Supplementary Material for "Leveraging Unlabeled Data for Superior ROC Curve Estimation via a Semiparametric Approach"

Menghua Zhang, Mengjiao Peng and Yong Zhou*

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

School of Statistics, Academy of Statistics and Interdisciplinary Sciences,

East China Normal University, Shanghai, China

December 30, 2024

The supplement is organized as follows. Section A contains the deferred simulation results. In Section B, we apply our method to the California Housing Dataset to analyze the relationship between the house price and its location. In Section C, we present additional simulation results. In Section D, we discuss how to construct semi-supervised estimators when both X and Y come from the same family. In Section E, we extend our results to the MAR setting, present additional simulation results and establish corresponding theoretical conclusions. In Section F, we provide the detailed proofs of Lemma 3.1, Theorem 3.1, Theorem 3.2 and Theorem 3.3.

*Co-corresponding Authors: Yong Zhou, *E-mail address: yzhou@amss.ac.cn* and Mengjiao Peng, *E-mail address: mjpeng@fem.ecnu.edu.cn*

A Simulation Results in the Main Paper

Owing to space constraints, the simulation results for Models (b) and (d), as described in Section 4 of the main paper, are provided in Tables S1 through S3 herein. Figure S1 presents the performance of the SS estimators across the entire ROC curve. Please refer to the Section 4 for more details.

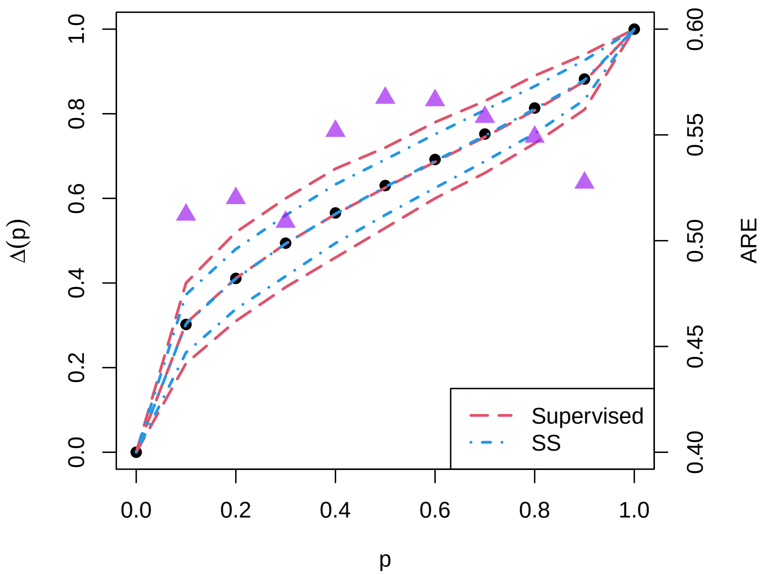

Figure S1: Average of estimators when $n = 100$, $N = 500$, $m = 200$ under setting (a). The black points are the true value. The red long dashed lines represent the mean and 95% confidence interval of the supervised estimator. The blue dotdash lines represent the mean and 95% confidence interval of the SS estimator. The purple triangles are ARE at different p.

Table S1: The results in setting (b) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	200	0	0.0078	0.0474	0.0456	0.0023	-	93.9	0.0077	0.0407	0.0395	0.0017	-	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200	0.0067	0.0431	0.0417	0.0019	17.72	93.5	0.0066	0.0362	0.0349	0.0013	21.49	94.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200	0.0057	0.0434	0.0417	0.0019	17.13	93.4	0.0056	0.0365	0.0349	0.0014	20.94	93.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	400	0.0062	0.0413	0.0403	0.0017	24.36	93.3	0.0063	0.0343	0.0332	0.0012	29.30	93.4
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	400	0.0055	0.0415	0.0403	0.0018	24.03	93.5	0.0056	0.0345	0.0332	0.0012	29.02	93.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	1000	0.0058	0.0393	0.0389	0.0016	31.50	94.1	0.0059	0.0323	0.0315	0.0011	37.49	94.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	1000	0.0054	0.0394	0.0389	0.0016	31.35	93.9	0.0056	0.0323	0.0315	0.0011	37.39	94.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	20000	0.0057	0.0390	0.0375	0.0015	32.83	93.6	0.0059	0.0313	0.0297	0.0010	40.96	92.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	20000	0.0057	0.0390	0.0375	0.0015	32.82	93.6	0.0059	0.0313	0.0297	0.0010	40.95	92.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200000	0.0058	0.0388	0.0374	0.0015	33.26	93.7	0.0060	0.0312	0.0296	0.0010	41.36	92.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200000	0.0058	0.0388	0.0374	0.0015	33.26	93.7	0.0060	0.0312	0.0296	0.0010	41.36	92.7
$\widehat{\Delta}_{\mathcal{L}}$	500	0	0.0080	0.0395	0.0388	0.0016	-	94.4	0.0070	0.0315	0.0310	0.0010	-	94.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500	0.0072	0.0374	0.0369	0.0014	10.72	95.2	0.0063	0.0288	0.0287	0.0009	16.43	94.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500	0.0064	0.0376	0.0369	0.0015	10.54	95.0	0.0055	0.0290	0.0287	0.0009	16.55	94.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	1000	0.0071	0.0368	0.0363	0.0014	13.61	94.7	0.0062	0.0282	0.0278	0.0008	20.17	94.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	1000	0.0066	0.0369	0.0363	0.0014	13.52	94.7	0.0057	0.0282	0.0278	0.0008	20.31	94.4
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	2500	0.0072	0.0360	0.0356	0.0013	17.07	94.8	0.0064	0.0273	0.0269	0.0008	24.52	94.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	2500	0.0070	0.0360	0.0356	0.0013	17.04	94.8	0.0061	0.0274	0.0269	0.0008	24.62	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	50000	0.0070	0.0352	0.0350	0.0013	20.89	95.0	0.0061	0.0261	0.0261	0.0007	30.97	93.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	50000	0.0070	0.0352	0.0350	0.0013	20.89	95.0	0.0061	0.0261	0.0261	0.0007	30.97	93.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500000	0.0070	0.0351	0.0349	0.0013	20.94	94.7	0.0061	0.0261	0.0261	0.0007	31.17	93.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500000	0.0070	0.0351	0.0349	0.0013	20.94	94.7	0.0061	0.0261	0.0261	0.0007	31.17	93.2

Table S2: The results of Model (c) with different (m, n, N), including the Bias , SD, SE, MSE, ARE (%) and CP (%).

			m = 100						m = 200					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	200	0	0.0002	0.0345	0.0364	0.0012	-	96.5	0.0003	0.0345	0.0359	0.0012	-	95.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200	0.0002	0.0256	0.0273	0.0007	45.03	95.8	0.0001	0.0252	0.0266	0.0006	46.52	95.6
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200	0.0002	0.0258	0.0273	0.0007	44.27	95.7	0.0001	0.0254	0.0266	0.0006	45.67	95.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	400	0.0008	0.0231	0.0235	0.0005	55.25	95.8	0.0008	0.0225	0.0226	0.0005	57.27	95.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	400	0.0008	0.0232	0.0235	0.0005	54.84	95.7	0.0008	0.0226	0.0226	0.0005	56.81	95.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	1000	0.0003	0.0184	0.0189	0.0003	71.44	94.9	0.0004	0.0177	0.0178	0.0003	73.61	94.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	1000	0.0003	0.0185	0.0189	0.0003	71.27	94.8	0.0004	0.0178	0.0178	0.0003	73.45	94.6
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	20000	0.0000	0.0118	0.0131	0.0001	88.22	96.5	0.0002	0.0108	0.0115	0.0001	90.23	94.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	20000	0.0000	0.0119	0.0131	0.0001	88.22	96.5	0.0002	0.0108	0.0115	0.0001	90.23	94.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200000	0.0001	0.0116	0.0127	0.0001	88.71	95.3	0.0001	0.0105	0.0110	0.0001	90.75	94.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200000	0.0001	0.0116	0.0127	0.0001	88.71	95.3	0.0001	0.0105	0.0110	0.0001	90.75	94.3
$\widehat{\Delta}_{\mathcal{L}}$	500	0	0.0003	0.0236	0.0238	0.0006	-	94.5	0.0006	0.0229	0.0231	0.0005	-	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500	0.0003	0.0173	0.0183	0.0003	46.27	95.1	0.0006	0.0165	0.0174	0.0003	47.77	95.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500	0.0003	0.0174	0.0183	0.0003	45.47	95.1	0.0006	0.0167	0.0174	0.0003	47.01	95.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	1000	0.0002	0.0154	0.0161	0.0002	57.25	95.2	0.0001	0.0147	0.0150	0.0002	58.67	95.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	1000	0.0002	0.0155	0.0161	0.0002	56.78	95.2	0.0000	0.0148	0.0150	0.0002	58.25	94.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	2500	0.0004	0.0127	0.0134	0.0002	71.01	95.4	0.0001	0.0118	0.0121	0.0001	73.58	95.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	2500	0.0004	0.0127	0.0134	0.0002	70.85	95.4	0.0001	0.0118	0.0121	0.0001	73.43	95.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	50000	0.0001	0.0089	0.0103	0.0001	85.74	97.5	0.0002	0.0075	0.0086	0.0001	89.17	97.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	50000	0.0001	0.0089	0.0103	0.0001	85.74	97.5	0.0002	0.0075	0.0086	0.0001	89.16	97.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500000	0.0001	0.0087	0.0101	0.0001	86.52	97.7	0.0002	0.0073	0.0083	0.0001	89.95	97.6
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500000	0.0001	0.0087	0.0101	0.0001	86.52	97.7	0.0002	0.0073	0.0083	0.0001	89.95	97.6

Table S3: The results of Model (d) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	200	0	0.0020	0.0527	0.0518	0.0028	-	94.0	0.0016	0.0454	0.0444	0.0021	-	93.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200	0.0025	0.0489	0.0477	0.0024	13.75	94.1	0.0024	0.0411	0.0395	0.0017	17.86	93.4
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200	0.0028	0.0491	0.0477	0.0024	12.85	94.0	0.0027	0.0414	0.0395	0.0017	16.84	93.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	400	0.0026	0.0469	0.0463	0.0022	20.48	94.1	0.0024	0.0389	0.0378	0.0015	26.37	93.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	400	0.0028	0.0471	0.0463	0.0022	19.95	94.1	0.0026	0.0391	0.0378	0.0015	25.77	93.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	1000	0.0018	0.0466	0.0448	0.0022	21.87	93.7	0.0016	0.0373	0.0359	0.0014	32.57	94.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	1000	0.0019	0.0466	0.0448	0.0022	21.64	93.5	0.0017	0.0373	0.0359	0.0014	32.35	94.0
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	20000	0.0017	0.0455	0.0434	0.0021	25.50	93.1	0.0017	0.0361	0.0341	0.0013	36.61	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	20000	0.0017	0.0455	0.0434	0.0021	25.49	93.1	0.0017	0.0361	0.0341	0.0013	36.60	93.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200000	0.0017	0.0454	0.0433	0.0021	25.71	93.2	0.0017	0.0360	0.0340	0.0013	37.04	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200000	0.0017	0.0454	0.0433	0.0021	25.71	93.2	0.0017	0.0360	0.0340	0.0013	37.04	93.1
$\widehat{\Delta}_{\mathcal{L}}$	500	0	0.0001	0.0462	0.0444	0.0021	-	94.0	0.0008	0.0355	0.0353	0.0013	-	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500	0.0002	0.0443	0.0425	0.0020	7.84	93.3	0.0005	0.0331	0.0328	0.0011	13.10	94.4
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500	0.0005	0.0445	0.0425	0.0020	7.10	93.3	0.0003	0.0332	0.0328	0.0011	12.40	94.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	1000	0.0003	0.0433	0.0418	0.0019	12.26	93.5	0.0009	0.0318	0.0320	0.0010	19.63	94.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	1000	0.0001	0.0434	0.0418	0.0019	11.80	93.4	0.0008	0.0319	0.0320	0.0010	19.21	94.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	2500	0.0003	0.0433	0.0412	0.0019	12.26	93.0	0.0010	0.0316	0.0311	0.0010	20.59	94.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	2500	0.0002	0.0433	0.0412	0.0019	12.04	92.9	0.0009	0.0317	0.0311	0.0010	20.41	94.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	50000	0.0005	0.0425	0.0405	0.0018	15.39	93.5	0.0012	0.0307	0.0302	0.0009	25.08	94.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	50000	0.0005	0.0425	0.0405	0.0018	15.38	93.5	0.0012	0.0307	0.0302	0.0009	25.07	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500000	0.0005	0.0425	0.0405	0.0018	15.19	93.4	0.0012	0.0307	0.0302	0.0009	25.15	94.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500000	0.0005	0.0425	0.0405	0.0018	15.19	93.4	0.0012	0.0307	0.0302	0.0009	25.15	94.3

B California Housing Dataset

In this section, we apply our proposed methods to the California housing prices dataset (Pace and Barry, 1997) with the aim of investigating whether houses located near the ocean are more expensive than those located inland. Consider inland houses (`INLAND`) as the treatment group and those within a one-hour drive to the ocean (`<1 OCEAN`) as the control group with sample size $m = 9034$. We employ all the available labeled data to generate a ROC curve, which serves as the gold standard. Subsequently, we randomly select $n = 1082$ treatment samples as labeled data, and regard the remaining $N = 5414$ samples as unlabeled data to examine the performance of our SS estimators. Except for geographical information, we regress the remaining six explanatory variables, namely `housing median age`, `total rooms`, `total bedrooms`, `population`, `households`, and `median income`, against the housing price.

The same data cleaning techniques are applied to the dataset. Before conducting the analyses, several data cleaning techniques are applied to the dataset. The Box-Cox transformation parameter is $\lambda_{\mathrm{BC}} = -0.0202$ and the adjusted prices exhibit approximate normality. Similarly, we subjected the samples in the treatment group to the same Box-Cox transformation and standardize both the covariates and the transformed prices by using min-max normalization.

Table S4: The estimates of $\Delta(p)$, SE, EFF. and p VAL.

p	$\widehat{\Delta}_{\mathrm{gold}}(p)$	SE.	p VAL.	$\widehat{\Delta}_{\mathcal{L}}(p)$	SE.	p VAL.	$\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{S}}(p)$	SE.	EFF.	p VAL.	$\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{U}}(p)$	SE.	EFF.	p VAL.
0.1	0.0131	0.0014	< 0.01	0.0102	0.0031	< 0.01	0.0071	0.0029	6.8787	< 0.01	0.0066	0.0029	6.8787	< 0.01
0.2	0.0268	0.0020	< 0.01	0.0268	0.0049	< 0.01	0.0235	0.0042	14.0395	< 0.01	0.0233	0.0042	14.0395	< 0.01
0.3	0.0419	0.0026	< 0.01	0.0434	0.0062	< 0.01	0.0355	0.0057	9.2175	< 0.01	0.0344	0.0057	9.2175	< 0.01
0.4	0.0593	0.0031	< 0.01	0.0647	0.0076	< 0.01	0.0547	0.0075	0.1377	< 0.01	0.0527	0.0075	0.1377	< 0.01
0.5	0.0862	0.0037	< 0.01	0.0998	0.0092	< 0.01	0.0815	0.0091	0.9833	< 0.01	0.0779	0.0091	0.9833	< 0.01
0.6	0.1145	0.0042	< 0.01	0.1285	0.0103	< 0.01	0.1079	0.0098	5.0266	< 0.01	0.1043	0.0098	5.0266	< 0.01
0.7	0.1573	0.0050	< 0.01	0.1784	0.0119	< 0.01	0.1695	0.0105	11.4021	< 0.01	0.1682	0.0105	11.4021	< 0.01
0.8	0.2351	0.0061	< 0.01	0.2542	0.0136	< 0.01	0.2420	0.0136	0.1851	< 0.01	0.2396	0.0136	0.1851	< 0.01
0.9	0.3748	0.0070	< 0.01	0.3965	0.0153	< 0.01	0.3948	0.0153	0.0036	< 0.01	0.3945	0.0153	0.0036	< 0.01

The analysis results are primarily reported in Table S4. $\widehat{\Delta}_{\mathrm{gold}}$ denotes the golden standard estimator. All three estimators, namely $\widehat{\Delta}_{\mathcal{L}}(p)$, $\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{U}}(p)$, and $\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{S}}(p)$, behave similarly to $\widehat{\Delta}_{\mathrm{gold}}$ and consistently support the conclusion that houses near the ocean are significantly

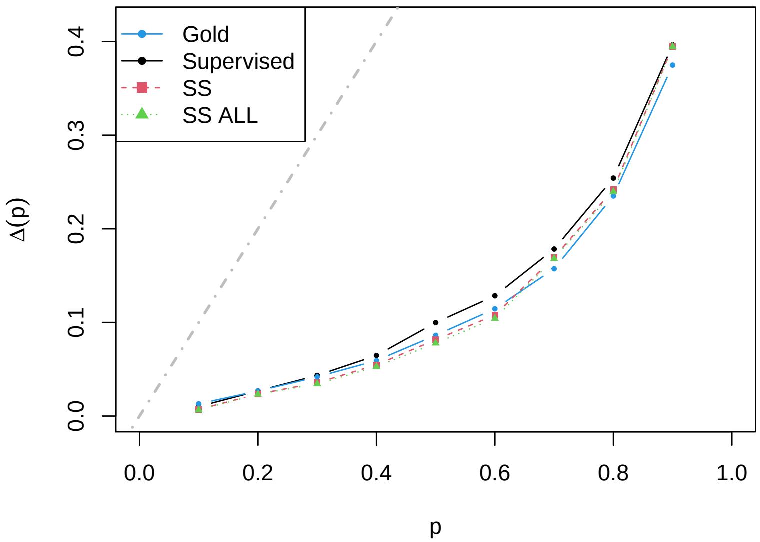

Figure S2: The ROC curve for the study of California Housing Prices dataset. The blue solid line is the golden standard estimator $\widehat{\Delta}_{\text{gold}}$. The black solid line is the supervised estimator $\widehat{\Delta}_{\mathcal{L}}(p)$. The red dash line is the first type of SS estimator $\tilde{\Delta}_{\lambda^{\text{opt}},\mathcal{U}}(p)$. The green dot line is the second type of SS estimator $\tilde{\Delta}_{\lambda^{\text{opt}},\mathcal{S}}(p)$

more expensive than inland houses at all quantile levels. As depicted in Figure S2, the ROC curves for all estimators lie below the diagonal line $\Delta(p) = p$ and the ROC curves of SS estimators are slightly lower than that of the supervised estimator, particularly when p deviates from 0 and 1. Furthermore, both SS estimators yield smaller values compared to the supervised estimator, indicating that SS estimators suggest more significant differences in house prices between coastal and inland areas. Overall, the SS estimators exhibit a closer alignment with the gold standard estimator $\widehat{\Delta}_{\text{gold}}$ than the supervised estimator. As shown in Figure S2, the red and green lines are notably closer to the blue line when compared to the black line. In detail, the average absolute differences between the three estimators (namely $\widehat{\Delta}_{\mathcal{L}}(p)$, $\tilde{\Delta}_{\lambda^{\text{opt}},\mathcal{U}}(p)$, and $\tilde{\Delta}_{\lambda^{\text{opt}},\mathcal{S}}(p)$) and $\widehat{\Delta}_{\text{gold}}$, are 0.0110, 0.0078 and 0.0086, respectively. All these facts suggest that our SS estimators outperform the supervised estimator. Additionally, it's worth noting that the SEs of the two SS estimators are either smaller or equal to the SE of the supervised estimator. The EFF. ranges from 0 to 14.0%, indicating that our

SS estimators are more efficient. These observations align seamlessly with our theoretical findings, providing further substantiation for the superior efficiency of the SS estimators.

C Additional Simulation Results

C.1 Compare different imputation methods when the dimension of Z increases

In this subsection, we conduct a comparative analysis of our single index model-based estimator against three other estimators across various settings as the dimensionality of $\mathbf{Z}$ increases. The simulation results are present in Tables S5. It is essential to note the following distinctions among the estimators:

(i) $\widehat{\Delta}_{\mathcal{L}}$ denotes the supervised estimator;

(ii) $\widehat{\Delta}^{\mathrm{lm}}_{\lambda,\mathcal{U}}$ and $\widehat{\Delta}^{\mathrm{lm}}_{\lambda,\mathcal{S}}$ represent two variants of weighted estimators that entail estimating missing labels for unlabeled data through linear regression;

(iii) $\widehat{\Delta}^{\mathrm{nw}}_{\lambda,\mathcal{U}}$ and $\widehat{\Delta}^{\mathrm{nw}}_{\lambda,\mathcal{S}}$ are two weighted estimators that replace our single index estimator with the original Nadaraya-Watson estimators;

(iv) $\tilde{\Delta}_{\lambda,\mathcal{U}}$ and $\tilde{\Delta}_{\lambda,\mathcal{S}}$ are our proposed single index model-based semi-supervised estimators.

Specifically, we examine settings (a), (e), and (f) at $p = 0.4$, where the label X is generated from a linear model under setting (a), setting (e) represents a nonlinear but single-index scenario, and setting (f) violates the single-index model assumption. In all the aforementioned settings, the variables in the last dimension follow a Bernoulli distribution with values of 0 and 1, Y follows a standard normal distribution, $\boldsymbol{\beta}$ is a d-dimensional vector whose elements are alternately composed of 1 and -1, and $\boldsymbol{\beta}_{-1}$ denotes $\boldsymbol{\beta}$ excluding the first element.

(a) $X = \mathbf{Z}'\boldsymbol{\beta} + \epsilon$, $Z_i \overset{\text{IID}}{\sim} N(0,1)$, $i = 1, \ldots, d-1$;

(e) $X = \sin(2\mathbf{Z}'\boldsymbol{\beta}) + \epsilon$, $Z_i \overset{\text{IID}}{\sim} \text{Unif}(0,1)$, $i = 1, \ldots, d-1$;

(f) $X = Z_1^2 + \sin(4\mathbf{Z}'_{-1}\boldsymbol{\beta}_{-1}) + \epsilon$, $Z_i \overset{\text{IID}}{\sim} \text{Unif}(0,1)$, $i = 1, \ldots, d-1$, where $\mathbf{Z}_{-1}$ denotes the covariate vector excluding Z_1.

We repeat the simulation 1000 times and summarize the results in Tables S5, including the bias, standard deviation (SD), mean squared error (MSE), and the asymptotic relative efficiency (ARE). This analysis aids in comprehending the behavior of each estimator across varying conditions and covariates dimensionalities.

The findings showcased in Table S5 offer compelling evidence of the consistently superior efficiency of our proposed single-index model-based (SIM) estimators in contrast to the supervised estimator, regardless of the true model and the dimensionality of the covariates. In setting (a), the linear estimators $\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{U}}$ and $\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{S}}$ exhibit optimal performance. However, when confronted with nonlinear true models, like settings (e) and (f), these linear estimators may succumb to substantial bias and variance, resulting in negative asymptotic relative efficiency (ARE). Notably, in setting (a), where linearity holds, our SIM estimators do not outperform the linear estimators but nevertheless showcase a competitive performance. Furthermore, our SIM estimators exhibit greater stability as the dimensionality of covariates increases, in comparison to the NW estimators without single index $\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{U}}$ and $\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{S}}$. In settings (a) and (e), where the true models adhere to the single index assumption, our proposed estimators $\tilde{\Delta}_{\lambda,\mathcal{U}}$ and $\tilde{\Delta}_{\lambda,\mathcal{S}}$ consistently outperform the NW estimators without single index across all scenarios. In setting (f), where the single index model assumption is violated, the NW estimators without single index exhibit slightly higher ARE than the SIM estimators only when the dimensionality is 3. Nevertheless, as the dimensionality increases, the performance of the NW estimators without single index deteriorates rapidly, which is a common phenomenon observed in all three settings.

Overall, our proposed SIM estimators offer reliable and highly efficient estimates across diverse scenarios. In contrast, the NW estimators without single index exhibits sensitivity to the dimensionality increase, resulting in a decline in their performance. Moreover, the linear

Table S5: The results of Model (a), (e) and (f) with $n = 100$ $N = 500$ and $m = 200$ under different dimensionality, including the Bias, SD, MSE and ARE (%).

		setting (a)				setting (e)				setting (f)			
d		BIAS	SD	MSE	ARE	BIAS	SD	MSE	ARE	BIAS	SD	MSE	ARE
3	$\widehat{\Delta}_{\mathcal{L}}$	-0.0030	0.0525	0.0028	-	-0.0153	0.0585	0.0036	-	-0.0074	0.0541	0.0030	-
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{U}}$	-0.0017	0.0287	0.0008	70.2	-0.0588	0.0693	0.0083	-126.1	0.0584	0.0701	0.0083	-178.8
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{S}}$	-0.0014	0.0288	0.0008	70.0	-0.0657	0.0725	0.0096	-162.3	0.0714	0.0770	0.0110	-269.4
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{U}}$	-0.0002	0.0393	0.0015	44.1	-0.0116	0.0509	0.0027	25.4	-0.0054	0.0408	0.0017	43.2
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{S}}$	-0.0001	0.0395	0.0016	43.7	-0.0113	0.0510	0.0027	25.3	-0.0054	0.0410	0.0017	42.8
	$\widehat{\Delta}_{\lambda,\mathcal{U}}$	-0.0022	0.0355	0.0013	54.2	-0.0115	0.0487	0.0025	31.3	-0.0048	0.0432	0.0019	36.7
	$\widehat{\Delta}_{\lambda,\mathcal{S}}$	-0.0021	0.0356	0.0013	54.0	-0.0111	0.0488	0.0025	31.3	-0.0046	0.0433	0.0019	36.6
5	$\widehat{\Delta}_{\mathcal{L}}$	0.0017	0.0515	0.0027	-	-0.0136	0.0580	0.0035	-	0.0022	0.0570	0.0032	-
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{U}}$	-0.0000	0.0254	0.0006	75.8	-0.0714	0.1088	0.0169	-377.8	0.0664	0.1041	0.0152	-369.1
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{S}}$	0.0002	0.0253	0.0006	75.8	-0.0804	0.1194	0.0207	-484.5	0.0787	0.1153	0.0195	-499.9
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{U}}$	0.0044	0.0426	0.0018	30.9	-0.0052	0.0525	0.0028	21.4	0.0031	0.0530	0.0028	13.3
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{S}}$	0.0045	0.0427	0.0018	30.5	-0.0049	0.0526	0.0028	21.2	0.0032	0.0531	0.0028	12.8
	$\widehat{\Delta}_{\lambda,\mathcal{U}}$	-0.0003	0.0330	0.0011	59.0	-0.0083	0.0466	0.0022	36.9	0.0030	0.0473	0.0022	30.7
	$\widehat{\Delta}_{\lambda,\mathcal{S}}$	-0.0003	0.0331	0.0011	58.8	-0.0079	0.0467	0.0022	36.9	0.0031	0.0473	0.0022	30.7
7	$\widehat{\Delta}_{\mathcal{L}}$	0.0017	0.0521	0.0027	-	-0.0152	0.0590	0.0037	-	0.0009	0.0579	0.0034	-
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{U}}$	0.0009	0.0227	0.0005	81.1	-0.1202	0.1297	0.0313	-742.8	0.0514	0.1071	0.0141	-320.8
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{S}}$	0.0010	0.0227	0.0005	81.0	-0.1382	0.1452	0.0402	-982.8	0.0611	0.1184	0.0177	-429.5
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{U}}$	0.0055	0.0480	0.0023	14.1	-0.0081	0.0592	0.0036	3.9	-0.0004	0.0607	0.0037	-9.9
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{S}}$	0.0055	0.0481	0.0023	14.0	-0.0080	0.0592	0.0036	3.8	-0.0004	0.0608	0.0037	-10.1
	$\widehat{\Delta}_{\lambda,\mathcal{U}}$	0.0008	0.0311	0.0010	64.4	-0.0099	0.0478	0.0024	35.9	-0.0009	0.0509	0.0026	22.7
	$\widehat{\Delta}_{\lambda,\mathcal{S}}$	0.0008	0.0312	0.0010	64.2	-0.0096	0.0479	0.0024	35.8	-0.0009	0.0509	0.0026	22.7
9	$\widehat{\Delta}_{\mathcal{L}}$	0.0018	0.0518	0.0027	-	-0.0094	0.0580	0.0035	-	0.0020	0.0562	0.0032	-
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{U}}$	0.0015	0.0222	0.0005	81.6	-0.1415	0.1279	0.0364	-952.4	0.0462	0.1018	0.0125	-295.2
	$\widehat{\Delta}^{\text{lm}}_{\lambda,\mathcal{S}}$	0.0015	0.0223	0.0005	81.5	-0.1666	0.1440	0.0485	-1302.8	0.0544	0.1122	0.0155	-391.7
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{U}}$	0.0047	0.0507	0.0026	3.4	-0.0015	0.0611	0.0037	-8.0	0.0014	0.0602	0.0036	-14.5
	$\widehat{\Delta}^{\text{nw}}_{\lambda,\mathcal{S}}$	0.0047	0.0508	0.0026	3.3	-0.0015	0.0611	0.0037	-8.0	0.0014	0.0602	0.0036	-14.6
	$\widehat{\Delta}_{\lambda,\mathcal{U}}$	0.0020	0.0300	0.0009	66.3	-0.0074	0.0451	0.0021	39.6	-0.0012	0.0513	0.0026	16.9
	$\widehat{\Delta}_{\lambda,\mathcal{S}}$	0.0020	0.0301	0.0009	66.2	-0.0073	0.0451	0.0021	39.6	-0.0012	0.0513	0.0026	16.8

estimators heavily rely on the assumption of linearity in the underlying model. Therefore, our SIM estimators provide a valuable alternative by delivering robust and efficient estimates without depending on restrictive assumptions about linearity or being adversely affected by the curse of dimensionality.

C.2 More simulation results for small n and large N

In this section, we display additional simulation results focusing on small n and large N. Tables S6 - S9 present simulation results for different scenarios with a fixed value of p at 0.4. In these scenarios, we examine a small labeled sample size (n) of either 50 or 100, while the unlabeled sample size (N) varies significantly from n to $1000n$.

Importantly, even when the labeled sample size n is as small as 50, our proposed semi-supervised estimators $\tilde{\Delta}_{\lambda,\mathcal{U}}$ and $\tilde{\Delta}_{\lambda,\mathcal{S}}$ still consistently outperform the supervised estimators $\widehat{\Delta}_{\mathcal{L}}$ across all scenarios, with the asymptotic relative efficiency (ARE) ranging from a minimum of 8.01% to a maximum of 46.60%. Furthermore, the increase in ARE as the unlabeled sample size N increases is consistent with the previous simulation results, indicating that the SS estimators achieve higher efficiency with larger unlabeled sample sizes.

Additionally, when both the labeled sample size (n) and the unlabeled sample size (N) are small, the partial imputed estimator $\tilde{\Delta}_{\lambda,\mathcal{U}}$ outperforms the fully imputed estimator $\tilde{\Delta}_{\lambda,\mathcal{S}}$ in terms of smaller SE and higher ARE. This discrepancy arises because the accuracy of imputation for individual unlabeled data points is inadequate when n is small, and the averaging effect among a small N is insufficient to mitigate this influence. However, this discrepancy gradually diminishes as the unlabeled sample size N increases. As the unlabeled sample size N becomes 100 or 1000 times larger than n, the proportion of the labeled part in the estimator becomes negligible, and the estimation inaccuracy can be compensated by the substantial sample size provided by the unlabeled data $\mathcal{U}$.

One limitation observed in the small sample setting is the tendency to underestimate the variance when the unlabeled sample size reaches $100n$ or $1000n$, leading to a coverage probability (CP) of approximately 90%. This occurs because, with a small labeled sample size, there are limited effective samples available for the imputation procedure.

Moreover, certain scenarios in our simulation exhibit a rise in bias. This increase in bias can be attributed to the limited number of labeled samples, which may not adequately capture the underlying data distribution during the imputation process for unlabeled data. However, it's important to note that while there may be a slight increase in bias, it remains relatively small overall. Conversely, the inclusion of unlabeled data leads to a considerable reduction in the variance of our estimators, consequently enhancing the MSE.

Furthermore, we present simulation results for setting (a) with $p = 0.8$ in Tables S10 - S11. The results are comparable to those obtained for $p = 0.4$.

Table S6: The results in setting (a) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

	n	N	**m = 100**						**m = 200**					
			BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	50	0	0.0016	0.0724	0.0742	0.0052	-	95.4	0.0017	0.0704	0.0722	0.0050	-	93.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50	0.0009	0.0643	0.0645	0.0041	21.30	94.4	0.0010	0.0619	0.0620	0.0038	22.80	94.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50	0.0010	0.0668	0.0645	0.0045	15.07	93.2	0.0008	0.0645	0.0620	0.0042	16.25	93.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	100	0.0004	0.0611	0.0608	0.0037	28.94	94.2	0.0001	0.0588	0.0581	0.0035	30.24	93.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	100	0.0016	0.0627	0.0608	0.0039	25.03	93.8	0.0013	0.0605	0.0581	0.0037	26.13	93.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	250	0.0004	0.0583	0.0568	0.0034	35.21	94.1	0.0007	0.0551	0.0537	0.0030	38.70	93.6
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	250	0.0010	0.0591	0.0568	0.0035	33.36	93.9	0.0013	0.0560	0.0537	0.0031	36.77	93.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	5000	0.0011	0.0547	0.0524	0.0030	42.99	92.9	0.0011	0.0516	0.0489	0.0027	46.29	91.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	5000	0.0011	0.0548	0.0524	0.0030	42.88	92.9	0.0011	0.0517	0.0489	0.0027	46.18	91.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50000	0.0010	0.0545	0.0522	0.0030	43.39	92.6	0.0011	0.0515	0.0486	0.0026	46.60	91.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50000	0.0011	0.0545	0.0522	0.0030	43.38	92.6	0.0011	0.0515	0.0486	0.0026	46.59	91.9
$\widehat{\Delta}_{\mathcal{L}}$	100	0	0.0029	0.0564	0.0554	0.0032	-	94.0	0.0030	0.0525	0.0526	0.0028	-	94.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100	0.0028	0.0466	0.0457	0.0022	31.58	94.1	0.0028	0.0422	0.0422	0.0018	35.42	94.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100	0.0025	0.0470	0.0457	0.0022	30.61	94.0	0.0026	0.0426	0.0422	0.0018	34.31	94.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	200	0.0027	0.0425	0.0419	0.0018	42.98	94.2	0.0027	0.0378	0.0382	0.0014	48.19	95.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	200	0.0025	0.0428	0.0419	0.0018	42.43	94.1	0.0026	0.0380	0.0382	0.0014	47.60	94.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	500	0.0024	0.0399	0.0378	0.0016	49.85	92.7	0.0022	0.0352	0.0336	0.0012	55.18	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	500	0.0023	0.0400	0.0378	0.0016	49.63	92.7	0.0021	0.0352	0.0336	0.0012	54.96	93.0
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	10000	0.0022	0.0353	0.0335	0.0013	60.68	93.2	0.0022	0.0297	0.0286	0.0009	68.05	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	10000	0.0022	0.0353	0.0335	0.0013	60.68	93.2	0.0022	0.0297	0.0286	0.0009	68.04	93.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100000	0.0020	0.0349	0.0332	0.0012	61.55	93.2	0.0020	0.0293	0.0282	0.0009	68.88	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100000	0.0020	0.0349	0.0332	0.0012	61.55	93.2	0.0020	0.0293	0.0282	0.0009	68.88	93.1

Table S7: The results in setting (b) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	50	0	0.0083	0.0707	0.0715	0.0051	-	93.5	0.0091	0.0680	0.0679	0.0047	-	93.4
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50	0.0041	0.0660	0.0637	0.0044	13.73	93.1	0.0046	0.0619	0.0596	0.0039	17.99	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50	0.0001	0.0678	0.0637	0.0046	9.45	92.3	0.0005	0.0636	0.0596	0.0040	14.09	92.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	100	0.0012	0.0634	0.0608	0.0040	20.70	93.0	0.0019	0.0594	0.0565	0.0035	24.98	92.6
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	100	0.0015	0.0646	0.0608	0.0042	17.72	92.3	0.0008	0.0605	0.0565	0.0037	22.25	92.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	250	0.0004	0.0616	0.0576	0.0038	25.30	92.1	0.0006	0.0574	0.0530	0.0033	29.93	91.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	250	0.0017	0.0622	0.0576	0.0039	23.76	91.9	0.0008	0.0580	0.0530	0.0034	28.55	91.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	5000	0.0018	0.0606	0.0545	0.0037	27.60	90.1	0.0008	0.0556	0.0496	0.0031	34.18	88.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	5000	0.0019	0.0606	0.0545	0.0037	27.51	90.1	0.0009	0.0557	0.0496	0.0031	34.10	88.6
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50000	0.0021	0.0604	0.0543	0.0036	28.10	89.9	0.0011	0.0556	0.0495	0.0031	34.33	88.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50000	0.0021	0.0604	0.0543	0.0036	28.09	89.9	0.0011	0.0556	0.0495	0.0031	34.32	88.7
$\widehat{\Delta}_{\mathcal{L}}$	100	0	0.0099	0.0563	0.0558	0.0033	-	93.4	0.0091	0.0500	0.0509	0.0026	-	94.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100	0.0088	0.0512	0.0494	0.0027	17.48	92.5	0.0083	0.0446	0.0439	0.0021	20.45	93.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100	0.0078	0.0515	0.0494	0.0027	17.02	92.3	0.0072	0.0448	0.0439	0.0021	20.43	93.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	200	0.0087	0.0487	0.0471	0.0024	25.11	92.7	0.0080	0.0415	0.0413	0.0018	30.77	94.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	200	0.0080	0.0489	0.0471	0.0025	24.89	92.7	0.0073	0.0417	0.0413	0.0018	30.81	93.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	500	0.0076	0.0468	0.0447	0.0022	31.16	92.6	0.0071	0.0397	0.0385	0.0016	36.97	92.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	500	0.0073	0.0469	0.0447	0.0023	31.05	92.6	0.0067	0.0398	0.0385	0.0016	37.00	92.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	10000	0.0078	0.0448	0.0423	0.0021	36.80	92.8	0.0071	0.0378	0.0356	0.0015	42.69	92.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	10000	0.0077	0.0448	0.0423	0.0021	36.79	92.8	0.0071	0.0378	0.0356	0.0015	42.70	92.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100000	0.0079	0.0447	0.0421	0.0021	37.08	92.2	0.0073	0.0377	0.0355	0.0015	43.04	93.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100000	0.0079	0.0447	0.0421	0.0021	37.08	92.2	0.0073	0.0377	0.0355	0.0015	43.04	93.0

Table S8: The results in setting (c) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	50	0	0.0011	0.0725	0.0713	0.0053	-	94.6	0.0009	0.0718	0.0710	0.0051	-	95.0
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50	0.0013	0.0681	0.0616	0.0046	11.96	92.6	0.0009	0.0670	0.0612	0.0045	12.76	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50	0.0010	0.0696	0.0616	0.0048	8.01	91.3	0.0008	0.0685	0.0612	0.0047	8.81	92.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	100	0.0019	0.0647	0.0578	0.0042	20.40	91.8	0.0014	0.0633	0.0573	0.0040	22.15	92.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	100	0.0017	0.0656	0.0578	0.0043	18.19	91.4	0.0012	0.0642	0.0573	0.0041	19.99	92.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	250	0.0009	0.0617	0.0535	0.0038	27.63	92.8	0.0003	0.0599	0.0529	0.0036	30.27	92.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	250	0.0008	0.0621	0.0535	0.0039	26.66	92.2	0.0003	0.0603	0.0529	0.0036	29.32	92.4
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	5000	0.0016	0.0593	0.0487	0.0035	33.12	90.2	0.0007	0.0575	0.0480	0.0033	35.85	90.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	5000	0.0016	0.0593	0.0487	0.0035	33.07	90.2	0.0007	0.0575	0.0480	0.0033	35.80	90.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50000	0.0017	0.0591	0.0484	0.0035	33.51	89.9	0.0007	0.0575	0.0477	0.0033	35.92	90.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50000	0.0017	0.0591	0.0484	0.0035	33.51	89.9	0.0007	0.0575	0.0477	0.0033	35.91	90.2
$\widehat{\Delta}_{\mathcal{L}}$	100	0	0.0027	0.0515	0.0508	0.0027	-	94.8	0.0025	0.0514	0.0504	0.0026	-	95.0
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100	0.0025	0.0382	0.0376	0.0015	44.74	95.3	0.0022	0.0379	0.0370	0.0014	45.57	94.6
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100	0.0025	0.0385	0.0376	0.0015	44.11	95.1	0.0022	0.0381	0.0370	0.0015	44.97	94.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	200	0.0011	0.0309	0.0320	0.0010	64.05	95.6	0.0011	0.0309	0.0313	0.0010	63.87	95.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	200	0.0011	0.0310	0.0320	0.0010	63.71	95.6	0.0011	0.0311	0.0313	0.0010	63.55	94.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	500	0.0009	0.0240	0.0251	0.0006	78.33	95.3	0.0008	0.0238	0.0243	0.0006	78.63	94.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	500	0.0009	0.0240	0.0251	0.0006	78.23	95.2	0.0008	0.0238	0.0243	0.0006	78.54	94.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	10000	0.0000	0.0159	0.0161	0.0003	90.52	92.3	0.0001	0.0154	0.0147	0.0002	91.04	89.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	10000	0.0000	0.0159	0.0161	0.0003	90.51	92.3	0.0001	0.0154	0.0147	0.0002	91.04	89.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100000	0.0002	0.0148	0.0154	0.0002	91.75	92.0	0.0002	0.0142	0.0140	0.0002	92.43	90.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100000	0.0002	0.0148	0.0154	0.0002	91.75	92.0	0.0002	0.0142	0.0140	0.0002	92.43	90.2

Table S9: The results in setting (d) with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	50	0	0.0027	0.0765	0.0799	0.0059	-	95.3	0.0023	0.0723	0.0753	0.0052	-	94.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50	0.0013	0.0706	0.0690	0.0050	14.90	94.5	0.0012	0.0661	0.0634	0.0044	16.45	93.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50	0.0007	0.0717	0.0690	0.0051	12.27	93.9	0.0007	0.0671	0.0634	0.0045	13.81	93.6
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	100	0.0007	0.0688	0.0650	0.0047	19.08	93.0	0.0013	0.0641	0.0589	0.0041	21.46	92.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	100	0.0003	0.0695	0.0650	0.0048	17.56	93.0	0.0009	0.0647	0.0589	0.0042	20.01	92.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	250	0.0010	0.0660	0.0606	0.0044	25.55	92.4	0.0011	0.0610	0.0539	0.0037	28.80	91.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	250	0.0012	0.0663	0.0606	0.0044	24.87	92.2	0.0013	0.0613	0.0539	0.0038	28.13	90.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	5000	0.0014	0.0627	0.0560	0.0039	32.84	90.8	0.0014	0.0575	0.0487	0.0033	36.66	89.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	5000	0.0014	0.0627	0.0560	0.0039	32.81	90.8	0.0014	0.0575	0.0487	0.0033	36.63	89.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50000	0.0010	0.0624	0.0558	0.0039	33.58	90.7	0.0012	0.0572	0.0483	0.0033	37.40	89.4
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50000	0.0010	0.0624	0.0558	0.0039	33.57	90.7	0.0012	0.0572	0.0483	0.0033	37.39	89.4
$\widehat{\Delta}_{\mathcal{L}}$	100	0	0.0036	0.0654	0.0625	0.0043	-	92.7	0.0037	0.0580	0.0566	0.0034	-	93.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100	0.0023	0.0609	0.0562	0.0037	13.47	92.2	0.0022	0.0521	0.0493	0.0027	19.45	92.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100	0.0020	0.0611	0.0562	0.0037	12.87	92.1	0.0020	0.0524	0.0493	0.0027	18.65	92.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	200	0.0006	0.0585	0.0539	0.0034	20.36	92.6	0.0005	0.0496	0.0466	0.0025	27.04	94.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	200	0.0004	0.0586	0.0539	0.0034	20.02	92.2	0.0004	0.0497	0.0466	0.0025	26.70	94.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	500	0.0000	0.0567	0.0515	0.0032	25.13	92.0	0.0003	0.0476	0.0437	0.0023	32.75	93.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	500	0.0001	0.0567	0.0515	0.0032	25.03	92.0	0.0003	0.0477	0.0437	0.0023	32.63	92.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	10000	0.0001	0.0549	0.0491	0.0030	29.69	91.5	0.0004	0.0450	0.0408	0.0020	39.95	92.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	10000	0.0001	0.0549	0.0491	0.0030	29.69	91.5	0.0004	0.0450	0.0408	0.0020	39.95	92.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100000	0.0002	0.0547	0.0489	0.0030	30.16	91.5	0.0003	0.0449	0.0406	0.0020	40.40	92.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100000	0.0002	0.0547	0.0489	0.0030	30.16	91.5	0.0003	0.0449	0.0406	0.0020	40.40	92.3

Table S10: The results in setting (a) at $p = 0.8$ with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	50	0	0.0055	0.0586	0.0588	0.0035	0.00	95.4	0.0052	0.0564	0.0571	0.0032	-	94.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50	0.0013	0.0536	0.0518	0.0029	17.08	92.4	0.0014	0.0508	0.0498	0.0026	19.50	93.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50	0.0069	0.0542	0.0518	0.0030	13.91	91.6	0.0071	0.0516	0.0498	0.0027	15.30	92.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	100	0.0032	0.0496	0.0491	0.0025	28.74	91.8	0.0036	0.0470	0.0469	0.0022	30.79	92.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	100	0.0069	0.0500	0.0491	0.0025	26.31	91.6	0.0074	0.0475	0.0469	0.0023	27.74	91.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	250	0.0063	0.0474	0.0462	0.0023	33.83	90.4	0.0066	0.0447	0.0440	0.0020	36.27	90.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	250	0.0082	0.0477	0.0462	0.0023	32.35	90.2	0.0085	0.0450	0.0440	0.0021	34.43	89.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	5000	0.0085	0.0453	0.0435	0.0021	38.70	88.8	0.0092	0.0424	0.0415	0.0019	41.38	89.9
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	5000	0.0086	0.0453	0.0435	0.0021	38.60	88.7	0.0093	0.0424	0.0415	0.0019	41.25	89.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	50	50000	0.0085	0.0451	0.0433	0.0021	39.14	89.2	0.0096	0.0419	0.0411	0.0018	42.41	89.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	50	50000	0.0085	0.0451	0.0433	0.0021	39.13	89.2	0.0096	0.0419	0.0411	0.0018	42.40	89.0
$\widehat{\Delta}_{\mathcal{L}}$	100	0	0.0062	0.0445	0.0442	0.0020	0.00	94.5	0.0055	0.0412	0.0418	0.0017	-	95.6
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100	0.0045	0.0377	0.0368	0.0014	28.69	93.2	0.0040	0.0337	0.0339	0.0012	33.14	94.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100	0.0034	0.0378	0.0368	0.0014	28.55	93.1	0.0030	0.0339	0.0339	0.0012	32.84	94.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	200	0.0040	0.0348	0.0339	0.0012	39.41	92.2	0.0033	0.0302	0.0307	0.0009	46.64	94.1
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	200	0.0033	0.0349	0.0339	0.0012	39.34	92.6	0.0026	0.0303	0.0307	0.0009	46.48	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	500	0.0030	0.0323	0.0307	0.0010	48.04	90.7	0.0022	0.0278	0.0273	0.0008	54.83	93.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	500	0.0026	0.0323	0.0307	0.0010	47.99	90.7	0.0019	0.0279	0.0273	0.0008	54.75	93.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	10000	0.0022	0.0289	0.0277	0.0008	58.49	91.5	0.0015	0.0237	0.0237	0.0006	67.41	91.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	10000	0.0022	0.0289	0.0277	0.0008	58.48	91.5	0.0015	0.0237	0.0237	0.0006	67.41	91.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	100	100000	0.0022	0.0285	0.0275	0.0008	59.44	91.4	0.0015	0.0233	0.0235	0.0005	68.45	92.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	100	100000	0.0022	0.0285	0.0275	0.0008	59.44	91.4	0.0015	0.0233	0.0235	0.0005	68.45	92.2

Table S11: The results in setting (a) at $p = 0.8$ with different (m, n, N), including the Bias, SD, SE, MSE, ARE (%) and CP (%).

			m = 100						**m = 200**					
	n	N	BIAS	SE	SD	MSE	ARE	CP	BIAS	SE	SD	MSE	ARE	CP
$\widehat{\Delta}_{\mathcal{L}}$	200	0	0.0059	0.0343	0.0343	0.0012	0.00	93.8	0.0058	0.0315	0.0313	0.0010	-	94.3
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200	0.0048	0.0297	0.0297	0.0009	25.10	94.2	0.0047	0.0261	0.0261	0.0007	31.33	94.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200	0.0037	0.0298	0.0297	0.0009	25.39	94.0	0.0037	0.0262	0.0261	0.0007	31.67	94.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	400	0.0040	0.0279	0.0280	0.0008	34.14	94.0	0.0038	0.0244	0.0242	0.0006	40.58	94.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	400	0.0033	0.0280	0.0280	0.0008	34.26	93.8	0.0032	0.0244	0.0242	0.0006	40.70	94.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	1000	0.0035	0.0262	0.0261	0.0007	42.22	94.6	0.0033	0.0222	0.0220	0.0005	50.88	93.7
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	1000	0.0032	0.0262	0.0261	0.0007	42.32	94.6	0.0030	0.0222	0.0220	0.0005	50.98	93.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	20000	0.0028	0.0242	0.0243	0.0006	51.03	93.4	0.0027	0.0199	0.0198	0.0004	60.60	93.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	20000	0.0028	0.0242	0.0243	0.0006	51.04	93.5	0.0027	0.0199	0.0198	0.0004	60.60	93.1
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	200	200000	0.0029	0.0239	0.0242	0.0006	52.06	93.7	0.0027	0.0196	0.0196	0.0004	61.62	93.3
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	200	200000	0.0029	0.0239	0.0242	0.0006	52.06	93.7	0.0027	0.0196	0.0196	0.0004	61.62	93.3
$\widehat{\Delta}_{\mathcal{L}}$	500	0	0.0050	0.0269	0.0268	0.0008	0.00	93.4	0.0047	0.0227	0.0227	0.0005	-	93.7
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500	0.0043	0.0245	0.0245	0.0006	17.44	94.5	0.0039	0.0200	0.0200	0.0004	23.01	94.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500	0.0035	0.0246	0.0245	0.0006	17.88	94.6	0.0031	0.0200	0.0200	0.0004	23.60	94.9
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	1000	0.0040	0.0239	0.0237	0.0006	22.08	93.5	0.0037	0.0190	0.0189	0.0004	30.59	95.0
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	1000	0.0035	0.0239	0.0237	0.0006	22.39	93.4	0.0031	0.0190	0.0189	0.0004	30.98	95.2
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	2500	0.0035	0.0230	0.0228	0.0005	27.62	93.6	0.0032	0.0178	0.0178	0.0003	39.03	94.5
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	2500	0.0033	0.0231	0.0228	0.0005	27.78	93.7	0.0029	0.0179	0.0178	0.0003	39.24	94.5
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	50000	0.0034	0.0219	0.0220	0.0005	34.41	94.5	0.0030	0.0164	0.0168	0.0003	48.43	93.8
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	50000	0.0034	0.0219	0.0220	0.0005	34.42	94.5	0.0030	0.0164	0.0168	0.0003	48.45	93.8
$\tilde{\Delta}_{\lambda,\mathcal{U}}$	500	500000	0.0033	0.0218	0.0219	0.0005	34.96	94.8	0.0030	0.0163	0.0167	0.0003	49.05	94.2
$\tilde{\Delta}_{\lambda,\mathcal{S}}$	500	500000	0.0033	0.0218	0.0219	0.0005	34.96	94.8	0.0030	0.0163	0.0167	0.0003	49.05	94.2

C.3 The performance of SS estimators for AUC and KS statistics

In this paper, we have introduced a more efficient SS estimation method for estimating ROC curves and have demonstrated the uniform properties of the estimators. Consequently, it is natural to anticipate improved performance in statistical summaries derived from our proposed SS estimators, such as AUC and Kolmogorov–Smirnov (KS) statistics.

To validate this assertion, we conduct additional simulation experiments, the results of which are presented in Tables S12 - S13. Recall that AUC $= \int_0^1 \Delta_0(p)dp$ and KS $= \max_p |\Delta_0(p) - p|$ (Pepe, 2003). To estimate the AUC and KS statistics, we start by estimating the ROC curve at $p = 0.1, 0.2, \ldots, 0.9$ using both the supervised estimator and the proposed SS estimators. Subsequently, the KS statistics are approximated directly as $\widehat{\text{KS}} = \max_{p \in 0.1,\ldots,0.9} |\widehat{\Delta}(p) - p|$. For AUC estimation, we utilize the trapezoidal rule to approximate the integration, i.e. $\widehat{\text{AUC}} = 0.1 \times \left\{ \sum_{k=1}^{9} \widehat{\Delta}(p) + 0.5 \right\}$. By substituting $\widehat{\Delta}(p)$ with $\widehat{\Delta}_{\mathcal{L}}(p)$, $\tilde{\Delta}_{\lambda,\mathcal{U}}(p)$ and $\tilde{\Delta}_{\lambda,\mathcal{S}}(p)$, we derive the corresponding estimators for AUC and KS.

Moreover, the simulations are conducted in setting (a) with m fixed at 200, where the labeled sample size n varies from 50 to 500 and the unlabeled sample size N is set as n/ρ with ρ taking the values of 1, 0.5, 0.2, 0.01, 0.001. The final results based on 1000 repetitions are detailed in Tables S12 - S13, including BIAS, SE, RMSE, and ARE (where ARE still gauges the reduction in MSE).

In general, our SS estimators for AUC and KS outperform the respective supervised estimators as anticipated. The ARE falls within a range of 20.62% to 78.49%, indicating a notable decrease in MSE. The trend of ARE concerning the sample size aligns with our earlier finding: greater amounts of unlabeled data lead to more precise estimators. Even with a small labeled sample size of $n = 50$, our SS estimators exhibit superior performance. Consequently, it is reasonable to anticipate that our SS estimators will also excel in estimating other summary statistics, like the partial Area Under the Curve (pAUC).

Table S12: Estimation of AUC in setting (a) with m fixed at 200, varying sample sizes (n) and sample ratios ($\rho = n/N$), including the Bias, SE, RMSE and ARE (%). n ranges from 50 to 500, and ρ takes values of 1, 0.5, 0.2, 0.01, and 0.001.

		n = 50				**n = 100**				**n = 200**				**n = 500**			
	ρ	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE
$\widehat{\mathrm{AUC}}_{\mathcal{L}}$	-	-0.0018	0.0474	0.0474	-	-0.0030	0.0355	0.0356	-	-0.0020	0.0274	0.0274	-	-0.0017	0.0202	0.0203	-
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{U}}$	1	-0.0002	0.0385	0.0385	34.16	-0.0026	0.0277	0.0278	39.15	-0.0016	0.0220	0.0221	35.20	-0.0014	0.0177	0.0177	23.80
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{S}}$	1	0.0015	0.0395	0.0395	30.81	-0.0023	0.0278	0.0279	38.82	-0.0013	0.0221	0.0221	34.89	-0.0012	0.0177	0.0177	23.56
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{U}}$	0.5	0.0004	0.0343	0.0343	47.72	-0.0024	0.0242	0.0243	53.33	-0.0014	0.0204	0.0205	44.16	-0.0015	0.0164	0.0165	34.25
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{S}}$	0.5	0.0016	0.0350	0.0350	45.67	-0.0021	0.0243	0.0244	53.18	-0.0012	0.0205	0.0205	43.98	-0.0013	0.0164	0.0165	34.14
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{U}}$	0.2	0.0014	0.0302	0.0303	59.33	-0.0018	0.0217	0.0218	62.48	-0.0012	0.0184	0.0185	54.56	-0.0013	0.0154	0.0155	41.69
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{S}}$	0.2	0.0020	0.0306	0.0306	58.41	-0.0017	0.0218	0.0218	62.43	-0.0011	0.0185	0.0185	54.48	-0.0012	0.0155	0.0155	41.64
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{U}}$	0.01	0.0018	0.0260	0.0260	69.96	-0.0016	0.0168	0.0169	77.50	-0.0006	0.0159	0.0159	66.50	-0.0011	0.0142	0.0143	50.65
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{S}}$	0.01	0.0018	0.0260	0.0260	69.91	-0.0016	0.0168	0.0169	77.50	-0.0006	0.0159	0.0159	66.49	-0.0011	0.0142	0.0143	50.65
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{U}}$	0.001	0.0017	0.0256	0.0256	70.82	-0.0015	0.0165	0.0165	78.49	-0.0007	0.0156	0.0156	67.41	-0.0011	0.0142	0.0142	51.00
$\widehat{\mathrm{AUC}}_{\lambda,\mathcal{S}}$	0.001	0.0017	0.0256	0.0256	70.81	-0.0015	0.0165	0.0165	78.49	-0.0007	0.0156	0.0156	67.41	-0.0011	0.0142	0.0142	51.00

Table S13: Estimation of KS statistics in setting (a) with m fixed at 200, varying sample sizes (n) and sample ratios ($\rho = n/N$), including the Bias, SE, RMSE and ARE (%). n ranges from 50 to 500, and ρ takes values of 1, 0.5, 0.2, 0.01, and 0.001.

		n = 50				**n = 100**				**n = 200**				**n = 500**			
	ρ	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE	BIAS	SE	RMSE	ARE
$\widehat{\mathrm{KS}}_{\mathcal{L}}$	-	0.0244	0.0656	0.0700	-	0.0135	0.0491	0.0509	-	0.0096	0.0388	0.0400	-	0.0051	0.0295	0.0300	-
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{U}}$	1	0.0233	0.0558	0.0605	25.26	0.0116	0.0393	0.0409	35.48	0.0082	0.0326	0.0336	29.37	0.0044	0.0263	0.0266	20.99
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{S}}$	1	0.0225	0.0577	0.0619	21.66	0.0114	0.0395	0.0411	34.93	0.0079	0.0327	0.0337	28.99	0.0040	0.0264	0.0267	20.62
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{U}}$	0.5	0.0233	0.0511	0.0562	35.57	0.0109	0.0354	0.0371	47.09	0.0075	0.0306	0.0314	38.06	0.0037	0.0244	0.0247	31.94
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{S}}$	0.5	0.0228	0.0524	0.0571	33.35	0.0107	0.0356	0.0371	46.92	0.0072	0.0306	0.0315	37.93	0.0035	0.0245	0.0247	31.75
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{U}}$	0.2	0.0231	0.0474	0.0527	43.22	0.0107	0.0327	0.0344	54.42	0.0073	0.0282	0.0291	46.86	0.0035	0.0237	0.0239	36.29
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{S}}$	0.2	0.0229	0.0480	0.0531	42.28	0.0106	0.0328	0.0344	54.36	0.0072	0.0282	0.0291	46.80	0.0033	0.0237	0.0239	36.22
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{U}}$	0.01	0.0226	0.0435	0.0490	50.93	0.0095	0.0273	0.0289	67.89	0.0073	0.0249	0.0259	57.85	0.0035	0.0221	0.0224	44.28
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{S}}$	0.01	0.0226	0.0436	0.0490	50.88	0.0095	0.0273	0.0289	67.89	0.0073	0.0249	0.0259	57.85	0.0035	0.0221	0.0224	44.28
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{U}}$	0.001	0.0225	0.0433	0.0488	51.38	0.0097	0.0270	0.0287	68.35	0.0073	0.0247	0.0258	58.46	0.0035	0.0221	0.0223	44.44
$\widehat{\mathrm{KS}}_{\lambda,\mathcal{S}}$	0.001	0.0225	0.0433	0.0488	51.38	0.0097	0.0270	0.0287	68.35	0.0073	0.0247	0.0258	58.46	0.0035	0.0221	0.0223	44.44

Now that we have established the uniform convergence property for the ROC curve, obtaining the convergence property for various summary statistics such as the AUC and KS statistics follows directly. In the subsequent corollary, we present the convergence result for the AUC.

Corollary C.1. *(i) Suppose that Assumption A2 is satisfied,* $\sqrt{n}\{\int_0^1 \widehat{\Delta}_{\mathcal{L}}(p)dp - \int_0^1 \Delta_0(p)dp\} \xrightarrow{d} \int_0^1 W_{\mathcal{L}}(p)dp$. *(ii) Under Assumptions A and B,* $\sqrt{n}\{\int_0^1 \tilde{\Delta}_{\lambda,\mathcal{U}}(p)dp - \int_0^1 \Delta_0(p)dp\} \xrightarrow{d} \int_0^1 W_{\lambda,\mathcal{U}}(p)dp$ *and* $\sqrt{n}\{\int_0^1 \tilde{\Delta}_{\lambda,\mathcal{S}}(p)dp - \int_0^1 \Delta_0(p)dp\} \xrightarrow{d} \int_0^1 W_{\lambda,\mathcal{S}}(p)dp$.

Proof. The continuous mapping theorem gives the conclusion directly. ∎

D When X and Y are both from the same parametric family

In this section, we discuss how to tackle with the situations that X and Y originate from the same parametric distribution family, such as location-scale families, but with differing parameters. This situation is frequently encountered in many practical applications. Our discussion primarily follows the methodology outlined in González-Manteiga et al. (2011). Now assume that

$$X = \mu_x(\mathbf{Z}) + \sigma_x(\mathbf{Z})\epsilon_x, \quad Y = \mu_y(\mathbf{V}) + \sigma_y(\mathbf{V})\epsilon_y,$$

where $\mu_x(\mathbf{Z}) = E(X|\mathbf{Z})$, $\sigma_x(\mathbf{Z}) = \mathrm{Var}(X|\mathbf{Z})$, $\mu_y(\mathbf{V}) = E(Y|\mathbf{V})$ and $\sigma_y(\mathbf{V}) = \mathrm{Var}(Y|\mathbf{V})$. Additionally, $\epsilon_x \perp X$ and $\epsilon_u \perp Y$. Let F and G denote the distribution functions of X and Y, respectively, and F_ϵ and G_ϵ denote the distribution functions of ϵ_x and ϵ_u, respectively. We consider two cases : (i) only $(X, \mathbf{Z})$ possesses the semi-supervised data framework; (ii) both $(X, \mathbf{Z})$ and $(Y, \mathbf{V})$ operate within a semi-supervised data framework.

First, we discuss case (i) that only X possesses the semi-supervised data framework.

$$\begin{aligned}\Delta(p) =& 1 - P\left(X \leq G^{-1}(1-p;\theta)\right) = 1 - P\left(\epsilon_x \leq \frac{G^{-1}(1-p) - \mu_x(\mathbf{Z})}{\sigma_x(\mathbf{Z})}\right) \\ =& 1 - E\left(E\left[\mathbb{I}\left\{\epsilon_x \leq \frac{G^{-1}(1-p) - \mu_x(\mathbf{Z})}{\sigma_x(\mathbf{Z})}\right\}\middle|\mathbf{Z}'\boldsymbol{\beta}\right]\right) \\ :=& 1 - E\left\{F_{\epsilon_x|\mathbf{Z}'\boldsymbol{\beta}}\left(\frac{G^{-1}(1-p) - \mu_x(\mathbf{Z})}{\sigma_x(\mathbf{Z})}\right)\right\}.\end{aligned}$$

Once an appropriate $\boldsymbol{\beta}$ is determined, we can proceed in the manner outlined by González-Manteiga et al. (2011) and use the kernel method to estimate $\mu_x(\mathbf{Z})$, $\sigma_x(\mathbf{Z})$, the conditional distribution of $\epsilon_x|\mathbf{Z}'\boldsymbol{\beta}$ and $G^{-1}(1-p)$. Concretely,

$$\widehat{\mu}_x(\mathbf{z}) = \frac{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})X_i}{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})}, \quad \widehat{\sigma}_x(\mathbf{z}) = \frac{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})\left\{X_i - \widehat{\mu}_x(\mathbf{Z}_i)\right\}^2}{\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})},$$

$$\widehat{\epsilon}_{xi} = \frac{X_i - \widehat{\mu}_x(\mathbf{Z}_i)}{\widehat{\sigma}_x(\mathbf{Z}_i)}, \quad \widehat{F}_{\epsilon_x|\mathbf{Z}'\boldsymbol{\beta}}(s) = \frac{\sum_{i=1}^n K_h(\mathbf{Z}_i'\boldsymbol{\beta} - \mathbf{Z}'\boldsymbol{\beta})\mathbb{I}\{\widehat{\epsilon}_{xi} \leq s\}}{\sum_{i=1}^n K_h(\mathbf{Z}_i'\boldsymbol{\beta} - \mathbf{Z}'\boldsymbol{\beta})}.$$

Different components of Y can be estimated in a similar manner. Subsequently, we can formulate the corresponding semi-supervised estimators as following:

$$\widehat{\Delta}_{\mathrm{ss}}(p) = 1 - \frac{\lambda}{n}\sum_{i=1}^n \mathbb{I}\left(\epsilon_{xi} \leq \frac{\widehat{G}^{-1}(1-p) - \widehat{\mu}_x(\mathbf{Z}_i)}{\widehat{\sigma}_x(\mathbf{Z}_i)}\right) - \frac{1-\lambda}{N}\sum_{i=n+1}^M \widehat{F}_{\epsilon_x|\mathbf{Z}_i'\boldsymbol{\beta}}\left(\frac{\widehat{G}^{-1}(1-p) - \widehat{\mu}_x(\mathbf{Z}_i)}{\widehat{\sigma}_x(\mathbf{Z}_i)}\right).$$

Next we consider case (ii) where both groups have semi-supervised data structure, i.e. for the control group we can also observe unlabeled data $(\mathbf{V})_{j=m+1}^{M'}$. We can use the additional unlabeled data to improve the estimation accuracy of $G(y)$ and consequently lead to more accurate $\widehat{G}^{-1}(1-p)$. Note that by the law of iterated expectation,

$$G(y) = P(Y \leq y) = P\left(\epsilon_y \leq \frac{y - \mu_y(\mathbf{V})}{\sigma_y(\mathbf{V})}\right) = E\left[\left\{\mathbb{I}\left(\epsilon_y \leq \frac{y - \mu_y(\mathbf{V})}{\sigma_y(\mathbf{V})}\right)\middle|\mathbf{V}'\boldsymbol{\gamma}\right\}\right].$$

Once $\boldsymbol{\gamma}$ is chosen and ϵ_y, $\mu_y(\cdot)$, $\sigma_y(\cdot)$ are estimated, analogously, we can construct an semi-supervised estimator for $G(y)$ as following:

$$\widehat{G}_{\mathrm{ss}}(y)=\frac{\lambda'}{m}\sum_{j=1}^{m}\mathbb{I}\left(\epsilon_{yj}\le\frac{y-\widehat{\mu}_y(\mathbf{V}_j)}{\widehat{\sigma}_y(\mathbf{V}_j)}\right)+\frac{1-\lambda'}{M'-m}\sum_{j=m+1}^{M'}\widehat{G}_{\epsilon_y|\mathbf{V}'\boldsymbol{\gamma}}\left(\frac{y-\widehat{\mu}_y(\mathbf{V}_j)}{\widehat{\sigma}_y(\mathbf{V}_j)}\right),$$

where

$$\widehat{G}_{\epsilon_y|\mathbf{V}'\boldsymbol{\gamma}}(s)=\frac{\sum_{j=1}^{m}K_h(\mathbf{V}_i'\boldsymbol{\gamma}-\mathbf{V}'\boldsymbol{\gamma})\mathbb{I}\{\widehat{\epsilon}_{yj}\le s\}}{\sum_{j=1}^{m}K_h(\mathbf{V}_j'\boldsymbol{\gamma}-\mathbf{V}'\boldsymbol{\gamma})}.$$

Thus, the corresponding semi-supervised estimator for $\Delta(p)$ is

$$\widehat{\Delta}_{\mathrm{ss}}'(p)=1-\frac{\lambda}{n}\sum_{i=1}^{n}\mathbb{I}\left(\epsilon_{xi}\le\frac{\widehat{G}_{\mathrm{ss}}^{-1}(1-p)-\widehat{\mu}_x(\mathbf{Z}_i)}{\widehat{\sigma}_x(\mathbf{Z}_i)}\right)-\frac{1-\lambda}{N}\sum_{i=n+1}^{M}\widehat{F}_{\epsilon_x|\mathbf{Z}_i'\boldsymbol{\beta}}\left(\frac{\widehat{G}_{\mathrm{ss}}^{-1}(1-p)-\widehat{\mu}_x(\mathbf{Z}_i)}{\widehat{\sigma}_x(\mathbf{Z}_i)}\right).$$

E Discussion on the Missing at Random (MAR) Case

In this section, we extend our estimator to accommodate the MAR setting within the SSL framework, allowing for scenarios where the missing rate approaches 1 as $n+N$ increases indefinitely. The MAR and SIM assumptions ensure that the conditional expectation of $q(X,\theta_0,p)$, given $\mathbf{Z}^\top\boldsymbol{\beta}_0$, is consistent across both the entire population and the labeled subset. Specifically, we have $E\{q(X,\theta_0,p)|\mathbf{Z}^\top\boldsymbol{\beta}_0\}=E\{q(X,\theta_0,p)|\mathbf{Z}^\top\boldsymbol{\beta}_0,\delta=1\}$, where δ denotes the missingness indicator. Leveraging this equivalence, we introduce a new kernel smoothing estimator for $Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0,p)$ utilizing only the labeled data. This estimator aligns with methodologies from Cheng (1994), Hu et al. (2012), and Wang et al. (2023), thereby reinforcing its validity and applicability. A key distinction of the new estimator under the MAR assumption within the SSL framework, compared to the original estimator, lies in its convergence rate similar to the findings in Zhang et al. (2023). Specifically, it converges at a rate of $O_p((Ma_M)^{-1/2})$, where $a_M^{-1}:=E\{\pi_M^{-1}(\mathbf{Z}_i)\}$ and $\pi_M(\mathbf{Z})=E(\delta|\mathbf{Z})$. By applying Jensen's inequality, we deduce that $Ma_M\le ME\{\pi_M(\mathbf{Z})\}=n$, indicating that within the SSL framework, the convergence rate under the MAR assumption is inherently slower than in the MCAR case. This finding represents a significant contribution of our research. Although the scenario where $\rho = \lim_{n,N\to\infty} n/N = 0$ is prevalent in the context of big data, traditional approaches to missing data do not typically consider this extreme situation, and no direct theoretical results exist. Our paper addresses this gap in the literature. Additionally, we focus on point estimation for a specific value of $p \in (0,1)$. Therefore, we omit the notation p in the subsequent discussion and proofs.

E.1 SS Estimator Under MAR Assumption

To illustrate, we first define some notations. Write δ as the indicator of missingness. $\delta_i = 1$ if X_i is observed and $\delta_i = 0$ if X_i is missing. Our observed data samples can be denoted as $\{(\delta_i, \delta_i X_i, \mathbf{Z}_i^\top)^\top\}_{i=1}^M$. Note that $\sum_{i=1}^M \delta_i = n$ and $n/M \to 0$ in probability as $n,\ M \to \infty$. Denote $\pi_M(\mathbf{Z}) = P(\delta = 1|\mathbf{Z})$ as the propensity score function and $\pi_M = E\{\pi_M(\mathbf{Z})\}$. Unlike the standard missing data problem, we allow the propensity score $\pi_M(\mathbf{Z})$ to decay to 0 uniformly on $\mathbf{Z}$ as the sample size M increases. Assume that $a_M^{-1} := E\{\pi_M^{-1}(\mathbf{Z})\} < \infty$ for each M and $Ma_M \to \infty$. Note that we allow δ, π_M to depend on M to enable a non-degenerate propensity score with $E(\delta) \to 0$ as $M \to \infty$. Consequently, the sequences $\{\delta_{M,i}\}_{M,i}$ and $\{\pi_M(X_i)\}_{M,i}$ form triangular arrays. For notational simplicity, we omit indicating the dependence of δ_M on M.

Recall that $q(X,\theta) = \mathbb{I}\big\{X > G^{-1}(1-p;\theta)\big\}$ and $Q(\mathbf{Z}^\top\boldsymbol{\beta},\theta) = E\{q(X,\theta)|\mathbf{Z}^\top\boldsymbol{\beta}\}$. In this section, we assume that the missing data mechanism follows MAR framework and that the single index model (SIM) assumption holds without loss of generality:

Assumption D1. (MAR) $\delta \perp X|\mathbf{Z}$.

Assumption D2. (SIM) There exists $\boldsymbol{\beta}_0$ such that $E\big\{q(X,\theta_0)|\mathbf{Z}\big\} = Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0)$ under the constraint $\|\boldsymbol{\beta}_0\| = 1$.

Next, we introduce a new potential estimator for Δ_0, grounded in these two primary assumptions, which aligns exactly with our previously developed estimator. Similarly, by applying the law of iterated expectation, we find that

$$\Delta_0 = E\big\{q(X,\theta_0)\big\} = E\big[E\big\{q(X,\theta_0)|\mathbf{Z}^\top\boldsymbol{\beta}_0\big\}\big] = E\big\{Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0)\big\}. \tag{A.1}$$

Once $Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0)$ can be estimated precisely, we can anticipate a precise estimator for Δ_0. Under the Assumptions D1 and D2, we can infer the following relationship:

$$Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0) = E\left\{q(X,\theta_0)|\mathbf{Z}\right\} = E\left\{q(X,\theta_0)|\mathbf{Z},\delta=1\right\} = E\left\{q(X,\theta_0)|\mathbf{Z}^\top\boldsymbol{\beta}_0,\delta=1\right\}, \tag{A.2}$$

where the second equality holds due to the MAR assumption and the last equality follows from the law of iterated expectations and Assumption D2. This relationship implies that the conditional expectation of $q(X,\theta_0)$ given $\mathbf{Z}^\top\boldsymbol{\beta}_0$ for the observed data ($\delta=1$) is equivalent to that of the entire population. This insight motivates the strategy of estimating the $\boldsymbol{\beta}_0$ using only the labeled samples. Specifically, the kernel smoothing estimator for $Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0) = E\big\{q(X,\theta_0)|\mathbf{Z}^\top\boldsymbol{\beta}_0,\delta=1\big\}$ is given by

$$\widehat{Q}(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0) = \frac{\sum_{i=1}^M \delta_i K_h(\mathbf{Z}_i^\top\boldsymbol{\beta}_0 - \mathbf{Z}^\top\boldsymbol{\beta}_0)q(X_i,\theta_0)}{\sum_{i=1}^M \delta_i K_h(\mathbf{Z}_i^\top\boldsymbol{\beta}_0 - \mathbf{Z}^\top\boldsymbol{\beta}_0)}.$$

Similar estimators have been employed in previous studies, such as those by Cheng (1994), Hu et al. (2012) and Wang et al. (2023), to address conditional expectation estimation problems under the traditional MAR assumption.

Without loss of generality, let us assume that the first n samples are labeled among the total M samples. Then the estimator $\widehat{Q}(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0)$ can be simplified as

$$\widehat{Q}(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0) = \frac{\sum_{i=1}^n K_h(\mathbf{Z}_i^\top\boldsymbol{\beta}_0 - \mathbf{Z}^\top\boldsymbol{\beta}_0)q(X_i,\theta_0)}{\sum_{i=1}^n K_h(\mathbf{Z}_i^\top\boldsymbol{\beta}_0 - \mathbf{Z}^\top\boldsymbol{\beta}_0)},$$

which is precisely the estimator we proposed previously. Given $\widehat{\theta}$ estimated using control group data, the $\boldsymbol{\beta}_0$ is estimated by minimizing the nonlinear least square loss function $\widehat{\boldsymbol{\beta}} = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\min}\ n^{-1}\sum_{i=1}^{n}\left\{q(X_i,\widehat{\theta}) - \widehat{Q}(\mathbf{Z}_i^\top\boldsymbol{\beta},\widehat{\theta})\right\}^2$. Subsequently, under the Assumption D1 (MAR), we propose a consistent estimator $\widehat{\Delta}_{ss}$ based on equation (A.1):

$$\widehat{\Delta}_{ss} = \frac{1}{M}\sum_{i=1}^{M}\widehat{Q}(\mathbf{Z}_i^\top\widehat{\boldsymbol{\beta}},\widehat{\theta}).$$

This estimator remains consistent regardless of whether we are operating under the traditional MAR framework or the SSL MAR framework, as it does not require the estimation of the propensity score function. However, the estimation of $\boldsymbol{\beta}_0$ can influence the asymptotic normality of $\widehat{\Delta}_{ss}$ in theoretical analysis. We consider the double core estimator to address this theoretical concern, an idea initially proposed by Hu et al. (2012). This estimator offers a potential solution to mitigate the impact of $\widehat{\boldsymbol{\beta}}$ on the asymptotic properties of $\widehat{\Delta}_{ss}$. To achieve this, we introduce an additional assumption:

Assumption D3. (SIM) There exists a vector $\boldsymbol{\gamma}_0$ such that $\pi_M(\mathbf{Z}) = \pi_M(\mathbf{Z}^\top\boldsymbol{\gamma}_0)$. Without loss of generality, we also assume that $\|\boldsymbol{\gamma}_0\| = 1$.

Similarly, we can estimate $\boldsymbol{\gamma}_0$ by minimizing the nonlinear least square loss function $\widehat{\boldsymbol{\gamma}} = \underset{\|\boldsymbol{\gamma}\|_2=1}{\arg\min}\ M^{-1}\sum_{i=1}^{M}\left\{\delta_i - \widehat{\pi}_M(\mathbf{Z}_i^\top\boldsymbol{\gamma})\right\}^2$, where $\widehat{\pi}_M(\mathbf{Z}^\top\boldsymbol{\gamma}) = \sum_{i=1}^{M}K_h(\mathbf{Z}_i^\top\boldsymbol{\gamma}-\mathbf{Z}^\top\boldsymbol{\gamma})\delta_i \Big/ \sum_{i=1}^{M}K_h(\mathbf{Z}_i^\top\boldsymbol{\gamma}-\mathbf{Z}^\top\boldsymbol{\gamma})$. In this framework, $\mathbf{Z}^\top\boldsymbol{\beta}_0$ and $\mathbf{Z}^\top\boldsymbol{\gamma}_0$ can be regarded as two "cores" that encapsulate the most significant information contained in $q(X,\theta_0)$ and δ, respectively. Building upon these two cores, we introduce our final estimator, termed the double core estimator, for Δ_0. Define $\boldsymbol{\kappa} = (\boldsymbol{\beta},\boldsymbol{\gamma})$ and $\widehat{\boldsymbol{\kappa}} = (\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\gamma}})$. Let $\mathbf{S} = \mathbf{Z}^\top\boldsymbol{\kappa}_0$ and $\mathbf{s} = \mathbf{z}^\top\boldsymbol{\kappa}_0$. The double core estimator is given by:

$$\widehat{\Delta}_{ss}^{dk} = \frac{1}{M}\sum_{i=1}^{M}\widehat{Q}(\mathbf{Z}_i^\top\widehat{\boldsymbol{\kappa}}),$$

where

$$\widehat{Q}(\mathbf{z}^\top\widehat{\boldsymbol{\kappa}}) = \frac{\sum_{i=1}^{M}\delta_i\mathbf{K}_h(\mathbf{Z}_i^\top\widehat{\boldsymbol{\kappa}} - \mathbf{z}^\top\widehat{\boldsymbol{\kappa}})q(X_i,\widehat{\theta})}{\sum_{i=1}^{M}\delta_i\mathbf{K}_h(\mathbf{Z}_i^\top\widehat{\boldsymbol{\kappa}} - \mathbf{z}^\top\widehat{\boldsymbol{\kappa}})}.$$

This estimator leverages the information encapsulated in both cores to provide a robust estimate of Δ_0 integrating the key insights from the observed outcomes and the missingness mechanism.

Next, we present the asymptotic properties of $\widehat{\Delta}_{ss}^{dk}$. The detailed proof of these properties can be found in Section E.3. Denote $\Sigma_M^{\mathrm{IF}} = \mathrm{Var}\big\{q(X,\theta_0)\big\} + E\big[\{1-\pi_M(\mathbf{S})\}\pi_M^{-1}(\mathbf{S})\big\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\big\}^2\big]$.

Assumption E. The following assumptions are required to ensure the consistency and asymptotic normality of $\widehat{\Delta}_{ss}^{dk}$:

E1. The density function $f(\mathbf{Z}^\top\boldsymbol{\kappa})$ is bounded below by a positive constant;

E2. $\widehat{\theta} - \theta_0 = O_p((Ma_M)^{-1/2})$ and $r_{\mathrm{mar}} = \lim_{m,M\to\infty} M/(m\Sigma_M^{\mathrm{IF}})$.

E3. (i) $c_M/\pi_M(\mathbf{Z}^\top\boldsymbol{\kappa}) = o_p(1)$ for any $\boldsymbol{\kappa}$ within its compact range; (ii) $\dot{\pi}_M(\mathbf{Z}^\top\boldsymbol{\kappa})$ is bounded;

E4. (i) There exists σ_ξ^2 such that $E\big[\big\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\big\}^2|\mathbf{S}\big] > \sigma_\xi^2 > 0$; (ii) For any $c > 0$, $M^{-1}\sum_{i=1}^M E\big[(\Sigma_M^{\mathrm{IF}})^{-1}\mathrm{IF}_i^2\mathbb{I}\big\{(\Sigma_M^{\mathrm{IF}})^{-1/2}|\mathrm{IF}_i| > cM^{-1/2}\big\}\big] = o_p(1)$; (iii) $Mh^4a_M^{-1} = o_p(1)$ and $Ma_M^2 \to \infty$;

Assumption E encompasses multiple mild conditions. Assumption E1 posits that the density function $f(\mathbf{Z}^\top\boldsymbol{\kappa})$ is bounded away from 0. Assumption E2 parallels Assumption A4 and imposes a mild requirement on the convergence rate of $\widehat{\theta}$. If this boundedness condition is not assumed, truncation techniques can be employed. Assumption E3 involves a mild condition on the convergence rate of the propensity score. Assumption E4 is required solely for establishing asymptotic normality.

Theorem E.1 (A concrete version of Theorem 3.4)**.** *Under Assumptions A2, A3, B1 - B3, **D** and **E**,* $\big(\Sigma_M^{IF}\big)^{-1/2}\sqrt{M}(\widehat{\Delta}_{ss}^{dk} - \Delta_0) \xrightarrow{d} N(0, \mathbf{I} + r_{mar}A(\theta_0)I^{-1}(\theta_0)A(\theta_0))$.

Remark E.1. It is important to highlight that $\Sigma_M^{\mathrm{IF}} \asymp a_M^{-1}$. Consequently, the convergence rate of $\widehat{\Delta}_{ss}^{dk} - \Delta_0$ is $O_p((Ma_M)^{-1/2})$, which differs from the estimator within the

MCAR SSL framework. Specifically, by applying Jensen's inequality, we infer that $Ma_M \leq ME\{\pi_M(\mathbf{Z})\} = n$. This indicates that the convergence rate under the MAR SSL framework is inherently slower than that observed in the MCAR SSL framework.

E.2 Numerical Results for MAR-SSL

In this subsection, we present some simulation results for our proposed SS estimator under the MAR setting. To simulate the decaying propensity score case, we refer to the offset logistic model, as proposed by Zhang et al. (2023), to generate the missing indicator :

$$\pi_M(\mathbf{Z}) = \frac{\exp\left\{\vec{\mathbf{Z}}^\top \boldsymbol{\gamma}_0 + \log(\pi_M)\right\}}{1 + \exp\left\{\vec{\mathbf{Z}}^\top \boldsymbol{\gamma}_0 + \log(\bar{\pi}_M)\right\}} := \operatorname{logit}\left\{\vec{\mathbf{Z}}^\top \boldsymbol{\gamma}_0 + \log(\pi_M)\right\}, \tag{A.3}$$

and $\delta \sim \text{Binomial}\big(1, \pi_M(\mathbf{Z})\big)$, where $\vec{\mathbf{Z}} = (1, \mathbf{Z}^\top)^\top$ and $\boldsymbol{\gamma}_0 \in \mathbb{R}^{d+1}$. Note that if $\pi_M = P(\delta = 1) \to 0$ as $M \to \infty$, then $\pi_M(\mathbf{Z}) \to 0$ as $M \to \infty$, thereby implying the possibility of a decaying propensity score. Specifically, according to Zhang et al. (2023), we consider the following data generation settings. The propensity score functions are generated from the models: (P1) representing the MCAR case, and (P2) representing the MAR case:

(P1). (MCAR) $\pi_M(\mathbf{Z}) = \pi_M$;

(P2). (MAR) $\pi_M(\mathbf{Z}) = \operatorname{logit}\{\vec{\mathbf{Z}}^\top \boldsymbol{\gamma} + \log_2(\pi_M)\}$,

where $\boldsymbol{\gamma} = \big(\gamma(1), 1, 1\big)^\top$ and $\gamma(1)$ is calibrated to ensure that $\pi_M = E\{\pi_M(\mathbf{Z})\}$. Specifically, $\gamma(1)$ is set to -2.910 to achieve $\pi_M = 0.01$ and -1.872 to achieve $\pi_M = 0.1$. The outcome X is generated from models (O1) or (O2) :

(O1) $X_i = \mathbf{Z}_i^\top \boldsymbol{\beta} + \epsilon_i$;

(O2) $X_i = 2\sin(\mathbf{Z}_i^\top \boldsymbol{\beta}) + \mathbf{Z}_i^\top \boldsymbol{\beta} + \epsilon_i$, where $\boldsymbol{\beta} = (1.1, -1.2)^\top$ and $\mathbf{Z} \overset{iid}{\sim} U(0,1)$. Besides, $Y \overset{iid}{\sim} N(0,1)$ and $m = 200$. We consider four settings : setting (a) : (P1) + (O1); setting (b) : (P2) + (O1); setting (c) : (P1) + (O2); setting (d) : (P2) + (O2). We compare our proposed estimator with four other estimators : the supervised estimator $\widehat{\Delta}_{sup}$, the oracle complete case estimator $\widehat{\Delta}_{cc}$, and two inverse probability weighting estimators, $\widehat{\Delta}_{ipw-c}$ and $\widehat{\Delta}_{ipw-m}$, with the propensity score model correctly specified and misspecified, respectively. For $\widehat{\Delta}_{ipw-m}$, the propensity score is estimated using a simple probit model. Next, we introduce the method to estimate the offset logistic propensity score for $\widehat{\Delta}_{ipw-c}$. To estimate $\boldsymbol{\gamma}_0$ in the offset logistic model (P2), we first provide an estimator for π_M, denoted as $\widehat{\pi}_M$. Naively, $\widehat{\pi}_M = \sum_{i=1}^{M} \delta_i / M$. Then we can estimate $\boldsymbol{\gamma}_0$ by minimizing the following loss function :

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} -\frac{1}{M}\sum_{i=1}^{M}\left[\delta_i \vec{\mathbf{Z}}_i^\top \boldsymbol{\gamma} - \log\left\{1 + \widehat{\pi}_M \exp(\vec{\mathbf{Z}}_i^\top \boldsymbol{\gamma})\right\}\right].$$

Consequently, the estimator for $\pi_M(\mathbf{Z})$ is $\widehat{\pi}_M(\mathbf{Z}) = \text{logit}\big\{\vec{\mathbf{Z}}\widehat{\boldsymbol{\gamma}} + \log(\widehat{\pi}_M)\big\}$. We consider four combinations of M and π_M, denoted as (M, π_M), namely (1000, 0.1), (5000, 0.1), (10000, 0.01) and (50000, 0.01). The experiment is conducted for 500 times and the results are presented in Tables S14 and S15 at $p = 0.4$.

Overall, our proposed estimator $\widehat{\Delta}_{ss}^{dk}$ consistently demonstrates superior performance relative to the supervised estimator $\widehat{\Delta}_{sup}$, irrespective of the missing data mechanism. It is evident that the oracle complete case estimator $\widehat{\Delta}_{cc}$ yields the best performance among all the estimators. However, our proposed estimator $\widehat{\Delta}_{ss}^{dk}$ generally approximates the performance of $\widehat{\Delta}_{cc}$ most closely. Under the MCAR setting, all the estimators are consistent and exhibit improved performance with increased sample sizes. But $\widehat{\Delta}_{cc}$, $\widehat{\Delta}_{ss}^{dk}$ and $\widehat{\Delta}_{ipw-m}$ are more efficient than the supervised estimator $\widehat{\Delta}_{sup}$ whereas $\widehat{\Delta}_{ipw-c}$ shows no improvement with a correctly specified propensity score model (Robins et al., 1994). Furthermore, $\widehat{\Delta}_{ss}^{dk}$ is more efficient than $\widehat{\Delta}_{ipw-m}$. Under the MAR setting, $\widehat{\Delta}_{sup}$ exhibits severe bias, with its performance deteriorating as the missing rate $(1-\pi_M)$ increases from 0.9 to 0.99. Conversely, the other four estimators remain consistent even under extremely high levels of missingness. For the two IPW estimators $\widehat{\Delta}_{ipw-c}$ with propensity model correctly specified performs better than $\widehat{\Delta}_{ipw-m}$, which is based on a misspecified propensity model. However, both IPW estimators generally exhibit larger standard deviations. $\widehat{\Delta}_{ipw-m}$ even displays a negative ARE. Our proposed SS estimator $\widehat{\Delta}_{ss}^{dk}$ not only maintains consistency but also achieves a smaller standard deviation, closely approximating the oracle estimator $\widehat{\Delta}_{cc}$. Under the Assumption D2, our proposed SS estimator $\widehat{\Delta}_{ss}^{dk}$ performs effectively in both MCAR and MAR cases.

Table S14: Simulation results under different missing data mechanisms using the (O1) model.

	(a) : (P1) + (O1)				**(b) : (P2) + (O1)**			
Est.	Bias	SD	MSE	ARE	Bias	SD	MSE	ARE
	$M = 1000, \pi_M = 0.1(M\pi_M = 100)$							
$\widehat{\Delta}_{sup}$	-0.0001	0.0534	0.0028	-	-0.0307	0.0635	0.0050	-
$\widehat{\Delta}_{ipw-m}$	0.0018	0.0428	0.0018	35.7	0.0077	0.0757	0.0058	-16.4
$\widehat{\Delta}_{ipw-c}$	-0.0001	0.0534	0.0028	0	0.0026	0.0676	0.0046	8.12
$\widehat{\Delta}_{cc}$	0.0000	0.0288	0.0008	70.8	0.0000	0.0288	0.0008	83.3
$\widehat{\Delta}_{ss}^{dk}$	0.0012	0.0393	0.0015	45.8	0.0021	0.0569	0.0032	34.9
	$M = 5000, \pi_M = 0.1(M\pi_M = 500)$							
$\widehat{\Delta}_{sup}$	-0.0037	0.0353	0.0013	-	-0.0333	0.0394	0.0027	-
$\widehat{\Delta}_{ipw-m}$	-0.0027	0.0309	0.0010	23.4	0.0023	0.0421	0.0018	33.3
$\widehat{\Delta}_{ipw-c}$	-0.0037	0.0353	0.0013	0	-0.0029	0.0389	0.0015	42.8
$\widehat{\Delta}_{cc}$	-0.0028	0.0287	0.0008	33.9	-0.0028	0.0287	0.0008	68.7
$\widehat{\Delta}_{ss}^{dk}$	-0.0032	0.0307	0.0009	24.5	-0.0032	0.0362	0.0013	50.5
	$M = 10000, \pi_M = 0.01(M\pi_M = 100)$							
$\widehat{\Delta}_{sup}$	-0.0008	0.0553	0.0031	-	-0.0324	0.0564	0.0042	-
$\widehat{\Delta}_{ipw-m}$	0.0022	0.0426	0.0018	40.7	0.0085	0.0670	0.0046	-7.61
$\widehat{\Delta}_{ipw-c}$	-0.0008	0.0553	0.0031	0	0.0028	0.0614	0.0038	11.0
$\widehat{\Delta}_{cc}$	0.0037	0.0284	0.0008	73.1	0.0037	0.0284	0.0008	80.6
$\widehat{\Delta}_{ss}^{dk}$	0.0019	0.0399	0.0016	47.9	0.0036	0.0499	0.0025	40.9
	$M = 50000, \pi_M = 0.01(M\pi_M = 500)$							
$\widehat{\Delta}_{sup}$	0.0011	0.0347	0.0012	-	-0.0347	0.0358	0.0025	-
$\widehat{\Delta}_{ipw-m}$	0.0005	0.0301	0.0009	24.6	0.0065	0.0381	0.0015	39.9
$\widehat{\Delta}_{ipw-c}$	0.0011	0.0347	0.0012	0	0.0011	0.0358	0.0013	48.5
$\widehat{\Delta}_{cc}$	0.0000	0.0265	0.0007	41.6	0.0000	0.0265	0.0007	71.8
$\widehat{\Delta}_{ss}^{dk}$	0.0004	0.0295	0.0009	27.5	0.0007	0.0321	0.0010	58.5

Table S15: Simulation results under different missing data mechanisms using the (O2) model.

Est.	**(c) : (P1) + (O2)** Bias	SD	MSE	ARE	**(d) : (P2) + (O2)** Bias	SD	MSE	ARE
	$M = 1000, \pi_M = 0.1 (M\pi_M = 100)$							
$\widehat{\Delta}_{sup}$	-0.0001	0.0494	0.0024	-	-0.0304	0.0599	0.0045	-
$\widehat{\Delta}_{ipw-m}$	0.0019	0.0342	0.0012	51.9	0.0107	0.0775	0.0061	-35.7
$\widehat{\Delta}_{ipw-c}$	-0.0001	0.0494	0.0024	0	0.0028	0.0677	0.0046	-1.54
$\widehat{\Delta}_{cc}$	-0.0004	0.0168	0.0003	88.4	-0.0004	0.0168	0.0003	93.7
$\widehat{\Delta}_{ss}^{dk}$	0.0008	0.0253	0.0006	73.8	0.0023	0.0459	0.0021	53.2
	$M = 5000, \pi_M = 0.1 (M\pi_M = 500)$							
$\widehat{\Delta}_{sup}$	-0.0026	0.0243	0.0006	-	-0.0312	0.0277	0.0017	-
$\widehat{\Delta}_{ipw-m}$	-0.0014	0.0169	0.0003	51.6	0.0079	0.0338	0.0012	31.1
$\widehat{\Delta}_{ipw-c}$	-0.0026	0.0243	0.0006	0	-0.0008	0.0299	0.0009	48.7
$\widehat{\Delta}_{cc}$	-0.0018	0.0125	0.0002	73.3	-0.0018	0.0125	0.0002	90.9
$\widehat{\Delta}_{ss}^{dk}$	-0.0020	0.0146	0.0002	63.4	-0.0018	0.0199	0.0004	77.2
	$M = 10000, \pi_M = 0.01 (M\pi_M = 100)$							
$\widehat{\Delta}_{sup}$	-0.0016	0.0516	0.0027	-	-0.0312	0.0518	0.0036	-
$\widehat{\Delta}_{ipw-m}$	0.0020	0.0324	0.0011	60.5	0.0085	0.0623	0.0039	-8.29
$\widehat{\Delta}_{ipw-c}$	-0.0016	0.0516	0.0027	0	0.0004	0.0557	0.0031	15.2
$\widehat{\Delta}_{cc}$	0.0007	0.0116	0.0001	95.0	0.0007	0.0116	0.0001	96.3
$\widehat{\Delta}_{ss}^{dk}$	0.0014	0.0233	0.0005	79.6	0.0004	0.0387	0.0015	58.9
	$M = 50000, \pi_M = 0.01 (M\pi_M = 500)$							
$\widehat{\Delta}_{sup}$	0.0009	0.0245	0.0006	-	-0.0331	0.0248	0.0017	-
$\widehat{\Delta}_{ipw-m}$	0.0002	0.0161	0.0003	57.1	0.0102	0.0283	0.0009	47.4
$\widehat{\Delta}_{ipw-c}$	0.0009	0.0245	0.0006	0	0.0019	0.0256	0.0007	61.5
$\widehat{\Delta}_{cc}$	-0.0007	0.0100	0.0001	83.3	-0.0007	0.0100	0.0001	94.1
$\widehat{\Delta}_{ss}^{dk}$	0.0000	0.0123	0.0002	75.0	0.0004	0.0168	0.0003	83.6

E.3 Theoretical Results for MAR Estimator

Lemma E.1. *Under Assumptions A2, A3, B, D and E1 - E3,* $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = O_p((M\pi_M)^{-1/2})$ *and* $\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0 = O_p((M\pi_M)^{-1/2})$ *as* $M \to \infty$.

Proof. Since Assumption E2 holds, $\widehat{\theta}$ is consistent. Let

$$
\begin{aligned}
\widehat{J}_M(\boldsymbol{\beta}, \theta) =& \frac{1}{M}\sum_{i=1}^{M}\delta_i\left\{q(X_i,\theta) - \widehat{Q}(\mathbf{Z}_i^\top\boldsymbol{\beta},\theta)\right\}^2, \quad J_M(\boldsymbol{\beta}, \theta) = \frac{1}{M}\sum_{i=1}^{M}\delta_i\left\{q(X_i,\theta) - Q(\mathbf{Z}_i^\top\boldsymbol{\beta},\theta)\right\}^2, \\
J(\boldsymbol{\beta}, \theta) =& E\left[\delta\left\{q(X,\theta) - Q(\mathbf{Z}^\top\boldsymbol{\beta},\theta)\right\}^2\right],
\end{aligned}
$$

and then $\widehat{\boldsymbol{\beta}} = \underset{\|\boldsymbol{\beta}\|_2=1}{\arg\min}\ \widehat{J}_M(\boldsymbol{\beta}, \widehat{\theta})$.

First, show that $\widehat{\boldsymbol{\beta}}$ is consistent. Let $R(\delta) = \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|>\delta} J(\boldsymbol{\beta},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0)$ for any $\delta > 0$. It is obvious that $R(\delta) > 0$. Write $\mathcal{J}_M = J(\widehat{\boldsymbol{\beta}},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0) + \widehat{J}_M(\boldsymbol{\beta}_0,\widehat{\theta}) - \widehat{J}_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})$. The event $\{|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0| > \delta\}$ implies that $J(\widehat{\boldsymbol{\beta}},\theta_0) > \inf_{|\boldsymbol{\beta}-\boldsymbol{\beta}_0|>\delta} J(\boldsymbol{\beta},\theta_0)$, and then $J(\widehat{\boldsymbol{\beta}},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0) > R(\delta)$. Since $\widehat{J}_M(\boldsymbol{\beta}_0,\widehat{\theta}) \geq \widehat{J}_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})$, we have $\mathcal{J}_M > R(\delta)$. Then $P\big(\big|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\big| > \delta\big) \leq P\big(\mathcal{J}_M - R(\delta) > 0\big)$. If $\mathcal{J}_M = o_p(1)$, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$.

Next we show that $\mathcal{J}_M = o_p(1)$. By the triangle inequality,

$$
\begin{aligned}
\mathcal{J}_M =& J(\widehat{\boldsymbol{\beta}},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0) + \widehat{J}_M(\boldsymbol{\beta}_0,\widehat{\theta}) - \widehat{J}_M(\widehat{\boldsymbol{\beta}},\widehat{\theta}) \\
\leq& \left|J(\widehat{\boldsymbol{\beta}},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0) + J_M(\boldsymbol{\beta}_0,\widehat{\theta}) - J_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})\right| + \left|\widehat{J}_M(\boldsymbol{\beta}_0,\widehat{\theta}) - J_M(\boldsymbol{\beta}_0,\widehat{\theta})\right| + \left|\widehat{J}_M(\widehat{\boldsymbol{\beta}},\widehat{\theta}) - J_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})\right| \\
:=& I_1 + I_2 + I_3.
\end{aligned}
$$

Subsequently, we analyze the sequence of the three elements individually. For I_1, we apply the triangle inequality once more and then proceed as follows:

$$
\begin{aligned}
I_1 =& \left|J(\widehat{\boldsymbol{\beta}},\theta_0) - J(\boldsymbol{\beta}_0,\theta_0) + J_M(\boldsymbol{\beta}_0,\widehat{\theta}) - J_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})\right| \\
\leq& \left|J_M(\widehat{\boldsymbol{\beta}},\widehat{\theta}) - J(\widehat{\boldsymbol{\beta}},\theta_0)\right| + \left|J_M(\boldsymbol{\beta}_0,\widehat{\theta}) - J(\boldsymbol{\beta}_0,\theta_0)\right| := I_{11} + I_{12}.
\end{aligned}
$$

Further, decompose I_{11} into three components

$$
\begin{aligned}
I_{11} \leq & \left| J_M(\widehat{\boldsymbol{\beta}}, \widehat{\theta}) - J(\widehat{\boldsymbol{\beta}}, \widehat{\theta}) - J_M(\widehat{\boldsymbol{\beta}}, \theta_0) + J(\widehat{\boldsymbol{\beta}}, \theta_0) \right| + \left| J(\widehat{\boldsymbol{\beta}}, \widehat{\theta}) - J(\widehat{\boldsymbol{\beta}}, \theta_0) \right| + \left| J_M(\widehat{\boldsymbol{\beta}}, \theta_0) - J(\widehat{\boldsymbol{\beta}}, \theta_0) \right| \\
:= & I_{111} + I_{112} + I_{113}.
\end{aligned}
$$

Similarly, we can demonstrate that the class of functions $\left\{\{q(\cdot, \theta) - Q(\cdot^\top \boldsymbol{\beta}, \theta)\}^2 : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta\right\}$ is Donsker. Consequently, the function class $\left\{\delta\{q(\cdot, \theta) - Q(\cdot^\top \boldsymbol{\beta}, \theta)\}^2 : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta\right\}$ is also Donsker. Therefore, $I_{111} = o_p(1)$ by equicontinuity, and the property of the GC class implies that $I_{113} = o_p(1)$. The consistency of $\widehat{\theta}$ ensures that $I_{112} = o_p(1)$. Consequently, $I_{11} = o_p(1)$. Similarly, we have $I_{12} = o_p(1)$. As a result, $I_1 = o_p(1)$.

To analyze I_2, consider the expression given by

$$
I_2 = \left| \widehat{J}_M(\boldsymbol{\beta}_0, \widehat{\theta}) - J_M(\boldsymbol{\beta}_0, \widehat{\theta}) \right| \leq \sup_{\boldsymbol{\beta}} \left| \widehat{J}_M(\boldsymbol{\beta}, \widehat{\theta}) - J_M(\boldsymbol{\beta}, \widehat{\theta}) \right| := \sup_{\boldsymbol{\beta}} \left| J_1(\boldsymbol{\beta}, \widehat{\theta}) - 2J_2(\boldsymbol{\beta}, \widehat{\theta}) \right|,
$$

where the functions are defined as follows:

$$
J_1(\boldsymbol{\beta}, \widehat{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \delta_i \left\{ Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) - \widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) \right\}^2,
$$

$$
J_2(\boldsymbol{\beta}, \widehat{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \delta_i \left\{ q(X_i, \widehat{\theta}) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) \right\} \left\{ \widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) \right\}.
$$

In accordance with the proof strategy outlined in Lemma F.4, we have the approximation $\widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) = \widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta_0) + O_p(\|\widehat{\theta} - \theta_0\|)$. According to Lemma F.1,

$$
\widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta_0) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta_0) = \frac{f(\mathbf{Z}_i^\top \boldsymbol{\beta}) E\left\{\delta q(X, \theta) | \mathbf{Z}_i^\top \boldsymbol{\beta}\right\} + O_p(c_M)}{f(\mathbf{Z}_i^\top \boldsymbol{\beta}) E(\delta | \mathbf{Z}_i^\top \boldsymbol{\beta}) + O_p(c_M)} - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta_0) = \frac{O_p(c_M)}{\pi_M(\mathbf{Z}_i^\top \boldsymbol{\beta})}.
$$

This indicates that under Assumption E3, $\widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta_0) = o_p(1)$. Consequently, it follows that $J_1(\boldsymbol{\beta}, \widehat{\theta}) = o_p(1)$ and $J_2(\boldsymbol{\beta}, \widehat{\theta}) = o_p(1)$. Therefore, it can be concluded that

$I_2 = o_p(1)$. By employing a similar line of reasoning, one can demonstrate that $I_3 = o_p(1)$. Thus far, we can deduce that $\mathcal{J}_M = o_p(1)$, which suggests the consistency of the estimator $\widehat{\boldsymbol{\beta}}$.

Subsequently, we examine the rate at which the estimator $\widehat{\boldsymbol{\beta}}$ converges to the true parameter $\boldsymbol{\beta}_0$. Before that, we need to determine the order of $\widehat{J}_M(\boldsymbol{\beta}, \widehat{\theta}) - J_M(\boldsymbol{\beta}, \widehat{\theta})$ and $\widehat{J}_M(\boldsymbol{\beta}, \theta_0) - J_M(\boldsymbol{\beta}, \theta_0)$. For any θ, we've known that $\widehat{J}_M(\boldsymbol{\beta}, \theta) - J_M(\boldsymbol{\beta}, \theta) = J_1(\boldsymbol{\beta}, \theta) - 2J_2(\boldsymbol{\beta}, \theta)$.

$$J_1(\boldsymbol{\beta}, \theta) = \frac{1}{M}\sum_{i=1}^{M} \frac{\delta_i}{\pi_M^2(\mathbf{Z}_i^\top \boldsymbol{\beta})} O_p(c_M^2) = O_p(a_M^{-1} c_M^2) = o_p((M a_M)^{-1/2}).$$

Next we employ the theory of U-statistics to determine the order of $J_2(\boldsymbol{\beta}, \widehat{\theta})$.

$$\begin{aligned}
J_2(\boldsymbol{\beta}) =& \frac{1}{M}\sum_{i=1}^{M} \delta_i \left\{ q(X_i, \theta) - Q(S_i, \widehat{\theta}) \right\} \frac{\frac{1}{M}\sum_{j=1}^{M} \delta_j K_h(S_j - S_i) \left\{ q(X_j, \theta) - Q(S_i, \widehat{\theta}) \right\}}{\frac{1}{M}\sum_{j=1}^{M} \delta_j K_h(S_j - S_i)} \\
=& \frac{1}{M}\sum_{i=1}^{M} \frac{1}{M}\sum_{j=1}^{M} \frac{\delta_i \delta_j K_h(S_j - S_i) \{q(X_i, \theta) - Q(S_i, \theta)\} \{q(X_j, \theta) - Q(S_i, \theta)\}}{\frac{1}{M}\sum_{j=1}^{M} \delta_j K_h(S_j - S_i)} \\
=& J_{21}(\boldsymbol{\beta}, \theta) + o_p(J_{21}(\boldsymbol{\beta}, \theta)),
\end{aligned}$$

where

$$J_{21}(\boldsymbol{\beta}, \theta) = \frac{1}{M}\sum_{i=1}^{M} \frac{1}{M}\sum_{j=1}^{M} \frac{\delta_i \delta_j K_h(S_j - S_i) \{q(X_i, \theta) - Q(S_i, \theta)\} \{q(X_j, \theta) - Q(S_i, \theta)\}}{\pi_M(S_i) f(S_i)}.$$

Further, $J_{21}(\boldsymbol{\beta}, \theta) = J_{211}(\boldsymbol{\beta}, \theta) + J_{212}(\boldsymbol{\beta}, \theta)$, where

$$J_{211}(\boldsymbol{\beta}, \theta) = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\delta_i \delta_j K_h(S_j - S_i) \{q(X_i, \theta) - Q(S_i, \theta)\} \{q(X_j, \theta) - Q(S_j, \theta)\}}{\pi_M(S_i) f(S_i)},$$

$$J_{212}(\boldsymbol{\beta}, \theta) = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\delta_i \delta_j K_h(S_j - S_i) \{q(X_i, \theta) - Q(S_i, \theta)\} \{Q(S_j, \theta) - Q(S_i, \theta)\}}{\pi_M(S_i) f(S_i)}.$$

For the sake of simplicity in notation, we will omit θ and substitute $\mathbf{Z}^\top \boldsymbol{\beta}$ with S in the following proof when doing so does not lead to confusion. It is easy to see that

$$
\begin{aligned}
2J_{211}(\boldsymbol{\beta}) =& \frac{2}{M^2}\sum_{i\neq j}\frac{\delta_i\delta_j K_h(S_j-S_i)\left\{q(X_i)-Q(S_i)\right\}\left\{q(X_j)-Q(S_j)\right\}}{\pi_M(S_i)f(S_i)} \\
=& \frac{1}{M^2}\sum_{i\neq j}\delta_i\delta_j K_h(S_j-S_i)\left\{q(X_i)-Q(S_i\right\}\left\{q(X_j)-Q(S_j)\right\}\left\{\frac{1}{\pi_M(S_i)f(S_i)}+\frac{1}{\pi_M(S_j)f(S_j)}\right\} \\
:=& \frac{M(M-1)}{2M^2}U_{M,1},
\end{aligned}
$$

where

$$
\begin{aligned}
U_{M,1} =& \frac{1}{M(M-1)/2}\sum_{i\neq j}\delta_i\delta_j K_h(S_j-S_i)\left\{q(X_i)-Q(S_i\right\}\left\{q(X_j)-Q(S_j)\right\} \\
& \times\left\{\frac{1}{\pi_M(S_i)f(S_i)}+\frac{1}{\pi_M(S_j)f(S_j)}\right\}. \\
:=& \frac{1}{M(M-1)/2}\sum_{i\neq j}(H_{ij,i}+H_{ij,j}).
\end{aligned}
$$

Define the Hájek projection of $U_{M,1}$ as

$$
\widehat{U}_{M,1} = \frac{1}{M}\sum_{i=1}^{M}E(H_{ij,i}+H_{ij,j}|\delta_i,\mathbf{Z}_i,X_i).
$$

We study the property of $E(H_{ij,i}+H_{ij,j}|\delta_i,\mathbf{Z}_i,X_i)$.

$$
\begin{aligned}
E(H_{ij,i}|\delta_i,\mathbf{Z}_i,X_i) =& \frac{\delta_i\left\{q(X_i)-Q(S_i\right\}}{\pi_M(S_i)f(S_i)}E\left[\delta_j K_h(S_j-S_i)\left\{q(X_j)-Q(S_j)\right\}|\mathbf{Z}_i\right] \\
=& \frac{\delta_i\left\{q(X_i)-Q(S_i\right\}}{\pi_M(S_i)f(S_i)}E\left(K_h(S_i-S_j)E\left[\delta_j\left\{q(X_j)-Q(S_j)\right\}|\mathbf{Z}_i,\mathbf{Z}_j\right]|\mathbf{Z}_i\right).
\end{aligned}
$$

Under the MAR assumption and the definition of $Q(S)$ as $Q(S)=E\big\{q(X)|S\big\}$, we have the following relationship: $E\big[\delta_j\{q(X_j)-Q(S_j)\}|\mathbf{Z}_i,\mathbf{Z}_j\big] = E(\delta_j|\mathbf{Z}_j)E\big\{q(X_j)-Q(S_j)|\mathbf{Z}_j\big\} = 0$. Then $E(H_{ij,i}|\delta_i,\mathbf{Z}_i,X_i)=0$. For $E(H_{ij,j}|\delta_i,\mathbf{Z}_i,X_i)$,the proof employs a similar methodology.

By utilizing the law of iterated expectations, we obtain:

$$
\begin{aligned}
E(H_{ij,j}|\delta_i, \mathbf{Z}_i, X_i) =& \delta_i \left\{q(X_i) - Q(S_i)\right\} E\left[\frac{\delta_j K_h(S_i - S_j)\left\{q(X_j) - Q(S_j)\right\}}{f(S_j)\pi_M(S_j)}\bigg|\mathbf{Z}_i\right] \\
=& \delta_i \left\{q(X_i) - Q(S_i)\right\} E\left(E\left[\frac{\delta_j K_h(S_i - S_j)\left\{q(X_j) - Q(S_j)\right\}}{f(S_j)\pi_M(S_j)}\bigg|\mathbf{Z}_i, \mathbf{Z}_j\right]\bigg|\mathbf{Z}_i\right) = 0.
\end{aligned}
$$

This result holds due to the MAR assumption and the definition of $Q(S)$, which implies:

$$
E\left[\frac{\delta_j K_h(S_i - S_j)\left\{q(X_j) - Q(S_j)\right\}}{f(S_j)\pi_M(S_j)}\bigg|\mathbf{Z}_i, \mathbf{Z}_j\right] = \frac{K_h(S_i - S_j)}{f(S_j)\pi_M(S_j)} E\left[\delta_j \left\{q(X_j) - Q(S_j)\right\}|\mathbf{Z}_j\right] = 0.
$$

As a result, $\widehat{U}_{M,1} = 0$ and then $J_{211}(\boldsymbol{\beta}) = o_p(M^{-1/2})$.

We can also use the same technique to determine its order.

$$
\begin{aligned}
2J_{212}(\boldsymbol{\beta}) =& \frac{2}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\delta_i\delta_j K_h(S_j - S_i)\left\{q(X_i) - Q(S_i)\right\}\left\{Q(S_j) - Q(S_i)\right\}}{\pi_M(S_i)f(S_i)} \\
:=& \frac{M(M-1)}{2M^2} U_{M,2},
\end{aligned}
$$

where $U_{M,2} = \left\{M(M-1)/2\right\}^{-1}\sum_{i\neq j}(H'_{ij,i} + H'_{ij,j})$,

$$
\begin{aligned}
H'_{ij,i} =& \delta_i\delta_j K_h(S_j - S_i)\frac{\left\{q(X_i) - Q(S_i)\right\}\left\{Q(S_j) - Q(S_i)\right\}}{\pi_M(S_i)f(S_i)}, \\
H'_{ij,j} =& \delta_i\delta_j K_h(S_j - S_i)\frac{\left\{q(X_j) - Q(S_j)\right\}\left\{Q(S_i) - Q(S_j)\right\}}{\pi_M(S_j)f(S_j)}.
\end{aligned}
$$

Define the Hájeck projection of $U_{M,2}$ as $\widehat{U}_{M,2} = M^{-1}\sum_{i=1}^{M} E(H'_{ij,i} + H'_{ij,j}|\delta_i, \mathbf{Z}_i, X_i)$. We just need to study the property of $E(H'_{ij,i} + H'_{ij,j}|\delta_i, \mathbf{Z}_i, X_i)$.

$$
\begin{aligned}
E(H'_{ij,i}|\delta_i, \mathbf{Z}_i, X_i) =& E\left[\frac{\delta_i\delta_j K_h(S_j - S_i)\left\{q(X_i,\theta,p) - Q(S_i,\theta,p)\right\}\left\{Q(S_j,\theta,p) - Q(S_i,\theta,p)\right\}}{\pi_M(S_i)f(S_i)}\bigg|\delta_i, \mathbf{Z}_i, X_i\right] \\
=& \frac{\delta_i\left\{q(X_i) - Q(S_i)\right\}}{\pi_M(S_i)f(S_i)} E\left[\delta_j K_h(S_j - S_i)\left\{Q(S_j) - Q(S_i)\right\}|\mathbf{Z}_i\right]
\end{aligned}
$$

By the law of iterated expectation and Taylor's expansion, we have

$$
\begin{aligned}
&E\left[\delta_j K_h(S_j - S_i)\left\{Q(S_j) - Q(S_i)\right\}|\mathbf{Z}_i\right] = E\left(E\left[\delta_j K_h(S_j - S_i)\left\{Q(S_j) - Q(S_i)\right\}|\mathbf{Z}_i, S_j\right]|\mathbf{Z}_i\right) \\
=&E\left[\pi_M(S_j)K_h(S_j - S_i)\left\{Q(S_j) - Q(S_i)\right\}|\mathbf{Z}_i\right] = \int \frac{1}{h}K\left(\frac{u - S_i}{h}\right)\pi_M(u)\left\{Q(u) - Q(S_i)\right\}f(u)du \\
=&\int K(v)\pi_M(S_i + hv)\left\{Q(S_i + hv) - Q(S_i)\right\}f(S_i + hv)dv \\
=&\int K(v)\left\{\pi_M(S_i) + \dot{\pi}_M(S_i)hv + \frac{1}{2}\ddot{\pi}_M(S_i)h^2v^2\right\}\left\{\dot{Q}(S_i)hv + \frac{1}{2}\ddot{\pi}_M(S_i)h^2v^2\right\} \\
&\qquad \times \left\{f(S_i) + \dot{f}(S_i)hv + \frac{1}{2}\ddot{f}(S_i)h^2v^2\right\}dv \\
=&\left\{\pi_M(S_i) + 1\right\}O(h^2).
\end{aligned}
$$

For $E(H'_{ij,j}|\delta_i, \mathbf{Z}_i, X_i)$, we use the SIM assumption and the MAR assumption and then

$$
\begin{aligned}
&E(H'_{ij,j}|\delta_i, \mathbf{Z}_i, X_i) \\
=&E\left[\left.\frac{\delta_i\delta_j K_h(S_j - S_i)\left\{q(X_j, \theta, p) - Q(S_j, \theta, p)\right\}\left\{Q(S_i, \theta, p) - Q(S_j, \theta, p)\right\}}{\pi_M(S_j)f(S_j)}\right|\delta_i, \mathbf{Z}_i, X_i\right] \\
=&\delta_i E\left[\left.\frac{\delta_j K_h(S_j - S_i)\left\{q(X_j, \theta, p) - Q(S_j, \theta, p)\right\}\left\{Q(S_i, \theta, p) - Q(S_j, \theta, p)\right\}}{\pi_M(S_j)f(S_j)}\right|\mathbf{Z}_i\right] \\
=&\delta_i E\left(E\left[\left.\frac{\delta_j K_h(S_j - S_i)\left\{q(X_j, \theta, p) - Q(S_j, \theta, p)\right\}\left\{Q(S_i, \theta, p) - Q(S_j, \theta, p)\right\}}{\pi_M(S_j)f(S_j)}\right|\mathbf{Z}_i, \mathbf{Z}_j\right]\middle|\mathbf{Z}_i\right) \\
=&\delta_i E\left(\frac{K_h(S_i - S_j)\left\{Q(S_i) - Q(S_j)\right\}}{\pi_M(S_i)f(S_j)}E\left[\delta_j\left\{q(X_j) - Q(S_j)\right\}|\mathbf{Z}_j\right]\middle|\mathbf{Z}_i\right) = 0.
\end{aligned}
$$

Consequently, we have $\widehat{U}_{M,2} = O(h^2)$. This leads to $J_{212}(\boldsymbol{\beta}) = o_p(M^{-1/2})$. As a result, it follows that $J_{21}(\boldsymbol{\beta}) = o_p(M^{-1/2})$. Therefore, $\widehat{J}_M(\boldsymbol{\beta}, \theta) - J_M(\boldsymbol{\beta}, \theta) = o_p(M^{-1/2})$ uniformly on $\boldsymbol{\beta}$.

Next we demonstrate the convergence rate of $\widehat{\boldsymbol{\beta}}$. Given that the class of functions $\big\{\delta\{q(\cdot, \theta) - Q(\cdot^\top\boldsymbol{\beta}, \theta)\}^2 : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta\big\}$ constitutes a Donsker class, the associated empirical process $G_M(\boldsymbol{\beta}, \theta) = \sqrt{M}\{J_M(\boldsymbol{\beta}, \theta) - EJ_M(\boldsymbol{\beta}, \theta)\}$ exhibits equicontinuity. This implies that $|G_M(\boldsymbol{\beta}, \widehat{\theta}) - G_M(\boldsymbol{\beta}, \theta_0)| = o_p(1)$. Since $E\big\{J_M(\boldsymbol{\beta}, \theta)\big\}$ is continuous in θ, uniformly over $\boldsymbol{\beta}$, we have

$$
\begin{aligned}
\widehat{J}_M(\boldsymbol{\beta},\widehat{\theta})-\widehat{J}_M(\boldsymbol{\beta},\theta_0)=&J_M(\boldsymbol{\beta},\widehat{\theta})-J_M(\boldsymbol{\beta},\theta_0)+o_p(M^{-1/2})\\
=&M^{-1/2}\{G_M(\boldsymbol{\beta},\theta)-G_M(\boldsymbol{\beta},\theta_0)\}+EJ_M(\boldsymbol{\beta},\widehat{\theta})-EJ_M(\boldsymbol{\beta},\theta_0)+o_p(M^{-1/2})\\
=&\nabla_\theta E\left\{J_M(\boldsymbol{\beta},\theta_0)\right\}(\widehat{\theta}-\theta_0)+o_p(M^{-1/2}),
\end{aligned}
$$

where the first equation holds because $\widehat{J}_M(\boldsymbol{\beta},\theta)-J_M(\boldsymbol{\beta},\theta)=o_p(M^{-1/2})$ uniformly over $\boldsymbol{\beta}$. Thus,

$$
\widehat{J}_M(\boldsymbol{\beta},\widehat{\theta})=\widehat{J}_M(\boldsymbol{\beta},\theta_0)+\nabla_\theta E\{J_M(\boldsymbol{\beta},\theta_0)\}(\widehat{\theta}-\theta_0)+o_p(M^{-1/2}) \tag{A.4}
$$

holds uniformly on $\boldsymbol{\beta}$.

By the definition of $\widehat{\boldsymbol{\beta}}$ and equation (A.4), the following estimating equation holds:

$$
\begin{aligned}
0=&\nabla_{\boldsymbol{\beta}}\widehat{J}_M(\widehat{\boldsymbol{\beta}},\widehat{\theta})=\nabla_{\boldsymbol{\beta}}J_M(\widehat{\boldsymbol{\beta}},\theta_0)+\nabla_{\boldsymbol{\beta}}\nabla_\theta E\left\{J_M(\widehat{\boldsymbol{\beta}},\theta_0)\right\}(\widehat{\theta}-\theta_0)+o_p(M^{-1/2})\\
=&\nabla_{\boldsymbol{\beta}}J_M(\boldsymbol{\beta}_0,\theta_0)+\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top}J_M(\boldsymbol{\beta}_0,\theta_0)(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(M^{-1/2}),
\end{aligned}
$$

which gives that $\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0=-\big\{\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top}J_M(\boldsymbol{\beta}_0,\theta_0)\big\}^{-1}\nabla_{\boldsymbol{\beta}}J_M(\boldsymbol{\beta}_0,\theta_0)+o_p(M^{-1/2})$. Perform a straightforward computation and subsequently

$$
\begin{aligned}
\nabla_{\boldsymbol{\beta}}J_M(\boldsymbol{\beta}_0,\theta_0)=&-\frac{2}{M}\sum_{i=1}^{M}\delta_i\big\{q(X_i,\theta_0)-Q(S_i,\theta_0)\big\}\dot{Q}(S_i,\theta_0)\mathbf{Z}_i,\\
\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top}J_M(\boldsymbol{\beta}_0,\theta_0)=&\frac{2}{M}\sum_{i=1}^{M}\delta_i\mathbf{Z}_i\mathbf{Z}_i^\top\left[\dot{Q}^2(S_i,\theta_0)-\{q(X_i,\theta_0)-Q(S_i,\theta_0)\}\ddot{Q}(S_i,\theta_0)\right]
\end{aligned}
$$

By the LLN and given that $E\big\{q(X)-Q(S)|\mathbf{Z}\big\}=0$, we can deduce that

$$
\begin{aligned}
\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top}J_M(\boldsymbol{\beta}_0,\theta_0)=&\frac{2}{M}\sum_{i=1}^{M}\delta_i\mathbf{Z}_i\mathbf{Z}_i^\top\left[\dot{Q}^2(S_i,\theta_0)-\{q(X_i,\theta_0)-Q(S_i,\theta_0)\}\ddot{Q}(S_i,\theta_0)\right]\\
\xrightarrow{p}&2E\left\{\delta\mathbf{Z}\mathbf{Z}^\top\dot{Q}^2(S,\theta_0)\right\}.
\end{aligned}
$$

Given that $E\big\{\delta\mathbf{Z}\mathbf{Z}^\top\dot{Q}^2(S,\theta_0)\big\}=E(\delta)E\big\{\mathbf{Z}\mathbf{Z}^\top\dot{Q}^2(S,\theta_0)\big\}=O_p(\pi_M)$, we conclude that

$\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top} J_M(\boldsymbol{\beta}_0, \theta_0) = O_p(\pi_M)$. Next, we examine the second moment of $\nabla_{\boldsymbol{\beta}} J_M(\boldsymbol{\beta}_0, \theta_0)$:

$$E\left\{\nabla_{\boldsymbol{\beta}} J_M^2(\boldsymbol{\beta}_0, \theta_0)\right\} = \frac{4}{M} E\left[\delta\left\{q(X, \theta_0) - Q(S, \theta_0)\right\}^2 \dot{Q}^2(S, \theta_0)\mathbf{Z}\mathbf{Z}^\top\right] = O(\pi_M M^{-1}).$$

This implies that $\nabla_{\boldsymbol{\beta}} J_M(\boldsymbol{\beta}_0, \theta_0) = O_p(\pi_M^{1/2} M^{-1/2})$. Consequently,

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = -\left\{\nabla_{\boldsymbol{\beta}\boldsymbol{\beta}^\top} J_M(\boldsymbol{\beta}_0, \theta_0)\right\}^{-1} \nabla_{\boldsymbol{\beta}} J_M(\boldsymbol{\beta}_0, \theta_0) + o_p(M^{-1/2}) = O_p((\pi_M M)^{-1/2}).$$

The argument for $\widehat{\boldsymbol{\gamma}}$ follows a similar line of reasoning. Thus, the proof is complete. ∎

Lemma E.2. *Under Assumptions B, D and E1 - E3,* $\widehat{Q}(\mathbf{Z}_i^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) = \widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\kappa}_0, \theta_0) + \check{\mathbb{M}}(\mathbf{s})^\top (\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) + q_1(\mathbf{s}, \theta_0)^\top (\widehat{\theta} - \theta_0) + \pi_M^{-1}(\mathbf{s}) O_p(\|\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0\|_2^2)$*, where* $\check{Q}_0(\mathbf{s}) = E\big(\delta\mathbf{Z}|\mathbf{S} = \mathbf{s}\big)$, $\check{Q}_1(\mathbf{s}, \theta) = E\big\{q(X, \theta)\delta\mathbf{Z}|\mathbf{S} = \mathbf{s}\big\}$ *and* $\check{\mathbb{M}}(\mathbf{s}) = \pi_M^{-1}(\mathbf{s})\dot{Q}(\mathbf{s})\big\{\mathbf{z}\pi_M(\mathbf{s}) - \check{Q}_0(\mathbf{s})\big\}$.

Proof. Denote $\xi_{0M}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}) = M^{-1}\sum_{i=1}^M \delta_i \mathbf{K}_h\big\{(\mathbf{Z}_i - \mathbf{z})^\top \widehat{\boldsymbol{\kappa}}\big\}$ and $\xi_{1M}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) = M^{-1}\sum_{i=1}^M \delta_i \mathbf{K}_h\big\{(\mathbf{Z}_i - \mathbf{z})^\top \widehat{\boldsymbol{\kappa}}\big\} q(X_i, \widehat{\theta})$. Then $\widehat{Q}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) = \xi_{1M}(\widehat{\boldsymbol{\kappa}}, \widehat{\theta}) \big/ \xi_{0M}(\widehat{\boldsymbol{\kappa}})$. Employing the methodology outlined in Lemma F.4, we can demonstrate that:

$$\xi_{0M}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}) = \xi_{0M}(\mathbf{z}^\top \boldsymbol{\kappa}_0) + \nabla_{\mathbf{S}}\left\{-E(\delta\mathbf{Z}|\mathbf{s})f(\mathbf{s}) + \mathbf{z}E(\delta|\mathbf{s})f(\mathbf{s})\right\}(\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) + O_p(\|\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0\|^2),$$

$$\begin{aligned}\xi_{1M}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) =& \xi_{1M}(\mathbf{z}^\top \boldsymbol{\kappa}_0, \theta_0) + \nabla_{\mathbf{S}}\left[-E\left\{\delta\mathbf{Z}q(X, \theta_0)|\mathbf{s}\right\} f(\mathbf{s}) + \mathbf{z}E\left\{\delta q(X, \theta_0)|\mathbf{s}\right\} f(\mathbf{s})\right]^\top (\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) \\ &+ \nabla_\theta Q(\mathbf{s}, \theta_0)\pi_M(\mathbf{s})f(\mathbf{s})(\widehat{\theta} - \theta_0) + O_p(\|\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0\|^2).\end{aligned}$$

Based on Lemma E.1, we have $\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0 = O_p((Ma_M)^{-1/2})$. Utilizing the MAR assumption and the SIM assumption, we know that $E\big\{\delta\mathbf{Z}q(X, \theta_0)|\mathbf{S}\big\} = Q(\mathbf{S}, \theta_0)E(\delta\mathbf{Z}|\mathbf{S})$ and $E\big\{\delta q(X, \theta_0)|\mathbf{S}\big\} = Q(\mathbf{S})\pi_M(\mathbf{S})$. Then we have

$$\begin{aligned}\widehat{Q}(\mathbf{z}^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) =& \widehat{Q}(\mathbf{z}^\top \boldsymbol{\kappa}_0, \theta_0) + \dot{Q}(\mathbf{S}, \theta_0)\frac{\mathbf{z}E(\delta|\mathbf{s}) - E(\mathbf{Z}\delta|\mathbf{s})}{\pi_M(\mathbf{S})}(\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) \\ &+ \nabla_\theta Q(\mathbf{s}, \theta_0)(\widehat{\theta} - \theta_0) + \pi_M^{-1}(\mathbf{s})O_p(\|\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0\|_2^2).\end{aligned}$$

The proof has been completed. ∎

With a slight abuse of notation, we define

$$\mathcal{W}_i = q(X_i, \theta_0) - Q(\mathbf{S}_i, \theta_0), \quad B(\mathbf{S}_i, \mathbf{S}_j) = Q(\mathbf{S}_i, \theta_0) - Q(\mathbf{S}_j, \theta_0),$$
$$\varphi_M(\mathbf{S}_j) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{K}_h(\mathbf{S}_i - \mathbf{S}_j)\mathcal{W}_i, \quad \mathcal{B}_M(\mathbf{S}_j) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{K}_h(\mathbf{S}_i - \mathbf{S}_j)B(\mathbf{S}_i, \mathbf{S}_j).$$

Lemma E.3. *Under Assumptions D, E1, E3 (ii) and E4 (iii),*

$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)} = \frac{1}{M}\sum_{i=1}^{M} \frac{\delta_i \mathcal{W}_i}{\pi_M(\mathbf{S}_i)} + o_p(h^2),$$
$$\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)B(\mathbf{S}_j, \mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)} = O_p(h^2 a_M^{-1}).$$

Proof. By simple calculations,

$$2\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)} = \frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \left\{ \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)} + \frac{\delta_i \mathbf{K}_h(\mathbf{S}_i - \mathbf{S}_j)\mathcal{W}_i}{\pi_M(\mathbf{S}_j) f(\mathbf{S}_j)} \right\}$$
$$= \frac{M(M-1)/2}{M^2} U_M + \frac{2}{M}\sum_{i=1}^{M} \frac{\delta_i \mathbf{K}_h(0)\mathcal{W}_i}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)},$$

where $U_M = \frac{1}{M(M-2)/2}\sum_{i\neq j}^{M} H_{ij}$ and

$$H_{ij} = \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)} + \frac{\delta_i \mathbf{K}_h(\mathbf{S}_i - \mathbf{S}_j)\mathcal{W}_i}{\pi_M(\mathbf{S}_j) f(\mathbf{S}_j)}.$$

Note that U_M is a U statistic of degree 2.

Denote $\widehat{U}_M = \frac{2}{M}\sum_{i=1}^{M} h(\mathbf{Z}_i, \delta_i, X_i)$, where

$$h(\mathbf{Z}_i, \delta_i, X_i) = E\left\{ \frac{\delta_j \mathbf{K}_h(\mathbf{S}_j - \mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i) f(\mathbf{S}_i)} + \frac{\delta_i \mathbf{K}_h(\mathbf{S}_i - \mathbf{S}_j)\mathcal{W}_i}{\pi_M(\mathbf{S}_j) f(\mathbf{S}_j)} \middle| \mathbf{Z}_i, \delta_i, X_i \right\}.$$

By the MAR assumption, $\delta \perp q(X, \theta_0)|\mathbf{Z}$ and the SIM assumption, we have $E(\delta_j \mathcal{W}_j|\mathbf{S}_j) =$

$E\big\{E(\delta_j\mathcal{W}_j|\mathbf{Z}_j)|\mathbf{S}_j\big\} = E\big[E\big\{\delta_j q(X_j,\theta_0)|\mathbf{Z}_j\big\} - E\big\{\delta_j|\mathbf{Z}_j\big\}Q(\mathbf{S}_j,\theta_0)|\mathbf{S}_j\big] = 0$. Subsequently, we have $E\big\{h(\mathbf{Z}_i,\delta_i,X_i)\big\} = 0$, since

$$\begin{aligned}E\left\{\frac{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}\right\} =&E\left[E\left\{\frac{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}\middle|\mathbf{S}_i,\mathbf{S}_j\right\}\right]\\ =&E\left\{\frac{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}E\left(\delta_j\mathcal{W}_j|\mathbf{S}_i,\mathbf{S}_j\right)\right\} = 0.\end{aligned}$$

Next we try to simplify $h(\mathbf{Z}_i,\delta_i,X_i)$. Denote

$$\begin{aligned}h_1(\mathbf{Z}_i,\delta_i,X_i) =&E\left\{\frac{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\},\\ h_2(\mathbf{Z}_i,\delta_i,X_i) =&E\left\{\frac{\delta_i\mathbf{K}_h(\mathbf{S}_i-\mathbf{S}_j)\mathcal{W}_i}{\pi_M(\mathbf{S}_j)f(\mathbf{S}_j)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\}.\end{aligned}$$

By using the law of iterated expectation repeatedly and the fact that $E(\delta_j\mathcal{W}_j|\mathbf{S}_j) = 0$, we have

$$\begin{aligned}h_1(\mathbf{Z}_i,\delta_i,X_i) =&E\left\{\frac{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\} = \frac{1}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}E\left\{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j\big|\mathbf{Z}_i,\delta_i,X_i\right\}\\ =&\frac{1}{\pi_M(S_i)f(S_i)}E\left\{\delta_jK_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j\big|\mathbf{Z}_i\right\} = \frac{1}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}E\left[E\left\{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j\big|\mathbf{Z}_i,\mathbf{S}_j\right\}|\mathbf{Z}_i\right]\\ =&\frac{1}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}E\left\{K_h(\mathbf{S}_j-\mathbf{S}_i)E\left(\delta_j\mathcal{W}_j|\mathbf{S}_j\right)\big|\mathbf{Z}_i\right\} = 0.\end{aligned}$$

Next we try to simplify $h_2(\mathbf{Z}_i,\delta_i,X_i)$:

$$\begin{aligned}h_2(\mathbf{Z}_i,\delta_i,X_i) =&E\left\{\frac{\delta_i\mathbf{K}_h(\mathbf{S}_i-\mathbf{S}_j)\mathcal{W}_i}{\pi_M(\mathbf{S}_j)f(\mathbf{S}_j)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\} = \delta_i\mathcal{W}_iE\left\{\frac{\mathbf{K}_h(\mathbf{S}_i-\mathbf{S}_j)}{\pi_M(\mathbf{S}_j)f(\mathbf{S}_j)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\}\\ =&\delta_i\mathcal{W}_i\int_{-c}^{c}\frac{1}{h}\mathbf{K}\left(\frac{\mathbf{S}_i-u}{h}\right)\frac{1}{\pi_M(u)}du = \delta_i\mathcal{W}_i\int_{\mathbf{S}_{\frac{i-c}{h}}}^{\mathbf{S}_{\frac{i-c}{h}}}\mathbf{K}(v)\frac{1}{\pi_M(\mathbf{S}_i-hv)}dv\\ =&\delta_i\mathcal{W}_i\int_{\mathbf{S}_{\frac{i-c}{h}}}^{\mathbf{S}_{\frac{i-c}{h}}}K(v)\frac{1}{\pi_M(\mathbf{S}_i)-\dot{\pi}_M(\mathbf{S}_i)hv+\frac{1}{2}\ddot{\pi}_M(\mathbf{S}_i)h^2v^2+O(h^3)}dv\\ =&\delta_i\mathcal{W}_i\int_{\mathbf{S}_{\frac{i-c}{h}}}^{\mathbf{S}_{\frac{i-c}{h}}}K(v)\frac{1}{\pi_M(\mathbf{S}_i)}\left\{1+\frac{\dot{\pi}_M(\mathbf{S}_i)}{\pi_M(\mathbf{S}_i)}hv-\frac{\ddot{\pi}_M(\mathbf{S}_i)h^2v^2+O(h^3)}{\pi_M(\mathbf{S}_i)}\right\}dv\end{aligned}$$

$$=\frac{\delta_i\mathcal{W}_i}{\pi_M(\mathbf{S}_i)}+\frac{\delta_i\mathcal{W}_i}{\pi_M^2(\mathbf{S}_i)}\ddot{\pi}_M(\mathbf{S}_i)\int_{\frac{\mathbf{S}_{i-c}}{h}}^{\frac{\mathbf{S}_{i-c}}{h}}K(v)v^2dvh^2+\frac{\delta_i\mathcal{W}_i}{\pi_M^2(\mathbf{S}_i)}O(h^3)$$

where the fifth equality holds due to the Taylor's expansion theorem, the properties of kernel function $K(\cdot)$ and Assumption E3 (ii). Then

$$\frac{1}{M}\sum_{i=1}^{M}\frac{\delta_i\mathcal{W}_i}{\pi_M^2(\mathbf{S}_i)}\ddot{\pi}_M(\mathbf{S}_i)\int_{\frac{\mathbf{S}_{i-c}}{h}}^{\frac{\mathbf{S}_{i-c}}{h}}K(v)v^2dv=E\left\{\frac{\delta_i\mathcal{W}_i}{\pi_M^2(\mathbf{S}_i)}\ddot{\pi}_M(\mathbf{S}_i)\int_{\frac{\mathbf{S}_{i-c}}{h}}^{\frac{\mathbf{S}_{i-c}}{h}}K(v)v^2dv\right\}+o_p(1)=o_p(1),$$

where the second equality holds due to the fact that $E(\delta_i\mathcal{W}_i|\mathbf{S}_i)=0$. Hence, we have

$$\widehat{U}_M=\frac{2}{M}\sum_{i=1}^{M}\frac{\delta_i\mathcal{W}_i}{\pi_M(\mathbf{S}_i)}+o_p(h^2).$$

By the projection method of U-statistics and Lemma F.5, $U_M=\widehat{U}_M+o_p(M^{-1/2})$.

$$\begin{aligned}&2\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\frac{\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\\&=\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M}\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\left\{\frac{\delta_jB(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}+\frac{\delta_iB(\mathbf{S}_i,\mathbf{S}_j)}{f(\mathbf{S}_j)\pi_M(\mathbf{S}_j)}\right\}=\frac{M(M-1)/2}{M^2}V_M,\end{aligned}$$

where

$$V_M=\frac{2}{M(M-1)}\sum_{i\neq j}^{M}\tilde{H}_{ij}=\frac{2}{M(M-1)}\sum_{i\neq j}^{M}\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\left\{\frac{\delta_jB(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}+\frac{\delta_iB(\mathbf{S}_i,\mathbf{S}_j)}{f(\mathbf{S}_j)\pi_M(\mathbf{S}_j)}\right\}.$$

Note that V_M is a U statistic of degree 2.

Denote

$$\widehat{V}_M=\frac{2}{M}\sum_{i=1}^{M}\tilde{h}(\mathbf{Z}_i,\delta_i,X_i)=\frac{2}{M}\sum_{i=1}^{M}E\left[\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\left\{\frac{\delta_jB(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}+\frac{\delta_iB(\mathbf{S}_i,\mathbf{S}_j)}{f(\mathbf{S}_j)\pi_M(\mathbf{S}_j)}\right\}\middle|\mathbf{Z}_i,\delta_i,X_i\right]$$

$$
=:\frac{2}{M}\sum_{i=1}^{M}\left\{\tilde{h}_1(\mathbf{Z}_i,\delta_i,X_i)+\tilde{h}_2(\mathbf{Z}_i,\delta_i,X_i)\right\}.
$$

First, we calculate the expectation of $\tilde{h}(\mathbf{Z}_i,\delta_i,X_i)$:

$$
\begin{aligned}
&E\left\{\tilde{h}(\mathbf{Z}_i,\delta_i,X_i)\right\}\\
=&E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{\delta_j B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\right\}=E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}E(\delta_j|\mathbf{S}_i,\mathbf{S}_j)\right\}\\
=&E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\pi_M(\mathbf{S}_j)\right\}=E\left[\pi_M(\mathbf{S}_j)E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\middle|\mathbf{S}_j\right\}\right]\\
=&E\left\{\pi_M(\mathbf{S}_j)\int\frac{1}{h}\mathbf{K}\left(\frac{\mathbf{S}_j-u}{h}\right)\frac{Q(\mathbf{S}_j)-Q(u)}{\pi_M(u)}du\right\}=E\left\{\pi_M(\mathbf{S}_j)\int\mathbf{K}(v)\frac{Q(\mathbf{S}_j)-Q(\mathbf{S}_j-hv)}{\pi_M(\mathbf{S}_j-hv)}dv\right\}\\
=&E\left\{\pi_M(\mathbf{S}_j)\int\mathbf{K}(v)\frac{\dot{Q}(\mathbf{S}_j)hv+O(h^2)}{\pi_M(\mathbf{S}_j)-\dot{\pi}_M(\mathbf{S}_j)hv+O(h^2)}dv\right\}=O\left(h^3a_M^{-1}\right).
\end{aligned}
$$

Next we turn to $\tilde{h}_1(\mathbf{Z}_i,\delta_i,X_i)$ and $\tilde{h}_2(\mathbf{Z}_i,\delta_i,X_i)$.

$$
\begin{aligned}
\tilde{h}_1(\mathbf{Z}_i,\delta_i,X_i)=&E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{\delta_j B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\middle|\mathbf{Z}_i,\delta_i,X_i\right\}\\
=&E\left[E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{\delta_j B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\middle|\mathbf{Z}_i,\delta_i,X_i,\mathbf{S}_j\right\}\middle|\mathbf{Z}_i,\delta_i,X_i\right]\\
=&E\left\{\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{B(\mathbf{S}_j,\mathbf{S}_i)}{f(\mathbf{S}_i)\pi_M(\mathbf{S}_i)}\pi_M(\mathbf{S}_j)\middle|\mathbf{Z}_i,\delta_i,X_i\right\}\\
=&f^{-1}(\mathbf{S}_i)\pi_M^{-1}(\mathbf{S}_i)\int\frac{1}{h}\mathbf{K}\left(\frac{u-\mathbf{S}_i}{h}\right)\left\{Q(u)-Q(\mathbf{S}_i)\right\}\pi_M(u)f(u)du\\
=&f^{-1}(\mathbf{S}_i)\pi_M^{-1}(\mathbf{S}_i)\int\mathbf{K}(v)\left\{Q(\mathbf{S}_i+hv)-Q(\mathbf{S}_i)\right\}\pi_M(\mathbf{S}_i+hv)f(\mathbf{S}_i+hv)du\\
=&\pi_M^{-1}(\mathbf{S}_i)O_p(h^2),
\end{aligned}
$$

and similarly

$$
\begin{aligned}
\tilde{h}_2(\mathbf{Z}_i,\delta_i,X_i)=&E\left[\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\frac{\delta_i B(\mathbf{S}_i,\mathbf{S}_j)}{f(\mathbf{S}_j)\pi_M(\mathbf{S}_j)}\middle|\mathbf{Z}_i,\delta_i,X_i\right]=\delta_i\int\frac{1}{h}\mathbf{K}\left(\frac{u-\mathbf{S}_i}{h}\right)\frac{Q(\mathbf{S}_i)-Q(u)}{f(u)\pi_M(u)}f(u)du\\
=&\delta_i\int\mathbf{K}(v)\frac{Q(\mathbf{S}_i)-Q(\mathbf{S}_i+hv)}{\pi_M(\mathbf{S}_i+hv)}dv=\frac{\delta_i}{\pi_M(\mathbf{S}_i)}O_p(h^2).
\end{aligned}
$$

Then by the definition of a_M^{-1}, $\widehat{V}_M = O_p(h^2 a_M^{-1}) = o_p((a_M M)^{-1/2})$ by Assumption E4 (iii). ∎

Denote

$$\mathrm{IF}_i = q(X_i, \theta_0) - \Delta_0 + \frac{\delta_i - \pi_M(\mathbf{S}_i)}{\pi_M(\mathbf{S}_i)} \{q(X_i, \theta_0) - Q(\mathbf{S}_i, \theta_0)\},$$

$$\Sigma_M^{\mathrm{IF}} = \mathrm{Var}\{q(X, \theta_0)\} + E\left[\frac{1 - \pi_M(\mathbf{S})}{\pi_M(\mathbf{S})} \{q(X, \theta_0) - Q(\mathbf{S}, \theta_0)\}^2\right].$$

Proof of Theorem E.1. First, derive the linear expansion of the estimator denoted as $\widehat{\Delta}_{ss}^{dk}$.

$$\begin{aligned}
&\widehat{\Delta}_{ss}^{dk} - \frac{1}{M}\sum_{i=1}^{M} Q(\mathbf{Z}_i^\top \boldsymbol{\kappa}_0, \theta_0) \\
=&\frac{1}{M}\sum_{i=1}^{M}\left\{\widehat{Q}(\mathbf{Z}_i^\top \widehat{\boldsymbol{\kappa}}, \widehat{\theta}) - \widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\kappa}_0, \theta_0)\right\} + \frac{1}{M}\sum_{i=1}^{M}\left\{\widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\kappa}_0, \theta_0) - Q(\mathbf{Z}_i^\top \boldsymbol{\kappa}_0, \theta_0)\right\} := I_1 + I_2.
\end{aligned}$$

By the Lemma E.2, we've shown that

$$I_1 = \frac{1}{M}\sum_{i=1}^{M}\left\{\check{\mathbb{M}}^\top(\mathbf{S}_i)(\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0) + q_1(\mathbf{S}_i)^\top(\widehat{\theta} - \theta_0) + \pi_M^{-1}(\mathbf{S}_i) O_p(\|\widehat{\boldsymbol{\kappa}} - \boldsymbol{\kappa}_0\|_2^2)\right\}.$$

Under the SIM Assumption, which posits that $E(\mathbf{Z}\delta|\mathbf{S}) = E(\mathbf{Z}|\mathbf{S})E(\delta|\mathbf{S})$, we have the following expression:

$$\begin{aligned}
\frac{1}{M}\sum_{i=1}^{M}\check{\mathbb{M}}(\mathbf{S}_i) =&E\left\{\check{\mathbb{M}}(\mathbf{S})\right\} + o_p(1) = E\left\{\dot{Q}(\mathbf{S}, \theta_0)\frac{\mathbf{Z}E(\delta|\mathbf{S}) - E(\mathbf{Z}\delta|\mathbf{S})}{\pi_M(\mathbf{S})}\right\} + o_p(1) \\
=&E\left\{\dot{Q}(\mathbf{S}, \theta_0)\frac{E(\mathbf{Z}|\mathbf{S})E(\delta|\mathbf{S}) - E(\mathbf{Z}\delta|\mathbf{S})}{\pi_M(\mathbf{S})}\right\} + o_p(1) = o_p(1),
\end{aligned}$$

The final equality is justified by the law of iterated expectations. Subsequently, under Assumption E4 (iii), we have

$$I_1 = A(\theta_0)^\top(\widehat{\theta} - \theta_0) + O_p((a_M \pi_M M)^{-1}) = A(\theta_0)^\top(\widehat{\theta} - \theta_0) + o_p((a_M M)^{-1/2}). \qquad \text{(A.5)}$$

Next, we use the theory of U statistic to decompose I_2. Note the fact that $\pi_M(s)Q_0(s) = E(\delta|S)E\{\mathbf{Z}|S\}$. By the definition of $\widehat{Q}(\cdot)$,

$$
\begin{aligned}
I_2 =& \frac{1}{M}\sum_{i=1}^{M}\Big\{\widehat{Q}(\mathbf{Z}_i^\top\boldsymbol{\kappa}_0,\theta_0) - Q(\mathbf{Z}_i^\top\boldsymbol{\kappa}_0,\theta_0)\Big\} \\
=& \frac{1}{M}\sum_{i=1}^{M}\frac{M^{-1}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\{q(X_j,\theta_0)-Q(\mathbf{S}_i,\theta_0)\}}{M^{-1}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)} \\
=& \frac{1}{M}\sum_{i=1}^{M}\frac{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)} + \frac{1}{M}\sum_{i=1}^{M}\frac{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)B(\mathbf{S}_j,\mathbf{S}_i)}{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)} \\
:=& I_{21} + I_{22}.
\end{aligned}
$$

By Lemma F.1, $M^{-1}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i) = f(\mathbf{S}_i)\pi_M(\mathbf{S}_i) + O_p(c_M)$, where $c_M = h^2 + \big\{\log h^{-1}/(Mh)\big\}^{1/2}$. Then I_{21} can be transformed into

$$
I_{21} = \frac{1}{M}\sum_{i=1}^{M}\frac{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)} + \frac{1}{M}\sum_{i=1}^{M}\frac{\frac{1}{M}\sum_{j=1}^{M}\delta_j\mathbf{K}_h(\mathbf{S}_j-\mathbf{S}_i)\mathcal{W}_j}{\pi_M(\mathbf{S}_i)f(\mathbf{S}_i)}O_p(c_M) := I_{211} + I_{212}.
$$

Since $O_p(c_M) = o_p(1)$, it follows that $I_{212} = o_p(I_{211})$. Consequently, our primary task is to ascertain the order I_{211}. According to Lemma E.3, $I_{211} = M^{-1}\sum_{i=1}^{M}\delta_i\mathcal{W}_i\pi_M^{-1}(\mathbf{S}_i) + o_p(h^2)$. Then $I_{21} = M^{-1}\sum_{i=1}^{M}\delta_i\mathcal{W}_i\pi_M^{-1}(\mathbf{S}_i) + o_p\big((Ma_M)^{-1/2}\big)$.

For I_{22}, the argument proceeds in a similar fashion and by Lemma E.3, we have $I_{22} = o_p((Ma_M)^{-1/2})$. Consequently, this leads to the expression:

$$
I_2 = \frac{1}{M}\sum_{i=1}^{M}\frac{\delta_i\mathcal{W}_i}{\pi_M(\mathbf{S}_i)} + o_p((Ma_M)^{-1/2}). \tag{A.6}
$$

By integrating equations (A.5) and (A.6), we derive:

$$
\widehat{\Delta}_{ss}^{dk} = \frac{1}{M}\sum_{i=1}^{M}Q(\mathbf{S}_i,\theta_0) + A(\theta_0)^\top(\widehat{\theta}-\theta_0) + \frac{1}{M}\sum_{i=1}^{M}\frac{\delta_i\mathcal{W}_i}{\pi_M(\mathbf{S}_i)} + o_p((Ma_M)^{-1/2}).
$$

By substituting $\mathcal{W}_i$ with its defined expression, we can subsequently derive the linear expansion of $\widehat{\Delta}_{ss}^{dk} - \Delta_0$:

$$\begin{aligned}
\widehat{\Delta}_{ss}^{dk} - \Delta_0 =& \frac{1}{M}\sum_{i=1}^{M}\{Q(\mathbf{S}_i,\theta_0) - \Delta_0\} + \frac{1}{M}\sum_{i=1}^{M}\frac{\delta_i}{\pi_M(\mathbf{S}_i)}\{q(X_i,\theta_0) - Q(\mathbf{S}_i,\theta_0)\} \\
&+ A(\theta_0)^\top(\widehat{\theta} - \theta_0) + o_p((Ma_M)^{-1/2}) \\
=& \frac{1}{M}\sum_{i=1}^{M}\left[q(X_i,\theta_0) - \Delta_0 + \frac{\delta_i - \pi_M(\mathbf{S}_i)}{\pi_M(\mathbf{S}_i)}\{q(X_i,\theta_0) - Q(\mathbf{S}_i,\theta_0)\}\right] \\
&+ A(\theta_0)^\top(\widehat{\theta} - \theta_0) + o_p((Ma_M)^{-1/2}).
\end{aligned}$$

Second, based on the linear expansion form, we establish the asymptotic normality for our estimator $\widehat{\Delta}_{ss}^{dk}$. Denote

$$\mathrm{IF}_i = q(X_i,\theta_0) - \Delta_0 + \frac{\delta_i - \pi_M(\mathbf{S}_i)}{\pi_M(\mathbf{S}_i)}\{q(X_i,\theta_0) - Q(\mathbf{S}_i,\theta_0)\}.$$

It is straightforward to verify that $E(\mathrm{IF}_i) = 0$. For the second moment, we have:

$$\begin{aligned}
\Sigma_M^{\mathrm{IF}} = E(\mathrm{IF}^2) =& \mathrm{Var}\{q(X,\theta_0)\} + E\left[\frac{\{\delta - \pi_M(\mathbf{S})\}^2}{\pi_M^2(\mathbf{S})}\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\}^2\right] \\
=& \mathrm{Var}\{q(X,\theta_0)\} + E\left[\frac{1 - \pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\}^2\right],
\end{aligned}$$

where the final equality is justified by the MAR assumption, the identity $\delta^2 = \delta$ and the SIM assumption for $\pi_M(\mathbf{Z})$. To control the order of $M^{-1}E(\mathrm{IF}^2)$, we enforce uniform lower and upper bounds for $E\big[\big\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\big\}^2|\mathbf{S}\big]$. By the definition of ROC curve, it is easy to see that both $q(X,\theta_0)$ and $Q(\mathbf{S},\theta_0)$ are upper bounded by 1. Thus, $\big\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\big\}^2 \le 4$ and $\mathrm{Var}\big\{q(X,\theta_0)\big\} \le 4$. By Assumption E4 (i), we have $\mathrm{Var}\{q(X,\theta_0)\} = E\big(E\big[\big\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\big\}^2|\mathbf{S}\big] + E\big[\big\{Q(\mathbf{S},\theta_0) - \Delta_0\big\}^2|\mathbf{S}\big]\big) \ge \sigma_\xi^2$. Use the MAR assumption ,

$$E\left[\frac{1 - \pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\{q(X,\theta_0) - Q(\mathbf{S},\theta_0)\}^2\right] \ge E\left[\frac{1 - \pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\right]\sigma_{\xi,1}^2,$$

$$E\left[\frac{1-\pi_M(S)}{\pi_M(\mathbf{S})}\{q(X,\theta_0)-Q(\mathbf{S},\theta_0)\}^2\right] \le 4E\left[\frac{1-\pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\right].$$

Recall that by definition $a_M^{-1} = E\big\{\pi_M^{-1}(\mathbf{Z})\big\}$. Subsequently, we have

$$a_M E(\mathrm{IF}^2) \ge a_M\left[\sigma_{\xi,1}^2 + E\left\{\frac{1-\pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\right\}\sigma_\xi^2\right] = \sigma_\xi^2 > 0,$$

$$a_M E(\mathrm{IF}^2) \le a_M\left[4 + 4E\left\{\frac{1-\pi_M(\mathbf{S})}{\pi_M(\mathbf{S})}\right\}\right] = 4 < \infty.$$

Consequently, $\Sigma_M^{\mathrm{IF}} \asymp a_M^{-1}$. Then we have $E\big\{(Ma_M)^{1/2}M^{-1}\sum_{i=1}^M \mathrm{IF}_i\big\}^2 = O_p(1)$. Under Assumption E4 (ii), by the Lindeberg-Feller theorem, $(\Sigma_M^{\mathrm{IF}})^{-1/2}M^{-1/2}\sum_{i=1}^M \mathrm{IF}_i \xrightarrow{d} N(0,1)$. By Assumption E2 that $r_{\mathrm{mar}} = \lim_{m,M\to\infty} M/(m\Sigma_M^{\mathrm{IF}})$. Then we have

$$(\Sigma_M^{\mathrm{IF}})^{-1/2}\sqrt{M}(\widehat{\theta}-\theta_0) = (\Sigma_M^{\mathrm{IF}})^{-1/2}\sqrt{\frac{M}{m}}\sqrt{m}(\widehat{\theta}-\theta_0) \to N(0, r_{mar}I^{-1}(\theta_0)).$$

Denote $\Sigma_M^{\mathrm{mar}} = \Sigma_M^{\mathrm{IF}} + r_{\mathrm{mar}}I^{-1}(\theta_0)$. The asymptotic normality for $\widehat{\Delta} - \Delta_0$ is

$$\left(\Sigma_M^{\mathrm{IF}}\right)^{-1/2}\sqrt{M}(\widehat{\Delta}_{ss}^{dk} - \Delta_0) \xrightarrow{d} N(0, 1 + r_{mar}A^2(\theta_0)I^{-1}(\theta_0)).$$

Then the proof has been completed. ∎

F Main Steps of Proof

In this section, we sketch the main steps in the proofs.

Write $f(s)$ is the density of $S_i = \mathbf{Z}_i^\top \boldsymbol{\beta}$; $s = \mathbf{z}^\top \boldsymbol{\beta}$, $Q(s,\theta,p) = E\{q(X_i,\theta,p)|S_i = s\}$, $q_1(s,\theta,p) = \nabla_\theta Q(s,\theta_0,p)$, $Q_0(s) = E(\mathbf{Z}_i|S_i = s)$ and $Q_1(s,\theta,p) = E\{q(X_i,\theta,p)\mathbf{Z}_i|S_i = s\}$. Denote $\widehat{f}_n(\mathbf{z}^\top\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n K_h\left\{(\mathbf{Z}_i - \mathbf{z})^\top\boldsymbol{\beta}\right\}$. Recall that

$$
\begin{aligned}
\widehat{Q}(\mathbf{Z}_j^\top\boldsymbol{\beta},\theta,p) =& n^{-1}\sum_{i=1}^n K_h\left\{(\mathbf{Z}_i - \mathbf{Z}_j)^\top\boldsymbol{\beta}\right\} q\left(X_i,\theta,p\right) \Big/ \widehat{f}_n(\mathbf{Z}_j^\top\boldsymbol{\beta}),\\
\tilde{Q}(\mathbf{Z}_j^\top\boldsymbol{\beta},\theta,p) =& n^{-1}\sum_{i=1}^n K_h\left\{(\mathbf{Z}_i - \mathbf{Z}_j)^\top\boldsymbol{\beta}\right\} q\left(X_i,\theta,p\right) \Big/ \tilde{f}_n(\mathbf{Z}_j^\top\boldsymbol{\beta}),
\end{aligned}
$$

where $\tilde{f}_n(\mathbf{z}^\top\boldsymbol{\beta}) = \widehat{f}_n(\mathbf{z}^\top\boldsymbol{\beta})\mathbb{I}\big\{\widehat{f}_n(\mathbf{z}^\top\boldsymbol{\beta}) \geq w_n\big\} + w_n\mathbb{I}\big\{\widehat{f}_n(\mathbf{z}^\top\boldsymbol{\beta}) < w_n\big\}$. The supervised estimator is $\widehat{\Delta}_{\mathcal{L}}(p) = n^{-1}\sum_{i=1}^n q(X_i,\widehat{\theta},p)$. Since we've assumed that $\widehat{\theta}$ is estimated by MLE, $\widehat{\theta} - \theta_0 = O_p(m^{-1/2})$.

For convenience, we define some notations. Denote $\widehat{\zeta}_i(p) = \mathbb{I}\big\{\widehat{f}_n(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}(p)) > w_n\big\}$ and $\zeta_i(p) = \mathbb{I}\big\{f(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(p)) > w_n\big\}$. Define ϵ_n such that $w_n^{-1}\epsilon_n \to 0$ and $\epsilon_n \asymp c_n$. Write $\zeta_i'(p) = \mathbb{I}\big\{f(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(p)) > w_n + \epsilon_n\big\}$, $J_i^0(p) = \mathbb{I}\big\{\widehat{f}_n(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}(p)) > w_n, f(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(p)) \leq w_n + \epsilon_n\big\}$ and $J_i^1(p) = \mathbb{I}\big\{\widehat{f}_n(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}(p)) \leq w_n, f(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(p)) > w_n + \epsilon_n\big\}$. When n is sufficiently large, $J_i^1(p) = 0$ since the set $\big\{\widehat{f}_n(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}}(p)) \leq w_n, f(\mathbf{Z}_i^\top\boldsymbol{\beta}_0(p)) > w_n + \epsilon_n\big\}$ is empty.

For the sake of the convenience, write $\Omega_1(p) = C(\boldsymbol{\beta}_0,\theta_0,p)^\top\mathbf{A}_0^{-1}\Sigma_{\boldsymbol{\beta}}\big(\mathbf{A}_0^{-1}\big)^\top C(\boldsymbol{\beta}_0,\theta_0,p)$, $\Omega_2(p) = C(\boldsymbol{\beta}_0,\theta_0,p)^\top\mathbf{A}_0^{-1}\Sigma_{q,\boldsymbol{\beta}}(p)$ and $\Omega_3(p) = C(\boldsymbol{\beta}_0,\theta_0,p)^\top\mathbf{A}_0^{-1}\Sigma_{Q,\boldsymbol{\beta}}(p)$, $\gamma(X) = C(\boldsymbol{\beta}_0,\theta_0,p)^\top\mathbf{A}_0^{-1}\alpha(X)$. Then $\Omega_1(p) = E\big\{\gamma^2(X)\big\}$, $\Omega_2(p) = E\big\{q(X,\theta_0,p)\gamma(X)\big\}$ and $\Omega_3(p) = E\big\{Q(S,\theta_0,p)\gamma(X)\big\}$.

Lemma F.1. *Suppose that the second derivatives of $f(x)$ are continuous and bounded. Let $E\{g(X,u)|U = u\}$ be continuous and twice differentiable at u and $E\big\{|g(X,U)|^2\big\} < \infty$. If the second derivative of $f(x)$ are continuous and bounded, then as $n \to \infty$, we have*

$$
\sup_{u\in\mathcal{U}}\left|\frac{1}{n}\sum_{i=1}^n K_h(U_i - u)\left(\frac{U_i - u}{h}\right)^k g(X_i,U_i) - f(u)E\{g(X,u)|U = u\}\mu_k\right| = O(c_n) \qquad a.s.
$$

where $\mathcal{U}$ *is the support of* U, $\mu_k = \int u^k \mathcal{K}(u)du$ *and* $c_n = h^2 + \{\log h^{-1}/(nh)\}^{1/2}$.

Proof. The proof may be constructed along the lines of Xia and Li (1999, Lemma A.2). ∎

Proof of Proposition 3.1. Consider the inequality

$$\sup_{0\le p\le 1}\left|\widehat{\Delta}_{\mathcal{L}}(p) - \Delta_0(p)\right| = \sup_{0\le p\le 1}\left|F_n G^{-1}(1-p,\widehat{\theta}) - FG^{-1}(1-p,\theta_0)\right|$$
$$\le \sup_{0\le p\le 1}\left|F_n G^{-1}(1-p,\widehat{\theta}) - FG^{-1}(1-p,\widehat{\theta})\right| + \sup_{0\le p\le 1}\left|FG^{-1}(1-p,\theta_0) - FG^{-1}(1-p,\widehat{\theta})\right|.$$

For the first term, we have $\sup\limits_{0\le p\le 1}|F_n G_m^{-1}(1-p,\widehat{\theta}) - FG_m^{-1}(1-p,\widehat{\theta})| \le \sup\limits_{x}|F_n(x) - F(x)| \to 0$ almost surely, which is guaranteed by the Glivenko–Cantelli theorem.

The second term can be written as below, using a Taylor's expansion,

$$\sup_{0\le p\le 1}\left|FG^{-1}(1-p,\theta_0) - FG^{-1}(1-p,\widehat{\theta})\right| = \sup_{0\le p\le 1}\left|\frac{fG^{-1}(1-p,\theta_n)}{gG^{-1}(1-p,\theta_n)}(\widehat{\theta}-\theta_0)\right|$$
$$= \sup_{0\le p\le 1}\left|A(\theta_n,p)(\widehat{\theta}-\theta_0)\right| \le \sup_{0\le p\le 1}|A(\theta_n,p)|\,|\widehat{\theta}-\theta_0| \xrightarrow{p} 0$$

if we assume that the density ratio $\sup\limits_{p\in(0,1)} A(\theta,p)$ is bounded and θ_n is between θ_0 and $\widehat{\theta}$.

Now that we've shown that each element of the maximum converges uniformly to 0 in probability, the proof is concluded. ∎

Proof of Proposition 3.2. In order to prove the weak convergence,

$$\sqrt{n}\left\{\widehat{\Delta}_{\mathcal{L}}(p) - \Delta_0(p)\right\} = \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} q(X_i,\widehat{\theta},p) - \Delta_0(p)\right\}$$
$$=\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n} q(X_i,\widehat{\theta},p) - E\left\{q(X,\widehat{\theta},p)\right\}\right] + \sqrt{n}\left[E\left\{q(X,\widehat{\theta},p)\right\} - \Delta_0(p)\right]$$
$$:= I_1(p) + I_2(p).$$

For $I_2(p)$, we use a Taylor's expansion, and then $I_2(p) = \sqrt{n}\big[E\{q(X,\widehat{\theta},p)\} - \Delta_0(p)\big] =$

$\nabla_\theta E\big\{q(X,\theta_0,p)\big\}^\top \sqrt{n}(\widehat{\theta}-\theta_0)+o_p(1)$. As a maximum likelihood estimator, $\sqrt{n}(\widehat{\theta}-\theta_0) \xrightarrow{d} N(0, I^{-1}(\theta_0))$ and does not rely on p. Since $\widehat{\theta}$ does not depend on p, it is easy to show that $\nabla_\theta E\big\{q(X,\theta_0,p)\big\}^\top \sqrt{m}(\widehat{\theta}-\theta_0)$ satisfies the stochastic equicontinuity. By the Donsker's theorem, $\big\{I_2(p) : p \in (0,1)\big\}$ converges weakly to a Gaussian process $\sqrt{r}W_3$ with zero mean and covariance function $\nabla_\theta E\big\{q(X,\theta_0,p_1)\big\}^\top I^{-1}(\theta_0)\nabla_\theta E\big\{q(X,\theta_0,p_2)\big\}$.

Since we've assumed that $\big\{G(1-p,\theta) : p \in (0,1), \theta \in \Theta\big\}$ is a VC class, $\big\{q(\cdot,\theta,p) : p \in (0,1), \theta \in \Theta\big\}$ is also a VC class and then Donsker(This conclusion comes from problem 9 on page 151 of Van Der Vaart et al. (1996)). By the asymptotic equicontinuity,

$$\frac{1}{n}\sum_{i=1}^n q(X_i,\widehat{\theta},p) - E\left\{q(X,\widehat{\theta},p)\right\} - \frac{1}{n}\sum_{i=1}^n q(X_i,\theta_0,p) + \Delta_0(p) = o_p(n^{-1/2})$$

holds uniformly for $p \in (0,1)$. Then $I_1(p)$ can be written as below

$$\begin{aligned}
I_1(p) =& \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n q(X_i,\widehat{\theta},p) - E\left\{q(X,\widehat{\theta},p)\right\}\right] \\
=& \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n q(X_i,\widehat{\theta},p) - E\left\{q(X,\widehat{\theta},p)\right\} - \frac{1}{n}\sum_{i=1}^n q(X_i,\theta_0,p) + \Delta_0(p)\right] \\
&+ \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n q(X_i,\theta_0,p) - \Delta_0(p)\right] \\
=& \sqrt{n}\left[\frac{1}{n}\sum_{i=1}^n q(X_i,\theta_0,p) - \Delta_0(p)\right] + o_p(1).
\end{aligned}$$

Since $\big\{q(\cdot,\theta_0,p) : p \in (0,1)\big\}$ is a measurable VC class with finite second moment, the Donsker's theorem guarantees that $\big\{I_1(p) : p \in (0,1)\big\}$ converges in distribution to a Gaussian process W_1 with zero mean and covariance given by $\mathrm{Cov}\{q(X,\theta_0,p_1), q(X,\theta_0,p_2)\} = \Delta_0(p_1 \wedge p_2) - \Delta_0(p_1)\Delta_0(p_2)$.

As a result, the sequence $\big\{\sqrt{n}\big\{\widehat{\Delta}_{\mathcal{L}}(p) - \Delta_0(p)\big\} : p \in (0,1)\big\}$ converges in distribution to a Gaussian process $W_{\mathcal{L}} := W_1 + \sqrt{r}W_3$, and W_1 is independent of W_3. ∎

Lemma F.2. *Under Assumptions A3 and B4,*

$$\frac{1}{M}\sum_{j=1}^{M}(1-\widehat{\zeta}_j(p))w_n^{-1}\left\{f(S_j)-w_n\right\}Q(S_j,\theta_0,p)=o_p(n^{-1/2}), \tag{A.7}$$

$$\frac{1}{n}\sum_{i=1}^{n}(1-\widehat{\zeta}_i(p))^2w_n^{-2}\left\{f(S_i)-w_n\right\}^2Q^2(S_i,\theta_0,p)=o_p(n^{-1/2}) \tag{A.8}$$

uniformly on $p\in(0,1)$.

Proof. We only prove (A.7), and then (A.8) can be proved by similar argument. Recall that $\widehat{\zeta}_j(p)-\zeta_j(p)'=J_j^0(p)-J_j^1(p)$ and $J_j^1(p)=0$ when n is sufficiently large for any p. Then $1-\widehat{\zeta}_j(p)=1-\zeta_j'(p)+\zeta_j'(p)-\widehat{\zeta}_j(p)=1-\zeta_j'(p)-J_j^0(p)$ when n is sufficiently large. Note that $M^{-1}\sum_{j=1}^{M}(1-\widehat{\zeta}_j(p))w_n^{-1}\big\{f(S_j)-w_n\big\}Q(S_j,\theta_0,p)=A_1(p)-A_2(p)$, where $A_1(p)=M^{-1}\sum_{j=1}^{M}(1-\zeta_j'(p))w_n^{-1}\left\{f(S_j)-w_n\right\}Q(S_j,\theta_0,p)$ and $A_2(p)=M^{-1}\sum_{j=1}^{M}J_j^0(p)w_n^{-1}\big\{f(S_j)-w_n\big\}Q(S_j,\theta_0,p)$.

Denote $\mu_n(p)=E\big[(1-\zeta_j'(p))w_n^{-1}\big\{f(S_j)-w_n\big\}Q(S_j,\theta_0,p)\big]$, $\sigma_n^2(p)=E\big[(1-\zeta_j'(p))w_n^{-1}\big\{f(S_j)-w_n\big\}Q(S_j,\theta_0,p)\big]^2$. Note that $\big|\big\{f(S_j)-w_n\big\}w_n^{-1}\big|\le 1$ and $\big\{f(S_j)-w_n\big\}^2w_n^{-2}\le 4$ under the event $\big\{f(S_j)\le w_n+\epsilon_n\big\}$ uniformly on $p\in(0,1)$ and the fact that $Q(S_j,\theta_0,p)=E\big\{q(X_j,\theta_0,p)|S_j\big\}\le 1$. Then we have $\mu_n(p)\le E\big\{(1-\zeta_j'(p))\big\}$, $\sigma_n^2(p)\le 4E\big\{(1-\zeta_j'(p))\big\}$. Assumption B4 gives that $\sup_{p\in(0,1)}E\big\{\sqrt{n}(1-\zeta_j'(p))\big\}=o_p(1)$. Consequently, $\mu_n(p)=o_p(n^{-1/2})$ and $\sigma_n^2(p)=o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Thus, $A_1(p)=\mu_n(p)+O_p(\sigma_n(p)n^{-1/2})=o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Similarly, we can prove that $A_2(p)=o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. The proof has been completed. ∎

Lemma F.3. *Under* **Assumptions A − C**, $\sup_{p\in(0,1)}\tilde{\boldsymbol{\beta}}(p)-\boldsymbol{\beta}_0(p)=O_p(n^{-1/2})$ as $n\to\infty$. Further,

$$\tilde{\boldsymbol{\beta}}(p)-\boldsymbol{\beta}_0(p)=\mathbf{A}_0^{-1}(p)\frac{1}{n}\sum_{i=1}^{n}\alpha(X_i,\mathbf{Z}_i,\boldsymbol{\beta}_0,\theta_0,p)+o_p(n^{-1/2})$$

uniformly on $p\in(0,1)$, where $\alpha(X_i,\mathbf{Z}_i,\boldsymbol{\beta}_0,\theta_0,p)=\big\{q(X_i,\theta_0,p)-Q(S_i,\theta_0,p)\big\}\dot{Q}(S_i,\theta_0,p)\mathbf{Z}_i$ and $\mathbf{A}_0(p)=E\big(\mathbf{Z}_i\mathbf{Z}_i^\top\big[\dot{Q}^2(S_i,\theta_0,p)-\big\{q(X_i,\theta_0,p)-Q(S_i,\theta_0,p)\big\}\ddot{Q}(S_i,\theta_0,p)\big]\big)$.

Proof. Since Assumption A4 holds, $\widehat{\theta}$ is consistent. Let

$$
\begin{aligned}
\tilde{J}_n(\boldsymbol{\beta}, \theta, p) =& \frac{1}{n} \sum_{i=1}^{n} \left\{ q(X_i, \theta, p) - \tilde{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta, p) \right\}^2, \\
J_n(\boldsymbol{\beta}, \theta, p) =& \frac{1}{n} \sum_{i=1}^{n} \left\{ q(X_i, \theta, p) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta, p) \right\}^2, \\
J(\boldsymbol{\beta}, \theta, p) =& E \left\{ q(X_i, \theta, p) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \theta, p) \right\}^2.
\end{aligned}
$$

Denote $\tilde{\boldsymbol{\beta}}(p) = \arg\min_{\|\boldsymbol{\beta}\|_2 = 1} \tilde{J}_n(\boldsymbol{\beta}, \widehat{\theta}, p)$. For the sake of simplicity in notation, we will omit the parameter p in functions J, J_n and $\tilde{J}_n$ in the following text.

First, show that $\tilde{\boldsymbol{\beta}}(p)$ is consistent uniformly on $p \in (0,1)$. Let $R(\delta, p) = \inf_{|\boldsymbol{\beta}(p) - \boldsymbol{\beta}_0(p)| > \delta} J(\boldsymbol{\beta}(p), \theta_0) - J(\boldsymbol{\beta}_0(p), \theta_0)$. It is obvious that $\inf_{p \in (0,1)} R(\delta, p) > 0$. Write $\mathcal{J}_n(p) = J(\tilde{\boldsymbol{\beta}}(p), \theta_0) - J(\boldsymbol{\beta}_0(p), \theta_0) + \tilde{J}_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) - \tilde{J}_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta})$. For any fixed $p \in (0,1)$, the event $\{|\tilde{\boldsymbol{\beta}}(p) - \boldsymbol{\beta}_0(p)| > \delta\}$ implies that $J(\tilde{\boldsymbol{\beta}}(p), \theta_0) > \inf_{|\boldsymbol{\beta}(p) - \boldsymbol{\beta}_0(p)| > \delta} J(\boldsymbol{\beta}(p), \theta_0)$, and then $J(\tilde{\boldsymbol{\beta}}(p), \theta_0) - J(\boldsymbol{\beta}_0(p), \theta_0) > R(\delta, p)$. Since $\tilde{J}_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) \geq \tilde{J}_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta})$, we have $\mathcal{J}_n(p) > R(\delta, p)$. Then

$$
P\left(\sup_{p \in (0,1)} \left| \tilde{\boldsymbol{\beta}}(p) - \boldsymbol{\beta}_0(p) \right| > \delta \right) \leq P\left(\sup_{p \in (0,1)} \mathcal{J}_n(p) - R(\delta, p) > 0 \right) = P\left(\sup_{p \in (0,1)} \mathcal{J}_n(p) > \inf_{p \in (0,1)} R(\delta, p) \right).
$$

If $\sup_{p \in (0,1)} \mathcal{J}_n(p) = o_p(1)$, $\tilde{\boldsymbol{\beta}}(p) - \boldsymbol{\beta}_0(p) = o_p(1)$ uniformly on $p \in (0,1)$.

Next we show that $\sup_{p \in (0,1)} \mathcal{J}_n(p) = o_p(1)$.

$$
\begin{aligned}
\sup_{p \in (0,1)} \mathcal{J}_n(p) =& \sup_{p \in (0,1)} J(\tilde{\boldsymbol{\beta}}(p), \theta_0) - J(\boldsymbol{\beta}_0(p), \theta_0) + \tilde{J}_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) - \tilde{J}_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta}) \\
\leq& \sup_{p \in (0,1)} \left| J(\tilde{\boldsymbol{\beta}}(p), \theta_0) - J(\boldsymbol{\beta}_0(p), \theta_0) + J_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) - J_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta}) \right| \\
& + \sup_{p \in (0,1)} \left| \tilde{J}_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) - J_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) \right| + \sup_{p \in (0,1)} \left| \tilde{J}_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta}) - J_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta}) \right| \\
:=& I_1 + I_2 + I_3.
\end{aligned}
$$

Use the triangle inequality again and then

$$
\begin{aligned}
I_1 =& \sup_{p\in(0,1)} \left| J(\tilde{\boldsymbol{\beta}}(p),\theta_0) - J(\boldsymbol{\beta}_0(p),\theta_0) + J_n(\boldsymbol{\beta}_0(p),\widehat{\theta}) - J_n(\tilde{\boldsymbol{\beta}}(p),\widehat{\theta}) \right| \\
\leq& \sup_{p\in(0,1)} \left| J_n(\tilde{\boldsymbol{\beta}}(p),\widehat{\theta}) - J(\tilde{\boldsymbol{\beta}}(p),\theta_0) \right| + \sup_{p\in(0,1)} \left| J_n(\boldsymbol{\beta}_0(p),\widehat{\theta}) - J(\boldsymbol{\beta}_0(p),\theta_0) \right| \\
:=& I_{11} + I_{12}.
\end{aligned}
$$

Decompose I_{11} into three parts

$$
\begin{aligned}
I_{11} \leq& \sup_{p\in(0,1)} \left| J_n(\tilde{\boldsymbol{\beta}}(p),\widehat{\theta}) - J(\tilde{\boldsymbol{\beta}}(p),\widehat{\theta}) - J_n(\tilde{\boldsymbol{\beta}}(p),\theta_0) + J(\tilde{\boldsymbol{\beta}}(p),\theta_0) \right| \\
&+ \sup_{p\in(0,1)} \left| J(\tilde{\boldsymbol{\beta}}(p),\widehat{\theta}) - J(\tilde{\boldsymbol{\beta}}(p),\theta_0) \right| + \sup_{p\in(0,1)} \left| J_n(\tilde{\boldsymbol{\beta}}(p),\theta_0) - J(\tilde{\boldsymbol{\beta}}(p),\theta_0) \right| \\
:=& I_{111} + I_{112} + I_{113}.
\end{aligned}
$$

Since $\left\{q(\cdot,\theta,p) : \theta \in \Theta, p \in (0,1)\right\}$ is a VC class, $\left\{Q(\cdot^{\top}\boldsymbol{\beta},\theta,p) : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta, p \in (0,1)\right\}$ is a Donsker class (Grunewalder, 2018). $\left\{q(\cdot,\theta,p) - Q(\cdot^{\top}\boldsymbol{\beta},\theta,p) : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta, p \in (0,1)\right\}$ is a Donsker. The boundedness of $q(X,\theta,p)$ guarantees that $\left\{\{q(\cdot,\theta,p) - Q(\cdot^{\top}\boldsymbol{\beta},\theta,p)\}^2 : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta, p \in (0,1)\right\}$ is a Donsker. (Theorem 2.10.6 in Van Der Vaart et al. (1996)). Thus, $I_{111} = o_p(1)$ by the equicontinuity and the property of GC class gives that $I_{113} = o_p(1)$. The consistency of $\widehat{\theta}$ guarantees $I_{112} = o_p(1)$. Consequently, $I_{11} = o_p(1)$. Similarly, we have $I_{12} = o_p(1)$. As a result, $I_1 = o_p(1)$.

Notice that

$$
\begin{aligned}
I_2 =& \sup_{p\in(0,1)} \left| \tilde{J}_n(\boldsymbol{\beta}_0(p),\widehat{\theta}) - J_n(\boldsymbol{\beta}_0(p),\widehat{\theta}) \right| \\
\leq& \sup_{p\in(0,1)} \sup_{\boldsymbol{\beta}} \left| \tilde{J}_n(\boldsymbol{\beta},\widehat{\theta},p) - J_n(\boldsymbol{\beta},\widehat{\theta},p) \right| := \sup_{p\in(0,1)} \sup_{\boldsymbol{\beta}} \left| J_1(\boldsymbol{\beta},p) - 2J_2(\boldsymbol{\beta},p) \right|,
\end{aligned}
$$

where $J_1(\boldsymbol{\beta},p) = n^{-1}\sum_{i=1}^n \left\{Q(\mathbf{Z}_i^{\top}\boldsymbol{\beta},\widehat{\theta},p) - \tilde{Q}(\mathbf{Z}_i^{\top}\boldsymbol{\beta},\widehat{\theta},p)\right\}^2$ and $J_2(\boldsymbol{\beta},p) = n^{-1}\sum_{i=1}^n \left\{q(X_i,\widehat{\theta},p) -\right.$

$Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p)\}\big\{\tilde{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p)\big\}$. Notice that $\tilde{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p) = \mathbb{I}\{\widehat{f}_n(\mathbf{Z}_i^\top \boldsymbol{\beta}) < w_n\}\big\{w_n^{-1}\widehat{f}_n(\mathbf{Z}_i^\top \boldsymbol{\beta}) - 1\big\}\widehat{Q}(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p)$. With the fact that $\big|q(X_i, \widehat{\theta}, p) - Q(\mathbf{Z}_i^\top \boldsymbol{\beta}, \widehat{\theta}, p)\big| \leq 2$, we can attain that $\sup_{p\in(0,1)} J_1(\boldsymbol{\beta}, p) = o_p(n^{-1/2})$ and $\sup_{p\in(0,1)} J_2(\boldsymbol{\beta}, p) = o_p(n^{-1/2})$ by following the proof line in Lemma F.2. Thus,

$$I_2 = \sup_{p\in(0,1)} \Big|\tilde{J}_n(\boldsymbol{\beta}_0(p), \widehat{\theta}) - J_n(\boldsymbol{\beta}_0(p), \widehat{\theta})\Big| = o_p(n^{-1/2}). \tag{A.9}$$

Similarly, we can show that

$$I_3 = \sup_{p\in(0,1)} \Big|\tilde{J}_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta}) - J_n(\tilde{\boldsymbol{\beta}}(p), \widehat{\theta})\Big| = o_p(n^{-1/2}). \tag{A.10}$$

Consequently, $\mathcal{J}_n = o_p(1)$. The proof of consistency is completed.

Second, we establish that $\sup_{p\in(0,1)} \big|\tilde{\boldsymbol{\beta}}(p) - \boldsymbol{\beta}_0(p)\big| = O_p(n^{-1/2})$. Notice that we've proven $\sup_{p\in(0,1)} \Big|\tilde{J}_n(\boldsymbol{\beta}(p), \widehat{\theta}) - J_n(\boldsymbol{\beta}(p), \widehat{\theta})\Big| = o_p(n^{-1/2})$ and $\sup_{p\in(0,1)} \Big|\tilde{J}_n(\boldsymbol{\beta}(p), \theta_0) - J_n(\boldsymbol{\beta}(p), \theta_0)\Big| = o_p(n^{-1/2})$ in the above. Since $\{\{q(\cdot, \theta, p) - Q(\cdot^\top \boldsymbol{\beta}, \theta, p)\}^2 : \boldsymbol{\beta} \in \mathcal{B}, \theta \in \Theta, p \in (0,1)\}$ is a Donsker, the empirical process $G_n(\boldsymbol{\beta}, \theta, p) = \sqrt{n}\{J_n(\boldsymbol{\beta}, \theta) - EJ_n(\boldsymbol{\beta}, \theta)\}$ satisfies equicontinuity, i.e. $\sup_{p\in(0,1)} |G_n(\boldsymbol{\beta}, \widehat{\theta}, p) - G_n(\boldsymbol{\beta}, \theta_0, p)| = o_p(1)$. Since $E\big\{J_n(\boldsymbol{\beta}, \theta)\big\}$ is continuous in θ, uniformly on $\boldsymbol{\beta}$ and $p \in (0,1)$,

$$\begin{aligned}
&\tilde{J}_n(\boldsymbol{\beta}, \widehat{\theta}) - \tilde{J}_n(\boldsymbol{\beta}, \theta_0) = J_n(\boldsymbol{\beta}, \widehat{\theta}) - J_n(\boldsymbol{\beta}, \theta_0) + o_p(n^{-1/2}) \\
=& n^{-1/2}\{G_n(\boldsymbol{\beta}, \theta, p) - G_n(\boldsymbol{\beta}, \theta_0, p)\} + EJ_n(\boldsymbol{\beta}, \widehat{\theta}) - EJ_n(\boldsymbol{\beta}, \theta_0) + o_p(n^{-1/2}) \\
=& \nabla_\theta E\{J_n(\boldsymbol{\beta}, \theta_0)\}(\widehat{\theta} - \theta_0) + o_p(n^{-1/2}).
\end{aligned}$$

Thus,

$$\tilde{J}_n(\boldsymbol{\beta}, \widehat{\theta}) = \tilde{J}_n(\boldsymbol{\beta}, \theta_0) + \nabla_\theta E\{J_n(\boldsymbol{\beta}, \theta_0)\}(\widehat{\theta} - \theta_0) + o_p(n^{-1/2}) \tag{A.11}$$

holds uniformly on $\boldsymbol{\beta}$ and $p \in (0,1)$.

By the definition of $\tilde{\boldsymbol{\beta}}$,(A.9), (A.10) and (A.11), the following estimating equation holds uniformly on $p \in (0,1)$:

$$
\begin{aligned}
0 =& \nabla_{\boldsymbol\beta} \tilde J_n(\tilde{\boldsymbol\beta}, \widehat\theta) = \nabla_{\boldsymbol\beta} J_n(\tilde{\boldsymbol\beta}, \theta_0) + \nabla_{\boldsymbol\beta}\nabla_\theta E\big\{J_n(\tilde{\boldsymbol\beta}, \theta_0)\big\}(\widehat\theta - \theta_0) + o_p(n^{-1/2}) \\
=& \nabla_{\boldsymbol\beta} J_n(\boldsymbol\beta_0, \theta_0) + \nabla_{\boldsymbol\beta\boldsymbol\beta^\top} J_n(\boldsymbol\beta_0, \theta_0)(\tilde{\boldsymbol\beta} - \boldsymbol\beta_0) + o_p(n^{-1/2}),
\end{aligned}
$$

which gives that $\tilde{\boldsymbol\beta}(p) - \boldsymbol\beta_0(p) = -\big\{\nabla_{\boldsymbol\beta\boldsymbol\beta^\top} J_n(\boldsymbol\beta_0, \theta_0)\big\}^{-1} \nabla_{\boldsymbol\beta} J_n(\boldsymbol\beta_0, \theta_0) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. Conduct some simple calculation and then

$$
\begin{aligned}
\nabla_{\boldsymbol\beta} J_n(\boldsymbol\beta_0, \theta_0) =& -\frac{2}{n}\sum_{i=1}^n \big\{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\big\}\dot Q(S_i, \theta_0, p)\mathbf{Z}_i, \\
\nabla_{\boldsymbol\beta\boldsymbol\beta^\top} J_n(\boldsymbol\beta_0, \theta_0) =& \frac{2}{n}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i^\top \Big[\dot Q^2(S_i, \theta_0, p) - \{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\}\, \ddot Q(S_i, \theta_0, p)\Big] \\
=& \frac{1}{n}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i^\top \nabla_{ss}\{q(X_i, \theta_0, p) - Q(s, \theta_0, p)\}^2 |_{s=S_i}
\end{aligned}
$$

Under Assumption B2 (ii) and the boundedness of $\mathbf{Z}$, we have that$\{\mathbf{Z}_i\mathbf{Z}_i^\top\big[\dot Q^2(S_i, \theta_0, p) - \{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\}\ddot Q(S_i, \theta_0, p)\big] : p \in (0,1)\}$ is a GC class, then by the Glivenko–Cantelli theorem, uniformly on $p \in (0,1)$,

$$
\begin{aligned}
\nabla_{\boldsymbol\beta\boldsymbol\beta^\top} J_n(\boldsymbol\beta_0, \theta_0) =& \frac{2}{n}\sum_{i=1}^n \mathbf{Z}_i\mathbf{Z}_i^\top \Big[\dot Q^2(S_i, \theta_0, p) - \{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\}\, \ddot Q(S_i, \theta_0, p)\Big] \\
\xrightarrow{p}& 2E\Big(\mathbf{Z}_i\mathbf{Z}_i^\top \Big[\dot Q^2(S_i, \theta_0, p) - \{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\}\, \ddot Q(S_i, \theta_0, p)\Big]\Big) := 2\mathbf{A}_0(p).
\end{aligned}
$$

Finally, $\tilde{\boldsymbol\beta} - \boldsymbol\beta_0 = \mathbf{A}_0^{-1} n^{-1}\sum_{i=1}^n \alpha(X_i, \mathbf{Z}_i, \boldsymbol\beta_0, \theta_0, p) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. The fact that $\boldsymbol\beta_0$ is the minimum point of $J(\boldsymbol\beta, \theta_0)$ guarantees that $E\big\{\alpha(X, \mathbf{Z}, \boldsymbol\beta_0, \theta_0, p)\big\} = 0$. Then we can verify that $\tilde{\boldsymbol\beta} - \boldsymbol\beta = O_p(n^{-1/2})$ uniformly on $p \in (0,1)$. ∎

Lemma F.4. *Under Assumptions A – B,* $\widehat Q(\mathbf{z}^\top\tilde{\boldsymbol\beta}, \widehat\theta, p) = \widehat Q(\mathbf{z}^\top\boldsymbol\beta_0, \theta_0, p) - \mathbb{M}(\mathbf{z}^\top\boldsymbol\beta, p)^\top(\tilde{\boldsymbol\beta} - \boldsymbol\beta_0) + q_1(\mathbf{z}^\top\boldsymbol\beta, \theta_0, p)(\widehat\theta - \theta_0) + o_p(n^{-1/2})$ *uniformly on* $p \in (0,1)$*, where* $\mathbb{M}(s,p) = \dot Q_1(s, \theta_0, p) - \mathbf{z}\dot Q(s, \theta_0, p) - Q(s, \theta_0, p)\dot Q_0(s,p) + f^{-1}(s)\dot f(s)\big\{Q_1(s, \theta_0, p) - Q(s, \theta_0, p)Q_0(s,p)\big\}$.

Proof. Denote $\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol\beta}, \widehat\theta, p) = n^{-1}\sum_{i=1}^n K_h\big\{(\mathbf{Z}_i - \mathbf{z})^\top\tilde{\boldsymbol\beta}\big\}q(X_i, \widehat\theta, p)$ and $\xi_{0n}(\mathbf{z}^\top\tilde{\boldsymbol\beta}) = n^{-1}\sum_{i=1}^n K_h\big\{(\mathbf{Z}_i -$

$\mathbf{z})^\top\tilde{\boldsymbol{\beta}}\big\}$. Then $\widehat{Q}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\theta,p)=\xi_{1n}(\tilde{\boldsymbol{\beta}},\theta,p)/\xi_{0n}(\tilde{\boldsymbol{\beta}})$. Through Taylor's expansion, we have

$$\begin{aligned}\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)=&\frac{1}{n}\sum_{i=1}^n\Big\{K_h(S_i-s)+\dot{K}_h(S_i-s)(\mathbf{Z}_i-\mathbf{z})^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\Big\}q(X_i,\widehat{\theta},p)+o_p(n^{-1/2})\\=&\xi_{11n}(\widehat{\theta},p)+\xi_{12n}(\widehat{\theta},p)^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+o_p(n^{-1/2})\end{aligned}$$

uniformly on $p\in(0,1)$, where $S_i=\mathbf{Z}_i^\top\boldsymbol{\beta}_0$, $s=\mathbf{z}^\top\boldsymbol{\beta}_0$, $\dot{K}_h(s)=\partial K_h(s)/\partial s$, $\xi_{11n}(\widehat{\theta},p)=n^{-1}\sum_{i=1}^n K_h(S_i-s)q(X_i,\widehat{\theta},p)$ and $\xi_{12n}(\widehat{\theta},p)=n^{-1}\sum_{i=1}^n\dot{K}_h(S_i-s)(\mathbf{Z}_i-\mathbf{z})q(X_i,\widehat{\theta},p)$. Denote $\xi_{11}(\theta,p)=E\big\{K_h(S-s)q(X,\theta,p)\big\}$ and $\xi_{12}(\theta,p)=E\big\{\dot{K}_h(S_i-s)(\mathbf{Z}_i-\mathbf{z})q(X_i,\theta,p)\big\}$. Then $\xi_{11n}(\widehat{\theta},p)=\big\{\xi_{11n}(\widehat{\theta},p)-\xi_{11}(\widehat{\theta},p)-\xi_{11n}(\theta_0,p)+\xi_{11}(\theta_0,p)\big\}+\big\{\xi_{11}(\widehat{\theta},p)+\xi_{11n}(\theta_0,p)-\xi_{11}(\theta_0,p)\big\}$. Note that $\{q(\cdot,\theta,p):\theta\in\Theta,p\in(0,1)\}$ and $\{K_h(\cdot^\top\boldsymbol{\beta}-\mathbf{z}^\top\boldsymbol{\beta}):\boldsymbol{\beta}\in\mathcal{B},z\in\mathcal{Z},h>0\}$ are both VC classes. By Lemma 7.19 in Sen (2018) we have that the covering number of the function class $\big\{K_h(\cdot^\top\boldsymbol{\beta}-\mathbf{z}^\top\boldsymbol{\beta})q(\cdot,\theta,p):\theta\in\Theta,z\in\mathcal{Z},\boldsymbol{\beta}\in\mathcal{B},h>0\big\}$ grows at a polynomial rate. Theorem 11.6 in Sen (2018) gives that this function class is a Donsker class. Thus, we have $\xi_{11n}(\widehat{\theta},p)-\xi_{11}(\widehat{\theta},p)-\xi_{11n}(\theta_0,p)+\xi_{11}(\theta_0,p)=o_p(n^{-1/2})$ uniformly on $p\in(0,1)$ by the equicontinuity of the empirical process. Assumptions A2 and A4 ensures that $\xi_{11}(\theta,p)$ is continuous in θ and $\widehat{\theta}-\theta_0=O_p(n^{-1/2})$. As a result, we can use Taylor's expansion directly to obtain that $\xi_{11}(\widehat{\theta},p)=\xi_{11}(\theta_0,p)+\nabla_\theta\xi_{11}(\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})$ uniformly for $p\in(0,1)$. Thus, $\xi_{11n}(\widehat{\theta},p)=\xi_{11n}(\theta_0,p)+\nabla_\theta\xi_{11}(\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})$ uniformly. Use the same skill and we can obtain that $\xi_{12n}(\widehat{\theta},p)=\xi_{12n}(\theta_0,p)+\nabla_\theta\xi_{12}(\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})$ uniformly for $p\in(0,1)$. Combine $\xi_{11n}(\widehat{\theta},p)$ and $\xi_{12n}(\widehat{\theta},p)$ and then we derive $\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)=\xi_{11n}(\theta_0,p)+\nabla_\theta\xi_{11}(\theta_0,p)(\widehat{\theta}-\theta_0)+\xi_{12n}^\top(\theta_0,p)(\tilde{\boldsymbol{\beta}}(p)-\boldsymbol{\beta}_0(p))+o_p(n^{-1/2})$ uniformly. Notice that $\xi_{11}(\theta_0,p)$ can be further decomposed. By expanding the expectation into an integral and utilizing Taylor's expansion, we can obtain $\xi_{11}(\theta_0,p)=f(s)Q(s,\theta_0,p)+O(h^2)$ uniformly on $p\in(0,1)$.

Denote $D_{11}(s,p)=Q(s,\theta_0,p)f(s)$, $D_{12}(s,p)=Q_1(s,\theta_0,p)f(s)$, $\dot{D}_{11}(s,p)=\partial D_{11}(s,p)/\partial s$ and $\dot{D}_{12}(s,p)=\partial D_{12}(s,p)/\partial s$. Due to Lemma F.1, the following equations hold uniformly on s and p, $f(s)Q(s,\theta,p) = n^{-1}\sum_{i=1}^{n} K_h(S_i - s)q(X_i,\theta,p) + O_p(c_n)$ and $f(s)Q_1(s,\theta,p) = n^{-1}\sum_{i=1}^{n} K_h(S_i - s)q(X_i,\theta,p)\mathbf{Z}_i + O_p(c_n)$. Then we can show that $\xi_{12n}(\theta_0,p) = \mathbf{z}\dot{D}_{11}(s,p) - \dot{D}_{12}(s,p) + O_p(c_n)$. Then uniformly on $p \in (0,1)$,

$$
\begin{aligned}
\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p) =& \xi_{11n}(\theta_0,p) + \nabla_\theta\xi_{11}(\theta_0,p)(\widehat{\theta}-\theta_0) + \xi_{12n}^\top(\theta_0,p)(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(n^{-1/2}) \\
=& \xi_{1n}(\mathbf{z}^\top\boldsymbol{\beta}_0,\theta_0,p) + f(s)\nabla_\theta Q(s,\theta_0,p)(\widehat{\theta}-\theta_0) \\
&+ \left\{\mathbf{z}\dot{D}_{11}(s,p) - \dot{D}_{12}(s,p)\right\}^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(n^{-1/2}).
\end{aligned}
$$

Similarly, we can derive

$$
\begin{aligned}
\xi_{0n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},p) =& \frac{1}{n}\sum_{i=1}^{n} K_h\left\{(\mathbf{Z}_i-\mathbf{z})^\top\tilde{\boldsymbol{\beta}}\right\} \\
=& \xi_{0n}(\mathbf{z}^\top\boldsymbol{\beta}_0) + \frac{1}{n}\sum_{i=1}^{n}\dot{K}_h(S_i - s)(\mathbf{Z}_i-\mathbf{z})^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(n^{-1/2}) \\
=& \xi_{0n}(\mathbf{z}^\top\boldsymbol{\beta}_0,p) - \left\{\dot{D}_{01}(s,p) - \mathbf{z}\dot{f}(s)\right\}^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + o_p(n^{-1/2})
\end{aligned}
$$

uniformly on $p \in (0,1)$, where $D_{01}(s,p) = Q_0(s,p)f(s)$ and $\dot{D}_{01}(s,p) = \partial D_{01}(s,p)/\partial s$. Substituting the expansions of $\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)$ and $\xi_{0n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},p)$ into $\widehat{Q}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)$ yields

$$
\begin{aligned}
\widehat{Q}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p) = \frac{\xi_{1n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)}{\xi_{0n}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},p)} \\
= \widehat{Q}(\mathbf{z}^\top\boldsymbol{\beta}_0,\theta_0,p) - \mathbb{M}(s)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + q_1(s,\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(n^{-1/2})
\end{aligned}
$$

uniformly on $p \in (0,1)$, where $\mathbb{M}(s,p) = \dot{Q}_1(s,\theta_0,p) - \mathbf{z}\dot{Q}(s,\theta_0,p) - Q(s,\theta_0,p)\dot{Q}_0(s,p) + \dot{f}(s)\big\{Q_1(s,\theta_0,p) - Q(s,\theta_0,p)Q_0(s,p)\big\}\big/f(s)$, $s = \mathbf{z}^\top\boldsymbol{\beta}_0$. The proof has been completed. ∎

Denote

$$
\mathcal{W}_i^p = q(X_i,\theta_0,p) - Q(S_i,\theta_0,p), \quad B^p(S_i,S_j) = Q(S_i,\theta_0,p) - Q(S_j,\theta_0,p),
$$

$$\varphi_n^p(S_j) = \frac{1}{n}\sum_{i=1}^n K_h(S_i - S_j)\mathcal{W}_i^p, \quad \mathcal{B}_n^p(S_j) = \frac{1}{n}\sum_{i=1}^n K_h(S_i - S_j)B^p(S_i, S_j).$$

Lemma F.5. *Assume that $X_1, \ldots, X_n$ are iid samples from the same distribution F and $\theta(p) = E\{h(X_1, \ldots, X_r; p)\}$ is the parameter of interest, where h is permutation symmetric in its r arguments. Denote $U_n(p)$ as the U statistic of $\theta(p)$ and $\widehat{U}_n(p)$ as the Hájek projection of $U_n(p)$. If $E\{h^2(X_1, \ldots, X_r; p)\} < \infty$ and $\zeta_1(p) = Var\{h_1(X_1; p)\} > 0$, then $\sup_{p\in(0,1)} \left|U_n(p) - \theta(p) - \widehat{U}_n(p)\right| = o_p(n^{-1/2})$, where $h_1(X_1; p) = E\big\{h(X_1, \ldots, X_r; p)|X_1\big\}$.*

Proof. The proof follows the line from Van der Vaart (2000). The Hájek projection of $U_n(p)$ is $\widehat{U}_n(p) = \sum_{i=1}^n E\{U_n(p)|X_i\} - (n-1)\theta(p)$. By some simple calculation, $E\{\widehat{U}_n(p)\} - \theta(p) = 0$ and $\mathrm{Var}(\widehat{U}_n(p)) = r^2\zeta_1(p)/n$ for all $p \in (0, 1)$. Let

$$R_n(p) = \frac{U_n(p) - E\{U_n(p)\}}{\sqrt{\mathrm{Var}\{U_n(p)\}}} - \frac{\widehat{U}_n(p) - E\{\widehat{U}_n(p)\}}{\sqrt{\mathrm{Var}\{\widehat{U}_n(p)\}}}.$$

It is easy to obtain that $E\{R_n(p)\} = 0$ uniformly in p. Since $\mathrm{Cov}\{U_n(p), \widehat{U}_n(p)\} = \mathrm{Var}\{\widehat{U}_n(p)\}$ holds all the time, by some simple calculation

$$\begin{aligned}\mathrm{Var}\left\{R_n(p)\right\} =& E\left\{R_n^2(p)\right\} = 2 - 2\frac{\mathrm{Cov}\left\{U_n(p), \widehat{U}_n(p)\right\}}{\sqrt{\mathrm{Var}\left\{U_n(p)\right\}\mathrm{Var}\left\{\widehat{U}_n(p)\right\}}} \\ =& 2 - 2\frac{\mathrm{Var}\left\{\widehat{U}_n(p)\right\}}{\sqrt{\mathrm{Var}\left\{U_n(p)\right\}\mathrm{Var}\left\{\widehat{U}_n(p)\right\}}} = 2 - 2\sqrt{\frac{\mathrm{Var}\left\{\widehat{U}_n(p)\right\}}{\mathrm{Var}\left\{U_n(p)\right\}}}.\end{aligned}$$

If we have

$$\sup_{p\in(0,1)} \frac{\mathrm{Var}\{\widehat{U}_n(p)\}}{\mathrm{Var}\{U_n(p)\}} \to 1, \tag{A.12}$$

then $\sup_{p\in(0,1)} \text{Var}\{R_n(p)\} \to 0$. By the Chebychev's inequality,

$$P\left\{\sup_{p\in(0,1)} R_n(p) > \epsilon\right\} \le \epsilon^{-2}E\left\{\sup_{p\in(0,1)} R_n(p)\right\}^2 \le \epsilon^{-2}\sup_{p\in(0,1)} E\{R_n(p)\}^2 \to 0.$$

Consequently, $\sup_{p\in(0,1)} R_n(p) = o_p(1)$. Then we can obtain that $U_n(p) - \widehat{U}_n(p) = o_p(n^{-1/2})$ uniformly on p.

Next we check that condition (A.12) holds. We've known that $\text{Var}\{\widehat{U}_n(p)\} = r^2\zeta_1(p)/n$. Let $\zeta_k(p) = \text{Cov}\{h(X_{i1},\ldots,X_{ir};p), h(X_{i'_1},\ldots,X_{i'_r};p)\}$ with the restriction that k variables are in common. Then

$$\begin{aligned}\text{Var}\left\{U_n(p)\right\} =& \frac{1}{(C_n^r)^2}\sum_{k=0}^{r} C_n^r C_r^k C_{n-r}^{r-k}\zeta_k(p) = \sum_{k=1}^{r}\frac{C_r^k C_{n-r}^{r-k}}{C_n^r}\zeta_k(p)\\ =&\sum_{k=1}^{r}\frac{r!^2}{k!(r-k)!^2}\frac{(n-r)(n-r-1)\cdots(n-2r+k+1)}{n(n-1)\cdots(n-r+1)}\zeta_k(p)\end{aligned}$$

In the above summation, the dominating term is the first term $r^2\zeta_1(p)/n = O(n^{-1})$ and the second term is $O(n^{-2})$. Subsequently,

$$\sup_{p\in(0,1)}\frac{\text{Var}\{\widehat{U}_n(p)\}}{\text{Var}\{U_n(p)\}} - 1 = \sup_{p\in(0,1)}\frac{\text{Var}\{\widehat{U}_n(p)\} - \text{Var}\{U_n(p)\}}{\text{Var}\{U_n(p)\}} = O(n^{-1}).$$

Thus, we've verified the condition. ∎

Lemma F.6. *Under Assumptions A1, A5 and B2 - B4,*

$$\frac{1}{M}\sum_{j=1}^{M}\frac{\varphi_n^p(S_j)}{f(S_j)} = \frac{1}{n}\sum_{j=1}^{n}\mathcal{W}_j^p\left\{1+O(h^2)\right\}, \tag{A.13}$$

$$\frac{1}{M}\sum_{j=1}^{M}\frac{\mathcal{B}_n^p(S_j)}{f(S_j)} = O(h^2) \tag{A.14}$$

uniformly on $p \in (0,1)$.

Proof. First, we focus on (A.13). It's easy to see that for all $p \in (0,1)$

$$\frac{1}{M}\sum_{j=1}^{M}\frac{\varphi_n^p(S_j)}{f(S_j)} = \frac{1}{nM}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{K_h(S_i - S_j)\mathcal{W}_j^p}{f(S_j)} = \frac{N}{M}A_1 + \frac{n-1}{M}A_2 + \frac{1}{M}A_3,$$

where

$$\begin{aligned} A_1 =& \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=n+1}^{M}\frac{K_h(S_i - S_j)\mathcal{W}_i^p}{f(S_j)}, \quad A_2 = \frac{1}{n(n-1)}\sum_{i,j=1, i\neq j}^{n}\frac{K_h(S_i - S_j)\mathcal{W}_i^p}{f(S_j)}, \\ A_3 =& \frac{1}{n}\sum_{i=1}^{n}\frac{K_h(0)\mathcal{W}_i^p}{f(S_i)}. \end{aligned}$$

Obviously, A_1, A_2 and A_3 are all U-statistics with mean being 0. Then by the projection method of U-statistics and Lemma F.5, we can show that $A_1 = n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p + o_p(M^{-1/2})$, $A_2 = n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p + o_p(n^{-1/2})$ and $A_3 = o_p(A_1 + A_2) = o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. Therefore, we obtain $M^{-1}\sum_{j=1}^{M}\varphi_n^p(S_j)f^{-1}(S_j) = n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p + o_p(M^{-1/2})$ uniformly on $p \in (0,1)$.

Next, consider (A.14).

$$\frac{1}{M}\sum_{j=1}^{M}\frac{\mathcal{B}_n^p(S_j)}{f(S_j)} = \frac{1}{nM}\sum_{i=1}^{n}\sum_{j=1}^{M}\frac{K_h(S_i - S_j)B^p(S_i, S_j)}{f(S_j)} = \frac{N}{M}B_1 + \frac{n-1}{M}B_2,$$

where

$$B_1 = \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=n+1}^{M}\frac{K_h(S_i - S_j)B^p(S_i, S_j)}{f(S_j)}, \quad B_2 = \frac{1}{n(n-1)}\sum_{i,j=1, i\neq j}^{n}\frac{K_h(S_i - S_j)B^p(S_i, S_j)}{f(S_j)}.$$

Analogously, the projection method and Lemma F.5 give that $B_1 = O(h^2) + o_p(M^{-1/2}) = o_p(n^{-1/2})$ and $B_2 = O(h^2) + o_p(M^{-1/2}) = o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. As a result, $M^{-1}\sum_{j=1}^{M}\mathcal{B}_n(S_j)/f(S_j) = O(h^2)$ uniformly on $p \in (0,1)$. The proof has been completed. ∎

Lemma F.7. *Under Assumptions A1, A5 and B2 - B5,*

$$\frac{1}{M}\sum_{j=1}^{M}(1-\zeta_j')\frac{\varphi_n^p(S_j)}{f(S_j)} = o_p(n^{-1/2})$$

uniformly on $p \in (0,1)$

Proof. In the following, we sketch the proof. Note that

$$\frac{1}{nM}\sum_{i=1}^{n}\sum_{j=1}^{M}K_h(S_i-S_j)\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)} = \frac{n-1}{M}D_1 + \frac{N}{M}D_2 + \frac{1}{M}D_3,$$

where

$$D_1 = \frac{1}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}K_h(S_i-S_j)\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)}, \quad D_2 = \frac{1}{nN}\sum_{i=1}^{n}\sum_{j=n+1}^{M}K_h(S_i-S_j)\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)},$$

$$D_3 = \frac{1}{n}\sum_{i=1}^{n}K_h(0)\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_i)}.$$

For D_1, denote

$$U_n = \frac{2}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}K_h(S_i-S_j)\left\{\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)} + \frac{(1-\zeta_i')\mathcal{W}_j^p}{f(S_i)}\right\}.$$

U_n is a U statistic. Write

$$H_{ij}(p) = K_h(S_i-S_j)\left\{\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)} + \frac{(1-\zeta_i')\mathcal{W}_j^p}{f(S_i)}\right\}.$$

By the law of iterated expectation, for any $p \in (0,1)$, we have that

$$E\left\{K_h(S_i-S_j)\frac{(1-\zeta_j')\mathcal{W}_i^p}{f(S_j)}\right\} = E\left\{K_h(S_i-S_j)\frac{1-\zeta_j'}{f(S_j)}E(\mathcal{W}_i^p|S_i,S_j)\right\} = 0.$$

Then $E\{H_{ij}(p)\} = 0$. Write

$$H(p, \mathbf{Z}_j, X_j) = E(H_{ij}(p)|\mathbf{Z}_j, X_j) = \mathcal{W}_j^p E\left\{K_h(S_i - S_j)\frac{1-\zeta_i'}{f(S_i)}\middle|\mathbf{Z}_j, X_j\right\}.$$

And it is obvious that $E\{H(p, \mathbf{Z}_j, X_j)\} = 0$.

Next we calculate the moments of $H(p, \mathbf{Z}_j, X_j)$ and $H_{ij}(p)$. Write $\sigma_q^2(S_i, p) = E\left\{(\mathcal{W}_i^p)^2|S_i\right\}$.

$$\begin{aligned}
\xi_1(p) =&\text{Var}\{H(p, \mathbf{Z}_j, X_j)\} = E\left\{(\mathcal{W}_j^p)^2\left[\int_{-c}^{c} K(t)\mathbb{I}\{f(S_j + ht) \le w_n + \epsilon_n\}dt\right]^2\right\} \\
\le& E\left[(\mathcal{W}_j^p)^2\int_{-c}^{c}\mathbb{I}\{f(S_j + ht) \le w_n + \epsilon_n\}dt\right]\int_{-c}^{c} K^2(t)dt \\
=& \int_{-c}^{c} E\left[\sigma_q^2(S_j, p)\mathbb{I}\{f(S_j + ht) \le w_n + \epsilon_n\}\right]dt\int_{-c}^{c} K^2(t)dt
\end{aligned}$$

where the first inequality is due to the Cauchy-Schwarz inequality. Since we've assumed $P(f(S_j) \le w_n + \epsilon_n) \to 0$ for any $\boldsymbol{\beta}$, then $\xi_1(p) \to 0$ uniformly on $p \in (0, 1)$ as $n \to \infty$.

Similarly, under the Assumption B5 that $\sup_{p\in(0,1)} E\big\{\sigma_q^2(S_i, p)f^{-1}(S_i)\big\} < \infty$,

$$\begin{aligned}
\xi_2(p) =& E\left\{K_h^2(S_i - S_j)\frac{(1-\zeta_j')\mathcal{W}_i^2}{f^2(S_j)}\right\} = \frac{1}{h}\int_{-c}^{c} K^2(t)E\left[\mathcal{W}_i^2\frac{\mathbb{I}\{f(S_i + ht) \le w_n + \epsilon_n\}}{f(S_i + ht)}\right]dt \\
\le& \frac{1}{h}\int_{-c}^{c} K^2(t)E\left\{\frac{\mathcal{W}_i^2}{f(S_i + ht)}\right\}dt = \frac{1}{h}E\left\{\frac{\sigma_q^2(S_i)}{f(S_i)}\right\}\int_{-c}^{c} K^2(t)dt\{1 + O(h)\} \\
\le& \frac{1}{h}E\left\{\frac{\sigma_q^2(S_i)}{f(S_i)}\right\}\int_{-c}^{c} K^2(t)dt\{1 + O(h)\} = O(h^{-1}),
\end{aligned}$$

uniformly on $p \in (0, 1)$. The projection of U_n is $\widehat{U}_n = 2\sum_{j=1}^n H_1(\mathbf{Z}_j, X_j)\big/n$. By Lemma F.5, $U_n(p) = \widehat{U}_n(p) + o_p(n^{-1/2})$ uniformly on $p \in (0, 1)$. Since $E(\widehat{U}_n) = 0$ and $\xi_1(p) \to 0$ uniformly on $p \in (0, 1)$, $\sup_{p\in(0,1)} \widehat{U}_n(p) = o_p(n^{-1/2})$. Therefore, $\sup_{p\in(0,1)} U_n(p) = o_p(n^{-1/2})$ and $\sup_{p\in(0,1)} D_1 = o_p(n^{-1/2})$. Analogously, we can show that $D_2 = o_p(n^{-1/2})$ uniformly on $p \in (0, 1)$. And $D_3 = o_p(D_1) = o_p(n^{-1/2})$ uniformly on $p \in (0, 1)$. Consequently, $B_2 = o_p(n^{-1/2})$ uniformly on $p \in (0, 1)$. ∎

Proof of Lemma 3.1. In Lemma F.3, we've proven $\sup_{p\in(0,1)} \tilde{\boldsymbol{\beta}}(p)-\boldsymbol{\beta}_0(p)=O_p(n^{-1/2})$. Here we only show how to decompose $\tilde{\Delta}_{\lambda,\mathcal{S}}(\boldsymbol{\beta},\theta,p)$. By the definition,

$$\tilde{\Delta}_{\lambda,\mathcal{S}}(p) =\frac{\lambda}{n}\sum_{i=1}^{n}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p)+\frac{1-\lambda}{N}\sum_{i=n+1}^{M}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p),$$

which is the weighted sum of two means. Thus, we focus on the decomposition of first part $n^{-1}\sum_{i=1}^{n}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p)$ and follow the same technique to tackle the second part. Denote $\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p)=n^{-1}\sum_{j=1}^{n}\{\tilde{Q}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)-Q(\mathbf{Z}_j^\top\boldsymbol{\beta}_0,\theta_0,p)\}$. Then $n^{-1}\sum_{i=1}^{n}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p)=\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p)+n^{-1}\sum_{i=1}^{n}Q(\mathbf{Z}_i^\top\boldsymbol{\beta}_0;\theta_0,p)$.

First, we show that

$$\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p) =A_1(p)+A_2(p)-\frac{1}{n}\sum_{j=1}^{n}\mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{n}\sum_{j=1}^{n}q_1(S_j,\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2}) \tag{A.15}$$

uniformly on $p\in(0,1)$, where $S_j=\mathbf{Z}_j^\top\boldsymbol{\beta}_0$, $\mathbb{M}(s)$ is defined in Lemma F.4, $A_1(p)=n^{-1}\sum_{j=1}^{n}\widehat{\zeta}_j\{\tilde{Q}(S_j,\theta_0,p)-Q(S_j,\theta_0,p)\}$ and $A_2(p)=n^{-1}\sum_{j=1}^{n}(1-\widehat{\zeta}_j)\{(nw_n)^{-1}\sum_{i=1}^{n}K_h(S_i-S_j)q(X_i,\theta_0,p)-Q(S_j,\theta_0,p)\}$. By the definition of $\tilde{f}_n$,

$$\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p) =\frac{1}{n}\sum_{j=1}^{n}\Big\{\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\widehat{Q}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)-Q(\mathbf{Z}_j^\top\boldsymbol{\beta}_0,\theta_0,p)\Big\}=\mathcal{L}_1(p)+\mathcal{L}_2(p), \tag{A.16}$$

where

$$\mathcal{L}_1(p)=\frac{1}{n}\sum_{j=1}^{n}\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\{\widehat{Q}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)-Q(\mathbf{Z}_j^\top\boldsymbol{\beta}_0,\theta_0,p)\},$$

$$\mathcal{L}_2(p)=\frac{1}{n}\sum_{j=1}^{n}\{\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})-1\}Q(\mathbf{Z}_j^\top\boldsymbol{\beta}_0,\theta_0,p).$$

Due to Lemma F.4, we have $\widehat{Q}(\mathbf{z}^\top\tilde{\boldsymbol{\beta}},\widehat{\theta},p)=\widehat{Q}(s,\theta_0,p)-\mathbb{M}(s)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+q_1(s,\theta_0,p)(\widehat{\theta}-$

$\theta_0) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$, where $s = \mathbf{z}^\top \boldsymbol{\beta}_0$. Substituting $\widehat{Q}(\cdot)$ into $\mathcal{L}_1(p)$ gives

$$\begin{aligned}\mathcal{L}_1(p) =& \frac{1}{n}\sum_{j=1}^n \widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) \left\{\widehat{Q}(S_j,\theta_0,p) - Q(S_j,\theta_0,p)\right\} \\ &- \frac{1}{n}\sum_{j=1}^n \mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \frac{1}{n}\sum_{j=1}^n q_1(S_j,\theta_0,p)(\widehat{\theta} - \theta_0) \\ &- \frac{1}{n}\sum_{j=1}^n \left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) - 1\right\}\mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ &+ \frac{1}{n}\sum_{j=1}^n \left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) - 1\right\} q_1(S_j,\theta_0,p)(\widehat{\theta} - \theta_0) + o_p(n^{-1/2})\end{aligned} \tag{A.17}$$

uniformly on $p \in (0,1)$. Next we show that the last two terms of the above equation are of order $o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. By the definition of $\tilde{f}_n$,

$$\frac{1}{n}\sum_{j=1}^n \left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) - 1\right\}\mathbb{M}(S_j) = \frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)\left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})w_n^{-1} - 1\right\}\mathbb{M}(S_j).$$

Under the event $\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) \le w_n\}$, $\sup_{p\in(0,1)} |w_n^{-1}\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) - 1| \le 1$. Jensen's inequality and Cauchy-Schwarz inequality give

$$\begin{aligned}&\sup_{p\in(0,1)}\left|\frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)\left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})w_n^{-1} - 1\right\}\mathbb{M}(S_j)\right| \\ &\le \sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)\left|\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})w_n^{-1} - 1\right| |\mathbb{M}(S_j)| \le \sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)|\mathbb{M}(S_j)| \\ &\le \sup_{p\in(0,1)}\sqrt{\frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)\frac{1}{n}\sum_{j=1}^n \mathbb{M}^2(S_j)} \le \sup_{p\in(0,1)}\sqrt{\frac{1}{n}\sum_{j=1}^n (1-\widehat{\zeta}_j)}\sup_{p\in(0,1)}\sqrt{\frac{1}{n}\sum_{j=1}^n \mathbb{M}^2(S_j)} \\ &\le \left\{\sup_{p\in(0,1)}\sqrt{P(f(S_j) \le w_n + \epsilon_n)} + o_p(1)\right\} O_p(1) \to 0 \quad \text{as } n \to \infty,\end{aligned}$$

where the last convergence is guaranteed by Assumption B5. Thus,

$$\sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^n \left\{\widehat{f}_n(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top \tilde{\boldsymbol{\beta}}) - 1\right\}\mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = o_p(n^{-1/2}). \tag{A.18}$$

Similarly, we can show that

$$\sup_{p\in(0,1)} \frac{1}{n}\sum_{j=1}^{n}\left\{\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})-1\right\}q_1(S_j,\theta_0)(\widehat{\theta}-\theta_0)=o_p(n^{-1/2}). \tag{A.19}$$

(A.17), (A.18) and (A.19) give

$$\begin{aligned}\mathcal{L}_1(p)=&\frac{1}{n}\sum_{j=1}^{n}\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\left\{\widehat{Q}(S_j,\theta_0,p)-Q(S_j,\theta_0,p)\right\}\\&-\frac{1}{n}\sum_{j=1}^{n}\mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{n}\sum_{j=1}^{n}q_1(S_j,\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})\end{aligned}$$

uniformly on $p\in(0,1)$. Using Taylor's expansion theorem, we can replace $\tilde{\boldsymbol{\beta}}$ in $\widehat{f}_n(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})$ with $\boldsymbol{\beta}_0$ while keeping the orders of $\mathcal{L}_1(p)$ and $\mathcal{L}_2(p)$ unchanged. This leads to

$$\begin{aligned}\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p)=&\frac{1}{n}\sum_{j=1}^{n}\left\{\widehat{f}_n(\mathbf{Z}_j^\top\boldsymbol{\beta}_0)\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\widehat{Q}(S_j,\theta_0,p)-Q(S_j,\theta_0,p)\right\}\\&-\frac{1}{n}\sum_{j=1}^{n}\mathbb{M}(S_j)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\frac{1}{n}\sum_{j=1}^{n}q_1(S_j,\theta_0,p)(\widehat{\theta}-\theta_0)+o_p(n^{-1/2})\end{aligned}$$

uniformly on $p\in(0,1)$. By the definition of $\tilde{f}_n(\cdot)$, it is easy to derive that

$$n^{-1}\sum_{j=1}^{n}\{\widehat{f}_n(\mathbf{Z}_j^\top\boldsymbol{\beta}_0)\tilde{f}_n^{-1}(\mathbf{Z}_j^\top\tilde{\boldsymbol{\beta}})\widehat{Q}(S_j,\theta_0,p)-Q(S_j,\theta_0,p)\}=A_1(p)+A_2(p).$$

As a result, we've proven equation (A.15).

Second, show that $A_1(p)=n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p+o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Recall that

$$\begin{aligned}&\mathcal{W}_i^p=q(X_i,\theta_0,p)-Q(S_i,\theta_0,p),\quad B^p(S_i,S_j)=Q(S_i,\theta_0,p)-Q(S_j,\theta_0,p),\\&\varphi_n^p(S_j)=\frac{1}{n}\sum_{i=1}^{n}K_h(S_i-S_j)\mathcal{W}_i^p,\quad \mathcal{B}_n^p(S_j)=\frac{1}{n}\sum_{i=1}^{n}K_h(S_i-S_j)B^p(S_i,S_j).\end{aligned}$$

Then $A_1(p)=A_{11}(p)+A_{12}(p)$, where $A_{11}(p)=n^{-1}\sum_{j=1}^{n}\widehat{\zeta}_j\widehat{f}_n^{-1}(S_j)\varphi_n^p(S_j)$ and $A_{12}(p)=$

$n^{-1}\sum_{j=1}^{n}\widehat{\zeta}_j\widehat{f}_n^{-1}(S_j)\mathcal{B}_n^p(S_j)$. Then $A_{11}(p)=n^{-1}\sum_{j=1}^{n}\varphi_n^p(S_j)f^{-1}(S_j)+B_1(p)-B_2(p)$, where $B_1(p)=n^{-1}\sum_{j=1}^{n}\left\{\widehat{\zeta}_j\widehat{f}_n^{-1}(S_j)f(S_j)-\zeta_j'\right\}\varphi_n^p(S_j)f^{-1}(S_j)$ and $B_2(p)=n^{-1}\sum_{j=1}^{n}\left\{1-\zeta_j'\right\}\varphi_n^p(S_j)f^{-1}(S_j)$. Follow the proof of Lemma F.6 and we can also prove that $n^{-1}\sum_{j=1}^{n}\varphi_n^p(S_j)f^{-1}(S_j)=n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p+o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Consequently, $A_{11}(p)=n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p+B_1(p)-B_2(p)+o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Furthermore, $B_1(p)=B_{11}(p)+B_{12}(p)$, where $B_{11}(p)=n^{-1}\sum_{j=1}^{n}(\widehat{\zeta}_j-\zeta_j')\widehat{f}_n^{-1}(S_j)\varphi_n^p(S_j)$ and $B_{12}(p)=n^{-1}\sum_{j=1}^{n}\zeta_j'\{\widehat{f}_n^{-1}(S_j)-f^{-1}(S_j)\}\varphi_n^p(S_j)$. Recall that $\widehat{\zeta}_j-\zeta_j'=J_j^0-J_j^1$ and $J_j^1=0$ when n is sufficiently large. Consequently, $B_{11}(p)=n^{-1}\sum_{j=1}^{n}J_j^0\widehat{f}_n^{-1}(S_j)\varphi_n^p(S_j)$ when n is large. Further, $B_{11}(p)=B_{111}(p)+B_{112}(p)$, where $B_{111}(p)=n^{-1}\sum_{j=1}^{n}J_j^0f^{-1}(S_j)\varphi_n^p(S_j)$ and $B_{112}(p)=n^{-1}\sum_{j=1}^{n}J_j^0\{\widehat{f}_n^{-1}(S_j)-f^{-1}(S_j)\}\varphi_n^p(S_j)$. Under the event J_j^0, $w_n-\epsilon_n\le f(S_j)\le w_n+\epsilon_n$ when n is large enough. Then by the uniform convergence rate of empirical process, $\sup_{p\in(0,1)}B_{111}(p)=o_p(n^{-1/2})$ and $\sup_{p\in(0,1)}B_{112}(p)=o_p(n^{-1/2})$. Thus, $\sup_{p\in(0,1)}B_{11}(p)=o_p(n^{-1/2})$. Similarly, $\sup_{p\in(0,1)}B_{12}(p)=o_p(n^{-1/2})$. In Lemma F.7, we've shown that $\sup_{p\in(0,1)}B_2(p)=o_p(n^{-1/2})$. As a result, $A_{11}(p)=n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p+o_p(n^{-1/2})$ uniformly on $p\in(0,1)$. Analogously, we can show that $\sup_{p\in(0,1)}A_{12}(p)=o_p(n^{-1/2})$. Therefore, $A_1(p)=n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p+o_p(n^{-1/2})$ uniformly on $p\in(0,1)$.

Third, show that $\sup_{p\in(0,1)}A_2(p)=o_p(n^{-1/2})$. Observe that

$$
\begin{aligned}
A_2(p)=&\frac{1}{n}\sum_{j=1}^{n}(1-\widehat{\zeta}_j)\left\{w_n^{-1}\frac{1}{n}\sum_{i=1}^{n}K_h(S_i-S_j)q(X_i,\theta_0,p)-Q(S_j,\theta_0,p)\right\}\\
=&\frac{1}{n}\sum_{j=1}^{n}(1-\widehat{\zeta}_j)w_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}K_h(S_i-S_j)q(X_i,\theta_0,p)-f(S_j)Q(S_j,\theta_0,p)\right\}\\
&+\frac{1}{n}\sum_{j=1}^{n}(1-\widehat{\zeta}_j)w_n^{-1}\left\{f(S_j)-w_n\right\}Q(S_j,\theta_0,p)=A_{21}(p)+A_{22}(p).
\end{aligned}
$$

In Lemma F.2, we've proven that $\sup_{p\in(0,1)}A_{22}(p)=o_p(n^{-1/2})$. Next we show that $\sup_{p\in(0,1)}A_{21}(p)=o_p(n^{-1/2})$. Notice that $A_{21}(p)=A_{211}(p)+A_{212}(p)$, where

$$
A_{211}(p)=\frac{1}{n}\sum_{j=1}^{n}(1-\zeta_j')w_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}K_h(S_i-S_j)q(X_i,\theta_0,p)-f(S_j)Q(S_j,\theta_0,p)\right\},
$$

$$A_{212}(p) = \frac{1}{n}\sum_{j=1}^{n}(\zeta_j' - \widehat{\zeta}_j)w_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}K_h(S_i - S_j)q(X_i,\theta_0,p) - f(S_j)Q(S_j,\theta_0,p)\right\}.$$

Apply Lemma F.1 and Assumption B5 again and we have

$$\begin{aligned}\sup_{p\in(0,1)} A_{211}(p) &\leq \sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^{n}(1-\zeta_j')w_n^{-1}\cdot \sup_{p\in(0,1)}\left|\frac{1}{n}\sum_{i=1}^{n}K_h(S_i - S_j)q(X_i,\theta_0,p) - f(S_j)Q(S_j,\theta_0,p)\right| \\ &= \sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^{n}(1-\zeta_j')w_n^{-1}\cdot O_p(c_n) = o_p(n^{-1/2}).\end{aligned}$$

Recall that $\widehat{\zeta}_j - \zeta_j' = J_j^0 - J_j^1$ and $J_j^1 = 0$ when n is sufficiently large.

$$\begin{aligned}&\sup_{p\in(0,1)} A_{212}(p) \\ &= \sup_{p\in(0,1)}\frac{1}{n}\sum_{j=1}^{n}J_j^0 w_n^{-1}\left\{\frac{1}{n}\sum_{i=1}^{n}K_h(S_i - S_j)q(X_i,\theta_0,p) - f(S_j)Q(S_j,\theta_0,p)\right\} \\ &\leq \sup_{p\in(0,1)}\sup_{s}\left|\frac{1}{n}\sum_{i=1}^{n}K_h(S_i - S_j)q(X_i,\theta_0,p) - f(S_j)Q(S_j,\theta_0,p)\right|\sup_{p\in(0,1)} w_n^{-1}\frac{1}{n}\sum_{j=1}^{n}J_j^0 \\ &= O_p(c_n w_n^{-1})O_p(n^{-1/2}) = o_p(n^{-1/2}).\end{aligned}$$

Thus, $\sup_{p\in(0,1)} A_{21}(p) = o_p(n^{-1/2})$. Consequently, $\sup_{p\in(0,1)} A_2(p) = o_p(n^{-1/2})$.

So far, we have proven $\mathcal{L}(\tilde{\boldsymbol{\beta}},\widehat{\theta},p) = n^{-1}\sum_{i=1}^{n}\mathcal{W}_i^p - C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta} - \theta_0) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$, where $C(\boldsymbol{\beta}_0,\theta_0,p) = E\{\mathbb{M}(S_1)\}$ and $A(\theta_0,p) = E\{q_1(S_1,\theta_0,p)\}$. Thus, $n^{-1}\sum_{i=1}^{n}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p) = n^{-1}\sum_{j=1}^{n}q(X_i,\theta_0,p) - C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta} - \theta_0) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. Follow the same steps and we can obtain that

$$\begin{aligned}\frac{1}{N}\sum_{i=n+1}^{M}\tilde{Q}(\mathbf{Z}_i^\top\tilde{\boldsymbol{\beta}};\widehat{\theta},p) =& \frac{1}{n}\sum_{i=1}^{n}q(X_i,\theta_0,p) - \frac{1}{n}\sum_{i=1}^{n}Q(S_i,\theta_0,p) + \frac{1}{N}\sum_{i=n+1}^{M}Q(S_i,\theta_0,p) \\ &- C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta} - \theta_0) + o_p(n^{-1/2})\end{aligned}$$

uniformly on $p \in (0,1)$. Consequently,

$$\begin{aligned}\tilde{\Delta}_{\lambda,\mathcal{S}}(p) =& \frac{1}{n}\sum_{i=1}^{n} q(X_i,\theta_0,p) - \frac{1-\lambda}{n}\sum_{i=1}^{n} Q(S_i,\theta_0,p) + \frac{1-\lambda}{N}\sum_{i=n+1}^{M} Q(S_i,\theta_0,p)\\ &- C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(n^{-1/2})\end{aligned}$$

uniformly on $p \in (0,1)$.

Notice that

$$C(\boldsymbol{\beta}_0,\theta_0,p) = f(s)\left\{Q_1(s,p) - Q_0(s,p)Q(s,p)\right\}\Big|_b^a, \tag{A.20}$$

where a and b are the boundary of the support set of $\mathbf{Z}^\top\boldsymbol{\beta}_0$. In detail,

$$\begin{aligned}
&C(\boldsymbol{\beta},\theta,p)\\
=&E\left[\dot{Q}_1(S,\theta_0,p) - \mathbf{Z}\dot{Q}(S,\theta_0,p) - Q(S,\theta_0,p)\dot{Q}_0(S,p) + \frac{\dot{f}(S)}{f(S)}\left\{Q_1(S,\theta_0,p) - Q(S,\theta_0,p)Q_0(S,p)\right\}\right]\\
=&E\left[\dot{Q}_1(S,\theta_0,p) - E(\mathbf{Z}|S)\dot{Q}(S,\theta_0,p) - Q(S,\theta_0,p)\dot{Q}_0(S,p) + \frac{\dot{f}(S)}{f(S)}\left\{Q_1(S,\theta_0,p) - Q(S,\theta_0,p)Q_0(S,p)\right\}\right]\\
=&E\left[\dot{Q}_1(S,\theta_0,p) - Q_0(S,p)\dot{Q}(S,\theta_0,p) - Q(S,\theta_0,p)\dot{Q}_0(S,p) + \frac{\dot{f}(S)}{f(S)}\left\{Q_1(S,\theta_0,p) - Q(S,\theta_0,p)Q_0(S,p)\right\}\right]\\
=&E\left[\nabla_S\left\{Q_1(S,\theta_0,p) - Q_0(S,p)Q(S,\theta_0,p)\right\} + \frac{\dot{f}(S)}{f(S)}\left\{Q_1(S,\theta_0,p) - Q(S,\theta_0,p)Q_0(S,p)\right\}\right]\\
=&\int_b^a\left[\nabla_s\left\{Q_1(s,\theta_0,p) - Q_0(s,p)Q(s,\theta_0,p)\right\} + \frac{\dot{f}(s)}{f(s)}\left\{Q_1(s,\theta_0,p) - Q(s,\theta_0,p)Q_0(s,p)\right\}\right]f(s)ds\\
=&\int_b^a\left[\nabla_s\left\{Q_1(s,\theta_0,p) - Q_0(s,p)Q(s,\theta_0,p)\right\}f(s) + \dot{f}(s)\left\{Q_1(s,\theta_0,p) - Q(s,\theta_0,p)Q_0(s,p)\right\}\right]ds\\
=&\int_b^a\nabla_s\left[f(s)\left\{Q_1(s,\theta_0,p) - Q_0(s,p)Q(s,\theta_0,p)\right\}\right]ds\\
=&f(s)\left\{Q_1(s,\theta_0,p) - Q_0(s,p)Q(s,\theta_0,p)\right\}\Big|_b^a.
\end{aligned}$$

Under Assumption C1 or C2, $C(\boldsymbol{\beta}_0,\theta_0,p) = 0$. The proof has been completed. ∎

Proof of Theorem 3.1. Lemma 3.1 gives that

$$\begin{aligned}\tilde{\Delta}_{\lambda,\mathcal{S}}(p) =& \frac{1}{n}\sum_{i=1}^{n} q(X_i,\theta_0,p) - \frac{1-\lambda}{n}\sum_{i=1}^{n} Q(S_i,\theta_0,p) + \frac{1-\lambda}{N}\sum_{i=n+1}^{M} Q(S_i,\theta_0,p)\\ &- C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(n^{-1/2})\end{aligned}$$

uniformly on $p\in(0,1)$. Then

$$\begin{aligned}&\sup_{p\in(0,1)} |\tilde{\Delta}_{\lambda,\mathcal{S}}(p) - \Delta_0(p)|\\ \leq& \sup_{p\in(0,1)}\left|\frac{1}{n}\sum_{i=1}^{n} q(X_i,\theta_0,p) - \Delta_0(p)\right| + (1-\lambda)\sup_{p\in(0,1)}\left|\frac{1}{n}\sum_{i=1}^{n} Q(S_i,\theta_0,p) - \Delta_0(p)\right|\\ &+ (1-\lambda)\sup_{p\in(0,1)}\left|\frac{1}{N}\sum_{i=n+1}^{M} Q(S_i,\theta_0,p) - \Delta_0(p)\right| + \sup_{p\in(0,1)}\left|C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\right|\\ &+ \sup_{p\in(0,1)} A(\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(1).\end{aligned}$$

The Glivenko–Cantelli theorem ensures that $\sup_{p\in(0,1)} |n^{-1}\sum_{i=1}^{n} q(X_i,\theta_0,p) - \Delta_0(p)| = o_p(1)$, $\sup_{p\in(0,1)} |n^{-1}\sum_{i=1}^{n} Q(S_i,\theta_0,p) - \Delta_0(p)| = o_p(1)$ and $\sup_{p\in(0,1)} |N^{-1}\sum_{i=n+1}^{M} Q(S_i,\theta_0,p) - \Delta_0(p)| = o_p(1)$. The fourth term $\sup_{p\in(0,1)} |C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)| = o_p(1)$, because we've shown the uniform consistency of $\tilde{\boldsymbol{\beta}}(p)$ in Lemma F.3. Since $\widehat{\theta}$ are consistent and does not rely on p, $\sup_{p\in(0,1)} A(\theta_0,p)(\widehat{\theta}-\theta_0) = o_p(1)$. Thus, $\sup_{p\in(0,1)} |\tilde{\Delta}_{\lambda,\mathcal{S}}(p) - \Delta_0(p)| = o_p(1)$. The proof has been completed. ∎

Proof of Theorem 3.2. In the following we only give the proof of $\tilde{\Delta}_{\lambda,\mathcal{S}}(p)$, because the proof of $\tilde{\Delta}_{\lambda,\mathcal{U}}(p)$ is very similar.

By Lemma 3.1, we have

$$\begin{aligned}\tilde{\Delta}_{\lambda,\mathcal{S}}(p) =& \frac{1}{n}\sum_{i=1}^{n} q(X_i,\theta_0,p) - \frac{1-\lambda}{n}\sum_{i=1}^{n} Q(S_i,\theta_0,p) + \frac{1-\lambda}{N}\sum_{i=n+1}^{M} Q(S_i,\theta_0,p)\\ &- C(\boldsymbol{\beta}_0,\theta_0,p)^\top(\tilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) + A(\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(n^{-1/2})\end{aligned}$$

uniformly on $p \in (0,1)$. Lemma F.3 gives that $\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = \mathbf{A}_0^{-1} n^{-1} \sum_{i=1}^n \alpha(X_i, \mathbf{Z}_i, \boldsymbol{\beta}_0, \theta_0, p) + o_p(n^{-1/2})$ uniformly on $p \in (0,1)$. By simple computation, we have

$$\begin{aligned}\tilde{\Delta}_{\lambda,\mathcal{S}}(p) =& \frac{1}{n}\sum_{i=1}^{n}\left\{q(X_i,\theta_0,p) - (1-\lambda)Q(S_i,\theta_0,p) - C(\boldsymbol{\beta}_0,\theta_0,p)^\top \mathbf{A}_0^{-1}\alpha(X_i,\mathbf{Z}_i,\boldsymbol{\beta}_0,\theta_0,p)\right\} \\ &+ \frac{1-\lambda}{N}\sum_{j=n+1}^{M} Q(S_j,\theta_0,p) + A(\theta_0,p)(\widehat{\theta}-\theta_0) + o_p(n^{-1/2}).\end{aligned}$$

Denote $\mathrm{IF}_i(p,\lambda) = q(X_i,\theta_0,p) - (1-\lambda)Q(S_i,\theta_0,p) - C(\boldsymbol{\beta}_0,\theta_0,p)^\top \mathbf{A}_0^{-1}\alpha(X_i,\mathbf{Z}_i,\boldsymbol{\beta}_0,\theta_0,p)$ and $E\big\{\mathrm{IF}(p,\lambda)\big\} = \lambda\Delta_0(p)$. Let $\mathbb{G}_n(p) = \sqrt{n}\big\{\tilde{\Delta}_{\lambda,\mathcal{S}}(p) - \Delta_0(p)\big\}$, $\mathbb{G}_n^{(1)}(p) = n^{-1/2}\sum_{i=1}^n \big\{\mathrm{IF}_i(p,\lambda) - \lambda\Delta_0(p)\big\}$, $\mathbb{G}_N^{(2)}(p) = N^{-1/2}\sum_{j=n+1}^M \big\{Q(S_j,\theta_0,p) - \Delta_0(p)\big\}$ and $\mathbb{G}_m^{(3)}(p) = A(\theta_0,p)\sqrt{m}(\widehat{\theta}-\theta_0)$. Then $\mathbb{G}_n(p) = \mathbb{G}_n^{(1)}(p) + (1-\lambda)\sqrt{n/N}\mathbb{G}_N^{(2)}(p) + \sqrt{n/m}\mathbb{G}_m^{(3)}(p)$. Notice that the independence of $\mathcal{L}$, $\mathcal{U}$ and $\mathcal{Y}$ ensures that $\mathbb{G}_n^{(1)}(p)$, $\mathbb{G}_N^{(2)}(p)$ and $\mathbb{G}_m^{(3)}(p)$ are also mutually independent.

Next we give the asymptotic distribution of $\mathbb{G}_n^{(1)}(p)$, $\mathbb{G}_N^{(2)}(p)$ and $\mathbb{G}_m^{(3)}(p)$, respectively. By the Donsker's theorem, $\big\{\mathbb{G}_n^{(1)}(p) : p \in (0,1)\big\}$ converges in distribution to a Gaussian process W_1' with zero mean and covariance given by $\mathrm{Cov}\big\{\mathrm{IF}(p_1,\lambda), \mathrm{IF}(p_2,\lambda)\big\}$. Since $\big\{Q(\mathbf{Z}^\top\boldsymbol{\beta}_0,\theta_0,p) : p \in (0,1)\big\}$ is a Donsker, by the Donsker's theorem, $\big\{\mathbb{G}_N^{(2)}(p) : p \in (0,1)\big\}$ converges in distribution to a Gaussian process W_2 with zero mean and covariance $\mathrm{Cov}\big\{Q(S_1,\theta_0,p_1), Q(S_2,\theta_0,p_2)\big\}$. Since $\widehat{\theta}$ does not depend on p, it is easy to show that $\mathbb{G}_m^{(3)}(p)$ satisfies the stochastic equicontinuity. Then apply the Donsker's theorem again and $\big\{\mathbb{G}_m^{(3)}(p) : p \in (0,1)\big\}$ converges in distribution to a Gaussian process W_3, where its mean is zero and the covariance function is $A(\theta_0,p_1)I^{-1}(\theta_0)A(\theta_0,p_2)$. As a result, by the independence among the three part of data, the sequence of $\big\{\mathbb{G}_n(p) : p \in (0,1)\big\}$ converges in distribution to a new Gaussian process $W_{\lambda,\mathcal{S}}$, where $W_{\lambda,\mathcal{S}} = W_{1,\mathcal{S}} + (1-\lambda)\sqrt{\rho}W_2 + \sqrt{r}W_3$, where $\rho = \lim_{n,N\to\infty} n/N$ and $r = \lim_{n,m\to\infty} n/m$.

Denote $\Xi_1(p) = \mathrm{Var}\big\{q(X_i,\theta_0,p)\big\}$, $\Xi_2(p) = \mathrm{Var}\big\{Q(S_i,\theta_0,p\big\}$, $\Xi_3 = I^{-1}(\theta_0)$, $\Sigma_{\boldsymbol{\beta}}(p) = \mathrm{Cov}\big\{\alpha(X_1,Z_1,\boldsymbol{\beta}_0,\theta_0,p)\big\}$, $\Sigma_{q,\boldsymbol{\beta}}(p) = \mathrm{Cov}\big\{q(X_1,\theta_0,p), \alpha(X_1,Z_1,\boldsymbol{\beta}_0,\theta_0,p)\big\}$ and $\Sigma_{Q,\boldsymbol{\beta}}(p) = \mathrm{Cov}\big\{Q(S,p), \alpha(X,p)\big\}$. It is easy to check that $\mathrm{Cov}\big\{q(X_i,\theta_0,p), Q(S_i,\theta_0,p)\big\} = \Xi_2(p)$.

When p is fixed, we have $\mathbb{G}_n(p) \xrightarrow{d} N(0, \Xi_{\lambda,\mathcal{S}}(p))$, where $\Xi_{\lambda,\mathcal{S}}(p) = \Xi_1(p) - \eta(\lambda)\Xi_2(p) + r\Xi_3 + \Omega_{\lambda,\mathcal{S}}(p)$. The proof has been completed. ∎

Proof of Theorem 3.3. Denote $p(x, \mathbf{z})$ as the joint density function of $(X, \mathbf{Z})$. We follow the structure outlined in Chapter 4 of Tsiatis (2006) to calculate the semiparametric efficiency bound.

First, characterize the tangent space. The density function $p(x, \mathbf{z})$ can be re-expressed as $p(x, \mathbf{z}) = p(x|\mathbf{z})p(\mathbf{z})$, where $p(x|\mathbf{z})$ is the conditional density of X given $\mathbf{Z}$, $p(\mathbf{z})$ is the marginal density function of $\mathbf{Z}$. Consider a regular parametric submodel specified by γ. The joint density function for X and $\mathbf{Z}$ is given by $p_\gamma(x, \mathbf{z}) = p(x|\mathbf{z};\gamma)p(\mathbf{z};\gamma)$. Denote the true value of γ by γ_0. The score function is given by $s_\gamma(x, \mathbf{z}) = s(x|\mathbf{z};\gamma) + s(\mathbf{z};\gamma)$, where $s(x|\mathbf{z};\gamma) = d\log p(x|\mathbf{z};\gamma)/d\gamma$, $s(\mathbf{z};\gamma) = d\log p(\mathbf{z};\gamma)/d\gamma$. According to Theorem 4.5 in Tsiatis (2006), the corresponding parametric submodel tangent space is

$$\Gamma = \left\{ s(x|\mathbf{z};\gamma_0) + s(\mathbf{z};\gamma_0) : \int s(x|\mathbf{z};\gamma_0)p(x|\mathbf{z})dx = 0, \forall \mathbf{z} \text{ and } \int s(\mathbf{z};\gamma_0)p(\mathbf{z})d\mathbf{z} = 0 \right\}.$$

Let

$$G_n(X, \mathbf{Z}, p) = \frac{1}{n}\sum_{i=1}^{n} \{q(X_i, \theta_0, p) - Q(\mathbf{Z}_i, \theta_0, p)\} + \frac{1}{N}\sum_{j=n+1}^{M} Q(\mathbf{Z}_j, \theta_0, p) - \Delta_0(p).$$

Since $E\{Q(\mathbf{Z}_i, \theta_0, p) - \Delta_0(p)\} = 0$ and $E\{q(X_i, \theta_0, p) - Q(\mathbf{Z}_i, \theta_0, p)|\mathbf{Z}\} = 0$, we can see that $G_n(X, \mathbf{Z}, p)$ belongs to the space Γ. Thus, the projection of $G_n(X, \mathbf{Z}, p)$ onto Γ is itself. As a consequence of Theorem 4.3 in Tsiatis (2006), the optimal variance bound is $E\big\{G_n(X, \mathbf{Z}, p)\big\}^2$. The proof of Lemma 3.1 shows that

$$\tilde{\Delta}_{\lambda^{\text{opt}},\mathcal{U}}(p) - \Delta_0(p) = \frac{1}{n}\sum_{i=1}^{n} \{q(X_i, \theta_0, p) - Q(S_i, \theta_0, p)\} + \frac{1}{N}\sum_{j=n+1}^{M} Q(S_j, \theta_0, p) - \Delta_0(p) + o_p(n^{-1/2}).$$

$$\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{S}}(p)-\Delta_0(p)=\frac{1}{n}\sum_{i=1}^{n}\{q(X_i,\theta_0,p)-Q(S_i,\theta_0,p)\}+\frac{1}{N}\sum_{j=n+1}^{M}Q(S_j,\theta_0,p)-\Delta_0(p)+o_p(n^{-1/2}).$$

Obviously, if $E\{q(X,\theta_0,p)|\mathbf{Z}\} = Q(S,\theta_0,p)$, the asymptotic variance of $\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{U}}(p)$ and $\tilde{\Delta}_{\lambda^{\mathrm{opt}},\mathcal{S}}(p)$ achieve the efficiency bound. ∎

References

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American statistical association 89*(425), 81–87.

González-Manteiga, W., J. C. Pardo-Fernández, and I. v. Keilegom (2011). Roc curves in non-parametric location-scale regression models. *Scandinavian Journal of Statistics 38*(1), 169–184.

Grunewalder, S. (2018). Plug-in estimators for conditional expectations and probabilities. In *International Conference on Artificial Intelligence and Statistics*, pp. 1513–1521. PMLR.

Hu, Z., D. A. Follmann, and J. Qin (2012). Semiparametric double balancing score estimation for incomplete data with ignorable missingness. *Journal of the American Statistical Association 107*(497), 247–257.

Pace, R. K. and R. Barry (1997). Sparse spatial autoregressions. *Statistics & Probability Letters 33*(3), 291–297.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction.* Oxford university press.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Sen, B. (2018). A gentle introduction to empirical process theory and applications. *Lecture Notes, Columbia University 11*, 28–29.

Tsiatis, A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, Volume 3. Cambridge university press.

Van Der Vaart, A. W., J. A. Wellner, A. W. van der Vaart, and J. A. Wellner (1996). *Weak convergence*. Springer.

Wang, R., M. Su, and Q. Wang (2023). Distributed nonparametric regression imputation for missing response problems with large-scale data. *Journal of Machine Learning Research 24*(68), 1–52.

Zhang, Y., A. Chakrabortty, and J. Bradic (2023). Double robust semi-supervised inference for the mean: selection bias under mar labeling with decaying overlap. *Information and Inference: A Journal of the IMA 12*(3), 2066–2159.