Sufficient Dimension Reduction under Dimension-reduction-based Imputation with Predictors Missing at Random

Academy of Mathematics and Systems Science,

Chinese Academy of Sciences.

Supplementary Material

In this supplementary note, we provide the simulation results under models (3.2)–(3.3), a simulation study with a discrete response, and detailed proofs for Lemmas A.1–A.3.

S1 Simulation Results under Models (3.2)–(3.3)

Tables 5–12 give the simulation results under models (3.2)–(3.3), with the missingness mechanisms (3.4)–(3.5). These results indicate similar features to those of Tables 1–4, and show the superiority of our proposed DRI-SIR over other methods.

Table 5: Comparison of the median TCC of the SDR estimations for model (3.2) under the missingness mechanism in (3.4), with different p_1 and missing proportions (mp)

p_1	C_0	mp	Full-SIR	**DRI-SIR**	CC-SIR	AIPW-SIR	MAIPW-SIR	PI-SIR
3	2.4	19.77%	0.9025	**0.9073**	0.8442	0.8684	0.8838	0.7097
	1.1	34.92%	0.9025	**0.9008**	0.7758	0.8196	0.8341	0.6324
	0	50.19%	0.9025	**0.8850**	0.6773	0.6952	0.7066	0.6362
5	2.3	20.75%	0.9189	**0.9139**	0.8373	0.8582	0.8825	0.6018
	1.1	34.81%	0.9199	**0.9016**	0.7501	0.7684	0.7955	0.5571
	0	50.10%	0.9167	**0.8761**	0.6237	0.5677	0.6304	0.5766
10	2.2	19.91%	0.9239	**0.9061**	0.7754	0.7525	0.8522	0.4720
	0.9	35.64%	0.9236	**0.8720**	0.6443	0.4502	0.6535	0.4691
	-0.1	50.11%	0.9236	**0.7795**	0.5249	0.4512	0.4656	0.5489

Table 6: Comparison of the median TCC of the SDR estimations for model (3.2) under the missingness mechanism in (3.5), with different p_1 and missing proportions (mp)

p_1	C_0	mp	Full-SIR	**DRI-SIR**	CC-SIR	AIPW-SIR	MAIPW-SIR	PI-SIR
3	2.4	20.33%	0.9068	**0.9083**	0.8846	0.7308	0.8108	0.6259
	0.2	35.50%	0.9041	**0.8988**	0.8524	0.6285	0.6824	0.5805
	-1.7	50.60%	0.9041	**0.8881**	0.8088	0.5748	0.6416	0.5869
5	2	20.65%	0.9185	**0.9178**	0.8911	0.6442	0.8020	0.5380
	0.1	34.93%	0.9186	**0.9071**	0.8586	0.5143	0.6458	0.4468
	-1.5	49.73%	0.9168	**0.8907**	0.8141	0.4495	0.5522	0.3923
10	0.9	20.84%	0.9242	**0.9140**	0.8123	0.6206	0.8604	0.7843
	-0.3	34.94%	0.9236	**0.8971**	0.7179	0.4513	0.7725	0.5617
	-1.6	49.61%	0.9236	**0.8659**	0.6441	0.3513	0.5959	0.3912

Table 7: Distribution (in percentages) of the estimated structural dimension for model (3.2) under the missingness mechanism in (3.4) with different p_1 and missing proportions (mp)

p_1	Method	$\widehat{d}$ 1	2	> 2	1	2	> 2	1	2	> 2
		mp=19.77%			mp=34.92%			mp=50.19%		
3	Full-SIR	0.0260	**0.9740**	0.0000	0.0260	**0.9740**	0.0000	0.0260	**0.9740**	0.0000
	DRI-SIR	0.0240	**0.9760**	0.0000	0.0280	**0.9720**	0.0000	0.0340	**0.9660**	0.0000
	CC-SIR	0.1100	**0.8900**	0.0000	0.2860	**0.7140**	0.0000	0.8060	**0.1940**	0.0000
	AIPW-SIR	0.0700	**0.9000**	0.0300	0.1020	**0.8060**	0.0920	0.1640	**0.7020**	0.1340
	MAIPW-SIR	0.0660	**0.9220**	0.0120	0.0660	**0.8620**	0.0720	0.0740	**0.7000**	0.2260
	PI-SIR	0.3700	**0.5780**	0.0520	0.4120	**0.4700**	0.1180	0.2540	**0.5520**	0.1940
		mp=20.75%			mp=34.81%			mp=50.10%		
5	Full-SIR	0.0020	**0.9980**	0.0000	0.0040	**0.9960**	0.0000	0.0080	**0.9920**	0.0000
	DRI-SIR	0.0080	**0.9920**	0.0000	0.0100	**0.9900**	0.0000	0.0060	**0.9940**	0.0000
	CC-SIR	0.1580	**0.8420**	0.0000	0.5280	**0.4720**	0.0000	0.9980	**0.0020**	0.0000
	AIPW-SIR	0.0440	**0.9280**	0.0280	0.0720	**0.8120**	0.1160	0.1700	**0.5520**	0.2780
	MAIPW-SIR	0.0500	**0.9120**	0.0380	0.0400	**0.8420**	0.1180	0.0820	**0.5960**	0.3220
	PI-SIR	0.4520	**0.4840**	0.0640	0.4740	**0.4160**	0.1100	0.3980	**0.4720**	0.1300
		mp=19.91%			mp=35.64%			mp=50.11%		
10	Full-SIR	0.0340	**0.9660**	0.0000	0.0600	**0.9400**	0.0000	0.0460	**0.9540**	0.0000
	DRI-SIR	0.0240	**0.9760**	0.0000	0.0520	**0.9480**	0.0000	0.1080	**0.8880**	0.0040
	CC-SIR	0.6580	**0.3420**	0.0000	0.9840	**0.0160**	0.0000	1.0000	**0.0000**	0.0000
	AIPW-SIR	0.1280	**0.8160**	0.0560	0.2460	**0.4900**	0.2640	0.3100	**0.4740**	0.2160
	MAIPW-SIR	0.0560	**0.9060**	0.0380	0.0780	**0.7160**	0.2060	0.1040	**0.5120**	0.3840
	PI-SIR	0.5560	**0.4020**	0.0420	0.5820	**0.3500**	0.0680	0.5460	**0.4020**	0.0520

Table 8: Distribution (in percentages) of the estimated structural dimension for model (3.2) under the missingness mechanism in (3.5), with different p_1 and missing proportions (mp)

p_1	Method $\widehat{d}$	1	2	> 2	1	2	> 2	1	2	> 2
		mp=20.33%			mp=35.50%			mp=50.60%		
3	Full-SIR	0.0220	**0.9780**	0.0000	0.0280	**0.9720**	0.0000	0.0280	**0.9720**	0.0000
	DRI-SIR	0.0120	**0.9880**	0.0000	0.0160	**0.9840**	0.0000	0.0140	**0.9860**	0.0000
	CC-SIR	0.0180	**0.9820**	0.0000	0.0400	**0.9600**	0.0000	0.1500	**0.8500**	0.0000
	AIPW-SIR	0.0720	**0.6060**	0.3220	0.1220	**0.4820**	0.3960	0.1580	**0.4400**	0.4020
	MAIPW-SIR	0.0700	**0.7040**	0.2260	0.0640	**0.5680**	0.3680	0.1200	**0.4700**	0.4100
	PI-SIR	0.4140	**0.5080**	0.0780	0.4640	**0.4520**	0.0840	0.4860	**0.4300**	0.0840
		mp=20.65%			mp=34.93%			mp=49.73%		
5	Full-SIR	0.0080	**0.9920**	0.0000	0.0080	**0.9920**	0.0000	0.0180	**0.9820**	0.0000
	DRI-SIR	0.0040	**0.9960**	0.0000	0.0040	**0.9960**	0.0000	0.0100	**0.9900**	0.0000
	CC-SIR	0.0040	**0.9960**	0.0000	0.0180	**0.9820**	0.0000	0.1960	**0.8040**	0.0000
	AIPW-SIR	0.0900	**0.4980**	0.4120	0.1520	**0.3980**	0.4500	0.2080	**0.3560**	0.4360
	MAIPW-SIR	0.0400	**0.6960**	0.2640	0.0900	**0.4860**	0.4240	0.1180	**0.3800**	0.5020
	PI-SIR	0.4640	**0.4700**	0.0660	0.5200	**0.4280**	0.0520	0.4820	**0.4800**	0.0380
		mp=20.84%			mp=34.94%			mp=49.61%		
10	Full-SIR	0.0360	**0.9640**	0.0000	0.0460	**0.9540**	0.0000	0.0460	**0.9540**	0.0000
	DRI-SIR	0.0360	**0.9640**	0.0000	0.0460	**0.9540**	0.0000	0.0700	**0.9300**	0.0000
	CC-SIR	0.6220	**0.3780**	0.0000	0.9420	**0.0580**	0.0000	0.9980	**0.0020**	0.0000
	AIPW-SIR	0.0560	**0.5340**	0.4100	0.1020	**0.4560**	0.4420	0.1540	**0.4480**	0.3980
	MAIPW-SIR	0.0620	**0.9020**	0.0360	0.0400	**0.7980**	0.1620	0.0560	**0.5860**	0.3580
	PI-SIR	0.3780	**0.6120**	0.0100	0.6120	**0.3760**	0.0120	0.5980	**0.3740**	0.0280

Table 9: Comparison of the median TCC of the SDR estimations for model (3.3) under the missingness mechanism in (3.4), with different p_1 and missing proportions (mp)

p_1	C_0	mp	Full-SIR	**DRI-SIR**	CC-SIR	AIPW-SIR	MAIPW-SIR	PI-SIR
3	2.4	20.11%	0.9085	**0.9042**	0.8440	0.8700	0.8830	0.6913
	1.1	35.16%	0.9084	**0.8945**	0.7689	0.8168	0.8421	0.6422
	0	50.17%	0.9083	**0.8810**	0.6594	0.7082	0.6856	0.6299
5	2.4	19.91%	0.9024	**0.8978**	0.8120	0.8392	0.8607	0.6275
	1.1	34.94%	0.9054	**0.8845**	0.7019	0.7401	0.7573	0.5979
	0	50.00%	0.9042	**0.8618**	0.5874	0.5548	0.5532	0.6048
10	2.3	20.25%	0.9026	**0.8889**	0.7270	0.7045	0.8144	0.5578
	1.1	34.64%	0.9022	**0.8548**	0.5850	0.4336	0.6109	0.5664
	0	50.31%	0.9026	**0.7537**	0.4439	0.4286	0.3741	0.6208

Table 10: Comparison of the median TCC of the SDR estimations for model (3.3) under the missingness mechanism in (3.5), with different p_1 and missing proportions (mp)

p_1	C_0	mp	Full-SIR	**DRI-SIR**	CC-SIR	AIPW-SIR	MAIPW-SIR	PI-SIR
3	2.4	20.48%	0.9084	**0.9032**	0.8537	0.7333	0.8381	0.6432
	0.3	35.38%	0.9084	**0.8956**	0.8010	0.5831	0.6580	0.6143
	-1.7	50.14%	0.9084	**0.8819**	0.7258	0.5345	0.5700	0.6209
5	2	20.46%	0.9022	**0.8987**	0.8299	0.6151	0.8054	0.6214
	0	35.28%	0.9027	**0.8865**	0.7635	0.4800	0.6032	0.5793
	-1.7	50.55%	0.9017	**0.8663**	0.6744	0.4095	0.4888	0.5586
10	1.1	20.45%	0.8983	**0.8893**	0.7537	0.5980	0.8345	0.8129
	0.3	35.33%	0.8998	**0.8623**	0.6493	0.4515	0.7481	0.6140
	-1.6	50.87%	00.9015	**0.8297**	0.5761	0.3675	0.5909	0.4831

Table 11: Distribution (in percentages) of the estimated structural dimension for model (3.3) under the missingness mechanism in (3.4), with different p_1 and missing proportions (mp)

p_1	Method	$\widehat{d}$ 1	2	>2	1	2	>2	1	2	>2
		mp=20.11%			mp=35.16%			mp=50.17%		
3	Full-SIR	0.0000	**0.9880**	0.0120	0.0000	**0.9880**	0.0120	0.0000	**0.9860**	0.0140
	DRI-SIR	0.0000	**0.9880**	0.0120	0.0000	**0.9860**	0.0140	0.0020	**0.9760**	0.0220
	CC-SIR	0.0060	**0.9880**	0.0060	0.0720	**0.9280**	0.0000	0.6980	**0.3020**	0.0000
	AIPW-SIR	0.0040	**0.9160**	0.0800	0.0200	**0.8000**	0.1800	0.0560	**0.6260**	0.3180
	MAIPW-SIR	0.0020	**0.9620**	0.0360	0.0160	**0.7900**	0.1940	0.0560	**0.5580**	0.3860
	PI-SIR	0.2280	**0.6140**	0.1580	0.3040	**0.4980**	0.1980	0.2760	**0.4500**	0.2740
		mp=19.91%			mp=34.94%			mp=50.00%		
5	Full-SIR	0.0000	**0.9860**	0.0140	0.0000	**0.9900**	0.0100	0.0000	**0.9860**	0.0140
	DRI-SIR	0.0000	**0.9960**	0.0040	0.0000	**0.9840**	0.0160	0.0040	**0.9800**	0.0160
	CC-SIR	0.0080	**0.9900**	0.0020	0.2900	**0.7100**	0.0000	0.9880	**0.0120**	0.0000
	AIPW-SIR	0.0060	**0.8860**	0.1080	0.0240	**0.7260**	0.2500	0.1160	**0.5040**	0.3800
	MAIPW-SIR	0.0100	**0.9200**	0.0700	0.0420	**0.6700**	0.2880	0.0960	**0.4060**	0.4980
	PI-SIR	0.3460	**0.4820**	0.1720	0.4260	**0.4240**	0.1500	0.3540	**0.4380**	0.2080
		mp=20.25%			mp=34.64%			mp=50.31%		
10	Full-SIR	0.0060	**0.9920**	0.0020	0.0060	**0.9920**	0.0020	0.0060	**0.9920**	0.0020
	DRI-SIR	0.0020	**0.9960**	0.0020	0.0060	**0.9860**	0.0080	0.0360	**0.9160**	0.0480
	CC-SIR	0.1260	**0.8740**	0.0000	0.9140	**0.0860**	0.0000	1.0000	**0.0000**	0.0000
	AIPW-SIR	0.0200	**0.7980**	0.1820	0.1440	**0.4940**	0.3620	0.2260	**0.4500**	0.3240
	MAIPW-SIR	0.0160	**0.8820**	0.1020	0.0460	**0.6280**	0.3260	0.1260	**0.3700**	0.5040
	PI-SIR	0.3520	**0.5500**	0.0980	0.4400	**0.4560**	0.1040	0.3980	**0.5020**	0.1000

Table 12: Distribution (in percentages) of the estimated structural dimension for model (3.3) under the missingness mechanism in (3.5), with different p_1 and missing proportions (mp)

p_1	Method	$\widehat{d}$	1	2	>2	1	2	>2	1	2	>2
			mp=20.48%			mp=35.38%			mp=50.14%		
3	Full-SIR		0.0000	**0.9880**	0.0120	0.0000	**0.9880**	0.0120	0.0000	**0.9880**	0.0120
	DRI-SIR		0.0000	**0.9860**	0.0140	0.0000	**0.9840**	0.0160	0.0000	**0.9720**	0.0280
	CC-SIR		0.0000	**0.9840**	0.0160	0.0080	**0.9920**	0.0000	0.1100	**0.8900**	0.0000
	AIPW-SIR		0.0660	**0.6020**	0.3320	0.0940	**0.4300**	0.4760	0.1640	**0.3840**	0.4520
	MAIPW-SIR		0.0220	**0.7620**	0.2160	0.0720	**0.5180**	0.4100	0.1120	**0.4080**	0.4800
	PI-SIR		0.1860	**0.5240**	0.2900	0.2260	**0.4380**	0.3360	0.1640	**0.4400**	0.3960
			mp=20.46%			mp=35.28%			mp=50.55%		
5	Full-SIR		0.0000	**0.9860**	0.0140	0.0000	**0.9860**	0.0140	0.0000	**0.9840**	0.0160
	DRI-SIR		0.0000	**0.9940**	0.0060	0.0000	**0.9880**	0.0120	0.0000	**0.9860**	0.0140
	CC-SIR		0.0000	**0.9940**	0.0060	0.0480	**0.9520**	0.0000	0.4300	**0.5700**	0.0000
	AIPW-SIR		0.0680	**0.5280**	0.4040	0.1220	**0.4080**	0.4700	0.1320	**0.4880**	0.3800
	MAIPW-SIR		0.0320	**0.7160**	0.2520	0.0700	**0.5000**	0.4300	0.1020	**0.3980**	0.5000
	PI-SIR		0.2660	**0.4660**	0.2680	0.2860	**0.4660**	0.2480	0.2800	**0.4440**	0.2760
			mp=20.45%			mp=35.33%			mp=50.87%		
10	Full-SIR		0.0000	**0.9980**	0.0020	0.0040	**0.9960**	0.0000	0.0040	**0.9920**	0.0040
	DRI-SIR		0.0000	**0.9940**	0.0060	0.0000	**0.9940**	0.0060	0.0020	**0.9800**	0.0180
	CC-SIR		0.2960	**0.7040**	0.0000	0.7880	**0.2120**	0.0000	0.9980	**0.0020**	0.0000
	AIPW-SIR		0.0640	**0.5160**	0.4200	0.0700	**0.4920**	0.4380	0.1020	**0.4440**	0.4540
	MAIPW-SIR		0.0140	**0.9280**	0.0580	0.0240	**0.7580**	0.2180	0.0380	**0.5820**	0.3800
	PI-SIR		0.0860	**0.8860**	0.0280	0.2960	**0.6280**	0.0760	0.4460	**0.4680**	0.0860

S2 A simulation Study with A Discrete Response

We here consider one numerical example, with a discrete response Y that has four categories. We first generate $\mathbf{X}_i = (X_{1i}, \cdots, X_{pi})^T$ from a p-variate normal distribution, with mean $\mathbf{0}$ and covariance $0.3^{|k-l|}$ between any two components X_k and X_l of $\mathbf{X} = (X_1, \cdots, X_p)^T$, and then generate Y_i from a conditional multinomial distribution given $\mathbf{X}$, with parameters $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$, where

$$\begin{aligned}
\pi_1 &= \Pr(Y=1|\mathbf{X}) = \frac{\exp(-1.5+\beta_1^T\mathbf{X})}{1+\exp(-1.5+\beta_1^T\mathbf{X})} \\
\pi_2 &= \Pr(Y=2|\mathbf{X}) = 0.4(1-\pi_1) \\
\pi_3 &= \Pr(Y=3|\mathbf{X}) = 0.3(1-\pi_1) \\
\pi_4 &= \Pr(Y=4|\mathbf{X}) = 0.3(1-\pi_1)
\end{aligned} \tag{S2.1}$$

We set $p = 15$, $p_1 = 3, 5$, and 10 (the dimension of missing predictors), and $\beta_1 = (0.5 \times \mathbf{1}_{p_1-1}, \mathbf{0}_{p-p_1-2}, 0.5, -1, -1)^T$, where $\mathbf{1}_s$ and $\mathbf{0}_s$ denote $1 \times s$ vectors, with all elements being one and two, respectively. It is clear in this scenario that the CS is $\mathcal{S}_{Y|\mathbf{X}} = \text{Span}\{\beta_1\}$ and, hence, the true structural dimension is $d = 1$. In addition to $(\mathbf{X}_i, Y_i)$, we generate the missingness indicators δ_{ki}, with $k = 1, \cdots, p_1$ from the following missingness mechanism:

$$\begin{aligned}
\Pr(\delta_k = 1|\mathbf{X}_{obs}, Y = 1) &= \frac{\exp(c_0 + \gamma^T \mathbf{X}_{obs} + 0.7)}{1 + \exp(c_0 + \gamma^T \mathbf{X}_{obs} + 0.7)} \\
\Pr(\delta_k = 1|\mathbf{X}_{obs}, Y = 2) &= \frac{\exp(c_0 + \gamma^T \mathbf{X}_{obs} + 1.2)}{1 + \exp(c_0 + \gamma^T \mathbf{X}_{obs} + 1.2)} \\
\Pr(\delta_k = 1|\mathbf{X}_{obs}, Y = 3) &= \frac{\exp(c_0 + \gamma^T \mathbf{X}_{obs} + 0.9)}{1 + \exp(c_0 + \gamma^T \mathbf{X}_{obs} + 0.9)} \\
\Pr(\delta_k = 1|\mathbf{X}_{obs}, Y = 4) &= \frac{\exp(c_0 + \gamma^T \mathbf{X}_{obs})}{1 + \exp(c_0 + \gamma^T \mathbf{X}_{obs})}
\end{aligned} \tag{S2.2}$$

where $\mathbf{X}_{obs} = (X_{p_1+1}, \cdots, X_p)^T$ is always observed, c_0 is a scalar constant to control the missing proportions, and $\gamma = (-1, -1, -1, \mathbf{0}_{p-p_1-5}, 0.5, 0.5)^T$, with $\mathbf{0}_{p-p_1-5}$ being a $1 \times (p - p_1 - 5)$ zero vector.

The simulations are repeated 500 times, where each sample is of size $n = 400$. Three methods, including the Full-SIR, CC-SIR, and DRI-SIR, are compared, whereas the AIPW-SIR, MAIPW-SIR, and PI-SIR are not illustrated because it becomes more difficult or even impossible to specify correct parametric models for the involved conditional expectations in the presence of a discrete response.

We report the median TCCs under known $d = 1$ in Table 13, and the empirical distributions of the estimated structural dimension $\widehat{d}$ in Table 14, to evaluate the performance of these three methods. Table 13 shows that in most cases the proposed DRI-SIR can achieve high accuracy when estimating the CS, with a known structural dimension. To be specific, it performs much better than the CC-SIR does, and even shows comparable performance to that of the Full-SIR under small missing proportions. Table 14 reveals that the proposed DRI-SIR selects the true structural dimension with a probability tending to one if the missing proportion does not exceed 50%. These quantitative features confirm that our method still performs very well in the presence of a discrete response, which greatly expands the scope of applicability of our method.

Table 13: Comparison of the median TCC of the SDR estimations for model (S2.1) under the missingness mechanism in (S2.2), with different p_1 and missing proportions (mp)

p_1	C_0	mp	Full-SIR	**DRI-SIR**	CC-SIR
3	1.6	20.76%	0.9445	**0.9397**	0.9023
	0.3	35.84%	0.9445	**0.9299**	0.8507
	-0.8	50.85%	0.9445	**0.9125**	0.7642
5	1.6	20.88%	0.9481	**0.9379**	0.8960
	0.3	35.85%	0.9497	**0.9236**	0.8216
	-0.8	50.86%	0.9494	**0.8735**	0.7007
10	1.6	20.07%	0.9579	**0.9373**	0.8831
	0.3	35.40%	0.9570	**0.8964**	0.7884
	-0.8	50.81%	0.9565	**0.7248**	0.5572

Table 14: Distribution (in percentages) of the estimated structural dimension for model (S2.1) under the missingness mechanism in (S2.2), with different p_1 and missing proportions (mp)

p_1	Method	$\widehat{d}$	1	> 1	1	> 1	1	> 1
			mp=20.76%		mp=35.84%		mp=50.85%	
3	Full-SIR		**0.9840**	0.0160	**0.9840**	0.0160	**0.9840**	0.0160
	DRI-SIR		**0.9680**	0.0320	**0.9300**	0.0700	**0.7960**	0.2040
	CC-SIR		**0.8800**	0.1200	**0.7880**	0.2120	**0.8200**	0.1800
			mp=20.88%		mp=35.85%		mp=50.86%	
5	Full-SIR		**0.9960**	0.0040	**0.9900**	0.0100	**1.0000**	0.0000
	DRI-SIR		**0.9820**	0.0180	**0.9220**	0.0780	**0.7340**	0.2660
	CC-SIR		**0.9020**	0.0980	**0.8140**	0.1860	**0.9360**	0.0640
			mp=20.07%		mp=35.40%		mp=50.81%	
10	Full-SIR		**0.9980**	0.0020	**1.0000**	0.0000	**1.0000**	0.0000
	DRI-SIR		**0.9900**	0.0100	**0.9200**	0.0800	**0.5180**	0.4820
	CC-SIR		**0.9560**	0.0440	**0.9700**	0.0300	**1.0000**	0.0000

S3 Proofs for Lemmas A.1–A.3

We here focus on presenting the proofs of both Lemma A.1 (i) and Lemma A.2 (i). Because the proofs of Lemma A.1 (ii)–(iv) are similar to those of Lemma A.1 (i), and the proofs of both Lemma A.2 (ii) and Lemma A.3 are similar to those of Lemma A.2 (i), we omit the details here.

S3.1 Proof of Lemma A.1 (i)

Recalling that

$$\begin{aligned}\widehat{E}(X_k) &= \frac{1}{n}\sum_{i=1}^{n}\{\delta_{ki}X_{ki} + (1-\delta_{ki})\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki} + (1-\delta_{ki})\frac{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}\right\}\end{aligned} \tag{S3.1}$$

where $\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) = n^{-1}\sum_{j=1}^{n}K_h(\widehat{\Gamma}_k^T\mathbf{V}_j - \widehat{\Gamma}_k^T\mathbf{V}_i)\delta_{kj}X_{kj}$ and $\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) = n^{-1}\sum_{j=1}^{n}K_h(\widehat{\Gamma}_k^T\mathbf{V}_j - \widehat{\Gamma}_k^T\mathbf{V}_i)\delta_{kj}$. It is easy to see that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)} \\ =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) - \widehat{G}_k(\Gamma_k^T\mathbf{V}_i) + \widehat{G}_k(\Gamma_k^T\mathbf{V}_i) - G_k(\Gamma_k^T\mathbf{V}_i) + G_k(\Gamma_k^T\mathbf{V}_i)\} \\ &\times\left\{\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)} + \frac{g_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)} + \frac{1}{g_k(\Gamma_k^T\mathbf{V}_i)}\right\}\end{aligned}$$

Let

$$\begin{aligned}T_1 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) - \widehat{G}_k(\Gamma_k^T\mathbf{V}_i)\}\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)} \\ T_2 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) - \widehat{G}_k(\Gamma_k^T\mathbf{V}_i)\}\frac{g_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)} \\ T_3 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i) - \widehat{G}_k(\Gamma_k^T\mathbf{V}_i)\}\Big/g_k(\Gamma_k^T\mathbf{V}_i) \\ T_4 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i) - G_k(\Gamma_k^T\mathbf{V}_i)\}\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)} \\ T_5 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i) - G_k(\Gamma_k^T\mathbf{V}_i)\}\frac{g_k(\Gamma_k^T\mathbf{V}_i) - \widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)} \\ T_6 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i) - G_k(\Gamma_k^T\mathbf{V}_i)\}\Big/g_k(\Gamma_k^T\mathbf{V}_i)\end{aligned}$$

$$T_7 =\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}$$

$$T_8 =\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)}$$

$$T_9 =\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\Big/g_k(\Gamma_k^T\mathbf{V}_i)=\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})M_k(\Gamma_k^T\mathbf{V}_i)$$

Then, after simple algebraic calculation, we have

$$\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}=\sum_{i=1}^{9}T_i \tag{S3.2}$$

We first deal with those dominant terms including T_3 , T_6 , T_7, and T_8, and other terms can be handled in a similar way.

Following Cook and Li (2002), it is easy to find that $\widehat{\Gamma}_k$ obtained using the method mentioned in Subsection 2.2 is a $\sqrt{n}$ consistent estimator of Γ_k, namely $\parallel \widehat{\Gamma}_k-\Gamma_k \parallel = O_p(n^{-1/2})$. Then, under conditions 1–4, Lemmas 1–2 of Li, Zhu and Zhu (2011) yield that

$$\sup_{\|\widehat{\Gamma}_k-\Gamma_k\|\le Cn^{-1/2}}\sup_{\mathbf{V}\in R^q}\left|\left\{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V})-\widehat{G}_k(\Gamma_k^T\mathbf{V})\right\}-\left\{G_k(\widehat{\Gamma}_k^T\mathbf{V})-G_k(\Gamma_k^T\mathbf{V})\right\}\right|$$

$$=O(h^m+n^{-1}h^{-(r_k+1)}\log n)\ a.s \tag{S3.3}$$

which together with the Taylor expansion implies that

$$\begin{aligned}T_3 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{G_k(\widehat{\Gamma}_k^T\mathbf{V}_i)-G_k(\Gamma_k^T\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}+o_p(n^{-1/2})\\ =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{\left\{G_k^{(1)}(\Gamma_k^T\mathbf{V}_i)\otimes\mathbf{V}_i\right\}^T}{g_k(\Gamma_k^T\mathbf{V}_i)}\left\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\right\}+o_p(n^{-1/2})\end{aligned} \tag{S3.4}$$

as $nh^{2m} \to 0$ and $nh^{2(r_k+1)}/(\log n)^2 \to \infty$, where $G_k^{(1)}(\cdot)$ denotes the first-order derivative of $G_k(\cdot)$.

For T_7, observe that

$$\begin{aligned}
T_7 =& \frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\} \\
&\times\left\{\frac{1}{g_k(\Gamma_k^T\mathbf{V}_i)}+\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)}\right\} \\
&\times\left\{\frac{1}{g_k(\Gamma_k^T\mathbf{V}_i)}+\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)}+\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}\right\} \\
=& \frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{G_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)}\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\} \\
&+\frac{2}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)g_k^2(\Gamma_k^T\mathbf{V}_i)} \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}\frac{\big\{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\big\}^2}{\widehat{g}_k^2(\Gamma_k^T\mathbf{V}_i)g_k^2(\Gamma_k^T\mathbf{V}_i)} \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\frac{\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}^2}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)} \\
&+\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}^2\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{\widehat{g}_k^2(\Gamma_k^T\mathbf{V}_i)\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)g_k(\Gamma_k^T\mathbf{V}_i)} \\
:=& T_{71}+T_{72}+T_{73}+T_{74}+T_{75} \qquad \text{(S3.5)}
\end{aligned}$$

We next consider the term T_{72}. Let

$$\zeta_k(\delta_{ki},\mathbf{V}_i) = (1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)}$$

Then, we have

$$
\begin{aligned}
T_{72} =& \frac{2}{n}\sum_{i=1}^{n}\frac{\zeta_k(\delta_{ki},\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)} \times \frac{1}{1+\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}} \\
=& \frac{2}{n}\sum_{i=1}^{n}\frac{\zeta_k(\delta_{ki},\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}\Big\{1-\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}+R_n\Big(\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}\Big)\Big\} \\
=& \frac{2}{n}\sum_{i=1}^{n}\frac{\zeta_k(\delta_{ki},\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)} + \frac{2}{n}\sum_{i=1}^{n}\frac{\zeta_k(\delta_{ki},\mathbf{V}_i)\{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)\}}{g_k^2(\Gamma_k^T\mathbf{V}_i)} \\
&+ \frac{2}{n}\sum_{i=1}^{n}\frac{\zeta_k(\delta_{ki},\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}R_n\Big(\frac{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)}{g_k(\Gamma_k^T\mathbf{V}_i)}\Big) \\
:=& T_{721}+T_{722}+T_{723}
\end{aligned}
\tag{S3.6}
$$

where the second equation holds because the Taylor expansion of the function $1/(1+x)$ at zero is $1-x+R_n(x)$, with $R_n(x)$ being the Lagrange-type remainder. Furthermore, utilizing the conclusions

$$
\begin{aligned}
&\sup_{\|\widehat{\Gamma}_k-\Gamma_k\|\le Cn^{-1/2}}\sup_{\mathbf{V}\in R^q}\Big|\big\{\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V})-\widehat{g}_k(\Gamma_k^T\mathbf{V})\big\}-\big\{g_k(\widehat{\Gamma}_k^T\mathbf{V})-g_k(\Gamma_k^T\mathbf{V})\big\}\Big| \\
=& O(h^m+n^{-1}h^{-(r_k+1)}\log n)\ a.s
\end{aligned}
\tag{S3.7}
$$

and

$$
\sup_{\Gamma_k^T\mathbf{V}\in R^{r_k}}\Big|\widehat{g}_k(\Gamma_k^T\mathbf{V})-g_k(\Gamma_k^T\mathbf{V})\Big| = O(h^m+n^{-1/2}h^{-r_k}\log n) \tag{S3.8}
$$

which are derived from Lemmas 1–2 of Li, Zhu and Zhu (2011), it is straightforward to obtain that

$$
T_{721}=o_p(n^{-1/2}) \quad \text{and} \quad T_{722}=o_p(n^{-1/2}) \tag{S3.9}
$$

as $nh^{2m}\to 0$ and $nh^{2(r_k+1)}/(\log n)^2\to\infty$. In addition, for some $c_2>0$

such that $\inf_{\Gamma_k^T \mathbf{V}} g_k(\Gamma_k^T \mathbf{V}) \geq c_2$, it is easy to show that

$$\begin{aligned}
T_{723} =& \frac{2}{n}\sum_{i=1}^{n} \frac{\zeta_k(\delta_{ki}, \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)} R_n\Big(\frac{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) - g_k(\Gamma_k^T \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)}\Big) \\
&\times I\Big\{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) \geq \frac{1}{2} g_k(\Gamma_k^T \mathbf{V}_i)\ , g_k(\Gamma_k^T \mathbf{V}_i) \geq c_2\Big\} \\
&+ \frac{2}{n}\sum_{i=1}^{n} \frac{\zeta_k(\delta_{ki}, \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)} R_n\Big(\frac{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) - g_k(\Gamma_k^T \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)}\Big) \\
&\times I\Big\{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) < \frac{1}{2} g_k(\Gamma_k^T \mathbf{V}_i)\ , g_k(\Gamma_k^T \mathbf{V}_i) \geq c_2\Big\} \\
&+ \frac{2}{n}\sum_{i=1}^{n} \frac{\zeta_k(\delta_{ki}, \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)} R_n\Big(\frac{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) - g_k(\Gamma_k^T \mathbf{V}_i)}{g_k(\Gamma_k^T \mathbf{V}_i)}\Big) I\big\{g_k(\Gamma_k^T \mathbf{V}_i) < c_2\big\} \\
&:= T_{7231} + T_{7232} + T_{7233}
\end{aligned} \tag{S3.10}$$

For T_{7231}, by the following inequality

$$|R_n(x)| = \left|\frac{1}{1+x} - 1 + x\right| \leq 2x^2 \quad \text{for} \quad |x| \leq \frac{1}{2} \tag{S3.11}$$

together with (S3.7)–(S3.8) and conditions 3–4, we obtain

$$\begin{aligned}
T_{7231} \leq& \frac{4}{n}\sum_{i=1}^{n} \Big| \frac{\zeta_k(\delta_{ki}, \mathbf{V}_i)\{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) - g_k(\Gamma_k^T \mathbf{V}_i)\}^2}{g_k^3(\Gamma_k^T \mathbf{V}_i)} I\big\{g_k(\Gamma_k^T \mathbf{V}_i) \geq c_2\big\}\Big| \\
=& o_p(n^{-1/2})
\end{aligned} \tag{S3.12}$$

In addition, for any $\varepsilon > 0$, it can be shown that

$$\begin{aligned}
P(|T_{7232}| > \varepsilon) &\leq P\Big(\bigcup_{i=1}^{n} \Big\{\widehat{g}_k(\Gamma_k^T \mathbf{V}_i) < \frac{1}{2} g_k(\Gamma_k^T \mathbf{V}_i)\ , g_k(\Gamma_k^T \mathbf{V}_i) \geq c_2\Big\}\Big) \\
&\leq P\Big(\sup_{t \in R^{r_k}} \big|\widehat{g}_k(t) - g_k(t)\big| \geq \frac{1}{2} c_2\Big) \to 0
\end{aligned} \tag{S3.13}$$

and

$$P(|T_{7233}| > \varepsilon) \le P\Big(\bigcup_{i=1}^{n}\big\{g_k(\Gamma_k^T\mathbf{V}_i) < c_2\big\}\Big) \to 0 \tag{S3.14}$$

by condition 4. Then, based on (S3.10) and (S3.12)–(S3.14), we have

$$T_{723} = o_p(n^{-1/2}) \ . \tag{S3.15}$$

Therefore, using (S3.6), (S3.9), and (S3.15), it follows that

$$T_{72} = o_p(n^{-1/2}) \tag{S3.16}$$

In addition, using similar arguments to those for T_{72}, it is easy to show that T_{73} ,T_{74}, and T_{75} are all of the order $o_p(n^{-1/2})$. These, together with (S3.5), (S3.7), and the Taylor expansion. can prove that

$$\begin{aligned}
T_7 =&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{G_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)}\big\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}+o_p(n^{-1/2})\\
=&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{G_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)}\big\{g_k(\Gamma_k^T\mathbf{V}_i)-g_k(\widehat{\Gamma}_k^T\mathbf{V}_i)\big\}+o_p(n^{-1/2})\\
=&-\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{G_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)}\big\{g_k^{(1)}(\Gamma_k^T\mathbf{V}_i)\otimes\mathbf{V}_i\big\}^T\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}\\
&+o_p(n^{-1/2})
\end{aligned} \tag{S3.17}$$

as $nh^{2m} \to 0$ and $nh^{2(r_k+1)}/(\log n)^2 \to \infty$, where $g_k^{(1)}(\cdot)$ denotes the first-order derivative of $g_k(\cdot)$. Then, using (S3.4), (S3.17), and the law of large numbers, we have

$$
\begin{aligned}
T_3+T_7 =& E\left[(1-\delta_k)\frac{G_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}}{g_k(\Gamma_k^T\mathbf{V})}\right]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}\\
&- E\left[(1-\delta_k)\frac{M_k(\Gamma_k^T\mathbf{V})g_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}}{g_k(\Gamma_k^T\mathbf{V})}\right]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}\\
&+ o_p(n^{-1/2})\\
=&E[(1-\delta_k)M_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}+o_p(n^{-1/2})\\
=&E\left[(1-\delta_k)\frac{\partial\{M_k(\Gamma_k^T\mathbf{V})\}}{\partial\{\text{vec}(\Gamma_k))\}}\right]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}+o_p(n^{-1/2})\\
=&E\left[(1-\delta_k)\frac{\partial\{E(X_k|\mathbf{V})\}}{\partial\{\text{vec}(\Gamma_k))\}}\right]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}+o_p(n^{-1/2})\\
=&E\left[\frac{\partial\{E[(1-\delta_k)X_k|\mathbf{V},\delta_k]\}}{\partial\{\text{vec}(\Gamma_k))\}}\right]^T\left\{\text{vec}(\widehat{\Gamma}_k)-\text{vec}(\Gamma_k)\right\}+o_p(n^{-1/2})\\
=&o_p(n^{-1/2}) \qquad\qquad (\text{S3.18})
\end{aligned}
$$

where the second equation holds owing to the fact $G_k^{(1)}(\Gamma_k^T\mathbf{V}) = M_k^{(1)}(\Gamma_k^T\mathbf{V})g_k(\Gamma_k^T\mathbf{V}) + M_k(\Gamma_k^T\mathbf{V})g_k^{(1)}(\Gamma_k^T\mathbf{V})$, the fifth equation holds for the MAR assumption, and the last equation holds because $\partial\{E[(1-\delta_k)X_k|\mathbf{V},\delta_k]\}/\partial\{\text{vec}(\Gamma_k)\} = 0$ for any $\mathbf{V}\in R^q$.

For T_6, the standard U-statistics theory (Serfling, 2009) shows that

$$
T_6 = \frac{1}{n}\sum_{i=1}^n\left\{\frac{\delta_{ki}X_{ki}}{\pi_k(\Gamma_k^T\mathbf{V}_i)}-\delta_{ki}X_{ki}\right\}-E\{(1-\pi_k(\Gamma_k^T\mathbf{V}))M_k(\Gamma_k^T\mathbf{V})\}+o_p(n^{-1/2}) \qquad (\text{S3.19})
$$

For T_8, using similar arguments to those used in (S3.5) and the standard

U-statistics theory, we can show

$$T_8 = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})G_k(\Gamma_k^T\mathbf{V}_i)\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)} + o_p(n^{-1/2})$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}M_k(\Gamma_k^T\mathbf{V}_i) - \frac{\delta_{ki}M_k(\Gamma_k^T\mathbf{V}_i)}{\pi_k(\Gamma_k^T\mathbf{V}_i)}\right\} + E\{(1-\pi_k(\Gamma_k^T\mathbf{V}))M_k(\Gamma_k^T\mathbf{V})\}$$
$$+ o_p(n^{-1/2}) \tag{S3.20}$$

We examine the term T_5. Lemma A.1 of Zhu, Wang and Zhu (2012) implies that $n^{-1}\sum_{i=1}^{n}\left\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i)-G_k(\Gamma_k^T\mathbf{V}_i)\right\}^2 = O_p(h^m)$ and $n^{-1}\sum_{i=1}^{n}\left\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)\right\}^2 = O_p(h^m)$. Furthermore, using the Cauchy-Schwartz inequality, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i)-G_k(\Gamma_k^T\mathbf{V}_i)\right\}\left\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)\right\}$$
$$\leq \left[n^{-1}\sum_{i=1}^{n}\left\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i)-G_k(\Gamma_k^T\mathbf{V}_i)\right\}^2\right]^{1/2}\left[n^{-1}\sum_{i=1}^{n}\left\{\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)-g_k(\Gamma_k^T\mathbf{V}_i)\right\}^2\right]^{1/2}$$
$$= O_p(h^m) \tag{S3.21}$$

which together with similar arguments to those used in (S3.5) indicates that

$$T_5 = \frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_i)-G_k(\Gamma_k^T\mathbf{V}_i)\}\frac{g_k(\Gamma_k^T\mathbf{V}_i)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_i)}{g_k^2(\Gamma_k^T\mathbf{V}_i)} + o_p(n^{-1/2})$$
$$= o_p(n^{-1/2}) \tag{S3.22}$$

as $nh^{2m}\to 0$.

Moreover, utilizing similar arguments to those used in T_{72}, it can be proved that

$$T_1 = o_p(n^{-1/2})\ ,\ T_2 = o_p(n^{-1/2}), \quad \text{and} \quad T_4 = o_p(n^{-1/2}) \tag{S3.23}$$

Finally, using the equations (S3.2), (S3.18), (S3.19), (S3.20), (S3.22), and (S3.23), we derive

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}(1-\delta_{ki})\frac{\widehat{G}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}{\widehat{g}_k(\widehat{\Gamma}_k^T\mathbf{V}_i)}\\ =&\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\delta_{ki}}{\pi_k(\Gamma_k^T\mathbf{V}_i)}\{X_{ki}-M_k(\Gamma_k^T\mathbf{V}_i)\}-\delta_{ki}X_{ki}+M_k(\Gamma_k^T\mathbf{V}_i)\right\}+o_p(n^{-1/2})\end{aligned}$$

This together with (S3.1) completes the proof of Lemma A.1 (i). □

S3.2 Proof for Lemma A.2 (i)

Recalling that

$$\widehat{T}_k(Y_i) = \frac{n^{-1}\sum_{j=1}^{n}K_h(Y_j-Y_i)\big\{\delta_{kj}X_{kj}+(1-\delta_{kj})\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_j)\big\}}{\widehat{f}_0(Y_i)} = \frac{\widehat{S}_k(Y_i)}{\widehat{f}_0(Y_i)}$$

We can show

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_k(Y_i)\widehat{T}_l(Y_i)\\ =&\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{T}_k(Y_i)-T_k(Y_i)+T_k(Y_i)\big\}\big\{\widehat{T}_l(Y_i)-T_l(Y_i)+T_l(Y_i)\big\}\\ =&\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_k(Y_i)T_l(Y_i)+\frac{1}{n}\sum_{i=1}^{n}T_k(Y_i)\widehat{T}_l(Y_i)-\frac{1}{n}\sum_{i=1}^{n}T_k(Y_i)T_l(Y_i)+o_p(n^{-1/2})\end{aligned} \tag{S3.24}$$

where the last equation holds because the arguments similar to those used in Lemma A.1 of Zhu and Zhu (2007) yield that $n^{-1}\sum_{i=1}^{n}\big\{\widehat{T}_k(Y_i)-T_k(Y_i)\big\}$

$\times\big\{\widehat{T}_l(Y_i) - T_l(Y_i)\big\} = o_p(n^{-1/2})$. Then, it suffices to handle $n^{-1}\sum_{i=1}^n \widehat{T}_k(Y_i)T_l(Y_i)$.

Next, we divide the proof into three steps.

Step 1 : Let $\widehat{T}_k^*(Y_i)$ be $\widehat{T}_k(Y_i)$, with $\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_j)$ in it replaced by $M_k(\Gamma_k^T\mathbf{V}_j)$, for $j = 1, \cdots, n$, and write $\widehat{T}_k^*(Y_i) = \widehat{S}_k^*(Y_i)/\widehat{f}_0(Y_i)$. We first show that $n^{-1}\sum_{i=1}^n \widehat{T}_k^*(Y_i)T_l(Y_i)$ admits an asymptotically linear representation. Noting that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n \widehat{T}_k^*(Y_i)T_l(Y_i)\\
=&\frac{1}{n}\sum_{i=1}^n T_l(Y_i)\frac{\widehat{S}_k^*(Y_i)}{\widehat{f}_0(Y_i)}\\
=&\frac{1}{n}\sum_{i=1}^n T_l(Y_i)\Big\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i) + T_k(Y_i)f_0(Y_i)\Big\}\left\{\frac{f_0(Y_i)-\widehat{f}_0(Y_i)}{\widehat{f}_0(Y_i)f_0(Y_i)} + \frac{1}{f_0(Y_i)}\right\}\\
=&\frac{1}{n}\sum_{i=1}^n T_k(Y_i)T_l(Y_i) + \frac{1}{n}\sum_{i=1}^n T_l(Y_i)\Big\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i)\Big\}\frac{f_0(Y_i)-\widehat{f}_0(Y_i)}{\widehat{f}_0(Y_i)f_0(Y_i)}\\
&+\frac{1}{n}\sum_{i=1}^n \left\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i)\right\}\frac{T_l(Y_i)}{f_0(Y_i)} + \frac{1}{n}\sum_{i=1}^n T_k(Y_i)T_l(Y_i)\frac{f_0(Y_i)-\widehat{f}_0(Y_i)}{\widehat{f}_0(Y_i)}\\
:=&D_1 + D_2 + D_3 + D_4 \qquad \text{(S3.25)}
\end{aligned}$$

Next, we examine the D_i terms one by one.

Using Lemma A.1 of Zhu, Wang and Zhu (2012), we can derive that $n^{-1}\sum_{i=1}^n \left\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i)\right\}^2 = O_p(h^m)$. In addition, Lemma A.1 of Zhu and Zhu (2007) indicates that $n^{-1}\sum_{i=1}^n\{\widehat{f}_0(Y_i) - f_0(Y_i)\}^2 = O_p(h^m)$. Then, the similar arguments to those for (S3.21) yield that $n^{-1}\sum_{i=1}^n \big\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i)\big\}\{\widehat{f}_0(Y_i) - f_0(Y_i)\} = O_p(h^m)$, which together with similar arguments to those used in T_7 show that

$$\begin{aligned} D_2 =& \frac{1}{n}\sum_{i=1}^{n} T_l(Y_i)\Big\{\widehat{S}_k^*(Y_i) - T_k(Y_i)f_0(Y_i)\Big\}\frac{f_0(Y_i) - \widehat{f}_0(Y_i)}{f_0^2(Y_i)} + o_p(n^{-1/2}) \\ =& o_p(n^{-1/2}) \end{aligned} \tag{S3.26}$$

as $nh^{2m} \to 0$.

For D_3, using the standard U-statistics theory, it follows that

$$\begin{aligned} D_3 =& \frac{1}{n}\sum_{i=1}^{n} \big\{\delta_{ki}X_{ki} + (1-\delta_{ki})M_k(\Gamma_k^T\mathbf{V}_i)\big\}T_l(Y_i) - \frac{1}{n}\sum_{i=1}^{n} T_k(Y_i)T_l(Y_i) \\ & + \frac{1}{n}\sum_{i=1}^{n} E\Big\{\big[\delta_k X_k + (1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]\Big|Y = Y_i\Big\}T_l(Y_i) \\ & - E\Big\{\big[\delta_k X_k + (1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]T_l(Y)\Big\} + o_p(n^{-1/2}) \end{aligned} \tag{S3.27}$$

For D_4, using similar arguments to those used in T_7 and the standard U-statistics theory, we have

$$\begin{aligned} D_4 =& \frac{1}{n}\sum_{i=1}^{n} T_k(Y_i)T_l(Y_i)\frac{f_0(Y_i) - \widehat{f}_0(Y_i)}{f_0(Y_i)} + o_p(n^{-1/2}) \\ =& -\frac{1}{n}\sum_{i=1}^{n} T_k(Y_i)T_l(Y_i) + E\big\{T_k(Y)T_l(Y)\big\} + o_p(n^{-1/2}) \end{aligned} \tag{S3.28}$$

Therefore, using (S3.25)–(S3.28), it follows that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_k^*(Y_i)T_l(Y_i)\\
=&\frac{1}{n}\sum_{i=1}^{n}\left\{\delta_{ki}X_{ki}+(1-\delta_{ki})M_k(\Gamma_k^T\mathbf{V}_i)\right\}T_l(Y_i)-\frac{1}{n}\sum_{i=1}^{n}T_k(Y_i)T_l(Y_i)\\
&+\frac{1}{n}\sum_{i=1}^{n}E\Big\{\big[\delta_kX_k+(1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]\Big|Y=Y_i\Big\}T_l(Y_i) \qquad \text{(S3.29)}\\
&-E\Big\{\big[\delta_kX_k+(1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]T_l(Y)\Big\}+E\big\{T_k(Y)T_l(Y)\big\}+o_p(n^{-1/2})
\end{aligned}
$$

Step 2 : We examine $n^{-1}\sum_{i=1}^n\big\{\widehat{T}_k^*(Y_i)-\widehat{T}_k(Y_i)\big\}T_l(Y_i)$, which can be written as

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{T}_k^*(Y_i)-\widehat{T}_k(Y_i)\big\}T_l(Y_i)\\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})\big\{M_k(\Gamma_k^T\mathbf{V}_j)-\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_j)\big\}T_l(Y_i)}{\widehat{f}_0(Y_i)}\\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})\big\{M_k(\Gamma_k^T\mathbf{V}_j)-\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_j)\big\}T_l(Y_i)}{f_0(Y_i)}+o_p(n^{-1/2})\\
=&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})M_k(\Gamma_k^T\mathbf{V}_j)T_l(Y_i)}{f_0(Y_i)}\\
&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})\big\{\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}_j)-\widehat{M}_k(\Gamma_k^T\mathbf{V}_j)\big\}T_l(Y_i)}{f_0(Y_i)}\\
&-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})\widehat{M}_k(\Gamma_k^T\mathbf{V}_j)T_l(Y_i)}{f_0(Y_i)}+o_p(n^{-1/2})\\
:=&\ M_1-M_2-M_3+o_p(n^{-1/2}) \qquad \text{(S3.30)}
\end{aligned}
$$

where the second equation holds owing to similar arguments to those used in T_7 and the strong consistency of $\widehat{f}_0(\cdot)$. Next, we check the M_i terms one by one.

For M_1, the standard U-statistics theory shows that

$$\begin{aligned} M_1 =& \frac{1}{n}\sum_{i=1}^{n}\Big\{(1-\delta_{ki})T_l(Y_i)M_k(\Gamma_k^T\mathbf{V}_i) - E\big[(1-\delta_k)T_l(Y)M_k(\Gamma_k^T\mathbf{V})\big]\Big\} \\ &+ \frac{1}{n}\sum_{i=1}^{n} E\big[(1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big|Y=Y_i\big]T_l(Y_i) + o_p(n^{-1/2}) \end{aligned} \tag{S3.31}$$

For M_2, with $\| \widehat{\Gamma}_k - \Gamma_k \| = O_p(n^{-1/2})$, Lemmas 1–2 of Li, Zhu and Zhu (2011) imply that

$$\begin{aligned} & \sup_{\|\widehat{\Gamma}_k-\Gamma_k\|\le Cn^{-1/2}} \sup_{\mathbf{V}\in R^q} \Big|\{\widehat{M}_k(\widehat{\Gamma}_k^T\mathbf{V}) - \widehat{M}_k(\Gamma_k^T\mathbf{V})\} - \{M_k(\widehat{\Gamma}_k^T\mathbf{V}) - M_k(\Gamma_k^T\mathbf{V})\}\Big| \\ =\ & O(h^m + n^{-1}h^{-(r_k+1)}\log n)\ a.s \end{aligned}$$

which together with the Taylor expansion yields that

$$\begin{aligned} M_2 &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})T_l(Y_i)}{f_0(Y_i)} \times \big\{M_k(\widehat{\Gamma}_k^T\mathbf{V}_j) - M_k(\Gamma_k^T\mathbf{V}_j)\big\} + o_p(n^{-1/2}) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})T_l(Y_i)}{f_0(Y_i)}\big\{M_k^{(1)}(\Gamma_k^T\mathbf{V}_j)\otimes\mathbf{V}_j\big\}^T\big\{\mathrm{vec}(\widehat{\Gamma}_k) - \mathrm{vec}(\Gamma_k)\big\} \\ &\quad + o_p(n^{-1/2}) \end{aligned}$$

as $nh^{2m} \to 0$ and $nh^{2(r_k+1)}/(\log n)^2 \to \infty$. Furthermore, using the standard U-statistic theory, we have

$$
\begin{aligned}
M_2 = &\Big\{\frac{1}{n}\sum_{j=1}^{n}(1-\delta_{kj})T_l(Y_j)\big\{M_k^{(1)}(\Gamma_k^T\mathbf{V}_j)\otimes\mathbf{V}_j\big\}^T \\
&+\frac{1}{n}\sum_{j=1}^{n}E\big[(1-\delta_k)T_l(Y)M_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}|Y=Y_j\big]^T \\
&-E\big[(1-\delta_k)T_l(Y)M_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}\big]^T\Big\}\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}+o_p(n^{-1/2}) \\
= &E\big[(1-\delta_k)T_l(Y)M_k^{(1)}(\Gamma_k^T\mathbf{V})\otimes\mathbf{V}\big]^T\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}+o_p(n^{-1/2}) \\
= &E\big[(1-\delta_k)T_l(Y)\partial\{M_k(\Gamma_k^T\mathbf{V})\}/\partial\{\mathrm{vec}(\Gamma_k)\}\big]^T\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}+o_p(n^{-1/2}) \\
= &E\big[(1-\delta_k)T_l(Y)\partial\{E(X_k|\mathbf{V})\}/\partial\{\mathrm{vec}(\Gamma_k)\}\big]^T\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}+o_p(n^{-1/2}) \\
= &E\big[\partial\{E[(1-\delta_k)T_l(Y)X_k|\mathbf{V},\delta_k]\}/\partial\{\mathrm{vec}(\Gamma_k)\}\big]^T\big\{\mathrm{vec}(\widehat{\Gamma}_k)-\mathrm{vec}(\Gamma_k)\big\}+o_p(n^{-1/2}) \\
= &o_p(n^{-1/2}) \qquad\qquad \text{(S3.32)}
\end{aligned}
$$

where the second equation holds owing to the law of large numbers, the fifth equation holds for the MAR assumption, and the last equation holds because of the fact $\partial\{E[(1-\delta_k)T_l(Y)X_k|\mathbf{V},\delta_k]\}/\partial\{\mathrm{vec}(\Gamma_k)\} = 0$ for any $\mathbf{V}\in R^q$.

For M_3, observe that

$$
\begin{aligned}
M_3 = &\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})T_l(Y_i)\widehat{G}_k(\Gamma_k^T\mathbf{V}_j)}{f_0(Y_i)\widehat{g}_k(\Gamma_k^T\mathbf{V}_j)} \\
= &\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j-Y_i)(1-\delta_{kj})T_l(Y_i)}{f_0(Y_i)}\big\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_j)-G_k(\Gamma_k^T\mathbf{V}_j)+G_k(\Gamma_k^T\mathbf{V}_j)\big\} \\
&\times\Big\{\frac{g_k(\Gamma_k^T\mathbf{V}_j)-\widehat{g}_k(\Gamma_k^T\mathbf{V}_j)}{\widehat{g}_k(\Gamma_k^T\mathbf{V}_j)g_k(\Gamma_k^T\mathbf{V}_j)}+\frac{1}{g_k(\Gamma_k^T\mathbf{V}_j)}\Big\}
\end{aligned}
$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j - Y_i)(1-\delta_{kj})T_l(Y_i)}{f_0(Y_i)g_k^2(\Gamma_k^T\mathbf{V}_j)}\big\{\widehat{G}_k(\Gamma_k^T\mathbf{V}_j) - G_k(\Gamma_k^T\mathbf{V}_j)\big\}\big\{g_k(\Gamma_k^T\mathbf{V}_j) - \widehat{g}_k(\Gamma_k^T\mathbf{V}_j)\big\}$$

$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j - Y_i)(1-\delta_{kj})T_l(Y_i)G_k(\Gamma_k^T\mathbf{V}_j)}{f_0(Y_i)g_k^2(\Gamma_k^T\mathbf{V}_j)}\big\{g_k(\Gamma_k^T\mathbf{V}_j) - \widehat{g}_k(\Gamma_k^T\mathbf{V}_j)\big\}$$

$$+ \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{K_h(Y_j - Y_i)(1-\delta_{kj})T_l(Y_i)\widehat{G}_k(\Gamma_k^T\mathbf{V}_j)}{f_0(Y_i)g_k(\Gamma_k^T\mathbf{V}_j)} + o_p(n^{-1/2})$$

$$:= M_{31} + M_{32} + M_{33} + o_p(n^{-1/2}) \tag{S3.33}$$

where the third equation holds by similar arguments to those used in T_7. The expression (S3.21) directly implies that

$$M_{31} = o_p(n^{-1/2}) \tag{S3.34}$$

as $nh^{2m} \to 0$. In addition, the standard U-statistics theory shows that

$$M_{32} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ki}M_k(\Gamma_k^T\mathbf{V}_i)}{\pi_k(\Gamma_k^T\mathbf{V}_i)}E\big[(1-\delta_k)T_l(Y)\big|\Gamma_k^T\mathbf{V} = \Gamma_k^T\mathbf{V}_i\big] + o_p(n^{-1/2}) \tag{S3.35}$$

and

$$\begin{aligned} M_{33} =& \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ki}X_{ki}}{\pi_k(\Gamma_k^T\mathbf{V}_i)}E\big[(1-\delta_k)T_l(Y)\big|\Gamma_k^T\mathbf{V} = \Gamma_k^T\mathbf{V}_i\big] \\ &+ \frac{1}{n}\sum_{i=1}^{n}\Big\{(1-\delta_{ki})T_l(Y_i)M_k(\Gamma_k^T\mathbf{V}_i) - E\big[(1-\delta_k)T_l(Y)M_k(\Gamma_k^T\mathbf{V})\big]\Big\} \\ &+ \frac{1}{n}\sum_{i=1}^{n}E\big[(1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big|Y = Y_i\big]T_l(Y_i) + o_p(n^{-1/2}) \end{aligned} \tag{S3.36}$$

Then, the results (S3.33)–(S3.36) jointly yield that

$$
\begin{aligned}
M_3 =& \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_{ki}\big[X_{ki} - M_k(\Gamma_k^T \mathbf{V}_i)\big]}{\pi_k(\Gamma_k^T\mathbf{V}_i)} E\big[(1-\delta_k)T_l(Y)\big|\Gamma_k^T\mathbf{V} = \Gamma_k^T\mathbf{V}_i\big] \\
&+ \frac{1}{n}\sum_{i=1}^{n}\Big\{(1-\delta_{ki})T_l(Y_i)M_k(\Gamma_k^T\mathbf{V}_i) - E\big[(1-\delta_k)T_l(Y)M_k(\Gamma_k^T\mathbf{V})\big]\Big\} \\
&+ \frac{1}{n}\sum_{i=1}^{n} E\big[(1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big|Y=Y_i\big]T_l(Y_i) + o_p(n^{-1/2}) \qquad \text{(S3.37)}
\end{aligned}
$$

Therefore, using the results (S3.30)–(S3.32) and (S3.37), it follows that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{T}_k^*(Y_i) - \widehat{T}_k(Y_i)\big\}T_l(Y_i) \\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ki}\big[M_k(\Gamma_k^T\mathbf{V}_i) - X_{ki}\big]}{\pi_k(\Gamma_k^T\mathbf{V}_i)} E\big[(1-\delta_k)T_l(Y)\big|\Gamma_k^T\mathbf{V} = \Gamma_k^T\mathbf{V}_i\big] + o_p(n^{-1/2})
\end{aligned}
\qquad \text{(S3.38)}
$$

Step 3 : By combining the results in (S3.29) and (S3.38), we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_k(Y_i)T_l(Y_i) \\
=&\frac{1}{n}\sum_{i=1}^{n}\widehat{T}_k^*(Y_i)T_l(Y_i) - \frac{1}{n}\sum_{i=1}^{n}\big\{\widehat{T}_k^*(Y_i) - \widehat{T}_k(Y_i)\big\}T_l(Y_i) \\
=&\frac{1}{n}\sum_{i=1}^{n}\big\{\delta_{ki}X_{ki} + (1-\delta_{ki})M_k(\Gamma_k^T\mathbf{V}_i)\big\}T_l(Y_i) - \frac{1}{n}\sum_{i=1}^{n}T_k(Y_i)T_l(Y_i) \\
&+ \frac{1}{n}\sum_{i=1}^{n}E\Big\{\big[\delta_kX_k + (1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]\Big|Y=Y_i\Big\}T_l(Y_i) \qquad \text{(S3.39)} \\
&+ \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ki}\big[X_{ki} - M_k(\Gamma_k^T\mathbf{V}_i)\big]}{\pi_k(\Gamma_k^T\mathbf{V}_i)} E\big[(1-\delta_k)T_l(Y)\big|\Gamma_k^T\mathbf{V} = \Gamma_k^T\mathbf{V}_i\big] \\
&- E\Big\{\big[\delta_kX_k + (1-\delta_k)M_k(\Gamma_k^T\mathbf{V})\big]T_l(Y)\Big\} + E\big\{T_k(Y)T_l(Y)\big\} + o_p(n^{-1/2})
\end{aligned}
$$

Finally, by simple algebraic calculations based on the results (S3.24) and (S3.39), we complete the proof of Lemma A.2 (i). □

References

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics 30*, pp. 455-474.

Li, Q. and Racine, J. S. (2007). *Nonparametric econometrics : theory and practice.* Princeton University Press.

Li, L., Zhu, L. and Zhu, L. (2011). Inference on the primary parameter of interest with the aid of dimension reduction estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*, pp. 59-80.

Serfling, R. J. (2009). *Approximation theorems of mathematical statistics, Vol. 162.* John Wiley & Sons.

Wand, M. P. and Jones, M. C. (1994)). *Kernel Smoothing.* Crc Press.

Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimate of sliced inverse regression. *The Annals of Statistics 24*, pp. 1053-1068.

Zhu, L. P. and Zhu, L. X. (2007). On kernel method for sliced average variance estimation. *Journal of Multivariate Analysis 98*, pp. 970-991.

Zhu, L. P., Wang, T. and Zhu, L. X. (2012). Sufficient dimension reduction in regressions with missing predictors. *Statistica Sinica 22*, pp. 1611-1637.