 9 with $\varepsilon = \eta n^{-\iota}\tilde{a}_5^{-1}$ and combining (S0.13)-(S0.16), we obtain

$$\begin{aligned}
&\max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} d_{ijk}\Big| \ge \frac{4\delta}{n}\Big)\\
&\le \max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(1)} \ge \frac{\delta}{n}\Big) + \max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(2)} \ge \frac{\delta}{n}\Big)\\
&\quad + \max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(3)} \ge \frac{\delta}{n}\Big) + \max_{1\le j\le p}\max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\phi_{ijk}^{(4)} \ge \frac{\delta}{n}\Big)\\
&\le 8\exp\Big(-\frac{\delta^2}{576K_1K_2^2C_4L_n^{-1}n + 24K_2\delta}\Big).
\end{aligned}$$

As in (S0.10), we get

$$\begin{aligned}
&P\Big(\big|II_{n1}\big| > \frac{4(1+h_1)C_3^{-1}L_n^{3/2}\delta}{n}\Big)\\
&\le P\big(\big\|A_n^{-1}\big\| > (1+h_1)C_3^{-1}L_n^{1}\big) + \sum_{k=1}^{L_n} P\big(\big|d_{njk}\big| > 4\delta/n\big)\\
&\le 6L_n^2\exp(-h_2L_n^{-3}n) + 8L_n\exp\Big(-\frac{\delta^2}{c_{19}L_n^{-1}n + c_{20}\delta}\Big). \qquad \text{(S0.17)}
\end{aligned}$$

where $c_{19} = 576K_1K_2^2C_4$ and $c_{20} = 24K_2$.

Under the condition of the lemma that $\epsilon^{-1}L_n^{1/2-d} \to 0$, we can choose large n such that $C_6L_n^{1/2-d} + C_5L_n^{-2} \le \epsilon/2$. Combining (S0.11), (S0.12) and (S0.17) and taking

$(5C_8L_n^{3/2}+C_9L_n^{5/2})\delta/n=\epsilon/2$, we get

$$\begin{aligned}
&\sup_{t\in\mathcal{T}} P\big(\big|\widehat{m}_{1j,w}(t)-m_{1j,w}(t)\big|>\epsilon\big)\le P\big(|II_{n1}|+|II_{n2}|>\epsilon/2\big)\\
&\le P\Big(\big|II_{n2}\big|\ge\frac{(C_8L_n^{3/2}+C_9L_n^{5/2})\delta}{n}\Big)+P\Big(\big|II_{n1}\big|>\frac{4(1+h_1)C_3^{-1}L_n^{3/2}\delta}{n}\Big)\\
&\le 18L_n^2\exp(-h_2L_n^{-3}n)+(10L_n+6L_n^2)\exp\Big(-\frac{\delta^2}{\tilde{c}_7L_n^{-1}n+\tilde{c}_8\delta}\Big),
\end{aligned}$$

where $\tilde{c}_7=\max(c_{17},c_{19})$ and $\tilde{c}_8=\max(c_{18},c_{20})$.

Because the second result of the lemma can be proved as same as the proof of the third result. So, it suffices to show the third one. To this end, we may follow the above step with applying Lemmas 3, 5, 6 and part (ii) of Lemma 9. The main modification of the proof is about the derivation of inequality (S0.17), whereas the term $\mathbf{E}_{n,w}$ in (S0.11) is replaced with $\mathbf{E}_{n,w}=\frac{1}{n}\sum_{i=1}^n\mathbf{B}(T_i)[w_i^2(\widehat{F})I(\widetilde{Y}_i<\widehat{Q}_{\tau,Y})-w_i^2(F)I(\widetilde{Y}_i<Q_{\tau,Y})]$. Let $e_{ik}=\frac{1}{n}\sum_{i=1}^n B_k(T_i)[w_i^2(\widehat{F})I(\widetilde{Y}_i<\widehat{Q}_{\tau,Y})-w_i^2(F)I(\widetilde{Y}_i<Q_{\tau,Y})]$, $\varphi_{ik}^{(1)}=4\eta n^{-\iota}\tilde{b}_1^{-1}B_k(T_i)/(1-\tau)$, $\varphi_{ik}^{(2)}=B_k(T_i)I\{Q_{\tau,Y}-\eta n^{-\iota}\tilde{b}_2^{-1}<Y_i\le Q_{\tau,Y}-\eta n^{-\iota}\tilde{b}_2^{-1}\}$, $\varphi_{ik}^{(3)}=3B_k(T_i)I\{Q_{\tau,Y}-\eta n^{-\iota}\tilde{b}_3^{-1}<Z_i\le Q_{\tau,Y}-\eta n^{-\iota}\tilde{b}_3^{-1}\}$, $\varphi_{ijk}^{(4)}=2B_k(T_i)I\{F^{-1}(\tau-\eta n^{-\iota}c_f^{-1}\tilde{b}_4^{-1})\le Z<F^{-1}(\tau+\eta n^{-\iota}c_f^{-1}\tilde{b}_4^{-1})\}$ and denote $\mu_k^{(l)}=\mathbb{E}\{\varphi_{ik}^{(l)}\},l=1,2,3$, where $\tilde{b}_1,\tilde{b}_2,\tilde{b}_3,\tilde{b}_4$ are some positive constants. Note that under conditions (D1) and (D2), there exist positive constants M_{f2},M_{h2} such that $f_{Y|T}(y)\le M_{f1}<\infty$ and $h_{Z|T}(y)\le M_{h2}<\infty$ uniformly in a neighbourhood of $Q_{\tau,Y}$, and also we can choose η or large n such that $4\eta n^{-\iota}\tilde{b}_1^{-1}/(1-\tau)\le 1$. Thus, we have $\max_k\mu_k^{(1)}\le C_0\eta\tilde{b}_1^{-1}n^{-\iota}L_n^{-1}/(1-\tau)$,

$\max_k \mu_k^{(2)} \leq 4C_0\eta\tilde{b}_2^{-1}M_{f2}n^{-\iota}L_n^{-1}$, $\max_k \mu_k^{(3)} \leq 12C_0\eta\tilde{b}_3^{-1}M_{h2}n^{-\iota}L_n^{-1}$ and $|\varphi_{ik}^{(1)} - \mu_k^{(1)}| \leq 2$, $|\varphi_{ik}^{(2)} - \mu_k^{(2)}| \leq 2$, $|\varphi_{ik}^{(3)} - \mu_k^{(3)}| \leq 6$ as well as $\mathrm{Var}(\varphi_{ik}^{(1)}) \leq C_4L_n^{-1}$, $\mathrm{Var}(\varphi_{ik}^{(2)}) \leq C_4L_n^{-1}$, $\mathrm{Var}(\varphi_{ik}^{(3)}) \leq 9C_4L_n^{-1}$. Then, taking $\delta > M_7n^{1-\iota}L_n^{-1}$ with $M_7 = 2C_0\eta\max(\tilde{b}_1^{-1}/(1-\tau), 4\tilde{b}_2^{-1}M_{f2}, 12\tilde{b}_3^{-1}M_{h2}, 8\tilde{b}_4^{-1}M_{h2})$, it follows from Lemma 3 that

$$\max_{1\leq k\leq L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(1)} \geq \frac{\delta}{n}\Big) \leq \max_{1\leq k\leq L_n} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(1)} - \mu_k^{(1)}\Big| \geq \frac{\delta}{2n}\Big)$$
$$\leq 2\exp\Big(-\frac{\delta^2}{8C_4L_n^{-1}n + 8\delta/3}\Big), \tag{S0.18}$$

and

$$\max_{1\leq k\leq L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(2)} \geq \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{8C_4L_n^{-1}n + 8\delta/3}\Big), \tag{S0.19}$$
$$\max_{1\leq k\leq L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(3)} \geq \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{72C_4L_n^{-1}n + 8\delta}\Big). \tag{S0.20}$$

Also, similar to (S0.16), we get

$$\max_{1\leq k\leq L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(4)} \geq \frac{\delta}{n}\Big) \leq 2\exp\Big(-\frac{\delta^2}{32C_4L_n^{-1}n + 16\delta/3}\Big). \tag{S0.21}$$

Hence, taking $\tilde{b}_5^{-1} = \min(\tilde{b}_1^{-1}, \tilde{b}_2^{-1}, \tilde{b}_3^{-1}, c_f^{-1}\tilde{b}_4^{-1}, \tilde{b}_4^{-1})$, applying part (ii) of Lemma 9 with

$\varepsilon = \eta n^{-\iota}\tilde{b}_5^{-1}$ and combining (S0.18)-(S0.21), we obtain

$$\begin{aligned}
&\max_{1\le k\le L_n} P\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} e_{ik}\Big| \ge \frac{4\delta}{n}\Big)\\
&\le \max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ijk}^{(1)} \ge \frac{\delta}{n}\Big) + \max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(2)} \ge \frac{\delta}{n}\Big)\\
&\quad + \max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(3)} \ge \frac{\delta}{n}\Big) + \max_{1\le k\le L_n} P\Big(\frac{1}{n}\sum_{i=1}^{n}\varphi_{ik}^{(4)} \ge \frac{\delta}{n}\Big)\\
&\le 8\exp\Big(-\frac{\delta^2}{72C_4L_n^{-1}n + 8\delta}\Big).
\end{aligned}$$

Thus, the second result of the lemma is verified. Likewise, we can prove the second result. □

Appendix C: Proofs of Theorems 2.1-2.3 and 3.1-3.3

Proof of Theorem 2.1: Since the proof of this theorem may complete by following the proof of Theorem 1 in Liu, Li and Wu (2014), which consists of three steps. We here just give an outline of the proof.

Step 1. For any $\epsilon > 0$, derive an exponential upper bound of $\sup_{t\in\mathcal{T}} P\big(\big|\widehat{\rho}_j(t) - \rho_j(t)\big| > \epsilon\big)$. To this end, observe that

$$\widehat{\rho}_j(t) = \frac{\widehat{m}_{1j}(t) - \widehat{m}_{2j}(t)\widehat{m}_{4j}(t)}{\sqrt{\{\widehat{m}_{2j}(t) - [\widehat{m}_{2j}(t)]^2\}\{\widehat{m}_{3j}(t) - [\widehat{m}_{4j}(t)]^2\}}}$$

and recall the definition of $\rho(t)$. Iteratively applying Lemma 8, some tedious calculations yield that

$$\begin{aligned}&\sup_{t\in\mathcal{T}} P\big(\big|\widehat{\rho}_j(t)-\rho_j(t)\big|>\epsilon\big)\\&\le 1218L_n^2\exp(-h_2L_n^{-3}n)+(254L_n+456L_n^2)\exp\left(-\frac{\delta^2}{c_{19}L_n^{-1}n+c_{20}\delta}\right)\\&\hat{=}p_n(\epsilon)\ (\text{say})\end{aligned}\tag{S0.22}$$

for some positive constants c_{19}, c_{20}, where $\delta \asymp \epsilon n L_n^{-5/2}$.

Step 2. For any $\epsilon > 0$, derive the upper bound of $P\big(\max_{1\le j\le p}\big|\widehat{u}_j-u_j\big|>\epsilon\big)$. For this, define $u_{nj}=n^{-1}\sum_{i=1}^n\rho_j^2(T_i)$. Since $|\widehat{u}_j-u_j|\le|\widehat{u}_j-u_{nj}|+|u_{nj}-u_j|$, we have

$$P\big(\big|\widehat{u}_j-u_j\big|>\epsilon\big)\le P\big(\big|\widehat{u}_j-u_{nj}\big|>\epsilon/2\big)+P\big(\big|u_{nj}-u_j\big|>\epsilon/2\big),\tag{S0.23}$$

where the second term can be bounded by $2\exp(-n\epsilon^2/64)$ according to Lemma 4 and the first term

$$\begin{aligned}P\big(\big|\widehat{u}_j-u_{nj}\big|>\epsilon/2\big) &\le P\left(\frac{1}{n}\sum_{i=1}^n\left|\widehat{\rho}_j^2(T_i)-\rho_j^2(T_i)\right|>\epsilon/2\right)\\&\le \sum_{i=1}^n P\left(\left|\widehat{\rho}_j^2(T_i)-\rho_j^2(T_i)\right|>\epsilon/2\right)\\&\le n\sup_{t\in\mathcal{T}}P\big(\big|\widehat{\rho}_j^2(t)-\rho_j^2(t)\big|>\epsilon/2\big).\end{aligned}$$

Thus, by Lemmas 7, 8 and (S0.22), we obtain

$$\begin{aligned}
&\max_{1\le j\le p} P\big(\big|\widehat{u}_j - u_{nj}\big| > \epsilon/2\big)\\
&\le n[p_n(\epsilon/2) + p_n(\epsilon/8) + p_n(\epsilon/4)]\\
&\le 3654 n L_n^2 \exp(-h_2 L_n^{-3} n) + (762 L_n + 1368 L_n^2) n \exp\left(-\frac{\delta^2}{c_{21} L_n^{-1} n + c_{22}\delta}\right)
\end{aligned}$$

where c_{21}, c_{22} are positive constants, and $\delta \asymp \epsilon n L_n^{-5/2}$. Hence, under condition (C7), invoking (S0.23) and setting $\epsilon = C L_n n^{-2\kappa}$, we have

$$\begin{aligned}
&P\Big(\max_{1\le j\le p} \big|\widehat{u}_j - u_j\big| > C L_n n^{-2\kappa}\Big)\\
&\le \sum_{j=1}^{p} P\big(\big|\widehat{u}_j - u_j\big| > C L_n n^{-2\kappa}\big)\\
&\le O\big(pn\{L_n^2 \exp(-h_2 L_n^{-3} n) + L_n \exp(-C_{10} L_n^{-2} n^{1-4\kappa})\}\big),
\end{aligned} \tag{S0.24}$$

where C_{10} is some positive constant. This proves result (i).

Step 3. Prove result (ii) under condition (C6). Recall the definition of $\widehat{\mathcal{M}}$ and set $\nu_n = C L_n n^{-2\kappa}$. Then, it follows from condition (C6) that

$$\begin{aligned}
&P\big(\mathcal{M}_* \subset \widehat{\mathcal{M}}\big) \ge P\Big(\min_{j\in\mathcal{M}_*} \widehat{u}_j > \nu_n\Big) = P\Big(\min_{j\in\mathcal{M}_*} u_j - \min_{j\in\mathcal{M}_*} \widehat{u}_j < \min_{j\in\mathcal{M}_*} u_j - \nu_n\Big)\\
&\ge P\Big(\min_{j\in\mathcal{M}_*} (\widehat{u}_j - u_j) > \nu_n - \min_{j\in\mathcal{M}_*} u_j\Big) \ge P\Big(\min_{j\in\mathcal{M}_*} u_j - \max_{j\in\mathcal{M}_*} \big|\widehat{u}_j - u_j\big| > \nu_n\Big)\\
&= 1 - P\Big(\max_{j\in\mathcal{M}_*} \big|\widehat{u}_j - u_j\big| > \min_{j\in\mathcal{M}_*} u_j - \nu_n\Big) \ge 1 - P\Big(\max_{j\in\mathcal{M}_*} \big|\widehat{u}_j - u_j\big| > \nu_n\Big)\\
&\ge 1 - O\big(s_n n\{L_n^2 \exp(-h_2 L_n^{-3} n) + L_n \exp(-C_{10} L_n^{-2} n^{1-4\kappa})\}\big),
\end{aligned}$$

where the last inequality is due to result (i). Thus, the result (ii) follows. □

Proof of Theorem 2.2: According to the condition of Theorem 2.2, there exists some $\delta_0 > 0$ such that $\min_{j\in\mathcal{M}_*} u_j - \max_{j\notin\mathcal{M}_*} u_j = \delta_0$. Note that by Fatou's Lemma, it follows that

$$P\Big(\liminf_{n\to\infty}\Big\{\min_{j\in\mathcal{M}_*}\widehat{u}_j - \max_{j\notin\mathcal{M}_*}\widehat{u}_j\Big\} \le 0\Big) \le \lim_{n\to\infty} P\Big(\min_{j\in\mathcal{M}_*}\widehat{u}_j \le \max_{j\notin\mathcal{M}_*}\widehat{u}_j\Big). \qquad \text{(S0.25)}$$

For the term on the right hand side of last inequality, we have

$$\begin{aligned}
& P\Big(\min_{j\in\mathcal{M}_*}\widehat{u}_j \le \max_{j\notin\mathcal{M}_*}\widehat{u}_j\Big) \\
&= P\Big(\Big\{\min_{j\in\mathcal{M}_*}\widehat{u}_j - \min_{j\in\mathcal{M}_*} u_j\Big\} \le \Big\{\max_{j\notin\mathcal{M}_*}\widehat{u}_j - \max_{j\notin\mathcal{M}_*} u_j\Big\} - \delta_0\Big) \\
&\le P\Big(\min_{j\in\mathcal{M}_*}\big|\widehat{u}_j - u_j\big| + \max_{j\notin\mathcal{M}_*}\big|\widehat{u}_j - u_j\big| \ge \delta_0\Big) \\
&\le P\Big(\max_{1\le j\le p}\big|\widehat{u}_j - u_j\big| \ge \delta_0/2\Big),
\end{aligned}$$

where, similar to (S0.24), we can derive

$$\begin{aligned}
& P\Big(\max_{1\le j\le p}\big|\widehat{u}_j - u_j\big| \ge \delta_0/2\Big) \le \sum_{j=1}^{p} P\Big(\big|\widehat{u}_j - u_j\big| \ge \delta_0/2\Big) \\
&\le O\big(pn\{L_n^2\exp(-h_2L_n^{-3}n) + L_n\exp(-C_{11}\delta_0^2L_n^{-4}n)\}\big) + 2p\exp(-\delta_0^2 n/64) \\
&\le O\big(pL_n^2 n\exp(-C_{11}\delta_0^2L_n^{-4}n)\big)
\end{aligned}$$

for some positive constant C_{11}. The last inequality goes to 0 as long as $n \to \infty$ and

$\log p < C_{11}\delta_0^2 L_n^{-4} n - 2\log L_n - \log n$. This together with (S0.25) implies

$$P\Big(\liminf_{n\to\infty}\Big\{\min_{j\in\mathcal{M}_*}\widehat{u}_j - \max_{j\notin\mathcal{M}_*}\widehat{u}_j\Big\} > 0\Big) = 1.$$

Hence, the ranking consistency is proved. □

Proof of Theorem 2.3: By definition of μ_j and condition (C5), we have

$$\sum_{j=1}^{p} u_j = \sum_{j=1}^{p}\mathbb{E}\{\rho_j^2(T)\} \le (K_3K_4)^{-1}\sum_{j=1}^{p}\mathbb{E}\big([\mathrm{Cov}\{I(Y_i > Q_{\tau,Y}), X_j|T\}]^2\big) = (K_3K_4)^{-1}\mathbb{E}\big(\|\mathbf{b}\|^2\big),$$

which implies that the size of $\{j : u_j > \delta L_n n^{-2\kappa}\}$ cannot exceed $O\big(L_n^{-1}n^{2\kappa}\mathbb{E}\{\|\mathbf{b}\|^2\}\big)$ for any $\delta > 0$. Thus, it follows that on the set $\mathcal{A}_n = \Big\{\max_{1\le j\le p}\big|\widehat{u}_j - u_j\big| \le \delta L_n n^{-2\kappa}\Big\}$, the size of $\{j : \widehat{u}_j > 2\delta L_n n^{-2\kappa}\}$ cannot exceed the size of $\{j : u_j > \delta L_n n^{-2\kappa}\}$, which is bounded by $O\big(L_n^{-1}n^{2\kappa}\mathbb{E}\{\|\mathbf{b}\|^2\}\big)$. Then, taking $\delta = C/2$, we have

$$P\Big(\big|\widehat{\mathcal{M}}\big| \le O\big(n^{2\kappa}L_n^{-1}\mathbb{E}\{\|\mathbf{b}\|^2\}\big)\Big) \ge P\big(\mathcal{A}_n\big) \ge 1 - P\Big(\max_{1\le j\le p}\big|\widehat{u}_j - u_j\big| > CL_n n^{-2\kappa}/2\Big).$$

The desired conclusion follows from part (i) of Theorem 2.1. □

Proofs of Theorems 3.1-3.3: Because the proofs can be complete directly through following the steps of the proofs of Theorems 2.1-2.3. Hence, we omit details. □

Bibliography

Bühlmann, P. and Van De Geer, S. (2011). *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, New York.

Fan, J., Ma, Y. and Dai, W. (2014). Nonparametric independent screening in sparse ultra-high dimensional varying coefficient models. *Journal of the American Statistical Association*, **109**, 1270-1284.

He, X., Wang, L. and Hong, H. (2013). Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Annals of Statistics*, **41**, 342-369.

Liu, J., Li, R. and Wu, R. (2014). Feature selection for varying coefficient models with ultrahigh-dimensional covariates. *Journal of the American Statistical Association*, **109**, 266-274.

Ma, X. and Zhang, J. (2016). Robust model-free feature screening via quantile correlation. *Journal of Multivariate Analysis*, **143**, 472-480.

Serfling R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

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

van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.

Wang, H. J. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**, 1117-1128.

Wu, Y. and Yin, G. (2015). Conditional qunatile screening in ultrahigh-dimensional heterogeneous data. *Biometrika*, **102**, 65-76.

Xiaochao Xia (Corresponding author), College of Science, Huazhong Agricultural University, Wuhan, China.

E-mail: (xia_xiao_chao@126.com)

Jialiang Li, Department of Statistics and Applied Probability, National University of Singapore; Duke-NUS Graduate Medical School; Singapore Eye Research Institute.

E-mail: (stalj@nus.edu.sg)

Bo Fu, School of Data Science, Fudan University, Shanghai, China; Administrative Data Research Centre for England, University College London, UK.

E-mail: (fu@fudan.edu.cn)