New HSIC-based tests for independence between two stationary multivariate time series

Guochang Wang, Wai Keung Li and Ke Zhu

Jinan University and The University of Hong Kong

Supplementary Material

This supplementary material give some additional simulation studies in Section S1 and the proofs of lemmas and theorems in Sections S2 and S3.

S1 Additional simulation studies

To see the impact of the kernel functions k and l, we examine the performance of our HSIC-based test statistics when k and l are chosen as inverse multi-quadratics kernels with $\alpha = \beta = 1$.

Tables S.1-S.2 report the sizes and power of all examined HSIC-based tests. Compared with the results in Tables 1-2, the results in Tables S.1-S.2 imply that similar performance of our HSIC-based tests retains for the two different choices of kernels in most cases, but some differences may exist in some cases. This is consistent with the findings in Gretton et al. (2009).

Table S.1: Empirical sizes and power (×100) of all HSIC-based tests based on the models in (5.1), with k and l being chosen as inverse multi-quadratics kernels

	EGP 1						EGP 2						EGP 3					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	1.5	5.7	12.1	1.2	5.4	11.4	52.6	74.9	83.9	92.1	97.1	98.7	79.5	94.2	99.5	100	100	100
$S_{1n}(3)$	1.5	6.2	12.7	1.0	6.4	11.0	1.0	6.5	11.3	1.0	6.0	10.8	0.3	2.7	7.5	1.0	3.6	9.2
$S_{2n}(3)$	1.1	5.7	11.7	1.0	6.0	11.8	0.7	5.4	12.0	1.2	6.4	12.8	0.3	2.9	7.8	0.9	3.8	7.4
$J_{1n}(3)$	1.5	7.2	15.2	0.6	5.4	13.3	21.2	50.4	65.6	61.8	83.2	91.3	12.6	43.0	68.8	80.5	96.2	98.7
$J_{1n}(6)$	0.5	6.7	17.1	1.1	5.7	11.7	10.8	34.7	55.0	41.4	69.1	81.3	1.0	13.0	35.6	39.8	76.7	88.8
$J_{2n}(3)$	0.9	6.2	13.1	1.3	6.7	14.1	21.2	47.1	63.2	62.8	83.4	90.3	12.0	42.5	67.3	79.0	94.4	97.9
$J_{2n}(6)$	0.8	5.3	15.1	0.9	5.7	13.9	9.7	34.0	52.6	41.6	71.6.7	82.3	0.9	14.8	35.6	36.7	75.5	88.4

	EGP 4						EGP 5						EGP 6					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.4	3.6	8.3	0.5	3.5	7.5	20.0	44.6	62.1	78.9	93.8	97.3	41.2	66.1	79.1	36.8	64.3	76.3
$S_{1n}(3)$	0.5	2.7	8.5	0.9	4.0	8.9	0.4	3.0	7.7	0.6	5.0	10.5	0.4	3.0	7.9	0.5	3.1	7.8
$S_{2n}(3)$	73.9	91.9	96.4	99.3	100	100	0.4	3.8	7.4	0.7	4.9	11.1	0.3	4.4	8.2	0.4	3.1	7.6
$J_{1n}(3)$	0.1	0.7	3.3	0.2	1.6	5.0	0.8	3.8	7.4	30.2	62.4	77.9	2.5	16.6	33.0	7.6	25.3	41.9
$J_{1n}(6)$	0.0	0.0	3.3	0.0	0.6	2.7	0.0	1.1	8.2	10.8	35.4	55.2	0.0	2.8	11.3	1.7	12.4	25.5
$J_{2n}(3)$	10.5	40.6	62.6	75.0	94.2	97.4	0.7	7.5	19.8	30.4	61.3	76.2	3.0	14.9	33.6	5.6	23.6	38.8
$J_{2n}(6)$	0.6	11.9	32.2	37.1	70.8	87.1	0.1	1.7	7.4	10.2	35.6	55.2	0.0	3.5	12.0	1.8	10.3	23.4

Table S.2: Empirical sizes and power (×100) of all HSIC-based tests based on the models in (5.2), with k and l being chosen as inverse multi-quadratics kernels

	EGP 1						EGP 2						EGP 3					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.4	4.1	8.0	0.6	4.8	11.0	100	100	100	100	100	100	99.9	100	100	100	100	100
$S_{1n}(3)$	0.9	4.2	9.0	0.6	5.1	9.3	1.0	5.1	10.7	0.6	5.30	10.2	0.8	3.9	8.2	0.1	2.1	5.7
$S_{2n}(3)$	0.2	2.7	8.2	0.7	5.0	9.5	0.8	4.3	9.6	0.7	4.5	9.6	0.5	3.6	7.4	0.3	3.2	7.4
$J_{1n}(3)$	0.5	2.4	6.3	1.3	4.9	10.4	98.8	99.9	100	100	100	100	99.0	100	100	99.9	100	100
$J_{1n}(6)$	0.3	2.7	6.9	0.6	3.9	9.2	87.1	98.3	99.7	99.9	100	100	55.6	85.6	95.0	95.3	99.6	100
$J_{2n}(3)$	0.4	3.3	8.3	0.9	4.4	9.9	98.1	99.7	99.9	100	100	100	89.6	97.8	99.6	99.8	100	100
$J_{2n}(6)$	0.3	2.4	7.1	0.6	3.5	9.3	87.3	97.7	99.3	99.5	99.9	100.0	55.6	84.4	94.3	96.0	99.5	99.8

	EGP 4						EGP 5						EGP 6					
	$n = 100$			$n = 200$			$n = 100$			$n = 200$			$n = 100$			$n = 200$		
Tests	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
$S_{1n}(0)$	0.5	2.9	6.4	0.4	3.4	7.2	65.5	87.3	93.4	89.4	97.2	98.4	90.6	97.6	98.4	98.9	99.7	99.9
$S_{1n}(3)$	0.3	3.1	8.2	0.6	2.7	6.1	0.6	3.0	7.5	1.1	4.2	8.7	0.7	3.1	8.2	0.9	4.3	10.3
$S_{2n}(3)$	99.9	99.9	100	100	100	100	0.4	3.0	6.7	0.5	3.9	9.0	0.5	3.2	6.8	0.3	3.3	5.6
$J_{1n}(3)$	0.1	1.6	4.4	0.1	1.7	4.0	14.8	38.8	55.6	38.5	67.0	80.9	42.4	71.6	82.80	77.6	92.9	96.9
$J_{1n}(6)$	0.1	0.7	2.9	0.0	0.5	2.7	2.6	16.4	32.9	11.3	37.5	57.7	10.2	39.2	61.4	42.3	74.3	86.5
$J_{2n}(3)$	87.3	97.7	98.7	99.8	100	100	13.3	38.1	56.2	38.1	68.6	81.6	42.0	71.0	83.6	79.1	92.5	96.5
$J_{2n}(6)$	51.4	83.5	91.4	94.5	99.5	99.9	1.9	16.2	32.0	11.3	39.6	58.8	10.9	38.1	59.1	42.0	74.5	86.4

S2 Proofs of Theorems

This section provides the proofs of all theorems. To facilitate it, the results of V-statistics are needed below, and they can be found in Hoeffding (1948) and Lee (1990) for the i.i.d. case and Yoshihara (1976) and Denker and Keller (1983) for the mixing case.

Proof of Theorem 3.1. (i) By Lemmas 3.1 and S3.1,

$$N[S_{1n}(m)] = Z_{1n}(m) + o_p(1),$$

where

$$\begin{aligned} Z_{1n}(m) := {} & N[S_{1n}^{(0)}(m)] + \zeta_{1n}^T[NS_{1n}^{(11)}(m)] + \zeta_{2n}^T[NS_{1n}^{(12)}(m)] \\ & + \frac{1}{2}\zeta_{1n}^T[NS_{1n}^{(21)}(m)]\zeta_{1n} + \frac{1}{2}\zeta_{2n}^T[NS_{1n}^{(22)}(m)]\zeta_{2n} \\ & + [\sqrt{N}\zeta_{1n}]^T S_{1n}^{(23)}(m)[\sqrt{N}\zeta_{2n}]. \end{aligned}$$

For $a, b = 1, 2$, $S_{1n}^{(ab)}(m)$ is a degenerate V-statistic of order 1 by Lemma 3.2(ii), and hence $NS_{1n}^{(ab)}(m) = O_p(1)$. By Assumption 2.3, it follows that

$$\begin{aligned} Z_{1n}(m) &= N[S_{1n}^{(0)}(m)] + [\sqrt{N}\zeta_{1n}]^T S_{1n}^{(23)}(m)[\sqrt{N}\zeta_{2n}] + o_p(1) \\ &= N[S_{1n}^{(0)}(m)] + [\sqrt{N}\zeta_{1n}]^T \Lambda_m^{(23)}[\sqrt{N}\zeta_{2n}] + o_p(1), \end{aligned}$$

where the last equality holds by the law of large numbers for V-statistics. Hence, $Z_{1n}(m) \to_d \chi_m$ as $n \to \infty$ by (3.9), Lemma 3.3, and the continuous mapping theorem. This completes the proof of (i).

(ii) It follows by a similar argument as for (i). □

Proof of Theorem 3.2. (i) By Lemmas 3.1 and S3.2, we have

$$\sqrt{N}\left[S_{1n}(m)-\Lambda_m^{(0)}\right]=\overline{Z}_{1n}(m)+o_p(1), \tag{S2.1}$$

where $\Lambda_m^{(0)}=E[h_m^{(0)}(\eta_1^{(m)},\eta_2^{(m)},\eta_3^{(m)},\eta_4^{(m)})]>0$ and

$$\begin{aligned}\overline{Z}_{1n}(m):=&\sqrt{N}\left[S_{1n}^{(0)}(m)-\Lambda_m^{(0)}\right]+[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(11)}(m)+[\sqrt{N}\zeta_{2n}]^TS_{1n}^{(12)}(m)\\&+\frac{1}{2\sqrt{N}}\Big\{[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(21)}(m)[\sqrt{N}\zeta_{1n}]+[\sqrt{N}\zeta_{2n}]^TS_{1n}^{(22)}(m)[\sqrt{N}\zeta_{2n}]\\&+2[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(23)}(m)[\sqrt{N}\zeta_{2n}]\Big\}.\end{aligned}$$

First, since $S_{1n}^{(0)}(m)$ is a non-degenerate V-statistic under $H_1^{(m)}$, part (c) of Theorem 2 in Denker and Keller (1983) implies that

$$\sqrt{N}\left[S_{1n}^{(0)}(m)-\Lambda_m^{(0)}\right]=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}h_{1m}^{(0)}(\eta_i^{(m)})+o_p(1)=O_p(1), \tag{S2.2}$$

where $h_{1m}^{(0)}(x_1)=E[h_m^{(0)}(x_1,\eta_2^{(m)},\eta_3^{(m)},\eta_4^{(m)})]-\Lambda_m^{(0)}$. Second, by the law of large numbers for V-statistics and Assumption 2.3, it follows that

$$[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(11)}(m)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_{1i}\right]^T\Lambda_m^{(11)}+o_p(1)=O_p(1), \tag{S2.3}$$

$$[\sqrt{N}\zeta_{2n}]^TS_{1n}^{(12)}(m)=\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\pi_{2i}\right]^T\Lambda_m^{(12)}+o_p(1)=O_p(1), \tag{S2.4}$$

$$\begin{aligned}&\frac{1}{2\sqrt{N}}\Big\{[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(21)}(m)[\sqrt{N}\zeta_{1n}]+[\sqrt{N}\zeta_{2n}]^TS_{1n}^{(22)}(m)[\sqrt{N}\zeta_{2n}]\\&+2[\sqrt{N}\zeta_{1n}]^TS_{1n}^{(23)}(m)[\sqrt{N}\zeta_{2n}]\Big\}=o_p(1),\end{aligned} \tag{S2.5}$$

where $\Lambda_m^{(1s)} = E[h_m^{(1s)}(\eta_1^{(m)}, \eta_2^{(m)}, \eta_3^{(m)}, \eta_4^{(m)})]$ for $s = 1, 2$. By (S2.2)-(S2.5), $\overline{Z}_{1n}(m) = O_p(1)$, which together with (S2.1) implies that $n[S_{1n}(m)] \to \infty$ in probability as $n \to \infty$. This completes the proof of (i).

(ii) It follows by a similar argument as for (i). $\square$

Proof of Theorem 4.1. (i) By Assumptions 4.1 and 4.2(i), $\sqrt{N}\zeta_{sn}^* = O_p^*(1)$. Then, by (4.1)-(4.2), Assumption 4.2, and a similar argument as for Lemmas 3.2(ii)-(iii) and S3.1, we can show that

$$S_{1n}^{**}(m) = \sum_{j=1}^{\infty} \lambda_{jm}^* \left[\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Phi_{jm}^*(\widehat{\eta}_i^{(m*)}) \right] + [\sqrt{N}\zeta_{1n}^{*T}]\Lambda_m^{(23*)}[\sqrt{N}\zeta_{2n}^*] + o_p^*(1) = O_p^*(1). \tag{S2.6}$$

This completes the proof of (i).

(ii) It follows by a similar argument as for (i).

(iii) Let $\mathcal{T}_{1i}^* = \left(\left(\Phi_{jm}^*(\widehat{\eta}_i^{(m*)}) \right)_{j \geq 1, 0 \leq m \leq M} \right)^T$, $\mathcal{T}_{2i}^* = \left((\pi_{si}^{*T})_{1 \leq s \leq 2} \right)^T$, and

$$\mathcal{T}_n^* = \Big(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathcal{T}_{1i}^{*T}, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{T}_{2i}^{*T} \Big)^T,$$

where π_{si}^* is defined as in Assumption 4.1. Also, let $\mathcal{T}_i^* = (\mathcal{T}_{1i}^{*T}, \mathcal{T}_{2i}^{*T})^T$. As for Lemma 3.3, it is not hard to see that conditional on ϖ_n,

$$\mathcal{T}_n^* \to_d \mathcal{T}^* \tag{S2.7}$$

in probability as $n \to \infty$, where $\mathcal{T}^*$ is a multivariate normal distribution with covariance matrix $\overline{\mathcal{T}}^*$, and $\overline{\mathcal{T}}^* = \lim_{n \to \infty} E^*(\mathcal{T}_1^* \mathcal{T}_1^{*T}) = E(\mathcal{T}_1 \mathcal{T}_1^T) = \overline{\mathcal{T}}$

in probability by Assumption 4.2.

Next, by Lemma S3.4(i) and Corollary XI.9.4(a) in Dunford and Schwartz (1963, p.1090), we can get

$$|\lambda^*_{jm} - \lambda_{jm}| = o(1). \tag{S2.8}$$

Hence, the conclusion holds by (S2.6)-(S2.8), Lemma S3.4(ii), and the continuous mapping theorem. This completes the proof of (iii).

(iv) It follows by a similar argument as for (iii). □

S3 The remaining proofs

Proof of Lemma 3.1. Denote $\widehat{z}_{ijqr} = \widehat{k}_{ij}\widehat{l}_{qr}$. By Taylor's expansion,

$$\begin{aligned}\widehat{z}_{ijqr} &= z^{(0)}_{ijqr} + (\widehat{\eta}_{ijqr} - \eta_{ijqr})^T W_{ijqr} \\ &\quad + \frac{1}{2}(\widehat{\eta}_{ijqr} - \eta_{ijqr})^T H^{\dagger}_{ijqr}(\widehat{\eta}_{ijqr} - \eta_{ijqr}) \\ &= z^{(0)}_{ijqr} + (\widehat{\eta}_{ijqr} - \eta_{ijqr})^T W_{ijqr} \\ &\quad + \frac{1}{2}(\widehat{\eta}_{ijqr} - \eta_{ijqr})^T H_{ijqr}(\widehat{\eta}_{ijqr} - \eta_{ijqr}) + R^{(1)}_{ijqr}, \end{aligned} \tag{S3.1}$$

where $z^{(0)}_{ijqr} = k_{ij}l_{qr}$, $\widehat{\eta}_{ijqr} = (\widehat{\eta}^T_{1i}, \widehat{\eta}^T_{1j}, \widehat{\eta}^T_{2q+m}, \widehat{\eta}^T_{2r+m})^T$, $\eta_{ijqr} = (\eta^T_{1i}, \eta^T_{1j}, \eta^T_{2q+m}, \eta^T_{2r+m})^T$, $W_{ijqr} = W(\eta_{ijqr})$, $H_{ijqr} = H(\eta_{ijqr})$, $H^{\dagger}_{ijqr} = H(\eta^{\dagger}_{ijqr})$, $\eta^{\dagger}_{ijqr}$ lies between η_{ijqr} and $\widehat{\eta}_{ijqr}$, and

$$R^{(1)}_{ijqr} = (\widehat{\eta}_{ijqr} - \eta_{ijqr})^T \left(H^{\dagger}_{ijqr} - H_{ijqr}\right)(\widehat{\eta}_{ijqr} - \eta_{ijqr}).$$

Here, $W : \mathcal{R}^{d_1} \times \mathcal{R}^{d_1} \times \mathcal{R}^{d_2} \times \mathcal{R}^{d_2} \to \mathcal{R}^{(2d_1+2d_2)\times 1}$ such that

$$W(u,u',v,v') =$$
$$\Big(k_x(u,u')^T l(v,v'), k_y(u,u')^T l(v,v'), k(u,u') l_x(v,v')^T, k(u,u') l_y(v,v')^T\Big)^T,$$

and $H : \mathcal{R}^{d_1} \times \mathcal{R}^{d_1} \times \mathcal{R}^{d_2} \times \mathcal{R}^{d_2} \to \mathcal{R}^{2d_1+2d_2} \times \mathcal{R}^{2d_1+2d_2}$ such that

$$H(u,u',v,v') =$$
$$\begin{pmatrix} k_{xx}(u,u')l(v,v') & k_{xy}(u,u')l(v,v') & k_x(u,u')l_x(v,v')^T & k_x(u,u')l_y(v,v')^T \\ * & k_{yy}(u,u')l(v,v') & k_y(u,u')l_x(v,v')^T & k_y(u,u')l_y(v,v')^T \\ * & * & k(u,u')l_{xx}(v,v') & k(u,u')l_{xy}(v,v') \\ * & * & * & k(u,u')l_{yy}(v,v') \end{pmatrix}$$

is a symmetric matrix.

Next, let $\theta = (\theta_1^T, \theta_2^T)^T$ and $\widehat{\theta}_n = (\widehat{\theta}_{1n}^T, \widehat{\theta}_{2n}^T)^T$, and denote

$$G_{ijqr}(\theta) = \Big(g_{1i}(\theta_1)^T, g_{1j}(\theta_1)^T, g_{2q+m}(\theta_2)^T, g_{2r+m}(\theta_2)^T\Big)^T,$$

where $g_{st}(\theta_s)$ is defined as in Assumption 2.2. By Taylor's expansion again, we have

$$\widehat{\eta}_{ijqr} - \eta_{ijqr} = \overline{R}_{ijqr}^{(2)} + \frac{\partial G_{ijqr}(\theta^\dagger)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0), \tag{S3.2}$$

where $\overline{R}_{ijqr}^{(2)} = (\widehat{R}_{1i}(\widehat{\theta}_{1n})^T, \widehat{R}_{1j}(\widehat{\theta}_{1n})^T, \widehat{R}_{2q+m}(\widehat{\theta}_{2n})^T, \widehat{R}_{2r+m}(\widehat{\theta}_{2n})^T)^T$, $\widehat{R}_{st}(\theta_s)$ is defined as in Assumption 2.4, and $\theta^\dagger$ lies between θ_0 and $\widehat{\theta}_n$. For the second term in (S3.2), we rewrite it as

$$\frac{\partial G_{ijqr}(\theta^\dagger)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0) = \overline{R}_{ijqr}^{(3)} + \frac{\partial G_{ijqr}(\theta_0)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0), \tag{S3.3}$$

where $\overline{R}^{(3)}_{ijqr} = \big[\frac{\partial G_{ijqr}(\theta^\dagger)}{\partial\theta^T} - \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta^T}\big](\widehat{\theta}_n - \theta_0)$.

Now, by (S3.1)-(S3.3), it follows that

$$\widehat{z}_{ijqr} = z^{(0)}_{ijqr} + (\widehat{\theta}_n - \theta_0)^T z^{(1)}_{ijqr} + \frac{1}{2}(\widehat{\theta}_n - \theta_0)^T z^{(2)}_{ijqr}(\widehat{\theta}_n - \theta_0) + R_{ijqr}, \quad \text{(S3.4)}$$

where $z^{(1)}_{ijqr} = \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta} W_{ijqr}$, $z^{(2)}_{ijqr} = \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta} H_{ijqr} \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta^T}$, and $R_{ijqr} = R^{(1)}_{ijqr} + R^{(2)}_{ijqr} + R^{(3)}_{ijqr} + R^{(4)}_{ijqr}$ with

$$\begin{aligned}
R^{(2)}_{ijqr} &= \left(\overline{R}^{(2)}_{ijqr} + \overline{R}^{(3)}_{ijqr}\right)^T W_{ijqr}, \\
R^{(3)}_{ijqr} &= \frac{1}{2}\left(\overline{R}^{(2)}_{ijqr} + \overline{R}^{(3)}_{ijqr}\right)^T H_{ijqr}\left(\overline{R}^{(2)}_{ijqr} + \overline{R}^{(3)}_{ijqr}\right), \\
R^{(4)}_{ijqr} &= (\widehat{\theta}_n - \theta_0)^T \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta} H_{ijqr}\left(\overline{R}^{(2)}_{ijqr} + \overline{R}^{(3)}_{ijqr}\right).
\end{aligned}$$

By (S3.4), it entails that

$$\begin{aligned}
S_{1n}(m) = S^{(0)}_{1n}(m) &+ (\widehat{\theta}_n - \theta_0)^T S^{(1)}_{1n}(m) + \frac{1}{2}(\widehat{\theta}_n - \theta_0)^T S^{(2)}_{1n}(m)(\widehat{\theta}_n - \theta_0) \\
&+ R_{1n}(m), \qquad \text{(S3.5)}
\end{aligned}$$

where

$$S^{(p)}_{1n}(m) = \frac{1}{N^2}\sum_{i,j} z^{(p)}_{ijij} + \frac{1}{N^4}\sum_{i,j,q,r} z^{(p)}_{ijqr} - \frac{2}{N^3}\sum_{i,j,q} z^{(p)}_{ijiq}$$

for $p \in \{0, 1, 2\}$, and

$$R_{1n}(m) = \frac{1}{N^2}\sum_{i,j} R_{ijij} + \frac{1}{N^4}\sum_{i,j,q,r} R_{ijqr} - \frac{2}{N^3}\sum_{i,j,q} R_{ijiq} \qquad \text{(S3.6)}$$

is the remainder term.

Furthermore, simple algebra shows that

$$(\widehat{\theta}_n - \theta_0)^T z^{(1)}_{ijqr} = \zeta^T_{1n}\overline{k}_{ij} l_{qr} + \zeta^T_{2n} k_{ij}\overline{l}_{qr}, \qquad \text{(S3.7)}$$

$$(\widehat{\theta}_n - \theta_0)^T z_{ijqr}^{(2)}(\widehat{\theta}_n - \theta_0) = \zeta_{1n}^T \breve{k}_{ij} l_{qr} \zeta_{1n} + \zeta_{2n}^T k_{ij} \breve{l}_{qr} \zeta_{2n} + \zeta_{1n}^T \left(2\overline{k}_{ij}\bar{l}_{qr}^T\right) \zeta_{2n}, \tag{S3.8}$$

where $\overline{k}_{ij}$, $\bar{l}_{ij}$, $\breve{k}_{ij}$, and $\breve{l}_{ij}$ are defined in (3.1)-(3.4), respectively. Finally, the conclusion holds by (S3.5) and (S3.7)-(S3.8). This completes the proof. □

Proof of Lemma 3.2. Without loss of generality, we only prove the results for $m = 0$, under which $N = n$, and $\eta_t^{(0)}$ and $\varsigma_t^{(0)}$ are denoted by $\eta_t := (\eta_{1t}, \eta_{2t})$ and $\varsigma_t := \left(\eta_{1t}, \frac{\partial g_{1t}(\theta_{10})}{\partial \theta_1}, \eta_{2t}, \frac{\partial g_{2t}(\theta_{20})}{\partial \theta_2}\right)$, respectively, for notational ease.

(i) Denote $x_1 = (x_{11}, x_{21})$ for $x_{11} \in \mathcal{R}^{d_1}$ and $x_{21} \in \mathcal{R}^{d_2}$. Then, we rewrite

$$\begin{aligned} h_0^{(0)}(x_1, \eta_2, \eta_3, \eta_4) &= \frac{1}{4!}\left[\sum_{t=1,(u,v,w)}^{(2,3,4)} z_{1,uvw}^{(0)}(x_1) + \sum_{u=1,(t,v,w)}^{(2,3,4)} z_{2,tvw}^{(0)}(x_1) \right. \\ &\qquad \left. + \sum_{v=1,(t,u,w)}^{(2,3,4)} z_{3,tuw}^{(0)}(x_1) + \sum_{w=1,(t,u,v)}^{(2,3,4)} z_{4,tuv}^{(0)}(x_1)\right] \\ &=: \frac{1}{4!}\left[\Delta_1^{(0)} + \Delta_2^{(0)} + \Delta_3^{(0)} + \Delta_4^{(0)}\right], \end{aligned}$$

where

$$z_{1,uvw}^{(0)}(x_1) = k(x_{11}, \eta_{1u})\left[l(x_{21}, \eta_{2u}) + l(\eta_{2v}, \eta_{2w}) - 2l(x_{21}, \eta_{2v})\right],$$

$$z_{2,tvw}^{(0)}(x_1) = k(\eta_{1t}, x_{11})\left[l(\eta_{2t}, x_{21}) + l(\eta_{2v}, \eta_{2w}) - 2l(\eta_{2t}, \eta_{2v})\right],$$

$$z_{3,tuw}^{(0)}(x_1) = k(\eta_{1t}, \eta_{1u}) \left[l(\eta_{2t}, \eta_{2u}) + l(x_{21}, \eta_{2w}) - 2l(\eta_{2t}, x_{21}) \right],$$

$$z_{4,tuv}^{(0)}(x_1) = k(\eta_{1t}, \eta_{1u}) \left[l(\eta_{2t}, \eta_{2u}) + l(\eta_{2v}, x_{21}) - 2l(\eta_{2t}, \eta_{2v}) \right].$$

By the symmetry of k and l, the stationarity of η_{1t} and η_{2t}, and the independence of $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$ under H_0, simple algebra shows that

$$E\Delta_1^{(0)} = 6E\left[k(x_{11}, \eta_{11})\right] \times E\left[l(\eta_{21}, \eta_{22}) - l(x_{21}, \eta_{21})\right],$$

$$E\Delta_2^{(0)} = 6E\left[k(x_{11}, \eta_{11})\right] \times E\left[l(x_{21}, \eta_{21}) - l(\eta_{21}, \eta_{22})\right],$$

$$E\Delta_3^{(0)} = 6E\left[k(\eta_{11}, \eta_{12})\right] \times E\left[l(\eta_{21}, \eta_{22}) - l(x_{21}, \eta_{21})\right],$$

$$E\Delta_4^{(0)} = 6E\left[k(\eta_{11}, \eta_{12})\right] \times E\left[l(x_{21}, \eta_{21}) - l(\eta_{21}, \eta_{22})\right].$$

Hence, it follows that under H_0, $E[h_0^{(0)}(x_1, \eta_2, \eta_3, \eta_4)] = 0$ for all x_1. This completes the proof of (i).

(ii) We only consider the proof for the case that $a = b = 1$, since the proofs of other cases are similar. Denote $x_1 = (x_{11}, y_{11}, x_{21}, y_{21})$ for $x_{11} \in \mathcal{R}^{d_1}$, $y_{11} \in \mathcal{R}^{p_1 \times d_1}$, $x_{21} \in \mathcal{R}^{d_2}$, and $y_{21} \in \mathcal{R}^{p_2 \times d_2}$. Then, we rewrite

$$\begin{aligned} h_0^{(11)}(x_1, \varsigma_2, \varsigma_3, \varsigma_4) = \frac{1}{4!} \Bigg[& \sum_{t=1,(u,v,w)}^{(2,3,4)} z_{1,uvw}^{(11)}(x_1) + \sum_{u=1,(t,v,w)}^{(2,3,4)} z_{2,tvw}^{(11)}(x_1) \\ & + \sum_{v=1,(t,u,w)}^{(2,3,4)} z_{3,tuw}^{(11)}(x_1) + \sum_{w=1,(t,u,v)}^{(2,3,4)} z_{4,tuv}^{(11)}(x_1) \Bigg] \\ =: & \frac{1}{4!} \left[\Delta_1^{(11)} + \Delta_2^{(11)} + \Delta_3^{(11)} + \Delta_4^{(11)} \right], \end{aligned}$$

where

$$z_{1,uvw}^{(11)}(x_1) = \left[y_{11}k_x(x_{11},\eta_{1u}) + \frac{\partial g_{1u}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1u},x_{11})\right]$$

$$\times\left[l(x_{21},\eta_{2u}) + l(\eta_{2v},\eta_{2w}) - 2l(x_{21},\eta_{2v})\right],$$

$$z_{2,tvw}^{(11)}(x_1) = \left[\frac{\partial g_{1t}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1t},x_{11}) + y_{11}k_x(x_{11},\eta_{1t})\right]$$

$$\times\left[l(\eta_{2t},x_{21}) + l(\eta_{2v},\eta_{2w}) - 2l(\eta_{2t},\eta_{2v})\right],$$

$$z_{3,tuw}^{(11)}(x_1) = \left[\frac{\partial g_{1t}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1t},\eta_{1u}) + \frac{\partial g_{1u}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1u},\eta_{1t})\right]$$

$$\times\left[l(\eta_{2t},\eta_{2u}) + l(x_{21},\eta_{2w}) - 2l(\eta_{2t},x_{21})\right],$$

$$z_{4,tuv}^{(11)}(x_1) = \left[\frac{\partial g_{1t}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1t},\eta_{1u}) + \frac{\partial g_{1u}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1u},\eta_{1t})\right]$$

$$\times\left[l(\eta_{2t},\eta_{2u}) + l(\eta_{2v},x_{21}) - 2l(\eta_{2t},\eta_{2v})\right].$$

Here, we have used the fact that $k_y(c,d) = k_x(d,c)$ by the symmetry of k. By the stationarity of η_{1t} and η_{2t}, and the independence of $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$ under H_0, simple algebra shows that

$$E\Delta_1^{(11)} = -E\Delta_2^{(11)}$$

$$= \left\{y_{11}Ek_x(x_{11},\eta_{11}) + E\left[\frac{\partial g_{11}(\theta_{10})}{\partial\theta_1}k_x(\eta_{11},x_{11})\right]\right\}$$

$$\times\left[4El(\eta_{21},\eta_{22}) + 2El(\eta_{21},\eta_{23}) - 6El(x_{21},\eta_{21})\right],$$

$$E\Delta_3^{(11)} = -E\Delta_4^{(11)}$$

$$= 4E\left[\frac{\partial g_{11}(\theta_{10})}{\partial\theta_1}k_x(\eta_{11},\eta_{12}) + \frac{\partial g_{12}(\theta_{10})}{\partial\theta_1}k_x(\eta_{12},\eta_{11})\right]$$

$$
\begin{aligned}
&\times \left[El(\eta_{21}, \eta_{22}) - El(x_{21}, \eta_{21})\right] \\
&+ 2E\left[\frac{\partial g_{11}(\theta_{10})}{\partial \theta_1} k_x(\eta_{11}, \eta_{13}) + \frac{\partial g_{13}(\theta_{10})}{\partial \theta_1} k_x(\eta_{13}, \eta_{11})\right] \\
&\times \left[El(\eta_{21}, \eta_{23}) - El(x_{21}, \eta_{21})\right].
\end{aligned}
$$

Hence, it follows that under H_0, $E[h_0^{(11)}(x_1, \varsigma_2, \varsigma_3, \varsigma_4)] = 0$ for all x_1. This completes the proof of (ii).

(iii) Denote $x_1 = (x_{11}, y_{11}, x_{21}, y_{21})$ for $x_{11} \in \mathcal{R}^{d_1}$, $y_{11} \in \mathcal{R}^{p_1 \times d_1}$, $x_{21} \in \mathcal{R}^{d_2}$, and $y_{21} \in \mathcal{R}^{p_2 \times d_2}$. Then, we rewrite

$$
\begin{aligned}
h_0^{(23)}(x_1, \varsigma_2, \varsigma_3, \varsigma_4) &= \frac{1}{4!}\left[\sum_{t=1,(u,v,w)}^{(2,3,4)} z_{1,uvw}^{(23)}(x_1) + \sum_{u=1,(t,v,w)}^{(2,3,4)} z_{2,tvw}^{(23)}(x_1) \right. \\
&\quad \left. + \sum_{v=1,(t,u,w)}^{(2,3,4)} z_{3,tuw}^{(23)}(x_1) + \sum_{w=1,(t,u,v)}^{(2,3,4)} z_{4,tuv}^{(23)}(x_1)\right] \\
&=: \frac{1}{4!}\left[\Delta_1^{(23)} + \Delta_2^{(23)} + \Delta_3^{(23)} + \Delta_4^{(23)}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
z_{1,uvw}^{(23)}(x_1) &= \left[y_{11} k_x(x_{11}, \eta_{1u}) + \frac{\partial g_{1u}(\theta_{10})}{\partial \theta_1} k_x(\eta_{1u}, x_{11})\right] \\
&\quad \times \left[y_{21} l_x(x_{21}, \eta_{2u}) + \frac{\partial g_{2u}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2u}, x_{21}) + \frac{\partial g_{2v}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2v}, \eta_{2w}) \right. \\
&\quad \left. + \frac{\partial g_{2w}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2w}, \eta_{2v}) - 2y_{21} l_x(x_{21}, \eta_{2v}) - 2\frac{\partial g_{2v}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2v}, x_{21})\right], \\
z_{2,tvw}^{(23)}(x_1) &= \left[y_{11} k_x(x_{11}, \eta_{1t}) + \frac{\partial g_{1t}(\theta_{10})}{\partial \theta_1} k_x(\eta_{1t}, x_{11})\right] \\
&\quad \times \left[y_{21} l_x(x_{21}, \eta_{2t}) + \frac{\partial g_{2t}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2t}, x_{21}) + \frac{\partial g_{2v}(\theta_{20})}{\partial \theta_2} l_x(\eta_{2v}, \eta_{2w})\right.
\end{aligned}
$$

$$+\frac{\partial g_{2w}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2w},\eta_{2v})-2\frac{\partial g_{2t}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2t},\eta_{2v})-2\frac{\partial g_{2v}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2v},\eta_{2t})\Bigg],$$

$$\begin{aligned}z^{(23)}_{3,tuw}(x_1)=&\left[\frac{\partial g_{1t}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1t},\eta_{1u})+\frac{\partial g_{1u}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1u},\eta_{1t})\right]\\&\times\Bigg[y_{21}l_x(x_{21},\eta_{2w})+\frac{\partial g_{2t}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2t},\eta_{2u})+\frac{\partial g_{2u}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2u},\eta_{2t})\\&+\frac{\partial g_{2w}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2w},x_{21})-2y_{21}l_x(x_{21},\eta_{2t})-2\frac{\partial g_{2t}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2t},x_{21})\Bigg],\end{aligned}$$

$$\begin{aligned}z^{(23)}_{4,tuv}(x_1)=&\left[\frac{\partial g_{1t}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1t},\eta_{1u})+\frac{\partial g_{1u}(\theta_{10})}{\partial\theta_1}k_x(\eta_{1u},\eta_{1t})\right]\\&\times\Bigg[y_{21}l_x(x_{21},\eta_{2v})+\frac{\partial g_{2t}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2t},\eta_{2u})+\frac{\partial g_{2u}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2u},\eta_{2t})\\&+\frac{\partial g_{2v}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2v},x_{21})-2\frac{\partial g_{2t}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2t},\eta_{2v})-2\frac{\partial g_{2v}(\theta_{20})}{\partial\theta_2}l_x(\eta_{2v},\eta_{2t})\Bigg].\end{aligned}$$

By the stationarity of η_{1t} and η_{2t}, and the independence of $\{\eta_{1t}\}$ and $\{\eta_{2t}\}$ under H_0, simple algebra shows that

$$\begin{aligned}E\Delta^{(23)}_1=&-E\Delta^{(23)}_2\\=&\left\{y_{11}Ek_x(x_{11},\eta_{11})+E\left[\frac{\partial g_{11}(\theta_{10})}{\partial\theta_1}k_x(\eta_{11},x_{11})\right]\right\}\\&\times\Bigg\{-6y_{21}El_x(x_{21},\eta_{21})-6E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2}l_x(\eta_{21},x_{21})\right]\\&+4E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2}l_x(\eta_{21},\eta_{22})\right]+2E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2}l_x(\eta_{21},\eta_{23})\right]\\&+4E\left[\frac{\partial g_{22}(\theta_{20})}{\partial\theta_2}l_x(\eta_{22},\eta_{21})\right]+2E\left[\frac{\partial g_{23}(\theta_{20})}{\partial\theta_2}l_x(\eta_{23},\eta_{21})\right]\Bigg\},\end{aligned}$$

$$\begin{aligned}E\Delta^{(11)}_3=&-E\Delta^{(11)}_4+\Upsilon\\=&\,4E\left[\frac{\partial g_{11}(\theta_{10})}{\partial\theta_1}k_x(\eta_{11},\eta_{12})+\frac{\partial g_{12}(\theta_{10})}{\partial\theta_1}k_x(\eta_{12},\eta_{11})\right]\end{aligned}$$

$$\times \left\{ E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2} l_x(\eta_{21},\eta_{22})\right] - E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2} l_x(\eta_{21},x_{21})\right] + E\left[\frac{\partial g_{22}(\theta_{20})}{\partial\theta_2} l_x(\eta_{22},\eta_{21})\right] - y_{21} El_x(x_{21},\eta_{21}) \right\}$$
$$+ 2E\left[\frac{\partial g_{11}(\theta_{10})}{\partial\theta_1} k_x(\eta_{11},\eta_{13}) + \frac{\partial g_{13}(\theta_{10})}{\partial\theta_1} k_x(\eta_{13},\eta_{11})\right]$$
$$\times \left\{ E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2} l_x(\eta_{21},\eta_{23})\right] - E\left[\frac{\partial g_{21}(\theta_{20})}{\partial\theta_2} l_x(\eta_{21},x_{21})\right] + E\left[\frac{\partial g_{23}(\theta_{20})}{\partial\theta_2} l_x(\eta_{23},\eta_{21})\right] - y_{21} El_x(x_{21},\eta_{21}) \right\}.$$

Hence, it follows that under H_0, $E[h_0^{(23)}(x_1,\varsigma_2,\varsigma_3,\varsigma_4)] = \Upsilon$ for all x_1. This completes the proof of (iii). $\square$

Proof of Lemma 3.3. Let $\mathcal{F}_i = \sigma(\mathcal{F}_{1i},\mathcal{F}_{2i})$. Under H_0, it is not hard to see that $E(\mathcal{T}_{1i}|\mathcal{F}_{i-1}) = E(\mathcal{T}_{1i}) = 0$ by Lemma 3.2(i). Since $E(\mathcal{T}_{2i}|\mathcal{F}_{i-1}) = 0$ by Assumption 2.3, it follows that $E(\mathcal{T}_i|\mathcal{F}_{i-1}) = 0$. Moreover, by Assumptions 2.3 and 2.5, it is straightforward to see that $E\|\mathcal{T}_i\|^2 < \infty$. By the central limit theorem for martingale difference sequence (see Corollary 5.26 in White (2001)), it follows that $\mathcal{T}_n \to_d \mathcal{T}$ as $n\to\infty$, where $\mathcal{T}$ is a multivariate normal distribution with covariance matrix $\overline{\mathcal{T}} = \lim_{n\to\infty} var(\mathcal{T}_n) = E(\mathcal{T}_1\mathcal{T}_1^T)$. $\square$

Moreover, we introduce two lemmas to deal with the remainder term $R_{1n}(m)$ in Lemma 3.1.

Lemma S3.1. *Suppose Assumptions 2.1, 2.2(i) and 2.3-2.5 hold. Then,* *under H_0, $n\|R_{1n}(m)\| = o_p(1)$, where $R_{1n}(m)$ is defined as in (S3.6).*

Proof. As for the proof of Lemma 3.2, we only prove the result for $m = 0$. Rewrite $R_{1n}(0) = R_n^{(1)} + R_n^{(2)} + R_n^{(3)} + R_n^{(4)}$, where

$$R_n^{(d)} = \frac{1}{n^2}\sum_{i,j} R_{ijij}^{(d)} + \frac{1}{n^4}\sum_{i,j,q,r} R_{ijqr}^{(d)} - \frac{2}{n^3}\sum_{i,j,q} R_{ijiq}^{(d)}$$

for $d = 1, 2, 3, 4$, and $R_{ijqr}^{(d)}$ is defined as in (S3.4).

We first consider $R_n^{(1)}$. By (S3.2)-(S3.3), we can rewrite $R_{ijqr}^{(1)}$ as

$$\begin{aligned}
R_{ijqr}^{(1)} &= [\overline{R}_{ijqr}^{(2)}]^T (H_{ijqr}^{\dagger} - H_{ijqr}) \overline{R}_{ijqr}^{(2)} \\
&\quad + [\overline{R}_{ijqr}^{(3)}]^T (H_{ijqr}^{\dagger} - H_{ijqr}) \overline{R}_{ijqr}^{(3)} \\
&\quad + \left[(\widehat{\theta}_n - \theta_0)^T \frac{\partial G_{ijqr}(\theta_0)}{\partial \theta}\right] (H_{ijqr}^{\dagger} - H_{ijqr}) \left[\frac{\partial G_{ijqr}(\theta_0)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0)\right] \\
&\quad + 2[\overline{R}_{ijqr}^{(2)}]^T (H_{ijqr}^{\dagger} - H_{ijqr}) \overline{R}_{ijqr}^{(3)} \\
&\quad + 2[\overline{R}_{ijqr}^{(2)}]^T (H_{ijqr}^{\dagger} - H_{ijqr}) \left[\frac{\partial G_{ijqr}(\theta_0)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0)\right] \\
&\quad + 2[\overline{R}_{ijqr}^{(3)}]^T (H_{ijqr}^{\dagger} - H_{ijqr}) \left[\frac{\partial G_{ijqr}(\theta_0)}{\partial \theta^T}(\widehat{\theta}_n - \theta_0)\right] \\
&=: r_{1,ijqr}^{(1)} + r_{2,ijqr}^{(1)} + r_{3,ijqr}^{(1)} + r_{4,ijqr}^{(1)} + r_{5,ijqr}^{(1)} + r_{6,ijqr}^{(1)}. \qquad \text{(S3.9)}
\end{aligned}$$

Then, by (S3.9), we have $R_n^{(1)} = \sum_{d=1}^{6} \Delta_d^{(1)}$, where

$$\Delta_d^{(1)} = \frac{1}{n^2}\sum_{i,j} r_{d,ijij}^{(1)} + \frac{1}{n^4}\sum_{i,j,q,r} r_{d,ijqr}^{(1)} - \frac{2}{n^3}\sum_{i,j,q} r_{d,ijiq}^{(1)}.$$

For the first entry of $[H_{ijqr}^{\dagger} - H_{ijqr}]$, we have $\big\|k_{xx}(\widehat{\eta}_{1i}^{\dagger}, \widehat{\eta}_{1j}^{\dagger}) l(\widehat{\eta}_{2q}^{\dagger}, \widehat{\eta}_{2r}^{\dagger}) - k_{xx}(\eta_{1i}, \eta_{1j})\ l(\eta_{2q}, \eta_{2r})\big\| \leq C\big[\|\widehat{\eta}_{1i}^{\dagger} - \eta_{1i}\| + \|\widehat{\eta}_{1j}^{\dagger} - \eta_{1j}\| + \|\widehat{\eta}_{2q}^{\dagger} - \eta_{2q}\| + \|\widehat{\eta}_{2r}^{\dagger} -$

$\eta_{2r}\|]$ by Triangle's inequality and Assumption 2.5. Meanwhile, by Taylor's expansion and Assumptions 2.2(i) and 2.3, we can show that $\|\widehat{\eta}^{\dagger}_{st} - \eta_{1i}\| \leq \|\widehat{R}_{st}(\widehat{\theta}_{sn})\| + \|\widehat{\theta}_{sn} - \theta_{s0}\| \sup_{\theta_s} \left\|\frac{\partial g_{st}(\theta_s)}{\partial \theta_s}\right\| = \|\widehat{R}_{st}(\widehat{\theta}_{sn})\| + o_p(1)$, where $\widehat{R}_{st}(\theta_s)$ is defined as in Assumption 2.4, $o_p(1)$ holds uniformly in t due to the fact that $\sqrt{n}\|\widehat{\theta}_{sn} - \theta_{s0}\| = O_p(1)$ and

$$\frac{1}{\sqrt{n}} \max_{1\leq t\leq n} \sup_{\theta_s} \left\| \frac{\partial g_{st}(\theta_s)}{\partial \theta_s} \right\| = o_p(1) \tag{S3.10}$$

by Assumption 2.2(i). Hence, it follows that

$$\begin{aligned} &\left\|k_{xx}(\widehat{\eta}^{\dagger}_{1i}, \widehat{\eta}^{\dagger}_{1j}) l(\widehat{\eta}^{\dagger}_{2q}, \widehat{\eta}^{\dagger}_{2r}) - k_{xx}(\eta_{1i}, \eta_{1j}) l(\eta_{2q}, \eta_{2r})\right\| \\ &\quad \leq C\left[\|\widehat{R}_{1i}(\widehat{\theta}_{1n})\| + \|\widehat{R}_{1j}(\widehat{\theta}_{1n})\| + \|\widehat{R}_{2q}(\widehat{\theta}_{2n})\| + \|\widehat{R}_{2r}(\widehat{\theta}_{2n})\|\right] \\ &\qquad + o_p(1), \end{aligned} \tag{S3.11}$$

where $o_p(1)$ holds uniformly in i, j, q, r. Similarly, (S3.11) holds for other entries of $[H^{\dagger}_{ijqr} - H_{ijqr}]$. Note that

$$\|\overline{R}^{(2)}_{ijqr}\| \leq \|\widehat{R}_{1i}(\widehat{\theta}_{1n})\| + \|\widehat{R}_{1j}(\widehat{\theta}_{1n})\| + \|\widehat{R}_{2q}(\widehat{\theta}_{2n})\| + \|\widehat{R}_{2r}(\widehat{\theta}_{2n})\|. \tag{S3.12}$$

Using the inequality $(\sum_{d=1}^{4} |a_d|)^3 \leq C \sum_{d=1}^{4} |a_d|^3$, by Assumption 2.4 and (S3.11)-(S3.12), it is not hard to show that

$$n\|\Delta^{(1)}_1\| = O_p(1/n). \tag{S3.13}$$

Furthermore, by Taylor's expansion, Assumptions 2.2(i) and 2.3, and a similar argument as for (S3.10), it is straightforward to see that

$$\begin{aligned}\|\overline{R}^{(3)}_{ijqr}\| &\le \left\|\frac{\partial G_{ijqr}(\theta^\dagger)}{\partial\theta^T} - \frac{\partial G_{ijqr}(\theta_0)}{\partial\theta^T}\right\| \times \|\widehat{\theta}_n - \theta_0\| \\ &\le \left[2\max_{1\le t\le n}\sup_{\theta_1}\left\|\frac{\partial^2 g_{1t}(\theta_1)}{\partial\theta_1^2}\right\| + 2\max_{1\le t\le n}\sup_{\theta_2}\left\|\frac{\partial^2 g_{2t}(\theta_2)}{\partial\theta_2^2}\right\|\right] \times \|\widehat{\theta}_n - \theta_0\|^2 \\ &= o_p(1/\sqrt{n}),\end{aligned}$$

where $o_p(1)$ holds uniformly in i, j, q, r. As for (S3.13), it entails that $n\|\Delta_2^{(1)}\| = o_p(1)$. Similarly, we can show that $n\|\Delta_d^{(1)}\| = o_p(1)$ for $d = 3, 4, 5, 6$. Therefore, it follows that $n\|R_n^{(1)}\| = o_p(1)$. By the analogous arguments, we can also show that $n\|R_n^{(d)}\| = o_p(1)$ for $d = 3, 4$.

Next, we consider the remaining term $R_n^{(2)}$. Denote $r^{(2)}_{1,ijqr} := [\overline{R}^{(2)}_{ijqr}]^T W_{ijqr}$ and $r^{(2)}_{2,ijqr} := [\overline{R}^{(3)}_{ijqr}]^T W_{ijqr}$. Then, we can rewrite $R_n^{(2)} = \Delta_1^{(2)} + \Delta_2^{(2)}$, where

$$\Delta_d^{(2)} = \frac{1}{n^2}\sum_{i,j} r^{(2)}_{d,ijij} + \frac{1}{n^4}\sum_{i,j,q,r} r^{(2)}_{d,ijqr} - \frac{2}{n^3}\sum_{i,j,q} r^{(2)}_{d,ijiq}$$

for $d = 1, 2$. By Assumptions 2.2(i) and 2.3-2.5 and (S3.12), we have $n\|\Delta_1^{(2)}\| = O_p(1/n)$. Rewrite $\Delta_2^{(2)} = (\widehat{\theta}_{1n} - \theta_{10})^T\Delta_{21}^{(2)} + (\widehat{\theta}_{2n} - \theta_{20})^T\Delta_{22}^{(2)}$, where

$$\Delta_{2d}^{(2)} = \frac{1}{n^2}\sum_{i,j} r^{(2)}_{2d,ijij} + \frac{1}{n^4}\sum_{i,j,q,r} r^{(2)}_{2d,ijqr} - \frac{2}{n^3}\sum_{i,j,q} r^{(2)}_{2d,ijiq}$$

for $d = 1, 2$, with $r^{(2)}_{21,ijqr} = k^\dagger_{ij} l_{qr}$ and $r^{(2)}_{22,ijqr} = k_{ij} l^\dagger_{qr}$. Here,

$$k^\dagger_{ij} = \left[\frac{\partial g_{1i}(\theta_1^\dagger)}{\partial\theta_1} - \frac{\partial g_{1i}(\theta_{10})}{\partial\theta_1}\right] k_x(\eta_{1i}, \eta_{1j}) + \left[\frac{\partial g_{1j}(\theta_1^\dagger)}{\partial\theta_1} - \frac{\partial g_{1j}(\theta_{10})}{\partial\theta_1}\right] k_y(\eta_{1i}, \eta_{1j}),$$

$$l_{qr}^{\dagger} = \left[\frac{\partial g_{2q}(\theta_2^{\dagger})}{\partial \theta_2} - \frac{\partial g_{2q}(\theta_{20})}{\partial \theta_2}\right] l_x(\eta_{2q}, \eta_{2r}) + \left[\frac{\partial g_{2r}(\theta_2^{\dagger})}{\partial \theta_2} - \frac{\partial g_{2r}(\theta_{20})}{\partial \theta_2}\right] l_y(\eta_{2q}, \eta_{2r}).$$

By the mean value theorem, $k_{ij}^{\dagger} = (\theta_1^{\dagger} - \theta_{10})^T k_{ij}^{\S}$, where $k_{ij}^{\S}$ is defined explicitly, and it satisfies that

$$\Delta_{21}^{(2)\S} := \frac{1}{n^2}\sum_{i,j} k_{ij}^{\S} l_{ij} + \frac{1}{n^4}\sum_{i,j,q,r} k_{ij}^{\S} l_{qr} - \frac{2}{n^3}\sum_{i,j,q} k_{ij}^{\S} l_{iq} = O_p(1/n). \quad \text{(S3.14)}$$

Here, (S3.14) holds, since $\Delta_{21}^{(2)\S}$ under H_0 is a degenerate V-statistic by Assumptions 2.1 and 2.5 and a similar argument as for Lemma 3.2(ii). Note that $\Delta_{21}^{(2)} = (\theta_1^{\dagger} - \theta_{10})^T \Delta_{21}^{(2)\S}$ and $\|\theta_1^{\dagger} - \theta_{10}\| \leq \|\widehat{\theta}_{1n} - \theta_{10}\| = o_p(1)$. Therefore, it follows that $\sqrt{n}\|\Delta_{21}^{(2)}\| = o_p(1)$. Similarly, we can show that $\sqrt{n}\|\Delta_{22}^{(2)}\| = o_p(1)$, and it follows that $n\|R_n^{(2)}\| = o_p(1)$. This completes the proof. □

Lemma S3.2. *Suppose Assumptions 2.1-2.5 hold. Then,* $\sqrt{n}\|R_n(m)\| = o_p(1)$, *where* $R_n(m)$ *is defined as in (S3.6).*

Proof. The proof is the same as the one for Lemma S3.1, except that when H_0 does not hold, we can only have $\Delta_{21}^{(2)\S} = O_p(1)$ in (S3.14) by part (c) of Theorem 2 in Denker and Keller (1983). □

Let $\varsigma_{st} = \left(\eta_{st}, \frac{\partial g_{st}(\theta_{s0})}{\partial \theta_s}\right)$ for $s = 1, 2$. To prove Theorem 4.1, we need the following two lemmas.

Lemma S3.3. *Suppose Assumptions 2.1, 2.2(i) and 2.3-2.5 hold. Then, under H_0, for $\forall K_0 > 0$,*

$$(i)\ \sup_{\Omega_1}\left|\frac{1}{N^4}\sum_{q,q',r,r'} h_m^{(0)}\left(x_1, x_2, \left(\eta_{1q}, \eta_{2q'+m}\right), \left(\eta_{1r}, \eta_{2r'+m}\right)\right) - E[h_m^{(0)}(x_1, x_2, \eta_3^{(m)}, \eta_4^{(m)})]\right| = o_p(1),$$

where $\Omega_1 = \{(x_1, x_2) : \|x_s\| \le K_0$ for $s = 1, 2\}$;

$$(ii)\ \sup_{\Omega_2}\left|\frac{1}{N^4}\sum_{i',j',q',r'} h_m^{(23)}\left(\left(z_{11}, \varsigma_{2i'+m}\right), \left(z_{12}, \varsigma_{2j'+m}\right), \left(z_{13}, \varsigma_{2q'+m}\right), \left(z_{14}, \varsigma_{2r'+m}\right)\right) - E\left[h_m^{(23)}\left(\left(z_{11}, \varsigma_{21}\right), \left(z_{12}, \varsigma_{22}\right), \left(z_{13}, \varsigma_{23}\right), \left(z_{14}, \varsigma_{24}\right)\right)\right]\right| = o_p(1),$$

where $\Omega_2 = \{(z_{11}, z_{12}, z_{13}, z_{14}) : \|z_{1s}\| \le K_0$ for $s = 1, 2, 3, 4\}$;

$$(iii)\ \sup_{\Omega_3}\left|\frac{1}{N^4}\sum_{i,j,q,r} h_m^{(23)}\left(\left(\varsigma_{1i}, z_{21}\right), \left(\varsigma_{1j}, z_{22}\right), \left(\varsigma_{1q}, z_{23}\right), \left(\varsigma_{1r}, z_{24}\right)\right) - E\left[h_m^{(23)}\left(\left(\varsigma_{11}, z_{21}\right), \left(\varsigma_{12}, z_{22}\right), \left(\varsigma_{13}, z_{23}\right), \left(\varsigma_{14}, z_{24}\right)\right)\right]\right| = o_p(1),$$

where $\Omega_3 = \{(z_{21}, z_{22}, z_{23}, z_{24}) : \|z_{2s}\| \le K_0$ for $s = 1, 2, 3, 4\}$.

Proof. (*i*) Denote $x_1 = (x_{11}, x_{21})$ and $x_2 = (x_{12}, x_{22})$. Without loss of generality, we assume that $m = 0$. By the definition of $h_0^{(00)}$, it has 24 different terms, and we only give the proof for its first term. That is, we are going to show that

$$\frac{1}{N^4}\sum_{q,r,q',r'} \tilde{k}_{12}^{(0)}[\tilde{l}_{12}^{(0)} + l_{q'r'}^{(0)} - 2\tilde{l}_{1q'}^{(0)} - E(\tilde{l}_{12}^{(0)}) - E(l_{34}^{(0)}) + 2E(\tilde{l}_{13}^{(0)})]$$

$$= o_p(1), \tag{S3.15}$$

where $o_p(1)$ holds uniformly in Ω_1, $\tilde{k}_{12}^{(0)} = k(x_{11}, x_{12})$, $\tilde{l}_{12}^{(0)} = k(x_{21}, x_{22})$, $l_{q'r'}^{(0)} = k(\eta_{2q'}, \eta_{2r'})$, $\tilde{l}_{1q'}^{(0)} = k(x_{21}, \eta_{2q'})$, $l_{34}^{(0)} = k(\eta_{23}, \eta_{24})$, and $\tilde{l}_{13}^{(0)} = k(x_{21}, \eta_{23})$.

By the triangle's inequality, we have

$$\begin{aligned}
&\left|\frac{1}{N^4}\sum_{q,r,q',r'} \tilde{k}_{12}^{(0)}[\tilde{l}_{12}^{(0)} + l_{q'r'}^{(0)} - 2\tilde{l}_{1q'}^{(0)} - E(\tilde{l}_{12}^{(0)}) - E(l_{34}^{(0)}) + 2E(\tilde{l}_{13}^{(0)})]\right| \\
&= \left|\frac{\tilde{k}_{12}^{(0)}}{N^4}\sum_{q,r,q',r'} [l_{q'r'}^{(0)} - 2\tilde{l}_{1q'}^{(0)} - E(l_{34}^{(0)}) + 2E(\tilde{l}_{13}^{(0)})]\right| \\
&\leq \left|\frac{C}{N^2}\sum_{q',r'=1}^{n} [l_{q'r'}^{(0)} - E(l_{34}^{(0)})]\right| + \left|\frac{C}{N}\sum_{q'=1}^{n}[\tilde{l}_{1q'}^{(0)} - E(\tilde{l}_{13}^{(0)})]\right|.
\end{aligned}$$

Hence, it follows that (S3.15) holds by noting the fact that

$$\frac{1}{N^2}\sum_{q',r'=1}^{n} [l_{q'r'}^{(0)} - E(l_{34}^{(0)})] = o_p(1), \tag{S3.16}$$

$$\sup_{\Omega_1}\frac{1}{N}\sum_{q'=1}^{n}[\tilde{l}_{1q'}^{(0)} - E(\tilde{l}_{13}^{(0)})] = o_p(1), \tag{S3.17}$$

where (S3.16) holds by the law of large numbers for V-statistics, and (S3.17) holds by Assumption 2.5 and standard arguments for uniform convergence.

(ii) & *(iii)* The conclusions hold by similar arguments as for *(i)*. □

Lemma S3.4. *Suppose Assumptions 2.1, 2.2(i) and 2.3-2.5 hold. Then, under* H_0*,*

$$(i)\ \sup_{\Omega_1}\left|E^*\left[h_m^{(0)}\left(x_1, x_2, \widehat{\eta}_3^{(m*)}, \widehat{\eta}_4^{(m*)}\right)\right]\right.$$

$$-E\left[h_m^{(0)}\left(\eta_1^{(m)},\eta_2^{(m)},\eta_3^{(m)},\eta_4^{(m)}\right)\right]\Big| = o_p(1),$$

where Ω_1 *is defined as in Lemma S3.3(i);*

$$(ii)\ \left|\Lambda_m^{(23*)}-\Lambda_m^{(23)}\right| = o_p(1).$$

Proof. (i) First, it is straightforward to see that

$$\begin{aligned}
&E^*\left[h_m^{(0)}\left(x_1,x_2,\widehat{\eta}_3^{(m*)},\widehat{\eta}_4^{(m*)}\right)\right]\\
&=\frac{1}{N^4}\sum_{q,q',r,r'} h_m^{(0)}\left(x_1,x_2,\left(\widehat{\eta}_{1q},\widehat{\eta}_{2q'+m}\right),\left(\widehat{\eta}_{1r},\widehat{\eta}_{2r'+m}\right)\right)\\
&=\frac{1}{N^4}\sum_{q,q',r,r'} h_m^{(0)}\left(x_1,x_2,\left(\eta_{1q},\eta_{2q'+m}\right),\left(\eta_{1r},\eta_{2r'+m}\right)\right)+o_p(1), \qquad \text{(S3.18)}
\end{aligned}$$

where $o_p(1)$ holds uniformly in Ω_1 by Taylor's expansion and Assumptions 2.3 and 2.5. Then, the conclusion holds by (S3.18) and Lemma S3.3(i).

(ii) Define

$$\mathcal{H}(i,i',j,j',q,q',r,r') = h_m^{(23)}\left(\left(\varsigma_{1i},\varsigma_{2i'}\right),\left(\varsigma_{1j},\varsigma_{2j'}\right),\left(\varsigma_{1q},\varsigma_{2q'}\right),\left(\varsigma_{1r},\varsigma_{2r'}\right)\right).$$

By a similar argument as for (S3.18), we have

$$\Lambda_m^{(23*)}-\Lambda_m^{(23)} = \Xi_0 + o_p(1),$$

where

$$\Xi_0 = \frac{1}{N^8}\sum_{i,j,q,r,i',j',q',r'} \mathcal{H}(i,i'+m,j,j'+m,q,q'+m,r,r'+m)-\Lambda_m^{(23)}.$$

Rewrite

$$\Xi_0 := \frac{1}{N^4}\sum_{i,j,q,r}\Xi_{1,ijqr} + \frac{1}{N^4}\sum_{i,j,q,r}\Xi_{2,ijqr}, \tag{S3.19}$$

where

$$\begin{aligned}\Xi_{1,ijqr} &= \frac{1}{N^4}\sum_{i',j',q',r'}\mathcal{H}(i,i'+m,j,j'+m,q,q'+m,r,r'+m)\\ &\quad - E_{\varsigma_{21},\varsigma_{22},\varsigma_{23},\varsigma_{24}}\left[\mathcal{H}(i,1,j,2,q,3,r,4)\right],\\ \Xi_{2,ijqr} &= E_{\varsigma_{21},\varsigma_{22},\varsigma_{23},\varsigma_{24}}\left[\mathcal{H}(i,1,j,2,q,3,r,4)\right] - \Lambda_m^{(23)}.\end{aligned}$$

By Lemma S3.3(ii), $\Xi_{1,ijqr} = o_p(1)$ uniformly in i, j, q, r, and hence

$$\frac{1}{N^4}\sum_{i,j,q,r}\Xi_{1,ijqr} = o_p(1). \tag{S3.20}$$

Moreover, we can rewrite

$$\begin{aligned}\frac{1}{N^4}\sum_{i,j,q,r}\Xi_{2,ijqr} &= E_{\varsigma_{21},\varsigma_{22},\varsigma_{23},\varsigma_{24}}\big\{\mathcal{H}(i,1,j,2,q,3,r,4)\\ &\quad - E_{\varsigma_{11},\varsigma_{12},\varsigma_{13},\varsigma_{14}}\left[\mathcal{H}(1,1,2,2,3,3,4,4)\right]\big\},\end{aligned} \tag{S3.21}$$

where we have used the fact that under H_0,

$$\Lambda_m^{(23)} = E_{\varsigma_{21},\varsigma_{22},\varsigma_{23},\varsigma_{24}}E_{\varsigma_{11},\varsigma_{12},\varsigma_{13},\varsigma_{14}}\left[\mathcal{H}(1,1,2,2,3,3,4,4)\right].$$

By (S3.21), Lemma S3.3(iii), Assumptions 2.2(i) and 2.5, and the dominated convergence theorem, we can show that

$$\frac{1}{N^4}\sum_{i,j,q,r}\Xi_{2,ijqr} = o_p(1). \tag{S3.22}$$

Hence, the conclusion holds by (S3.19)-(S3.20) and (S3.22). □

Bibliography

DENKER, M. and KELLER, G. (1983) On U-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrsch. Verw. Gebiete* **64**, 505-522.

DUNFORD, N. and SCHWARTZ, J.T. (1963) *Linear Operators Part 2: Spectral Theory*. New York: Interscience.

GRETTON, A., FUKUMIZU,K., HARCHAOUI, Z. and SRIPERUMBUDUR, B. (2009) A fast, consistent kernel two sample test. *Advances in Neural Information Processiong Systems* **22**, 673-681.

HOEFFDING, W. (1948) A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**, 293-325.

LEE, A.J. (1990) *U-Statistics: Theory and Practice*. New York: Marcel Dekke.

WHITE, H. (2001) *Asymptotic Theory for Econometricians (rev. ed.)*. New York: Academic Press.

YOSHIHARA, K.I. (1976) Limiting behavior of U-statistics for stationary, absolutely regular processes. *Z. Wahrsch. Verw. Gebiete* **35**, 237-252.