INFERENCE FOR A TWO-STEP JOINT MODEL OF EXTREME QUANTILE AND EXPECTED SHORTFALL REGRESSION

Qingzhao Zhong[1], Jingyu Ji[2], Liujun Chen[3], Yanxi Hou[1], and Deyuan Li[4]

[1] *School of Data Science, Fudan University*

[2] *School of Statistics and Data Science, Capital University of Economics and Business*

[3] *School of Management, University of Science and Technology of China*

[4] *School of Management, Fudan University*

Supplementary Material

This supplement contains some experimental results for Sections 4 and 5, all the proofs of theoretical results in the main article, as well as some additional explanations.

S1 Experimental Results

This section provides a summary of the mean and standard error of $\{\mathrm{ISE}_j\}_{j=1}^{m}$ across all models applied in Section 4. Specifically, Tables S1 - S9 include the results with Student-t, Pareto and Fréchet errors with $\gamma = 0.2, 0.3,$

Corresponding authors: Jingyu Ji, E-mail: jingyuji@cueb.edu.cn. Liujun Chen, E-mail: ljchen22@ustc.edu.cn.

and 0.4, respectively. Figure S1 provides the results of the in-sample and out-of-sample conditional ES estimators of the weekly loss of DJI30 data.

S2 Technical Proofs

Denote $U_Y(\cdot)$ and $U_Y(\cdot|\mathbf{x})$ as the tail quantile function of $F_Y(\cdot)$ and $F_Y(\cdot|\mathbf{x})$, respectively, and we have $Q_Y(\tau) = U_Y(1/(1-\tau))$ and $Q_Y(\tau|\mathbf{x}) = U_Y(1/(1-\tau)|\mathbf{x})$ for any $\tau \in (0,1)$ and $\mathbf{x} \in \mathcal{X}$.

S2.1 Some Necessary Results

Lemma S1. *Under Assumption 1, we have that as $t \to \infty$,*

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \frac{U_Y(t\,|\,\mathbf{x})}{U_0(t)g(\mathbf{x})^\gamma} - 1 \right| = O(A(t)). \tag{S2.1}$$

Moreover, for any $s > 0$, we also have that as $t \to \infty$,

$$\sup_{\mathbf{x}\in\mathcal{X}} \left| \left(\frac{U_Y(ts|\mathbf{x})}{U_Y(t\,|\,\mathbf{x})} - s^\gamma \right) - s^\gamma \frac{s^\rho - 1}{\rho} A_1(t) \right| = o(A_1(t)). \tag{S2.2}$$

Lemma S2. *Under Assumption 1, for any $s > 0$, we have that as $t \to \infty$,*

$$\left(\frac{U_Y(ts)}{U_Y(t)} - s^\gamma \right) - s^\gamma \frac{s^\rho - 1}{\rho} A_1(t) = o(A_1(t)). \tag{S2.3}$$

The Lemmas S1 and S2 play a crucial role in deriving our main conclusions, which are based on Theorems 1 and 2 in Xu et al. (2022). To facilitate further analysis, we define the function $\psi_\tau(u) = \tau - I(u \le 0)$ and denote E_i as the expectation with respective to $\mathbf{X}_i = \mathbf{x}_i$.

Table S1: *Student-t errors* ($\gamma = 0.2$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	90	55	0.99	0.1567(0.0298)	0.1567(0.0275)	0.0474(0.0244)	0.0412(0.0233)
				0.995	0.1821(0.0414)	0.1678(0.0341)	0.0564(0.0301)	0.0495(0.0290)
				0.999	0.2796(0.0741)	0.1952(0.0505)	0.0795(0.0452)	0.0716(0.0440)
1000	0.5	91	56	0.99	0.1660(0.0315)	0.1522(0.0305)	0.0343(0.0230)	0.0339(0.0216)
				0.995	0.1909(0.0432)	0.1563(0.0377)	0.0390(0.0278)	0.0379(0.0260)
				0.999	0.2890(0.0762)	0.1662(0.0549)	0.0514(0.0421)	0.0488(0.0379)
1000	0.9	85	52	0.99	0.1710(0.0325)	0.1490(0.0323)	0.0286(0.0223)	0.0323(0.0214)
				0.995	0.1956(0.0443)	0.1496(0.0399)	0.0321(0.0271)	0.0347(0.0252)
				0.999	0.2937(0.0780)	0.1508(0.0576)	0.0424(0.0441)	0.0421(0.0360)
2000	0	113	68	0.99	0.1561(0.0224)	0.1634(0.0200)	0.0567(0.0213)	0.0507(0.0210)
				0.995	0.1811(0.0315)	0.1809(0.0256)	0.0705(0.0273)	0.0638(0.0267)
				0.999	0.2496(0.0640)	0.2237(0.0397)	0.1067(0.0427)	0.0991(0.0417)
2000	0.5	110	66	0.99	0.1655(0.0234)	0.1653(0.0224)	0.0481(0.0225)	0.0479(0.0215)
				0.995	0.1902(0.0330)	0.1781(0.0286)	0.0579(0.0285)	0.0573(0.0270)
				0.999	0.2577(0.0660)	0.2082(0.0442)	0.0827(0.0439)	0.0815(0.0416)
2000	0.9	113	68	0.99	0.1706(0.0281)	0.1630(0.0240)	0.0397(0.0220)	0.0437(0.0211)
				0.995	0.1949(0.0340)	0.1712(0.0307)	0.0462(0.0273)	0.0498(0.0260)
				0.999	0.2620(0.0669)	0.1898(0.0473)	0.0621(0.0408)	0.0651(0.0390)
3000	0	126	74	0.99	0.1570(0.0188)	0.1655(0.0168)	0.0618(0.0194)	0.0561(0.0192)
				0.995	0.1822(0.0256)	0.1866(0.0220)	0.0789(0.0253)	0.0725(0.0249)
				0.999	0.2483(0.0553)	0.2379(0.0354)	0.1236(0.0407)	0.1164(0.0398)
3000	0.5	136	80	0.99	0.1665(0.0201)	0.1694(0.0181)	0.0540(0.0197)	0.0532(0.0193)
				0.995	0.1913(0.0271)	0.1855(0.0235)	0.0664(0.0254)	0.0653(0.0246)
				0.999	0.2562(0.0566)	0.2233(0.0373)	0.0981(0.0401)	0.0965(0.0386)
3000	0.9	130	77	0.99	0.1716(0.0207)	0.1701(0.0190)	0.0488(0.0203)	0.0522(0.0197)
				0.995	0.1960(0.0279)	0.1832(0.0247)	0.0588(0.0260)	0.0621(0.0248)
				0.999	0.2605(0.0571)	0.2133(0.0393)	0.0835(0.0406)	0.0867(0.0386)

Table S2: *Student-t errors* ($\gamma = 0.3$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	83	51	0.99	0.0865(0.0320)	0.0877(0.0291)	0.0236(0.0262)	0.0278(0.0372)
				0.995	0.0928(0.0388)	0.0856(0.0349)	0.0319(0.0391)	0.0356(0.0503)
				0.999	0.2249(0.6183)	0.0837(0.0474)	0.0570(0.0845)	0.0599(0.0955)
1000	0.5	90	55	0.99	0.0919(0.0335)	0.0819(0.0300)	0.0336(0.0590)	0.0387(0.0605)
				0.995	0.0976(0.0411)	0.0739(0.0346)	0.0504(0.0936)	0.0544(0.0891)
				0.999	0.2433(0.7329)	0.0612(0.0435)	0.1107(0.2301)	0.1116(0.2015)
1000	0.9	90	55	0.99	0.0949(0.0345)	0.0765(0.0309)	0.0482(0.0829)	0.0481(0.0702)
				0.995	0.1000(0.0425)	0.0653(0.0345)	0.0756(0.1329)	0.0718(0.1085)
				0.999	0.2589(0.8490)	0.0490(0.0428)	0.1799(0.3368)	0.1643(0.2650)
2000	0	120	72	0.99	0.0813(0.0243)	0.0872(0.0231)	0.0158(0.0175)	0.0172(0.0220)
				0.995	0.0850(0.0325)	0.0867(0.0289)	0.0215(0.0261)	0.0222(0.0307)
				0.999	0.1292(0.0810)	0.0880(0.0417)	0.0384(0.0556)	0.0377(0.0601)
2000	0.5	117	70	0.99	0.0861(0.0254)	0.0865(0.0239)	0.0159(0.0216)	0.0207(0.0281)
				0.995	0.0892(0.0342)	0.0819(0.0295)	0.0229(0.0335)	0.0274(0.0405)
				0.999	0.1342(0.0832)	0.0735(0.0408)	0.0469(0.0768)	0.0509(0.0854)
2000	0.9	120	72	0.99	0.0886(0.0262)	0.0841(0.0249)	0.0205(0.0308)	0.0245(0.0334)
				0.995	0.0914(0.0353)	0.0763(0.0300)	0.0314(0.0491)	0.0343(0.0500)
				0.999	0.1384(0.0899)	0.0615(0.0387)	0.0714(0.1185)	0.0714(0.1143)
3000	0	125	74	0.99	0.0810(0.0197)	0.0857(0.0175)	0.0100(0.0087)	0.0135(0.0142)
				0.995	0.0831(0.0267)	0.0865(0.0228)	0.0137(0.0129)	0.0168(0.0189)
				0.999	0.1132(0.0934)	0.0905(0.0357)	0.0244(0.0274)	0.0268(0.0348)
3000	0.5	130	77	0.99	0.0857(0.0208)	0.0866(0.0185)	0.0098(0.0118)	0.0150(0.0168)
				0.995	0.0869(0.0281)	0.0834(0.0235)	0.0140(0.0183)	0.0190(0.0234)
				0.999	0.1178(0.1083)	0.0775(0.0352)	0.0279(0.0423)	0.0326(0.0474)
3000	0.9	158	93	0.99	0.0883(0.0214)	0.0843(0.0191)	0.0133(0.0168)	0.0157(0.0178)
				0.995	0.0891(0.0286)	0.0768(0.0238)	0.0206(0.0270)	0.0219(0.0267)
				0.999	0.1209(0.1238)	0.0615(0.0331)	0.0474(0.0646)	0.0458(0.0608)

Table S3: *Student-t errors* ($\gamma = 0.4$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	98	60	0.99	0.0428(0.0352)	0.0306(0.0202)	0.2036(0.2098)	0.2540(0.2759)
				0.995	0.0705(0.0826)	0.0277(0.0225)	0.3062(0.3263)	0.3666(0.4084)
				0.999	0.8699(4.6025)	0.0572(0.0842)	0.6719(0.7858)	0.7639(0.9209)
1000	0.5	85	52	0.99	0.0454(0.0356)	0.0269(0.0214)	0.2761(0.2859)	0.3375(0.3576)
				0.995	0.0769(0.1065)	0.0245(0.0262)	0.4261(0.4544)	0.5006(0.5428)
				0.999	0.8885(4.1348)	0.0743(0.1217)	1.0007(1.1618)	1.1183(1.3076)
1000	0.9	88	54	0.99	0.0470(0.0365)	0.0244(0.0210)	0.3544(0.3323)	0.4027(0.4074)
				0.995	0.0826(0.1284)	0.0236(0.0262)	0.5566(0.5346)	0.6129(0.6272)
				0.999	0.9316(4.0514)	0.0987(0.1434)	1.3627(1.4087)	1.4451(1.5682)
2000	0	117	70	0.99	0.0265(0.0162)	0.0247(0.0153)	0.1695(0.1551)	0.2071(0.1840)
				0.995	0.0389(0.0468)	0.0200(0.0149)	0.2525(0.2431)	0.2950(0.2740)
				0.999	0.4535(1.2541)	0.0371(0.0539)	0.5414(0.5924)	0.5983(0.6262)
2000	0.5	113	68	0.99	0.0281(0.0168)	0.0231(0.0155)	0.2165(0.1962)	0.2516(0.2243)
				0.995	0.0407(0.0481)	0.0175(0.0148)	0.3294(0.3121)	0.3692(0.3422)
				0.999	0.4662(1.2723)	0.0445(0.0716)	0.7488(0.7907)	0.8021(0.8276)
2000	0.9	112	67	0.99	0.0292(0.0172)	0.0222(0.0158)	0.2540(0.2225)	0.2889(0.2568)
				0.995	0.0419(0.0493)	0.0166(0.0149)	0.3913(0.3564)	0.4310(0.3957)
				0.999	0.4753(1.2323)	0.0532(0.0830)	0.9178(0.9222)	0.9721(0.9797)
3000	0	128	76	0.99	0.0222(0.0127)	0.0218(0.0124)	0.1560(0.1194)	0.1879(0.1434)
				0.995	0.0246(0.0195)	0.0160(0.0117)	0.2309(0.1864)	0.2665(0.2110)
				0.999	0.2322(0.4545)	0.0278(0.0346)	0.4881(0.4429)	0.5342(0.4675)
3000	0.5	127	75	0.99	0.0236(0.0135)	0.0213(0.0128)	0.1892(0.1441)	0.2209(0.1616)
				0.995	0.0262(0.0205)	0.0143(0.0115)	0.2847(0.2274)	0.3206(0.2418)
				0.999	0.2560(0.5235)	0.0309(0.0450)	0.6315(0.5594)	0.6787(0.5611)
3000	0.9	130	77	0.99	0.0244(0.0140)	0.0206(0.0130)	0.2198(0.1605)	0.2418(0.1727)
				0.995	0.0273(0.0225)	0.0132(0.0114)	0.3351(0.2555)	0.3583(0.2626)
				0.999	0.2742(0.5773)	0.0365(0.0521)	0.7672(0.6448)	0.7923(0.6314)

Table S4: *Pareto errors* ($\gamma = 0.2$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	91	56	0.99	0.1567(0.0298)	0.1566(0.0275)	0.0472(0.0244)	0.0406(0.0233)
				0.995	0.1821(0.0414)	0.1675(0.0341)	0.0560(0.0302)	0.0488(0.0290)
				0.999	0.2796(0.0741)	0.1946(0.0505)	0.0788(0.0452)	0.0705(0.0440)
1000	0.5	94	57	0.99	0.1660(0.0315)	0.1511(0.0305)	0.0332(0.0224)	0.0327(0.0213)
				0.995	0.1909(0.0432)	0.1547(0.0376)	0.0376(0.0270)	0.0363(0.0254)
				0.999	0.2890(0.0762)	0.1628(0.0549)	0.0493(0.0404)	0.0464(0.0368)
1000	0.9	85	52	0.99	0.1710(0.0325)	0.1490(0.0323)	0.0286(0.0223)	0.0323(0.0214)
				0.995	0.1956(0.0443)	0.1496(0.0399)	0.0321(0.0271)	0.0347(0.0252)
				0.999	0.2937(0.0780)	0.1508(0.0576)	0.0424(0.0441)	0.0421(0.0360)
2000	0	116	69	0.99	0.1561(0.0224)	0.1630(0.0201)	0.0559(0.0212)	0.0500(0.0209)
				0.995	0.1811(0.0315)	0.1801(0.0257)	0.0694(0.0271)	0.0628(0.0265)
				0.999	0.2496(0.0640)	0.2219(0.0397)	0.1047(0.0423)	0.0971(0.0413)
2000	0.5	109	65	0.99	0.1655(0.0234)	0.1655(0.0223)	0.0483(0.0225)	0.0483(0.0216)
				0.995	0.1902(0.0330)	0.1784(0.0285)	0.0583(0.0285)	0.0579(0.0272)
				0.999	0.2577(0.0660)	0.2088(0.0442)	0.0833(0.0440)	0.0824(0.0419)
2000	0.9	110	66	0.99	0.1706(0.0241)	0.1638(0.0235)	0.0406(0.0219)	0.0449(0.0213)
				0.995	0.1949(0.0340)	0.1276(0.0300)	0.0475(0.0272)	0.0514(0.0263)
				0.999	0.2621(0.0669)	0.1925(0.0462)	0.0642(0.0407)	0.0677(0.0393)
3000	0	131	77	0.99	0.1570(0.0188)	0.1657(0.0167)	0.0616(0.0190)	0.0555(0.0190)
				0.995	0.1822(0.0256)	0.1865(0.0217)	0.0784(0.0247)	0.0716(0.0245)
				0.999	0.2483(0.0553)	0.2372(0.0345)	0.1225(0.0395)	0.1148(0.0389)
3000	0.5	126	74	0.99	0.1665(0.0201)	0.1699(0.0182)	0.0555(0.0202)	0.0550(0.0197)
				0.995	0.1913(0.0271)	0.1868(0.0236)	0.0687(0.0261)	0.0680(0.0251)
				0.999	0.2562(0.0566)	0.2268(0.0377)	0.1024(0.0414)	0.1013(0.0397)
3000	0.9	130	77	0.99	0.1716(0.0207)	0.1701(0.0190)	0.0488(0.0203)	0.0522(0.0196)
				0.995	0.1960(0.0279)	0.1832(0.0247)	0.0588(0.0260)	0.0621(0.0248)
				0.999	0.2605(0.0571)	0.2133(0.0393)	0.0835(0.0406)	0.0867(0.0386)

Table S5: *Pareto errors* ($\gamma = 0.3$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	82	50	0.99	0.0699(0.0258)	0.0850(0.0249)	0.0214(0.0165)	0.0203(0.0165)
				0.995	0.0811(0.0346)	0.1018(0.0333)	0.0317(0.0241)	0.0287(0.0222)
				0.999	0.2005(0.6656)	0.1503(0.0547)	0.0662(0.0467)	0.0592(0.0427)
1000	0.5	90	55	0.99	0.0741(0.0275)	0.0891(0.0265)	0.0199(0.0166)	0.0199(0.0159)
				0.995	0.0856(0.0367)	0.1046(0.0348)	0.0292(0.0238)	0.0275(0.0213)
				0.999	0.2139(0.7764)	0.1488(0.0557)	0.0599(0.0450)	0.0546(0.0404)
1000	0.9	90	55	0.99	0.0765(0.0285)	0.0892(0.0279)	0.0186(0.0170)	0.0199(0.0173)
				0.995	0.0881(0.0381)	0.1031(0.0363)	0.0269(0.0239)	0.0267(0.0223)
				0.999	0.2246(0.8592)	0.1426(0.0575)	0.0538(0.0442)	0.0504(0.0400)
2000	0	115	69	0.99	0.0673(0.0183)	0.0778(0.0177)	0.0136(0.0096)	0.0125(0.0094)
				0.995	0.0758(0.0272)	0.0919(0.0241)	0.0211(0.0148)	0.0181(0.0130)
				0.999	0.1263(0.0961)	0.1338(0.0409)	0.0478(0.0318)	0.0411(0.0282)
2000	0.5	124	74	0.99	0.0713(0.0193)	0.0817(0.0188)	0.0122(0.0097)	0.0123(0.0096)
				0.995	0.0800(0.0286)	0.0946(0.0251)	0.0187(0.0147)	0.0172(0.0128)
				0.999	0.1312(0.0969)	0.1323(0.0416)	0.0417(0.0307)	0.0368(0.0265)
2000	0.9	125	75	0.99	0.0736(0.0199)	0.0826(0.0197)	0.0111(0.0097)	0.0122(0.0098)
				0.995	0.0821(0.0293)	0.0941(0.0261)	0.0167(0.0146)	0.0165(0.0126)
				0.999	0.1350(0.1071)	0.1278(0.0429)	0.0365(0.0300)	0.0332(0.0253)
3000	0	133	79	0.99	0.0670(0.0156)	0.0741(0.0151)	0.0102(0.0080)	0.0096(0.0078)
				0.995	0.0727(0.0234)	0.0864(0.0208)	0.0162(0.0124)	0.0139(0.0111)
				0.999	0.1069(0.0505)	0.1238(0.0362)	0.0383(0.0273)	0.0328(0.0247)
3000	0.5	130	77	0.99	0.0712(0.0166)	0.0777(0.0165)	0.0092(0.0083)	0.0100(0.0084)
				0.995	0.0766(0.0247)	0.0887(0.0227)	0.0144(0.0129)	0.0137(0.0115)
				0.999	0.1113(0.0532)	0.1219(0.0391)	0.0332(0.0279)	0.0296(0.0245)
3000	0.9	137	81	0.99	0.0740(0.0181)	0.0788(0.0171)	0.0085(0.0076)	0.0108(0.0088)
				0.995	0.0770(0.0244)	0.0884(0.0231)	0.0129(0.0117)	0.0139(0.0113)
				0.999	0.1086(0.0653)	0.1171(0.0392)	0.0286(0.0250)	0.0269(0.0223)

Table S6: *Pareto errors* ($\gamma = 0.4$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	92	56	0.99	0.0428(0.0352)	0.0305(0.0206)	0.2015(0.2139)	0.2532(0.2762)
				0.995	0.0705(0.0826)	0.0277(0.0232)	0.3032(0.3334)	0.3644(0.4094)
				0.999	0.8699(4.6025)	0.0573(0.0864)	0.6663(0.8065)	0.7572(0.9283)
1000	0.5	87	53	0.99	0.0454(0.0356)	0.0268(0.0213)	0.2807(0.2848)	0.3393(0.3538)
				0.995	0.0769(0.1065)	0.0245(0.0254)	0.4338(0.4534)	0.5044(0.5378)
				0.999	0.8885(4.1348)	0.0760(0.1202)	1.0218(1.1610)	1.1314(1.3001)
1000	0.9	97	59	0.99	0.0470(0.0365)	0.0232(0.0196)	0.3806(0.3460)	0.4088(0.3935)
				0.995	0.0826(0.1284)	0.0234(0.0262)	0.6001(0.5577)	0.6295(0.6128)
				0.999	0.9316(4.0514)	0.1080(0.1527)	1.4806(1.4770)	1.5118(1.5591)
2000	0	117	70	0.99	0.0265(0.0162)	0.0247(0.0153)	0.1695(0.1551)	0.2071(0.1840)
				0.995	0.0389(0.0468)	0.0199(0.0149)	0.2525(0.2431)	0.2950(0.2740)
				0.999	0.4535(1.2541)	0.0371(0.0539)	0.5414(0.5924)	0.5983(0.6262)
2000	0.5	111	66	0.99	0.0281(0.0168)	0.0231(0.0156)	0.2156(0.1950)	0.2561(0.2298)
				0.995	0.0407(0.0481)	0.0175(0.0149)	0.3278(0.3099)	0.3748(0.3493)
				0.999	0.4662(1.2723)	0.0442(0.0708)	0.7443(0.7842)	0.8107(0.8396)
2000	0.9	110	66	0.99	0.0292(0.0172)	0.0223(0.0157)	0.2525(0.2219)	0.2871(0.2536)
				0.995	0.0419(0.0493)	0.0166(0.0147)	0.3890(0.3555)	0.4278(0.3905)
				0.999	0.4753(1.2323)	0.0529(0.0824)	0.9121(0.9194)	0.9643(0.9656)
3000	0	122	72	0.99	0.0222(0.0127)	0.0215(0.0124)	0.1576(0.1228)	0.1920(0.1460)
				0.995	0.0246(0.0195)	0.0159(0.0118)	0.2335(0.1926)	0.2718(0.2150)
				0.999	0.2322(0.4545)	0.0283(0.0371)	0.4950(0.4621)	0.5444(0.4789)
3000	0.5	126	74	0.99	0.0236(0.0135)	0.0213(0.0127)	0.1888(0.1451)	0.2216(0.1612)
				0.995	0.0262(0.0205)	0.0143(0.0115)	0.2841(0.2291)	0.3211(0.2414)
				0.999	0.2560(0.5235)	0.0309(0.0456)	0.6300(0.5647)	0.6787(0.5612)
3000	0.9	129	76	0.99	0.0244(0.0140)	0.0206(0.0130)	0.2187(0.1597)	0.2426(0.1715)
				0.995	0.0273(0.0225)	0.0132(0.0114)	0.3333(0.2541)	0.3588(0.2606)
				0.999	0.2742(0.5773)	0.0361(0.0516)	0.7623(0.6406)	0.7915(0.6261)

Table S7: *Fréchet errors* ($\gamma = 0.2$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	80	49	0.99	0.2270(0.0224)	0.2477(0.0183)	0.1862(0.0252)	0.1718(0.0241)
				0.995	0.2646(0.0330)	0.2962(0.0229)	0.2344(0.0307)	0.2197(0.0291)
				0.999	0.3728(0.0693)	0.4121(0.0319)	0.3534(0.0411)	0.3396(0.0393)
1000	0.5	83	51	0.99	0.2413(0.0239)	0.2561(0.0193)	0.1856(0.0269)	0.1756(0.0255)
				0.995	0.2790(0.0347)	0.3020(0.0241)	0.2312(0.0326)	0.2210(0.0307)
				0.999	0.3853(0.0720)	0.4109(0.0337)	0.3434(0.0438)	0.3336(0.0413)
1000	0.9	86	53	0.99	0.2493(0.0247)	0.2597(0.0202)	0.1833(0.0281)	0.1765(0.0268)
				0.995	0.2866(0.0356)	0.3036(0.0252)	0.2269(0.0340)	0.2199(0.0320)
				0.999	0.3912(0.0737)	0.4073(0.0350)	0.3339(0.0457)	0.3272(0.0430)
2000	0	107	64	0.99	0.2262(0.0169)	0.2422(0.0147)	0.1790(0.0213)	0.1649(0.0213)
				0.995	0.2645(0.0246)	0.2898(0.0190)	0.2260(0.0263)	0.2117(0.0260)
				0.999	0.3475(0.0639)	0.4041(0.0275)	0.3432(0.0361)	0.3295(0.0356)
2000	0.5	116	69	0.99	0.2406(0.0178)	0.2535(0.0155)	0.1822(0.0224)	0.1717(0.0219)
				0.995	0.2786(0.0260)	0.2991(0.0198)	0.2274(0.0274)	0.2167(0.0265)
				0.999	0.3597(0.0658)	0.4074(0.0284)	0.3390(0.0373)	0.3287(0.0359)
2000	0.9	118	71	0.99	0.2486(0.0185)	0.2584(0.0164)	0.1818(0.0238)	0.1743(0.0224)
				0.995	0.2862(0.0267)	0.3023(0.0210)	0.2253(0.0291)	0.2177(0.0271)
				0.999	0.3659(0.0671)	0.4061(0.0301)	0.3323(0.0398)	0.3250(0.0371)
3000	0	136	80	0.99	0.2270(0.0134)	0.2409(0.0116)	0.1771(0.0170)	0.1623(0.0178)
				0.995	0.2639(0.0193)	0.2882(0.0151)	0.2238(0.0211)	0.2088(0.0216)
				0.999	0.3489 (0.0492)	0.4021(0.0223)	0.3405(0.0294)	0.3261(0.0295)
3000	0.5	136	80	0.99	0.2415(0.0143)	0.2521(0.0124)	0.1802(0.0182)	0.1689(0.0190)
				0.995	0.2782(0.0201)	0.2974(0.0161)	0.2250(0.0226)	0.2135(0.0229)
				0.999	0.3613(0.0510)	0.4053(0.0236)	0.3361(0.0314)	0.3249(0.0311)
3000	0.9	127	75	0.99	0.2495(0.0148)	0.2569(0.0131)	0.1794(0.0195)	0.1708(0.0199)
				0.995	0.2858(0.0206)	0.3004(0.0170)	0.2225(0.0242)	0.2136(0.0240)
				0.999	0.3676(0.0519)	0.4034(0.0252)	0.3287(0.0338)	0.3200(0.0328)

Table S8: *Fréchet errors* ($\gamma = 0.3$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	82	50	0.99	0.0685(0.0238)	0.0801(0.0237)	0.0174(0.0145)	0.0172(0.0147)
				0.995	0.0767(0.0320)	0.0941(0.0312)	0.0255(0.0209)	0.0234(0.0197)
				0.999	0.1901(0.3253)	0.1354(0.0506)	0.0527(0.0403)	0.0471(0.0373)
1000	0.5	85	52	0.99	0.0729(0.0253)	0.0822(0.0258)	0.0155(0.0145)	0.0174(0.0160)
				0.995	0.0809(0.0337)	0.0942(0.0337)	0.0223(0.0207)	0.0227(0.0205)
				0.999	0.2000(0.3456)	0.1294(0.0539)	0.0447(0.0394)	0.0420(0.0365)
1000	0.9	79	48	0.99	0.0753(0.0263)	0.0812(0.0271)	0.0150(0.0153)	0.0190(0.0200)
				0.995	0.0835(0.0348)	0.0913(0.0353)	0.0211(0.0213)	0.0238(0.0243)
				0.999	0.2093(0.3902)	0.1213(0.0562)	0.0407(0.0392)	0.0406(0.0389)
2000	0	115	69	0.99	0.0669(0.0194)	0.0765(0.0178)	0.0126(0.0096)	0.0122(0.0097)
				0.995	0.0727(0.0253)	0.0891(0.0239)	0.0192(0.0144)	0.0172(0.0135)
				0.999	0.1238(0.0833)	0.1271(0.0403)	0.0426(0.0298)	0.0374(0.0277)
2000	0.5	116	69	0.99	0.0712(0.0208)	0.0788(0.0191)	0.0108(0.0096)	0.0124(0.0104)
				0.995	0.0768(0.0268)	0.0895(0.0255)	0.0161(0.0142)	0.0165(0.0137)
				0.999	0.1295(0.0958)	0.1215(0.0424)	0.0348(0.0289)	0.0325(0.0264)
2000	0.9	115	69	0.99	0.0735(0.0215)	0.0790(0.0200)	0.0099(0.0102)	0.0128(0.0113)
				0.995	0.0789(0.0277)	0.0880(0.0265)	0.0144(0.0147)	0.0163(0.0145)
				0.999	0.1333(0.1074)	0.1152(0.0434)	0.0300(0.0284)	0.0297(0.0261)
3000	0	137	81	0.99	0.0665(0.0166)	0.0731(0.0143)	0.0093(0.0069)	0.0096(0.0083)
				0.995	0.0718(0.0226)	0.0842(0.0198)	0.0145(0.0108)	0.0134(0.0109)
				0.999	0.1023 (0.0493)	0.1183(0.0347)	0.0339(0.0242)	0.0298(0.0227)
3000	0.5	135	80	0.99	0.0707(0.0176)	0.0763(0.0158)	0.0081(0.0072)	0.0102(0.0095)
				0.995	0.0755(0.0237)	0.0858(0.0217)	0.0125(0.0111)	0.0134(0.0117)
				0.999	0.1067(0.0538)	0.1149(0.0374)	0.0284(0.0242)	0.0267(0.0220)
3000	0.9	130	77	0.99	0.0731(0.0180)	0.0768(0.0165)	0.0074(0.0071)	0.0108(0.0104)
				0.995	0.0776(0.0242)	0.0848(0.0225)	0.0112(0.0108)	0.0135(0.0124)
				0.999	0.1096(0.0583)	0.1095(0.0385)	0.0244(0.0231)	0.0246(0.0216)

Table S9: *Fréchet errors* ($\gamma = 0.4$): the mean and standard error (in the brackets) of ISEs.

n	r	k	$\widetilde{k}$	τ_n'	Method I	Method II	Method III	Method IV
1000	0	94	57	0.99	0.0539(0.0684)	0.0180(0.0241)	0.3569(0.3021)	0.4261(0.3564)
				0.995	0.1476(0.2207)	0.0328(0.0475)	0.4552(0.4315)	0.5304(0.4930)
				0.999	1.3986(6.5942)	0.0868(0.1440)	0.7198(0.8693)	0.8090(0.0393)
1000	0.5	82	50	0.99	0.0592(0.0781)	0.0206(0.0296)	0.4491(0.3948)	0.5372(0.4687)
				0.995	0.1582(0.2446)	0.0399(0.0604)	0.5966(0.5788)	0.6851(0.6630)
				0.999	1.5204(7.5863)	0.1177(0.1974)	0.9917(1.2509)	1.1169(1.3652)
1000	0.9	95	58	0.99	0.0623(0.0850)	0.0210(0.0309)	0.4858(0.4158)	0.5617(0.4611)
				0.995	0.1666(0.2622)	0.0415(0.0630)	0.6359(0.6089)	0.7223(0.6568)
				0.999	1.6380(8.4556)	0.1263(0.2066)	1.0898(1.3246)	1.1995(1.3766)
2000	0	114	68	0.99	0.0332(0.0347)	0.0161(0.0173)	0.4157(0.2596)	0.4869(0.3196)
				0.995	0.0926(0.1074)	0.0337(0.0371)	0.5409(0.3371)	0.6205(0.4458)
				0.999	0.3475(0.0639)	0.4041(0.0275)	0.3432(0.0361)	0.3295(0.0356)
2000	0.5	123	74	0.99	0.0355(0.0364)	0.0160(0.0182)	0.4561(0.2851)	0.5330(0.3432)
				0.995	0.0993(0.1142)	0.0342(0.0392)	0.5956(0.4144)	0.6834(0.4800)
				0.999	1.1102(2.8624)	0.1084(0.1308)	0.9980(0.8587)	1.1139(0.9456)
2000	0.9	118	71	0.99	0.0370(0.0379)	0.0175(0.0199)	0.5092(0.3166)	0.5846(0.3716)
				0.995	0.1028(0.1168)	0.0384(0.0431)	0.6721(0.4634)	0.7584(0.5244)
				0.999	1.1579(2.9399)	0.1267(0.1472)	1.1595(0.9808)	1.2746(1.0596)
3000	0	141	83	0.99	0.0248(0.0231)	0.0142(0.0146)	0.4246(0.2287)	0.4916(0.2713)
				0.995	0.0705(0.0658)	0.0318(0.0321)	0.5557(0.3345)	0.6303(0.3812)
				0.999	0.6236(1.2181)	0.1010(0.1070)	0.9206(0.6891)	1.0132(0.7463)
3000	0.5	136	80	0.99	0.0266(0.0245)	0.0144(0.0155)	0.4705(0.2558)	0.5540(0.3069)
				0.995	0.0749(0.0695)	0.0331(0.0345)	0.6195(0.3769)	0.7152(0.4334)
				0.999	0.6590(1.2924)	0.1111(0.1194)	1.0532(0.7983)	1.1801(0.8706)
3000	0.9	141	83	0.99	0.0277(0.0257)	0.0151(0.0166)	0.5052(0.2740)	0.5862(0.3245)
				0.995	0.0782(0.0731)	0.0351(0.0366)	0.6683(0.4031)	0.7613(0.4586)
				0.999	0.6894(1.3646)	0.1211(0.1271)	1.1542(0.8558)	1.2788(0.9269)

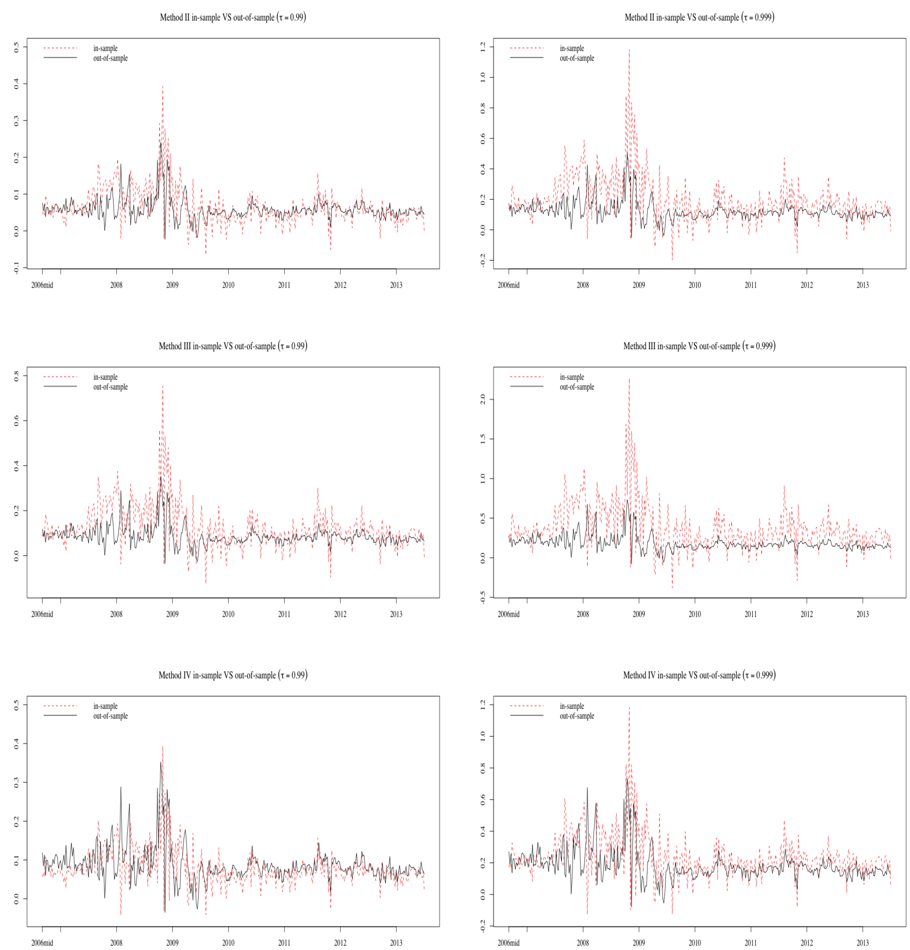

Figure S1: The results of the in-sample (red dashed curve) and out-of-sample (black solid curve) conditional ES estimators of Methods II, III, IV with $\tau_n' = 0.99$ (left panel) and $\tau_n' = 0.999$ (right panel).

Lemma S3. *For $\alpha \geq 1$, as $\tau \to 1$, we have that*

$$E_i \left|\psi_\tau\left(y_i - \boldsymbol{\beta}_0(\tau)^\top \mathbf{x}_i\right)\right|^\alpha = (1-\tau)^\alpha \tau + \tau^\alpha (1-\tau). \tag{S2.4}$$

Proof of Lemma S3. By a direct calculation, we can get

$$\begin{aligned}
& E_i \left|\psi_\tau\left(y_i - \boldsymbol{\beta}_0(\tau)^\top \mathbf{x}_i\right)\right|^\alpha \\
=& E_i \left|\left(\tau - I\left(y_i \leq F_Y^{-1}(\tau \mid \mathbf{x}_i)\right)\right)\right|^\alpha \\
=& E_i \left[E\left[\left|\tau - I\left(y_i \leq F_Y^{-1}(\tau \mid \mathbf{x})\right)\right|^\alpha \left| I\left(y_i \leq F_Y^{-1}(\tau \mid \mathbf{x}_i)\right)\right.\right]\right] \\
=& E_i \left[(1-\tau)^\alpha I\left(y_i \leq F_Y^{-1}(\tau \mid \mathbf{x}_i)\right) + \tau^\alpha I\left(y_i > F_Y^{-1}(\tau \mid \mathbf{x}_i)\right)\right] \\
=& (1-\tau)^\alpha \tau + \tau^\alpha (1-\tau).
\end{aligned}$$

□

Lemma S4. *Suppose $\alpha \geq 1$, $E[|Y_-|^\alpha] < \infty$, and $0 < \gamma < \frac{1}{\alpha}$. Then, we have that as $\tau \to 1$,*

$$E_i\left[(y_i - \boldsymbol{\beta}_0(\tau)^\top \mathbf{x}_i)_+^\alpha\right] = (1-\tau)(\boldsymbol{\beta}_0(\tau)^\top \mathbf{x}_i)^\alpha \frac{1}{\gamma} B(\alpha+1, 1/\gamma - \alpha)(1+o(1)). \tag{S2.5}$$

Proof of Lemma S4. By integration by parts and the change of variables, we have that,

$$
\begin{aligned}
&E_i\left[(y_i-\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)_+^\alpha\right]\\
=&\int_{\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i}^{\infty}(t-\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^\alpha\,dF_Y(t|\mathbf{x}_i)\\
=&\int_{\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i}^{\infty}1-F_Y(t|\mathbf{x}_i)\,d(t-\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^\alpha\\
=&\alpha\int_{\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i}^{\infty}[1-F_Y(t|\mathbf{x}_i)](t-\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^{\alpha-1}\,dt\\
=&\alpha(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^\alpha\int_1^{\infty}[1-F_Y(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i\cdot s|\mathbf{x}_i)](s-1)^{\alpha-1}\,ds\\
=&(1-\tau)\alpha(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^\alpha\int_1^{\infty}\frac{1-F_Y(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i\cdot s|\mathbf{x}_i)}{1-F_Y(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i|\mathbf{x}_i)}(s-1)^{\alpha-1}\,ds\\
=&(1-\tau)(\boldsymbol{\beta}_0(\tau)^\top\mathbf{x}_i)^\alpha\frac{1}{\gamma}B(\alpha+1,1/\gamma-\alpha)(1+o(1)),
\end{aligned}
$$

where the last step follows from a uniform convergence theorem for regular variation such as Proposition B.1.10 in de Haan and Ferreira (2006). □

Proposition S1. *Suppose Assumption 1 holds with* $0<\gamma<1/2$. *Then, we have that as* $n\to\infty$,

$$
\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0\left(1/\left(1-\tau_n\right)\right)}\right)\xrightarrow{d}N\left(\mathbf{0}_p,\gamma^2\left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1}E\left[\mathbf{X}\mathbf{X}^\top\right]\left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1}\right).
$$

Proof of Proposition S1. The proof follows a similar approach to the proof of Proposition 1 in the supplement of Xu et al. (2022), and thus is omitted here for brevity. □

S2.2 Proofs of Main Results

Proof of Proposition 1. We first note that the quantity $YI(Y \geq Q_Y(\tau|\mathbf{x}))$ is always positive when τ tends to 1. It is because $Q_Y(\tau|\mathbf{x}) \to \infty$ as $\tau \to 1$ due to the heavy-tailed assumption on Y. Then, by Fubini's theorem, we have that,

$$
\begin{aligned}
\frac{\mathrm{ES}_Y(\tau|\mathbf{x})}{Q_Y(\tau|\mathbf{x})} &= \frac{E[YI(Y \geq U_Y(1/(1-\tau)|\mathbf{x}))|\mathbf{x}]}{(1-\tau)U_Y(1/(1-\tau)|\mathbf{x})} \\
&= \frac{\int_0^\infty \mathbb{P}(YI(Y \geq U_Y(1/(1-\tau)|\mathbf{x})) \geq y|\mathbf{x})\,dy}{(1-\tau)U_Y(1/(1-\tau)|\mathbf{x})} \\
&= \frac{\int_0^\infty \mathbb{P}(Y \geq y \vee U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})\,dy}{(1-\tau)U_Y(1/(1-\tau)|\mathbf{x})} \\
&= \frac{\int_0^{U_Y(1/(1-\tau)|\mathbf{x})} \mathbb{P}(Y \geq U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})\,dy + \int_{U_Y(1/(1-\tau)|\mathbf{x})}^\infty \mathbb{P}(Y \geq y|\mathbf{x})\,dy}{(1-\tau)U_Y(1/(1-\tau)|\mathbf{x})} \\
&= 1 + \frac{1}{(1-\tau)U_Y(1/(1-\tau)|\mathbf{x})} \int_{U_Y(1/(1-\tau)|\mathbf{x})}^\infty \mathbb{P}(Y \geq y|\mathbf{x})\,dy \\
&= 1 + \frac{1}{1-\tau} \int_1^\infty \mathbb{P}(Y \geq yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})\,dy \\
&= 1 + \int_1^\infty \frac{1-F_Y(yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}{1-F_Y(U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}\,dy \\
&= 1 + \int_1^\infty \frac{1-F_Y(yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}{1-F_Y(U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})} - y^{-1/\gamma}\,dy + \int_1^\infty y^{-1/\gamma}\,dy \\
&= \int_1^\infty \frac{1-F_Y(yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}{1-F_Y(U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})} - y^{-1/\gamma}\,dy + \frac{1}{1-\gamma}.
\end{aligned}
$$

Then, it is sufficient to show that

$$
\sup_{\mathbf{x}\in\mathcal{X}} \left| \int_1^\infty \left[\frac{1-F_Y(yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}{1-F_Y(U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})} - y^{-1/\gamma} \right] dy \right| = O\left(A_1\left(\frac{1}{1-\tau} \right) \right). \tag{S2.6}
$$

Second, we reformulate (S2.2) in view of conditional distribution $F_Y(\cdot|\mathbf{x})$ as follows,

$$\sup_{\mathbf{x}\in\mathcal{X}}\left|\left(\frac{1-F_Y(ts|\mathbf{x})}{1-F_Y(t|\mathbf{x})}-s^{-1/\gamma}\right)-s^{-1/\gamma}\frac{s^{\rho/\gamma}-1}{\gamma\rho}A_1\left(\frac{1}{1-F_Y(t|\mathbf{x})}\right)+o\left(A_1\left(\frac{1}{1-F_Y(t|\mathbf{x})}\right)\right)\right|=o(1), \tag{S2.7}$$

implying that,

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{X}}\left|\int_1^\infty\left[\frac{1-F_Y(yU_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}{1-F_Y(U_Y(1/(1-\tau)|\mathbf{x})|\mathbf{x})}-y^{-1/\gamma}\right]dy\right|\\
=&\left|\int_1^\infty y^{-1/\gamma}\frac{y^{\rho/\gamma}-1}{\gamma\rho}A_1\left(\frac{1}{1-\tau}\right)+o\left(A_1\left(\frac{1}{1-\tau}\right)\right)dy\right|+o(1)\\
=&O\left(A_1\left(\frac{1}{1-\tau}\right)\right).
\end{aligned}$$

which completes this proof. □

Proof of Proposition 2. By recalling the closed form for $\hat{\boldsymbol{\theta}}_n(\tau_n)$ and $\boldsymbol{\theta}_0(\tau_n)$ given in (2.13) and (2.10), we have that

$$\begin{aligned}
&\hat{\boldsymbol{\theta}}_n(\tau_n)-\boldsymbol{\theta}_0(\tau_n)\\
=&\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)-\frac{1}{1-\tau_n}\left[(E[\mathbf{X}\mathbf{X}^\top])^{-1}-\left(\frac{1}{n}\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+\cdot\mathbf{X}\right]\\
&+\left(\frac{1}{n}\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\frac{1}{1-\tau_n}\left\{\frac{1}{n}\sum_{i=1}^n(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i-E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+\cdot\mathbf{X}\right]\right\}\\
&+\left(\frac{1}{n}\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\frac{1}{1-\tau_n}\left\{\frac{1}{n}\sum_{i=1}^n(y_i-\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i-\frac{1}{n}\sum_{i=1}^n(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i\right\}
\end{aligned}$$

which implies

$$
\begin{aligned}
&\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\theta}}_n(\tau_n)-\boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)\\
=&\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)-T_{n1}(\tau_n)+\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\cdot T_{n2}(\tau_n)\\
&+\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\cdot T_{n3}(\tau_n),
\end{aligned}
\tag{S2.8}
$$

where

$$
\begin{cases}
T_{n1}(\tau_n)=\frac{\sqrt{n(1-\tau_n)}}{(1-\tau_n)U_0(1/(1-\tau_n))}\left[(E[\mathbf{X}\mathbf{X}^\top])^{-1}-\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+\cdot\mathbf{X}\right],\\
T_{n2}(\tau_n)=\frac{1}{\sqrt{n(1-\tau_n)}U_0(1/(1-\tau_n))}\sum_{i=1}^n\left\{(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i-E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+\cdot\mathbf{X}\right]\right\},\\
T_{n3}(\tau_n)=\frac{1}{n(1-\tau_n)}\sum_{i=1}^n\sqrt{n(1-\tau_n)}\left(\frac{(y_i-\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i-(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i}{U_0(1/(1-\tau_n))}\right).
\end{cases}
$$

Part I: By central limit theorem, we have

$$
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top-E[\mathbf{X}\mathbf{X}^\top]\right)\xrightarrow{d}N(\mathbf{0}_p,\mathrm{Var}(\mathbf{X}\mathbf{X}^\top)),
$$

implying that as $n\to\infty$, there exists a absolute constant K such that,

$$
\begin{aligned}
T_{n1}(\tau_n)=&\frac{\sqrt{n(1-\tau_n)}}{(1-\tau_n)U_0(1/(1-\tau_n))}\left[(E[\mathbf{X}\mathbf{X}^\top])^{-1}-\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+\cdot\mathbf{X}\right]\\
=&K\sqrt{n}\left[(E[\mathbf{X}\mathbf{X}^\top])^{-1}-\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]\frac{(1-\tau_n)E\left[\mathbf{x}U_Y(1/(1-\tau_n)|\mathbf{x})\right]}{\sqrt{1-\tau_n}U_0(1/(1-\tau_n))}\\
=&K\sqrt{n}\left[(E[\mathbf{X}\mathbf{X}^\top])^{-1}-\left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}\right]\sqrt{1-\tau_n}E\left[\frac{U_Y(1/(1-\tau_n)|\mathbf{x})}{U_0(1/(1-\tau_n))}\cdot\mathbf{x}\right]\\
=&O_{\mathbb{P}}((1-\tau_n)^{1/2}),
\end{aligned}
$$

by noting that, using Lemma S4,

$$
\begin{aligned}
& E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \cdot \mathbf{X}\right] \\
=& E\left[E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \cdot \mathbf{X} \,\middle|\, \mathbf{X}=\mathbf{x}\right]\right] \\
=& E\left[\mathbf{x} E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \,\middle|\, \mathbf{X}=\mathbf{x}\right]\right] \\
=& (1-\tau_n)\frac{B(2,1/\gamma-1)}{\gamma} E\left[\mathbf{x} U_Y(1/(1-\tau_n)|\mathbf{x})\right].
\end{aligned}
$$

Part II: Now, we will show $T_{n2}(\tau_n)$ converges to a multivariate normal distribution. We fix a non-zero vector $\mathbf{c}=(c_1,c_2,\ldots,c_p)\in\mathbb{R}^p$ and define

$$
W_{ni} = \frac{(y_i-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+\mathbf{c}^\top \mathbf{x}_i - E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+\mathbf{c}^\top \mathbf{X}\right]}{\sqrt{n(1-\tau_n)}U_0(1/(1-\tau_n))}.
$$

We aim to apply the Lindeberg-Feller central limit theorem such that for each $\epsilon>0$, as $n\to\infty$,

$$
\sum_{i=1}^n E\left\{|W_{ni}|^2 I(|W_{ni}|>\epsilon)\right\} \to 0,
$$

and

$$
\sum_{i=1}^n \mathrm{Var}(W_{ni}) \to \sigma_0^2,
$$

for some $\sigma_0^2>0$.

Since $\gamma\in(0,1/2)$, there exists $0\le\delta\le\delta_0$ such that $\gamma<\frac{1}{2+\delta}<1/2$, then by using Markov's inequality and Lemma S4, as $n \to \infty$,

$$
\begin{aligned}
&\sum_{i=1}^{n} E\left\{|W_{ni}|^2 I(|W_{ni}|>\epsilon)\right\} \\
\leq& E\left[\sum_{i=1}^{n} \frac{E_i|W_{ni}|^{2+\delta}}{\epsilon^\delta}\right] \\
=&\frac{1}{\epsilon^\delta[n(1-\tau_n)]^{1+\frac{\delta}{2}}[U_0(1/(1-\tau_n))]^{2+\delta}} E\left[\sum_{i=1}^{n} E_i\left|(y_i-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+ \mathbf{c}^\top \mathbf{x}_i\right.\right. \\
&\left.\left.-E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \mathbf{c}^\top \mathbf{X}\right]\right|^{2+\delta}\right] \\
\leq&\frac{K}{\epsilon^\delta[n(1-\tau_n)]^{1+\frac{\delta}{2}}[U_0(1/(1-\tau_n))]^{2+\delta}} E\left[\sum_{i=1}^{n}\left\{E_i\left|(y_i-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+ \mathbf{c}^\top \mathbf{x}_i\right|^{2+\delta}\right.\right. \\
&\left.\left.+\ \left|E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \mathbf{c}^\top \mathbf{X}\right]\right|^{2+\delta}\right\}\right] \\
=&\frac{K}{\epsilon^\delta[n(1-\tau_n)]^{1+\frac{\delta}{2}}[U_0(1/(1-\tau_n))]^{2+\delta}} E\left[\sum_{i=1}^{n}\left\{|\mathbf{c}^\top \mathbf{x}_i|^{2+\delta}\cdot(1-\tau_n)(U_Y(1/(1-\tau_n)|\mathbf{x}_i))^{2+\delta}\right.\right. \\
&\left.\left.+\ (1-\tau_n)^{2+\delta}|E[\mathbf{c}^\top \mathbf{x} U_Y(1/(1-\tau_n)|\mathbf{x})]|^{2+\delta}\right\}\right] \\
\leq&\frac{K}{\epsilon^\delta[n(1-\tau_n)]^{\frac{\delta}{2}}} E\left[\frac{1}{n}\sum_{i=1}^{n}|\mathbf{c}^\top \mathbf{x}_i|^{2+\delta}\left(\frac{U_Y(1/(1-\tau_n)|\mathbf{x}_i)}{U_0(1/(1-\tau_n))}\right)^{2+\delta}\right] \\
&+\frac{K(1-\tau_n)^{1+\frac{\delta}{2}}}{\epsilon^\delta n^{\frac{\delta}{2}}}\left|E\left[\mathbf{c}^\top \mathbf{x}\frac{U_Y(1/(1-\tau_n)|\mathbf{x})}{U_0(1/(1-\tau_n))}\right]\right|^{2+\delta} \\
=&\frac{K}{\epsilon^\delta[n(1-\tau_n)]^{\frac{\delta}{2}}} E\left[\frac{1}{n}\sum_{i=1}^{n}|\mathbf{c}^\top \mathbf{x}_i|^{2+\delta} g(\mathbf{x}_i)^{\gamma(2+\delta)}\right]+\frac{K(1-\tau_n)^{1+\frac{\delta}{2}}}{\epsilon^\delta n^{\frac{\delta}{2}}}\left|E\left[\mathbf{c}^\top \mathbf{x} g(\mathbf{x})^\gamma\right]\right|^{2+\delta} \\
=&O\left([n(1-\tau_n)]^{-\delta/2}+(1-\tau_n)^{1+\delta/2}n^{-\delta/2}\right),
\end{aligned}
$$

where K is also a constant not related to $\mathbf{x}_i$ and the second inequality follows the Cr-inequality $|a+b|^r \leq C(|a|^r+|b|^r)$ with $C=\max\{1,2^{r-1}\}$ for any $a, b \in \mathbb{R}$, $r > 0$. Moreover, we can also find that

$$
\begin{aligned}
& E(W_{ni}) \\
=& \frac{1}{\sqrt{n(1-\tau_n)}U_0(1/(1-\tau_n))} E\left[E_i[(y_i - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+ \mathbf{c}^\top \mathbf{x}_i] - E\left[(Y - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \mathbf{c}^\top \mathbf{X} \,\middle|\, \mathbf{X} = \mathbf{x}\right]\right] \\
=& 0,
\end{aligned}
$$

and, by Lemma S4 again,

$$
\begin{aligned}
& \sum_{i=1}^n \mathrm{Var}(W_{ni}) \\
=& \frac{1}{n(1-\tau_n)U_0^2(1/(1-\tau_n))} E\left[\sum_{i=1}^n E_i\left\{(y_i - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+ \mathbf{c}^\top \mathbf{x}_i - E\left[(Y - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \mathbf{c}^\top \mathbf{X}\right]\right\}^2\right] \\
=& \frac{1}{n(1-\tau_n)U_0^2(1/(1-\tau_n))} E\left[\sum_{i=1}^n \left\{E_i[(y_i - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+^2 (\mathbf{c}^\top \mathbf{x}_i)^2]\right.\right. \\
& \left.\left. - \left(E\left[(Y - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{X})_+ \mathbf{c}^\top \mathbf{X}\right]\right)^2\right\}\right] \\
=& \frac{B(3, 1/\gamma - 2)(1 + o(1))}{\gamma} E\left[\frac{1}{n}\sum_{i=1}^n (\mathbf{c}^\top \mathbf{x}_i)^2 \left(\frac{U_Y(1/(1-\tau_n)|\mathbf{x}_i)}{U_0(1/(1-\tau_n))}\right)^2\right] \\
& - O((1-\tau_n)) E\left[(\mathbf{c}^\top \mathbf{x})^2 \left(\frac{U_Y(1/(1-\tau_n)|\mathbf{x})}{U_0(1/(1-\tau_n))}\right)^2\right] \\
=& \frac{B(3, 1/\gamma - 2)}{\gamma} \mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}]\mathbf{c} + o(1).
\end{aligned}
$$

Therefore, letting $\sigma_0^2 = \frac{B(3,1/\gamma-2)}{\gamma}\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}]\mathbf{c}$, we can conclude that $T_{n2}(\tau_n)$ converges weakly to a multivariate normal distribution with covariance matrix $\frac{B(3,1/\gamma-2)}{\gamma} E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}]$, that is,

$$
T_{n2}(\tau_n) \xrightarrow{d} N\left(\mathbf{0}_p, \frac{B(3, 1/\gamma - 2)}{\gamma} E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}]\right). \tag{S2.9}
$$

Part III: Define $h_i(t) = (y_i/U_0(1/(1-\tau_n)) - t^\top \mathbf{x}_i)_+ \cdot \mathbf{x}_i$, with derivative

$$h_i'(t) = -\mathbf{x}_i\mathbf{x}_i^\top I(y_i/U_0(1/(1-\tau_n)) \geq t^\top \mathbf{x}_i).$$

It is readily seen that, from Proposition S1, there exists a $\tilde{\boldsymbol{\beta}}$ between $\hat{\boldsymbol{\beta}}_n(\tau_n)/U_0(1/(1-\tau_n))$ and $\boldsymbol{\beta}_0(\tau_n)/U_0(1/(1-\tau_n))$ converging to $\boldsymbol{\beta}_0(\tau_n)/U_0(1/(1-\tau_n))$ such that,

$$\begin{aligned}
T_{n3}(\tau_n) =& \frac{1}{n(1-\tau_n)} \sum_{i=1}^n \sqrt{n(1-\tau_n)} \left(\frac{(y_i - \hat{\boldsymbol{\beta}}_n(\tau_n)^\top \mathbf{x}_i)_+ \cdot \mathbf{x}_i - (y_i - \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i)_+ \cdot \mathbf{x}_i}{U_0(1/(1-\tau_n))} \right) \\
=& \frac{1}{n(1-\tau_n)} \sum_{i=1}^n \sqrt{n(1-\tau_n)} \left\{ \left(\frac{y_i}{U_0(1/(1-\tau_n))} - \left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)}{U_0(1/(1-\tau_n))} \right)^\top \mathbf{x}_i \right)_+ \cdot \mathbf{x}_i \right. \\
& \left. - \left(\frac{y_i}{U_0(1/(1-\tau_n))} - \left(\frac{\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))} \right)^\top \mathbf{x}_i \right)_+ \cdot \mathbf{x}_i \right\} \\
=& -\frac{1}{n(1-\tau_n)} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top I(y_i/U_0(1/(1-\tau_n)) \geq \tilde{\boldsymbol{\beta}}^\top \mathbf{x}_i) \sqrt{n(1-\tau_n)} \left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)^\top - \boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))} \right).
\end{aligned}$$

Thus, (S2.8) can be reformulated as

$$\begin{aligned}
& \sqrt{n(1-\tau_n)} \left(\frac{\hat{\boldsymbol{\theta}}_n(\tau_n) - \boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))} \right) \\
=& \underbrace{\left[I_p - \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top \right)^{-1} \frac{1}{n(1-\tau_n)} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top I(y_i/U_0(1/(1-\tau_n)) \geq \tilde{\boldsymbol{\beta}}^\top \mathbf{x}_i) \right]}_{M_n(\tau_n)} \\
& \times \sqrt{n(1-\tau_n)} \left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n) - \boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))} \right) \\
& + \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top \right)^{-1} T_{n2}(\tau_n) + O_{\mathbb{P}}((1-\tau_n)^{1/2}),
\end{aligned}$$

where I_p denotes p-order identity matrix. Next, we show $M_n(\tau_n) = o_{\mathbb{P}}(1) I_p$.

Before this, we first note that

$$
\begin{aligned}
&E\left[\left|\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)-\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right|^2\right]\\
=&E\left[\left|\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right|^2\right]+\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)^2\\
&-2\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\\
=&E\left[\frac{1}{n^2(1-\tau_n)^2}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})^2I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\\
&+E\left[\frac{1}{n^2(1-\tau_n)^2}\sum_{i\neq j}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})(\mathbf{c}^\top\mathbf{x}_j\mathbf{x}_j^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)I(y_j/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_j)\right]\\
&+\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)^2-2\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}E_i\left[(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\right]\\
=&E\left[\frac{1}{n^2(1-\tau_n)^2}\sum_{i=1}^{n}E_i\left[(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})^2I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\right]\\
&+E\left[\frac{1}{n^2(1-\tau_n)^2}\sum_{i\neq j}E_i\left[(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\right.\\
&\left.\times E_j\left[(\mathbf{c}^\top\mathbf{x}_j\mathbf{x}_j^\top\mathbf{c})I(y_j/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_j)\right]\right]\\
&+\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)^2-2\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}E_i\left[(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})I(y_i/U_0(1/(1-\tau_n))\geq\tilde{\boldsymbol{\beta}}^\top\mathbf{x}_i)\right]\right]\\
=&\frac{1}{n(1-\tau_n)}E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})^2\right]+\frac{n-1}{n}E\left[\frac{1}{n(n-1)}\sum_{i\neq j}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})(\mathbf{c}^\top\mathbf{x}_j\mathbf{x}_j^\top\mathbf{c})\right]+\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)^2\\
&-2\left(\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right)E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{c}^\top\mathbf{x}_i\mathbf{x}_i^\top\mathbf{c})\right]+o(1)\\
=&O((n(1-\tau_n))^{-1}).
\end{aligned}
$$

Hence, for a ε, by Markov inequality, we have that as $n \to \infty$,

$$
\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}(\mathbf{c}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{c}) I(y_i/U_0(1/(1-\tau_n)) \geq \tilde{\boldsymbol{\beta}}^\top \mathbf{x}_i) - \mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right| \geq \varepsilon\right) \\
\leq & \frac{E\left[\left|\frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}(\mathbf{c}^\top \mathbf{x}_i \mathbf{x}_i^\top \mathbf{c}) I(y_i/U_0(1/(1-\tau_n)) \geq \tilde{\boldsymbol{\beta}}^\top \mathbf{x}_i) - \mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top]\mathbf{c}\right|^2\right]}{\varepsilon^2} \\
= & \frac{O((n(1-\tau_n))^{-1})}{\varepsilon^2} \\
= & o(1),
\end{aligned}
$$

which implies that as $n \to \infty$,

$$
\begin{aligned}
M_n(\tau_n) &= \left[I_p - \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top\right)^{-1} \frac{1}{n(1-\tau_n)}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i^\top I(y_i/U_0(1/(1-\tau_n)) \geq \tilde{\boldsymbol{\beta}}^\top \mathbf{x}_i)\right] \\
&= \left[I_p - \left(E[\mathbf{X}\mathbf{X}^\top] + o_{\mathbb{P}}(1)\right)^{-1}\left(E[\mathbf{X}\mathbf{X}^\top] + o_{\mathbb{P}}(1)\right)\right] \\
&= o_{\mathbb{P}}(1) I_p.
\end{aligned}
$$

This makes $T_{n2}(\tau_n)$ become the major term determining the asymptotic property for $\hat{\boldsymbol{\theta}}_n(\tau_n)$. Then, we can rewrite

$$
\begin{aligned}
&\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n) - \boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))}, \frac{\hat{\boldsymbol{\theta}}_n(\tau_n) - \boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)^\top \\
=&\begin{pmatrix} 1 & 0 \\ 0 & \left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1}\end{pmatrix}\begin{pmatrix}\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n) - \boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right) \\ T_{n2}(\tau_n)\end{pmatrix} + o_{\mathbb{P}}(1).
\end{aligned}
\tag{S2.10}
$$

Part IV: To find the joint normality, it remains to calculate the covariance of $\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n) - \boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)$ and $T_{n2}(\tau_n)$. By using Bahadur representation (we refer to Chapter 4.3 in Koenker (2005)), we have

$$\begin{aligned}
&\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)\\
=&\left(E\left[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right]\right)^{-1}\gamma\,\underbrace{\frac{1}{\sqrt{n(1-\tau_n)}}\sum_{i=1}^n\left(\tau_n-I(y_i\le\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)\right)\cdot\mathbf{x}_i}_{R_n(\tau_n)}+o_{\mathbb{P}}(1).
\end{aligned} \tag{S2.11}$$

Therefore, using Lemma S4 again, the covariance follows from that as $n\to\infty$,

$$\begin{aligned}
&\mathbf{c}^\top\mathrm{Cov}\left(R_n(\tau_n),T_{n2}(\tau_n)\right)\mathbf{c}\\
=&\frac{1}{n(1-\tau_n)U_0(1/(1-\tau_n))}\times\sum_{i=1}^n\mathrm{Cov}\left(\left(\tau_n-I(y_i\le\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)\right)(\mathbf{c}^\top\mathbf{x}_i),\right.\\
&\left.(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+(\mathbf{c}^\top\mathbf{x}_i)-E\left[(Y-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{X})_+(\mathbf{c}^\top\mathbf{X})\right]\right)+o(1)\\
=&\frac{1}{n(1-\tau_n)U_0(1/(1-\tau_n))}\times\sum_{i=1}^n\mathrm{Cov}\left(\tau_n(\mathbf{c}^\top\mathbf{x}_i),(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+(\mathbf{c}^\top\mathbf{x}_i)\right)\\
-&\frac{1}{n(1-\tau_n)U_0(1/(1-\tau_n))}\times\sum_{i=1}^n\mathrm{Cov}\left(I(y_i\le\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)(\mathbf{c}^\top\mathbf{x}_i),(y_i-\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i)_+(\mathbf{c}^\top\mathbf{x}_i)\right)+o(1)\\
=&\left\{\frac{\tau_n}{\gamma}B(2,1/\gamma-1)\mathbf{c}^\top\left[E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^\gamma]-E[\mathbf{X}]E[\mathbf{X}^\top g(\mathbf{X})^\gamma]\right]\mathbf{c}\right.\\
+&\left.\frac{\tau_nB(2,1/\gamma-1)}{\gamma}\mathbf{c}^\top E[\mathbf{X}]E[\mathbf{X}^\top g(\mathbf{X})^\gamma]\mathbf{c}\right\}+o(1)\\
=&\frac{B(2,1/\gamma-1)}{\gamma}\mathbf{c}^\top E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^\gamma]\mathbf{c}+o(1).
\end{aligned} \tag{S2.12}$$

Finally, combining the arguments in Proposition S1, (S2.9), (S2.10), and (S2.12) yields the final result that

$$\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))},\frac{\hat{\boldsymbol{\theta}}_n(\tau_n)-\boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)^\top \xrightarrow{d} N\left(\begin{pmatrix}\mathbf{0}_p\\ \mathbf{0}_p\end{pmatrix},\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}\right),$$

where

$$\begin{cases}\Sigma_{11} &= \gamma^2\left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1}E\left[\mathbf{X}\mathbf{X}^\top\right]\left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1},\\ \Sigma_{12} &= B(2,1/\gamma-1)\left(E\left[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right]\right)^{-1}E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma}]\left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1},\\ \Sigma_{21} &= \Sigma_{12}^\top,\\ \Sigma_{22} &= \frac{B(3,1/\gamma-2)}{\gamma}\left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1}E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}]\left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1}.\end{cases}$$

Thus, the proof is completed.

□

Proof of Theorem 1. This proof is a direct corollary of Proposition 2. It is straightforward to see that, by Lemma S1 and Proposition 1,

$$\begin{aligned}&\sqrt{n(1-\tau_n)}\left(\frac{\widehat{Q}_Y(\tau_n|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})}-1,\frac{\widehat{\mathrm{ES}}_Y(\tau_n|\mathbf{x})}{\mathrm{ES}_Y(\tau_n|\mathbf{x})}-1\right)^\top\\ =&\mathbf{x}^\top\sqrt{n(1-\tau_n)}\left(\frac{\hat{\boldsymbol{\beta}}_n(\tau_n)-\boldsymbol{\beta}_0(\tau_n)}{U_0(1/(1-\tau_n))},\frac{\hat{\boldsymbol{\theta}}_n(\tau_n)-\boldsymbol{\theta}_0(\tau_n)}{U_0(1/(1-\tau_n))}\right)^\top\\ &\times\begin{pmatrix}\left(\frac{U_Y(1/(1-\tau_n)|\mathbf{x})}{U_0(1/(1-\tau_n))}\right)^{-1} & 0\\ 0 & \left(\frac{U_Y(1/(1-\tau_n)|\mathbf{x})}{U_0(1/(1-\tau_n))}\right)^{-1}\left(\frac{\mathrm{ES}_Y(\tau|\mathbf{x})}{Q_Y(\tau|\mathbf{x})}\right)^{-1}\end{pmatrix}\\ \xrightarrow{d}&N\left(\begin{pmatrix}0\\0\end{pmatrix},\mathbf{x}^\top\begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{pmatrix}\mathbf{x}\right)\begin{pmatrix}g(\mathbf{x})^{-\gamma} & 0\\ 0 & (1-\gamma)g(\mathbf{x})^{-\gamma}\end{pmatrix}=N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}\sigma_{11}^2 & \sigma_{12}^2\\ \sigma_{21}^2 & \sigma_{22}^2\end{pmatrix}\right),\end{aligned}$$

where all elements can be computed as follows,

$$\begin{cases} \sigma_{11}^2 & = g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \Sigma_{11}\mathbf{x} \\ & = \gamma^2 g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1} E\left[\mathbf{X}\mathbf{X}^\top\right] \left(E\left[\mathbf{X}\mathbf{X}^\top g\left(\mathbf{X}\right)^{-\gamma}\right]\right)^{-1}\mathbf{x}, \\ \sigma_{12}^2 & = (1-\gamma)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \Sigma_{12}\mathbf{x} \\ & = (1-\gamma)B(2,1/\gamma-1)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \left(E\left[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right]\right)^{-1} E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma}] \left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1}\mathbf{x}, \\ \sigma_{21}^2 & = (1-\gamma)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \Sigma_{21}\mathbf{x} \\ & = (1-\gamma)B(2,1/\gamma-1)g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1} E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{\gamma}] \left(E\left[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{-\gamma}\right]\right)^{-1}\mathbf{x}, \\ \sigma_{22}^2 & = (1-\gamma)^2 g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \Sigma_{22}\mathbf{x} \\ & = \frac{(1-\gamma)^2 B(3,1/\gamma-2)}{\gamma} g(\mathbf{x})^{-2\gamma}\mathbf{x}^\top \left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1} E[\mathbf{X}\mathbf{X}^\top g(\mathbf{X})^{2\gamma}] \left(E[\mathbf{X}\mathbf{X}^\top]\right)^{-1}\mathbf{x}. \end{cases}$$

□

Proof of Proposition 3. **First**, for $i = 1$, note that $\tau_n = 1 - k/n$, from Theorem 1, we have

$$\frac{\widehat{Q}_Y(\tau_n \mid \mathbf{x})}{Q_Y(\tau_n \mid \mathbf{x})} = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

Then, combining (3.20), (3.18), and (2.13), we have that, by denoting symmetric matrix $M_n := \left(\frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top\right)^{-1}$,

$$
\begin{aligned}
&\frac{\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})}\\
=&1+\frac{1}{1-\tau_n}\frac{\left(\frac{1}{n}\sum_{i=1}^n(y_i-\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)I(y_i\geq\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)\cdot\mathbf{x}_i^\top\right)M_n\mathbf{x}}{\widehat{Q}_Y(\tau_n\,|\,\mathbf{x})}\\
=&1+\underbrace{\frac{1}{n(1-\tau_n)}\sum_{i=1}^n(y_i-\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}}_{P_n(\tau_n)}\left(1+O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right).
\end{aligned}
$$

Now we work on the term $P_n(\tau_n)$, which follows that

$$
\begin{aligned}
&P_n(\tau_n)\\
=&\frac{1}{n(1-\tau_n)}\sum_{i=1}^n(y_i-\hat{\boldsymbol{\beta}}_n(\tau_n)^\top\mathbf{x}_i)_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
=&\frac{1}{n(1-\tau_n)}\sum_{i=1}^n\left(\frac{y_i}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}-\frac{\widehat{Q}_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}\right)_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
=&\frac{1}{[n(1-\tau_n)]^{3/2}}\sum_{i=1}^n\sqrt{n(1-\tau_n)}\left\{\left(\frac{y_i}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}-\frac{\widehat{Q}_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}\right)_+-\left(\frac{y_i}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}-1\right)_+\right\}\\
&\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
&+\frac{1}{n(1-\tau_n)}\sum_{i=1}^n\left(\frac{y_i}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}-1\right)_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
=&\frac{-1}{[n(1-\tau_n)]^{3/2}}\sum_{i=1}^n\sqrt{n(1-\tau_n)}\left(\frac{\widehat{Q}_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}-1\right)I\left(\frac{y_i}{Q_Y(\tau_n\,|\,\mathbf{x}_i)}\geq\widetilde{Q}\right)\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
&+\frac{1}{n(1-\tau_n)}\sum_{i=1}^n\left(y_i-Q_Y(\tau_n\,|\,\mathbf{x}_i)\right)_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\\
:=&P_{n1}(\tau_n)+P_{n2}(\tau_n),
\end{aligned}
$$

where $\widetilde{Q}$ is an intermediate point between $\widehat{Q}_Y(\tau_n\,|\,\mathbf{x}_i)/Q_Y(\tau_n\,|\,\mathbf{x}_i)$ and 1, and it converges to 1 in probability as $n \to \infty$.

For the term $P_{n1}(\tau_n)$, using Theorem 1 again, a direct calculation yields,

$$
\begin{aligned}
E[P_{n1}(\tau_n)] &= \frac{O_{\mathbb{P}}(1)}{[n(1-\tau_n)]^{3/2}} E\left[\sum_{i=1}^{n} I\left(\frac{y_i}{Q_Y(\tau_n \,|\, \mathbf{x}_i)} \geq \widetilde{Q}\right) \cdot \mathbf{x}_i^\top M_n \mathbf{x} \frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right] \\
&= \frac{O_{\mathbb{P}}(1)}{[n(1-\tau_n)]^{3/2}} E\left[\sum_{i=1}^{n} E_i\left[I\left(\frac{y_i}{Q_Y(\tau_n \,|\, \mathbf{x}_i)} \geq \widetilde{Q}\right)\right] \mathbf{x}_i^\top M_n \mathbf{x} \frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right] \\
&= \frac{O_{\mathbb{P}}(1)}{\sqrt{n(1-\tau_n)}} E\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}_i \mathbf{x}_i^\top M_n \mathbf{x} \frac{1}{Q_Y(\tau_n \,|\, \mathbf{x})}\right] \\
&= \frac{O_{\mathbb{P}}(1)}{\sqrt{n(1-\tau_n)}} E\left[\boldsymbol{\beta}_0(\tau_n)^\top \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top M_n \mathbf{x} \frac{1}{Q_Y(\tau_n \,|\, \mathbf{x})}\right] \\
&= \frac{O_{\mathbb{P}}(1)}{\sqrt{n(1-\tau_n)}} E\left[\frac{\boldsymbol{\beta}_0(\tau_n)^\top \mathbf{x}}{Q_Y(\tau_n \,|\, \mathbf{x})}\right] \\
&= O_{\mathbb{P}}((n(1-\tau_n))^{-1/2}) \ (= O_{\mathbb{P}}(k^{-1/2})).
\end{aligned}
$$

Moreover, we also have that

$$
\begin{aligned}
\mathrm{Var}[P_{n1}(\tau_n)] =& \frac{O_{\mathbb{P}}(1)}{[n(1-\tau_n)]^{3}} \sum_{i=1}^{n} \mathrm{Var}\left(I\left(\frac{y_i}{Q_Y(\tau_n \,|\, \mathbf{x}_i)} \geq \widetilde{Q}\right) \cdot \mathbf{x}_i^\top M_n \mathbf{x} \frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right) \\
=& \frac{O_{\mathbb{P}}(1)}{[n(1-\tau_n)]^{3}} \left\{E\left[\sum_{i=1}^{n} E_i\left[I\left(\frac{y_i}{Q_Y(\tau_n \,|\, \mathbf{x}_i)} \geq \widetilde{Q}\right)\right] (\mathbf{x}_i^\top M_n \mathbf{x})^2 \left(\frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right)^2\right]\right. \\
&\left. - \sum_{i=1}^{n}\left(E\left[E_i\left[I\left(\frac{y_i}{Q_Y(\tau_n \,|\, \mathbf{x}_i)} \geq \widetilde{Q}\right)\right] \mathbf{x}_i^\top M_n \mathbf{x} \frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right]\right)^2\right\} \\
=& \frac{O_{\mathbb{P}}(1)}{[n(1-\tau_n)]^{2}} E\left[\frac{1}{n}\sum_{i=1}^{n} (\mathbf{x}_i^\top M_n \mathbf{x})^2 \left(\frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right)^2\right] \\
&- \frac{O_{\mathbb{P}}(1)}{n^3(1-\tau_n)} \sum_{i=1}^{n}\left(E\left[\mathbf{x}_i^\top M_n \mathbf{x} \frac{Q_Y(\tau_n \,|\, \mathbf{x}_i)}{Q_Y(\tau_n \,|\, \mathbf{x})}\right]\right)^2 \\
=& o_{\mathbb{P}}([n(1-\tau_n)]^{-2}) \ (= o_{\mathbb{P}}(k^{-2})),
\end{aligned}
$$

where the last step follows a direct use of (S2.1),

$$\begin{aligned}\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})} &= \frac{U_Y(1/(1-\tau_n)|\mathbf{x}_i)}{U_Y(1/(1-\tau_n)|\mathbf{x})}\\ &= \frac{U_Y(1/(1-\tau_n)|\mathbf{x}_i)/[U_0(1/(1-\tau_n))g(\mathbf{x}_i)^\gamma]}{U_Y(1/(1-\tau_n)|\mathbf{x})/[U_0(1/(1-\tau_n))g(\mathbf{x})^\gamma]}\,\frac{g(\mathbf{x}_i)^\gamma}{g(\mathbf{x})^\gamma}\\ &= O_{\mathbb{P}}(1)\cdot\left[\frac{g(\mathbf{x}_i)}{g(\mathbf{x})}\right]^\gamma.\end{aligned}$$

Thus, the above argument indicates that $P_{n1}(\tau_n) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$.

For the term $P_{n2}(\tau_n)$, it is straightforward to see that

$$\begin{aligned}E\left[P_{n2}(\tau_n)\right] &= E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^n (y_i - Q_Y(\tau_n\,|\,\mathbf{x}_i))_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right]\\ &= E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^n E_i\left[(y_i - Q_Y(\tau_n\,|\,\mathbf{x}_i))_+\right]\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right]\\ &= \frac{1}{\gamma}B(2,1/\gamma-1)E\left[\frac{1}{n(1-\tau_n)}\sum_{i=1}^n(1-\tau_n)\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}_i\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right](1+o(1))\\ &= \frac{1}{\gamma}B(2,1/\gamma-1)E\left[\boldsymbol{\beta}_0(\tau_n)^\top\frac{1}{n}\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right](1+o(1))\\ &= \frac{1}{\gamma}B(2,1/\gamma-1)E\left[\frac{\boldsymbol{\beta}_0(\tau_n)^\top\mathbf{x}}{Q_Y(\tau_n\,|\,\mathbf{x})}\right](1+o(1))\\ &= \frac{1}{\gamma}B(2,1/\gamma-1)(1+o(1)),\end{aligned}$$

as well as

$$
\begin{aligned}
&\operatorname{Var}[P_{n2}(\tau_n)]\\
=&\frac{1}{[n(1-\tau_n)]^2}\sum_{i=1}^{n}\left\{E\left[\left((y_i-Q_Y(\tau_n\,|\,\mathbf{x}_i))_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)^2\right]\right.\\
&\left.-\left(E\left[(y_i-Q_Y(\tau_n\,|\,\mathbf{x}_i))_+\cdot\mathbf{x}_i^\top M_n\mathbf{x}\frac{1}{Q_Y(\tau_n\,|\,\mathbf{x})}\right]\right)^2\right\}\\
=&\frac{1}{[n(1-\tau_n)]^2}\left\{E\left[\sum_{i=1}^{n}E_i\left[(y_i-Q_Y(\tau_n\,|\,\mathbf{x}_i))_+^2\right]\left(\frac{\mathbf{x}_i^\top M_n\mathbf{x}}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)^2\right]\right.\\
&\left.-\sum_{i=1}^{n}\left(E\left[E_i\left[(y_i-Q_Y(\tau_n\,|\,\mathbf{x}_i))_+\right]\left(\frac{\mathbf{x}_i^\top M_n\mathbf{x}}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)\right]\right)^2\right\}\\
=&\frac{K}{[n(1-\tau_n)]^2}\left\{E\left[\sum_{i=1}^{n}(1-\tau_n)Q_Y(\tau_n\,|\,\mathbf{x}_i)^2\left(\frac{\mathbf{x}_i^\top M_n\mathbf{x}}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)^2\right]\right.\\
&\left.-\sum_{i=1}^{n}\left(E\left[(1-\tau_n)Q_Y(\tau_n\,|\,\mathbf{x}_i)\left(\frac{\mathbf{x}_i^\top M_n\mathbf{x}}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)\right]\right)^2\right\}\\
=&\frac{K}{n(1-\tau_n)}E\left[\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i^\top M_n\mathbf{x})^2\left(\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)^2\right]-\frac{K}{n^2}\sum_{i=1}^{n}\left(E\left[(\mathbf{x}_i^\top M_n\mathbf{x})\left(\frac{Q_Y(\tau_n\,|\,\mathbf{x}_i)}{Q_Y(\tau_n\,|\,\mathbf{x})}\right)\right]\right)^2\\
=&O((n(1-\tau_n))^{-1})\ (=O(k^{-1})),
\end{aligned}
$$

where K is a uniform constant. This implies that $P_{n2}(\tau_n)=\frac{1}{1-\gamma}-1+o_{\mathbb{P}}(1)$ by noting $B(2,1/\gamma-1)=\frac{\gamma}{1-\gamma}-\gamma$.

Second, we only consider the case for $i=3$. Following the previous proof of Hill estimator in Example 5.1.5 in de Haan and Ferreira (2006), we can conclude that there exists a standard Brownian motion $W(\cdot)$ on $[0,1]$

such that,

$$\sqrt{k}(\hat{\gamma}-\gamma) \xrightarrow{\mathbb{P}} \gamma\left(-W(1)+\int_0^1 W(s)\frac{ds}{s}\right) =: \Gamma. \tag{S2.13}$$

Thus, Γ is a normal random variable with mean zero and variance γ^2. By applying the delta method, we conclude that

$$\sqrt{k}\left(\frac{1}{1-\hat{\gamma}}-\frac{1}{1-\gamma}\right)-\frac{\Gamma}{(1-\gamma)^2} \xrightarrow{\mathbb{P}} 0,$$

which means

$$\frac{1}{1-\hat{\gamma}} = \frac{1}{1-\gamma}+\frac{1}{\sqrt{k}}\frac{\Gamma}{(1-\gamma)^2}+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right). \tag{S2.14}$$

Then, we note that by (3.24),

$$\frac{\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})} = \frac{1}{1-\hat{\gamma}} = \frac{1}{1-\gamma}+\frac{1}{\sqrt{k}}\frac{\Gamma}{(1-\gamma)^2}+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).$$

Thus, the proof is completed. □

Proof of Theorem 2. **First**, we consider the case for $i=1$. Recall that $\tau_n = 1-k/n$ and we note the following decomposition,

$$\frac{\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} = d_n^{\hat{\gamma}-\gamma}\times\frac{\widehat{\mathrm{ES}}_Y(1-k/n|\mathbf{x})}{\mathrm{ES}_Y(1-k/n|\mathbf{x})}\times d_n^{\gamma}\frac{\mathrm{ES}_Y(1-k/n|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} := A_1\times A_2\times A_3.$$

For the term A_1, (S2.13) leads to that as $n\to\infty$,

$$\frac{\sqrt{k}}{\log d_n}(d_n^{\hat{\gamma}-\gamma}-1)-\Gamma \xrightarrow{\mathbb{P}} 0, \tag{S2.15}$$

implying that $A_1 = 1+\frac{\log d_n}{\sqrt{k}}\Gamma+o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{k}}\right)$.

For the term A_2, the $\widehat{\mathrm{ES}}_Y(1-k/n|\mathbf{x})$ part in binary normality of Theorem 1 shows that,

$$\sqrt{k}\left(\frac{\widehat{\mathrm{ES}}_Y(1-k/n|\mathbf{x})}{\mathrm{ES}_Y(1-k/n|\mathbf{x})}-1\right)\xrightarrow{d} N(0,\sigma_{22}^2), \tag{S2.16}$$

where σ_{22}^2 has already been displayed in Theorem 1. This indicates that $A_2 = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$.

For the term A_3, using Proposition 1 and Lemma S1, we have that,

$$A_3 = \frac{\mathrm{ES}_Y(1-k/n|\mathbf{x})}{Q_Y(1-k/n|\mathbf{x})}\frac{Q_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}d_n^{\gamma}\frac{Q_Y(1-k/n|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})} = 1+O\left(A_1\left(\frac{n}{k}\right)\right) = 1+o\left(\frac{1}{\sqrt{k}}\right). \tag{S2.17}$$

Combining all terms A_1, A_2 and A_3, it follows that,

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} &= \left(1+\frac{\log d_n}{\sqrt{k}}\Gamma + o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{k}}\right)\right)\times\left(1+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right)\times\left(1+o\left(\frac{1}{\sqrt{k}}\right)\right)\\ &= 1+\frac{\log d_n}{\sqrt{k}}\Gamma + o_{\mathbb{P}}\left(\frac{\log d_n}{k}\right)+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right),\end{aligned}$$

which completes the proof of this case by

$$\frac{\sqrt{k}}{\log d_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(1)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}-1\right) = \Gamma + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)+o_{\mathbb{P}}\left(\frac{1}{\log d_n}\right)\xrightarrow{d} N(0,\gamma^2).$$

Second, we only consider the case for $i=3$ while the case for $i=2$ is quite the same. We still note the following decomposition,

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} &= \frac{1-\gamma}{1-\hat{\gamma}}\times d_n^{\hat{\gamma}-\gamma}\times\frac{\widehat{Q}_Y(1-k/n|\mathbf{x})}{Q_Y(1-k/n|\mathbf{x})}\times d_n^{\gamma}\frac{Q_Y(1-k/n|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})}\times\frac{Q_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})(1-\gamma)}\\ &:= B_1\times B_2\times B_3\times B_4\times B_5.\end{aligned}$$

For the term B_1, a combination of Delta-method and (S2.13) indicates that,

$$\sqrt{k}\left(\frac{1-\gamma}{1-\hat{\gamma}}-1\right)+\frac{1}{\gamma-1}\Gamma \xrightarrow{\mathbb{P}} 0, \tag{S2.18}$$

which is equivalent to $B_1 = 1 + \frac{1}{\sqrt{k}}\frac{1}{1-\gamma}\Gamma + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$.

For the term B_2, we have already shown that $B_2 = 1 + \frac{\log d_n}{\sqrt{k}}\Gamma + o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{k}}\right)$.

For the term B_3, the $\widehat{Q}_Y(1-k/n|\mathbf{x})$ part in binary normality of Theorem 1 also shows that,

$$\sqrt{k}\left(\frac{\widehat{Q}_Y(1-k/n|\mathbf{x})}{Q_Y(1-k/n|\mathbf{x})}-1\right) \xrightarrow{d} N(0,\sigma_{11}^2), \tag{S2.19}$$

where σ_{11}^2 can been still found in Theorem 1. This indicates that $B_3 = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)$.

For the term B_4, it follows that, by Lemma S1,

$$B_4 = d_n^{\gamma}\frac{Q_Y(1-k/n|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})} = 1 + O\left(A_1\left(\frac{n}{k}\right)\right) = 1 + o\left(\frac{1}{\sqrt{k}}\right).$$

For the term B_5, using Proposition 1 again, we have that,

$$B_5 = \frac{Q_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})(1-\gamma)} = \left((1-\gamma)\frac{1}{1-\gamma} + O(A_1(1/(1-\tau_n')))\right)^{-1} = 1+o(1).$$

Combining all terms B_1, B_2, B_3, B_4, and B_5, it follows that,

$$
\begin{aligned}
&\frac{\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}\\
=&\left(1+\frac{1}{\sqrt{k}}\frac{1}{1-\gamma}\Gamma+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right)\left(1+\frac{\log d_n}{\sqrt{k}}\Gamma+o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{k}}\right)\right)\\
&\times\left(1+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right)\right)\left(1+o\left(\frac{1}{\sqrt{k}}\right)\right)(1+o(1))\\
=&1+\frac{\log d_n}{\sqrt{k}}\Gamma+\frac{1}{\sqrt{k}}\frac{\Gamma}{1-\gamma}+\frac{\log d_n}{k}\frac{\Gamma^2}{1-\gamma}+o_{\mathbb{P}}\left(\frac{\log d_n}{\sqrt{k}}\right)+o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right).
\end{aligned}
$$

Therefore, the proof is completed by noting that

$$
\frac{\sqrt{k}}{\log d_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(3)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}-1\right)=\Gamma+\frac{1}{\log d_n}\frac{\Gamma}{1-\gamma}+\frac{1}{\sqrt{k}}\frac{\Gamma^2}{1-\gamma}+O_{\mathbb{P}}(1)+o_{\mathbb{P}}\left(\frac{1}{\log d_n}\right)\xrightarrow{d}N(0,\gamma^2).
$$

□

Proof of Proposition 4. By (2.5) and (3.28), we have that

$$
Q_Y(\omega(\tau)|\mathbf{x})=Q_Y(\tau|\mathbf{x})+\frac{1}{1-\tau}E\left[(Y-Q_Y(\tau|\mathbf{x}))_+|\mathbf{X}=\mathbf{x}\right],
$$

which implies that by applying integration by part and change of variables,

$$
\begin{aligned}
\frac{Q_Y(\omega(\tau)|\mathbf{x})}{Q_Y(\tau|\mathbf{x})}&=1+\frac{1}{1-\tau}E\left[\left(\frac{Y}{Q_Y(\tau|\mathbf{x})}-1\right)_+\middle|\mathbf{X}=\mathbf{x}\right]\\
&=1+\frac{1}{1-\tau}\int_{Q_Y(\tau|\mathbf{x})}^{\infty}\frac{y}{Q_Y(\tau|\mathbf{x})}-1\,dF_Y(y|\mathbf{x})\\
&=1+\frac{1}{1-\tau}\int_1^{\infty}1-F_Y(tQ_Y(\tau|\mathbf{x})|\mathbf{x})\,dt\\
&=1+\int_1^{\infty}\frac{1-F_Y(tQ_Y(\tau|\mathbf{x})|\mathbf{x})}{1-F_Y(Q_Y(\tau|\mathbf{x})|\mathbf{x})}\,dt.
\end{aligned}
$$

Based on (S2.2), by using regular variation on both sides of above formulation, we have that, as $\tau \to 1$,

$$\lim_{\tau\to 1}\left(\frac{1-\tau}{1-\omega(\tau)}\right)^{\gamma} = \frac{1}{1-\gamma}.$$

This leads to the final result. $\square$

Before proving Proposition 5 and Theorem 3, we present Proposition S2 to state asymptotic property for $\widehat{Q}_Y(\hat{\omega}(\tau_n)|\mathbf{x})$ in (3.31). Notice that Proposition S2 follows from Theorem 1 in Ji et al. (2024) and Proposition 4 in our paper.

Proposition S2. *Suppose Assumption 1 hold with* $0 < \gamma < 1/2$. *If* $\widetilde{k} = o(k)$ *and* $\sqrt{\widetilde{k}}A_1(n/\widetilde{k}) = o(1)$ *as* $n \to \infty$, *then*

$$\sqrt{\widetilde{k}}\left(\frac{\widehat{Q}_Y(\hat{\omega}(\tau_n)|\mathbf{x})}{Q_Y(\omega(\tau_n)|\mathbf{x})} - 1\right) \xrightarrow{d} N\left(0, \sigma^2_{\gamma,g}(\mathbf{x})\right), \tag{S2.20}$$

where $\widehat{Q}_Y(\hat{\omega}(\tau_n)|\mathbf{x})$ *is estimated with the two-step procedure* (2.11) *and* (2.14) *by substituting the estimated level* $\hat{\omega}(\tau_n)$ *in* (3.29), *and*

$$\sigma^2_{\gamma,g}(\mathbf{x}) = \frac{\gamma^3}{(1-\gamma)g(\mathbf{x})^{2\gamma}}\mathbf{x}^T\left(E\left(\mathbf{X}\mathbf{X}^T g(\mathbf{X})^{-\gamma}\right)\right)^{-1} E\left(\mathbf{X}\mathbf{X}^T\right)\times\left(E\left(\mathbf{X}\mathbf{X}^T g(\mathbf{X})^{-\gamma}\right)\right)^{-1}\mathbf{x}.$$

Proof of Proposition 5. By (3.31), we have that,

$$\begin{aligned}
\frac{\widehat{\mathrm{ES}}^{(4)}_Y(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})} &= \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}}\frac{\widehat{Q}_Y\left(\hat{\omega}(\tau_n)|\mathbf{x}\right)}{\widetilde{Q}_Y\left(\tau_n'|\mathbf{x}\right)} \\
&= \frac{\widehat{Q}_Y\left(\hat{\omega}(\tau_n)|\mathbf{x}\right)}{Q_Y\left(\omega(\tau_n)|\mathbf{x}\right)} \times \frac{Q_Y\left(\omega(\tau_n)|\mathbf{x}\right)}{Q_Y\left(\tau_n|\mathbf{x}\right)} \times \frac{Q_Y\left(\tau_n|\mathbf{x}\right)}{\widehat{Q}_Y\left(\tau_n|\mathbf{x}\right)} \times \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}}\frac{\widehat{Q}_Y\left(\tau_n|\mathbf{x}\right)}{\widetilde{Q}_Y\left(\tau_n'|\mathbf{x}\right)} \\
&:= C_1 \times C_2 \times C_3 \times C_4.
\end{aligned}$$

For the term C_1, Proposition S2 indicates that, $C_1 = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\tilde{k}}}\right)$.

For the term C_2, it follows that, by (3.28),

$$C_2 = \frac{Q_Y\left(\omega(\tau_n)|\mathbf{x}\right)}{Q_Y\left(\tau_n|\mathbf{x}\right)} = \frac{Q_Y\left(1-(1-\tau_n)(1-\gamma)^{1/\gamma}\middle|\,\mathbf{x}\right)}{Q_Y(\tau_n|\mathbf{x})} = \frac{1}{1-\gamma} + o\left(\frac{1}{\sqrt{\widetilde{k}}}\right).$$

For the term C_3, the $\widehat{Q}_Y(1-\tilde{k}/n|\mathbf{x})$ part in binary normality of Theorem 1 also shows that,

$$\sqrt{\tilde{k}}\left(\frac{\widehat{Q}_Y(1-\tilde{k}/n|\mathbf{x})}{Q_Y(1-\tilde{k}/n|\mathbf{x})} - 1\right) \xrightarrow{d} N(0,\sigma_{11}^2), \tag{S2.21}$$

where σ_{11}^2 can been still found in Theorem 1. This indicates that $C_3 = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\tilde{k}}}\right)$.

For the term C_4, it follows that, by (3.18),

$$C_4 = \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \frac{\widehat{Q}_Y\left(\tau_n|\mathbf{x}\right)}{\widetilde{Q}_Y\left(\tau_n'|\mathbf{x}\right)} = \left(\frac{1-\tau_n}{1-\tau_n'}\right)^{\hat{\gamma}} \times \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{\hat{\gamma}} = 1.$$

Combining all terms C_1, C_2, C_3 and C_4, it follows that,

$$\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\widetilde{Q}_Y(\tau_n'\,|\,\mathbf{x})} = \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{k}}}\right)\right) \times \frac{1}{1-\gamma} \times \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{k}}}\right)\right) = \frac{1}{1-\gamma} + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{k}}}\right).$$

Then, the proof is completed. □

Proof of Theorem 3. Recall that $\widetilde{k} = n(1-\tau_n)$. Notice that,

$$\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} = \widetilde{d}_n^{\hat{\gamma}-\gamma} \times \frac{\widehat{Q}_Y(\hat{\omega}(\tau_n)|\mathbf{x})}{Q_Y(\omega(\tau_n)|\mathbf{x})} \times \widetilde{d}_n^{\gamma}\frac{\mathrm{ES}_Y(\tau_n|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} := D_1 \times D_2 \times D_3.$$

For term D_1, as discussed previously, we have $D_1 = 1 + \frac{\log \widetilde{d}_n}{\sqrt{k}}\Gamma + o_{\mathbb{P}}\left(\frac{\log \widetilde{d}_n}{\sqrt{k}}\right)$.

For term D_2, by Proposition S2, we have that as $n \to \infty$, $D_2 = 1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{\widetilde{k}}}}\right)$.

For term D_3, a similar argument to (S2.17) implies that

$$D_3 = \frac{\mathrm{ES}_Y(\tau_n|\mathbf{x})}{Q_Y(\tau_n|\mathbf{x})}\frac{Q_Y(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})}\widetilde{d}_n^{\gamma}\frac{Q_Y(\tau_n|\mathbf{x})}{Q_Y(\tau_n'|\mathbf{x})} = 1 + O\left(A_1\left(\frac{n}{\widetilde{\widetilde{k}}}\right)\right). \quad \text{(S2.22)}$$

Combining all terms D_1, D_2 and D_3, it follows that,

$$\begin{aligned}\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} &= \left(1 + \frac{\log \widetilde{d}_n}{\sqrt{k}}\Gamma + o_{\mathbb{P}}\left(\frac{\log \widetilde{d}_n}{\sqrt{k}}\right)\right) \times \left(1 + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{\widetilde{k}}}}\right)\right) \times \left\{1 + O\left(A_1\left(\frac{n}{\widetilde{\widetilde{k}}}\right)\right)\right\} \\ &= 1 + \frac{\log \widetilde{d}_n}{\sqrt{k}}\Gamma + O_{\mathbb{P}}\left(\frac{1}{\sqrt{\widetilde{\widetilde{k}}}}\right) + O\left(A_1\left(\frac{n}{\widetilde{\widetilde{k}}}\right)\right),\end{aligned}$$

thus,

$$\frac{\sqrt{k}}{\log \widetilde{d}_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} - 1\right) = \Gamma + O_{\mathbb{P}}\left(\frac{\sqrt{k}}{\log \widetilde{d}_n \sqrt{\widetilde{\widetilde{k}}}}\right) + O\left(\frac{\sqrt{k}}{\log \widetilde{d}_n}A_1\left(\frac{n}{\widetilde{\widetilde{k}}}\right)\right).$$

Recall that $\log \widetilde{d}_n \to \infty$, $\lim_{n\to\infty} \sqrt{\widetilde{k}}A_1(n/\widetilde{k}) = 0$, and $(\widetilde{k}/k)^{1/2} \log \widetilde{d}_n \to \infty$, as $n \to \infty$, we have that $\frac{\sqrt{k}}{\log \widetilde{d}_n}\frac{1}{\sqrt{\widetilde{\widetilde{k}}}} \to 0$ and $\frac{\sqrt{k}}{\log \widetilde{d}_n}A_1\left(\frac{n}{\widetilde{\widetilde{k}}}\right) \to 0$. Hence

$$\frac{\sqrt{k}}{\log \widetilde{d}_n}\left(\frac{\widehat{\mathrm{ES}}_Y^{(4)}(\tau_n'|\mathbf{x})}{\mathrm{ES}_Y(\tau_n'|\mathbf{x})} - 1\right) \xrightarrow{d} N(0, \gamma^2),$$

which completes the proof. $\square$

S3 Supplementary explanation for (2.5) and (2.10)

In this section, we provide additional explanation concerning the derivation of the closed-form expressions for $\mathrm{ES}_Y(\tau|\mathbf{x})$ and the corresponding regression coefficient $\boldsymbol{\theta}_0(\tau)$ as presented in (2.5) and (2.10). As explained in Gneiting (2011), the expected shortfall is not elicitable, which implies that it can not be expressed as a solution to a minimization problem. This characteristic poses a challenge in developing statistical inference methods for conditional ES within a regression framework. Although ES is not elicitable on its own, Fissler and Ziegel (2016) has shown that it is jointly elicitable with the quantile by using a class of strictly consistent joint loss functions. Given a level $\tau \in (0,1)$, the conditional quantile and ES are modeled by,

$$Q_Y(\tau|\mathbf{x}) = \boldsymbol{\beta}_0^\top \mathbf{x}, \text{ and } \mathrm{ES}_Y(\tau|\mathbf{x}) = \boldsymbol{\theta}_0^\top \mathbf{x}, \tag{S3.23}$$

where the coefficients depend on τ and we simplify them as $\boldsymbol{\beta}_0$ and $\boldsymbol{\theta}_0$ herein. The joint loss function proposed in Fissler and Ziegel (2016) gives,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; Y, \mathbf{x}) := &\left\{(1-\tau) - I(Y \geq \boldsymbol{\beta}^\top \mathbf{x})\right\}\left\{G_1(Y) - G_1(\boldsymbol{\beta}^\top \mathbf{x})\right\} \\ &- G_2(\boldsymbol{\theta}^\top \mathbf{x})\left\{\underbrace{(\boldsymbol{\beta}-\boldsymbol{\theta})^\top \mathbf{x} + \frac{(Y - \boldsymbol{\beta}^\top \mathbf{x})}{1-\tau} I(Y \geq \boldsymbol{\beta}^\top \mathbf{x})}_{\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; Y, \mathbf{x})}\right\} - \mathcal{G}_2(\boldsymbol{\theta}^\top \mathbf{x}),\end{aligned} \tag{S3.24}$$

where G_1 is an increasing and integrable function, $\mathcal{G}_2$ is a three times continuously differentiable function such that both $\mathcal{G}_2$ and its derivative $G_2 = \mathcal{G}_2'$ are strictly positive. Based on (S3.24), Dimitriadis and Bayer (2019) proposed an M-estimator, defined as the global minimum of any member of this joint loss functions class over some compact set $\boldsymbol{\Theta}$,

$$(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^\top \in \underset{\boldsymbol{\beta},\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; y_i, \mathbf{x}_i), \tag{S3.25}$$

where the sample $\{(y_i, \mathbf{x}_i)\}_{i=1}^n$ are drawn independently from the distribution of $(\mathbf{X}, Y)$. However, the objective function $\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}; Y, \mathbf{x})$ is non-differentiable and non-convex for any choice of the functions G_1 and G_2, which makes the optimization more challenging computationally.

In the model (S3.23), if our goal is to estimate ES, then $\boldsymbol{\beta}_0$ can be viewed as a nuisance parameter. Motivated by the idea of using Neyman-orthogonal scores to reduce sensitivity with respect to nuisance parameters (Neyman (1979), Chernozhukov et al. (2018)), Barendse (2020) proposed to employ a two-step procedure, bypassing the non-convexity optimization problems. In the first step, an estimator $\hat{\boldsymbol{\beta}}_n$ of $\boldsymbol{\beta}_0$ is obtained via standard quantile regression. The second step employs an orthogonal score with fitted quantiles to estimate $\boldsymbol{\theta}_0$. The key observation is as follows. Define

$$\begin{aligned}\Phi(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{X}) &= E[\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; Y, \mathbf{X})|\mathbf{X}] \\ &= \left(1 - \frac{P(Y \geq \boldsymbol{\beta}^\top \mathbf{X}|\mathbf{X})}{1-\tau}\right)\boldsymbol{\beta}^\top \mathbf{X} - \boldsymbol{\theta}^\top \mathbf{X} + E[Y|Y \geq \boldsymbol{\beta}^\top \mathbf{X}, \mathbf{X}],\end{aligned}$$

where $\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; Y, \mathbf{X})$ is defined in (S3.24). In fact, $\Phi(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{X})$ is a Neyman-orthogonal score function with respect to $\boldsymbol{\beta}$. It is because that, as shown in He et al. (2023), it satisfies $\frac{\partial}{\partial \boldsymbol{\beta}}\Phi(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{X})|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} = 0$ for any $\boldsymbol{\theta}$. This implies that the function is insensitive to small perturbations in $\boldsymbol{\beta}$. Besides, under model (S3.23), it follows that $\Phi(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0; \mathbf{X}) = 0$ almost surely over $\mathbf{X}$, as well as $E[\mathbf{X}\Phi(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0; \mathbf{X})] = 0$. Therefore, given true parameter $\boldsymbol{\beta}_0$, this property motivates us to define the true coefficient $\boldsymbol{\theta}_0$ via a least square method,

$$\begin{aligned}\boldsymbol{\theta}_0 &= \arg\min_{\boldsymbol{\theta}} E[\ell^2(\boldsymbol{\beta}_0, \boldsymbol{\theta}; Y, \mathbf{X})] \\ &= \arg\min_{\boldsymbol{\theta}} E[\phi_\tau(Y - \boldsymbol{\theta}^\top \mathbf{X}, Y - \boldsymbol{\beta}_0^\top \mathbf{X})],\end{aligned} \tag{S3.26}$$

where $\phi_\tau(u, v) = \left(u - v + \frac{v}{1-\tau} I(v \geq 0)\right)^2$. It is easy to see that $\phi_\tau(Y - \boldsymbol{\theta}^\top \mathbf{X}, Y - \boldsymbol{\beta}_0^\top \mathbf{X})$ is convex on $\boldsymbol{\theta}$ and taking derivative on the right-hand side of above objective function yields that

$$\begin{aligned}&-2E\left[\left((\boldsymbol{\beta}_0 - \boldsymbol{\theta}_0)^\top \mathbf{X} + \frac{1}{1-\tau}(Y - \boldsymbol{\beta}_0^\top \mathbf{X}) I(Y \geq \boldsymbol{\beta}_0^\top \mathbf{X})\right)\mathbf{X}\right] \\ =&-2E\left[E\left[\left((\boldsymbol{\beta}_0 - \boldsymbol{\theta}_0)^\top \mathbf{X} + \frac{1}{1-\tau}(Y - \boldsymbol{\beta}_0^\top \mathbf{X}) I(Y \geq \boldsymbol{\beta}_0^\top \mathbf{X})\right)\mathbf{X}\middle|\mathbf{X}\right]\right] \\ =&-2E\left[\mathbf{X}E\left[\left((\boldsymbol{\beta}_0 - \boldsymbol{\theta}_0)^\top \mathbf{X} + \frac{1}{1-\tau}(Y - \boldsymbol{\beta}_0^\top \mathbf{X}) I(Y \geq \boldsymbol{\beta}_0^\top \mathbf{X})\right)\middle|\mathbf{X}\right]\right] \\ =&-2E[\mathbf{X}\Phi(\boldsymbol{\beta}_0, \boldsymbol{\theta}_0; \mathbf{X})] = 0,\end{aligned}$$

which implies a closed-form expression for $\boldsymbol{\theta}_0$,

$$\boldsymbol{\theta}_0 = \boldsymbol{\beta}_0 + \frac{1}{1-\tau}(E[\mathbf{X}\mathbf{X}^\top])^{-1} \cdot E[(Y - \boldsymbol{\beta}_0^\top \mathbf{X})_+ \cdot \mathbf{X}], \tag{S3.27}$$

aligning with (2.10).

If we replace $\boldsymbol{\beta}_0^\top \mathbf{x}$ and $\boldsymbol{\theta}_0^\top \mathbf{x}$ by $Q_Y(\tau|\mathbf{x})$ and $\mathrm{ES}_Y(\tau|\mathbf{x})$, then (S3.26) can be reformulated as a optimization problem for $\mathrm{ES}_Y(\tau|\mathbf{x})$, that is,

$$\mathrm{ES}_Y(\tau|\mathbf{x}) = \arg\min_{e\in\mathbb{R}} E[\phi_\tau(Y-e, Y-Q_Y(\tau|\mathbf{x}))|\mathbf{X}=\mathbf{x}].$$

Taking derivative on the right-hand side yields that,

$$\mathrm{ES}_Y(\tau|\mathbf{x}) = Q_Y(\tau|\mathbf{x}) + \frac{1}{1-\tau}E[(Y-Q_Y(\tau|\mathbf{x}))_+|\mathbf{X}=\mathbf{x}],$$

which is consistent with (2.5).

Bibliography

de Haan, L., and Ferreira, A. (2006). *Extreme value theory: an introduction*, Springer.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *Econom. J.* **21**, C1-C68.

Ji, J., Hou, Y., and Li, D. (2024). Estimation of extreme conditional expectile via tail risk transition from intermediate quantile. *Stat. Its Interface*, forthcoming.

Koenker, R. (2005). *Quantile regression*, Cambridge University Press.

Neyman, J. (1979). $C(\alpha)$ tests and their use. *Sankhyā Ser. A*, 1-21.

Xu, W., Hou, Y., and Li, D. (2022). Prediction of extremal expectile based on regression models with heteroscedastic extremes. *J. Bus. Econ. Stat.* **40**, 522-536.