Supplementary Material for "Subgroup testing in change-plane models and its applications to medical data"

Xu Liu[1], Jian Huang[2], Yong Zhou[3], Feipeng Zhang[4] and Panpan Ren[1]

[1]*Shanghai University of Finance and Economics,* [2]*The Hong Kong Polytechnic University*

[3]East China Normal University and [4]Xi'an Jiaotong University

This Supplementary Material includes the detailed proofs of Theorems 1 – 5, Corollary 1 and the related Lemmas in Section S1. In Section S2, additional comparison of final sample performance for high-dimensional $\boldsymbol{Z}$ is reported by some simulation examples. Section S3 conducts additional comparison of final sample performance for low-dimensional $\boldsymbol{Z}$. In Section S4, case study on BRCA data is analysed to illustrate the performance of the proposed test.

S1 Proof of Theorems

We first give some useful notation for convenience of expression. For a vector $\boldsymbol{v} \in \mathbb{R}^d$ and a squared matrix $A = (a_{ij}) \in \mathbb{R}^{d\times d}$, denote by $\|\boldsymbol{v}\|$ the Euclidean norm of $\boldsymbol{v}$, by $\text{trace}(A) = \sum_{i=1}^d a_{ii}$ the trace of A; $\|\boldsymbol{v}\|_A^2 = \sum_{i,j} a_{ij}v_iv_j$ and $\boldsymbol{v}^{\otimes 2} = \boldsymbol{v}\boldsymbol{v}^{\mathrm{T}}$. Denote by $\|A\|_p = \sup\{\|A\boldsymbol{x}\|_p : \boldsymbol{x} \in \mathbb{R}^d, \|\boldsymbol{x}\|_p = 1\}$ the induced operator norm for a matrix $A = (a_{ij}) \in \mathbb{R}^{m\times d}$.

Denote by $\mathbf{P}$ the ordinary probability measure such that $\mathrm{E}f = \int f d\mathbf{P}$

for any measurable function f, by $\mathbb{P}_n$ the empirical measure of a sample of random elements from $\mathbf{P}$ such that $\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(\boldsymbol{V}_i)$, and by $\mathbb{G}_n$ the empirical process indexed by a class $\mathcal{F}$ of measurable functions such that $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - \mathbf{P})f$ for any $f \in \mathcal{F}$. Let $L^p(Q)$ be the space of all measurable functions f such that $\|f\|_{Q,p} := (Q|f|^p)^{1/p} < \infty$, where $p \in [1,\infty)$ and $(Q|f|^p)^{1/p}$ denotes the essential supremum when $p = \infty$. Let $N(\epsilon, \mathcal{F}, \|\cdot\|_{Q,2})$ be an ϵ-covering number of $\mathcal{F}$ with respect to the $L^2(Q)$ seminorm $\|\cdot\|_{Q,2}$, where $\mathcal{F}$ is a class of measure functions and Q is finite discrete.

Lemma 1. Let $\boldsymbol{X}$ be normally distributed with mean $\boldsymbol{\mu}$ and covariance Σ. Denote by $\phi(z)$ and $\Phi(z)$ the density function and cumulative distribution function of the standard normal distribution, respectively. For any nonzero vectors $\boldsymbol{a} \in \mathbb{R}^p$ and $\boldsymbol{b} \in \mathbb{R}^p$ and any numbers $\eta_a \in \mathbb{R}$ and $\eta_b \in \mathbb{R}$, we have

$$\begin{aligned}
\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)] =& \Phi\left(\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} - \eta_a}{(\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a})^{1/2}}\right), \\
\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)\mathbf{1}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} \geq \eta_b)] =& \int_{\frac{\eta_b - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}}{(\boldsymbol{b}^{\mathrm{T}}\Sigma\boldsymbol{b})^{1/2}}}^{+\infty} \phi(z)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} - \eta_a}{(\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a})^{1/2}} + \rho z\right)\right) dz,
\end{aligned}$$

where $\rho = \frac{\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{b}}{(\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}\boldsymbol{b}^{\mathrm{T}}\Sigma\boldsymbol{b})^{1/2}}$, and $\mathbf{1}(\cdot)$ denotes the indicator function. Moreover,

when $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} - \eta_a = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu} - \eta_b = 0$, we have

$$\begin{aligned}\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)] =& \frac{1}{2},\\ \mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)\mathbf{1}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} \geq \eta_b)] =& \frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).\end{aligned}$$

Proof of Lemma 1. Let $\mu_1 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu}$, $\mu_2 = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu}$, $\sigma_1^2 = \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}$, $\sigma_2^2 = \boldsymbol{b}^{\mathrm{T}}\Sigma\boldsymbol{b}$. Noting that the random variables $Z_1 = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} - \mu_1 \sim N(0, \sigma_1^2)$ and $Z_2 = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} - \mu_2 \sim N(0, \sigma_2^2)$, we have $Z = (Z_1, Z_2)^{\mathrm{T}} \sim N(\mathbf{0}, \Gamma)$ with

$$\Gamma = \begin{pmatrix} \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a} & \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{b} \\ \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{b} & \boldsymbol{b}^{\mathrm{T}}\Sigma\boldsymbol{b} \end{pmatrix}.$$

which implies that

$$\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)] = \mathrm{E}[\mathbf{1}(Z_1 \geq \eta_a - \mu_1)] = \int_{\eta_a-\mu_1}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{z_1^2}{2\sigma_1^2}\right\} dz_1 = \Phi\left(\frac{\mu_1 - \eta_a}{\sigma_1}\right),$$

and

$$\begin{aligned}&\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)\mathbf{1}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} \geq \eta_b)]\\ =&\mathrm{E}[\mathbf{1}(Z_1 \geq \eta_a - \mu_1)\mathbf{1}(Z_2 \geq \eta_b - \mu_2)]\\ =&\int_{\eta_b-\mu_2}^{+\infty}\int_{\eta_a-\mu_1}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2}Z^{\mathrm{T}}\Gamma^{-1}Z\right\} dz_1 dz_2\\ =&\int_{\eta_b-\mu_2}^{+\infty}\int_{\eta_a-\mu_1}^{+\infty} \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{z_1}{\sigma_1} - \rho\frac{z_2}{\sigma_2}\right)^2 - \frac{z_2^2}{2\sigma_2^2}\right\} dz_1 dz_2\\ =&\int_{\eta_b-\mu_2}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{z_2^2}{2\sigma_2^2}\right\}\Phi\left(-\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\eta_a-\mu_1}{\sigma_1} - \frac{\rho}{\sigma_2}z_2\right)\right) dz_2\\ =&\int_{\frac{\eta_b-\mu_2}{\sigma_2}}^{+\infty} \phi(z)\Phi\left(\frac{1}{\sqrt{1-\rho^2}}\left(\frac{\mu_1-\eta_a}{\sigma_1} + \rho z\right)\right) dz.\end{aligned}$$

When $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\mu} - \eta_a = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\mu} - \eta_b = 0$, it is straightforward that

$$\mathrm{E}[\mathbf{1}(\boldsymbol{a}^{\mathrm{T}}\boldsymbol{X} \geq \eta_a)\mathbf{1}(\boldsymbol{b}^{\mathrm{T}}\boldsymbol{X} \geq \eta_b)] = \int_0^{+\infty} \phi(z)\Phi\left(\frac{\rho}{\sqrt{1-\rho^2}}z\right) dz = \frac{1}{4} + \frac{1}{2\pi}\arctan\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).$$

Lemma 2. If Assumptions (A1)-(A3) hold, we have

$$\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = -J\mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0) + o_p(n^{-1/2}),$$

where $\boldsymbol{\alpha}_0$ is given in Assumption (A1), and $J = \left(\partial \mathbf{P}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha}^{\mathrm{T}}\right)^{-1} \in \mathbb{R}^{r\times r}$.

Proof of Lemma 2. Note that $\mathbf{P}[\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)] = \mathbb{P}_n\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0) - (\mathbb{P}_n - \mathbf{P})\{\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\}$. It suffices to prove that

$$(\mathbb{P}_n - \mathbf{P})\{\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\} = o_p(n^{-1/2}), \tag{S1.1}$$

and

$$\mathbf{P}[\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)] = J^{-1}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) + o(n^{-1/2}). \tag{S1.2}$$

Let $\mathcal{U}_{\alpha_0} = \{\boldsymbol{\alpha} \in \Theta_\alpha : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| = o(1)\}$ and $\mathcal{F}_{1\alpha} = \cup_{1\leq j\leq p}\mathcal{F}_{1\alpha}^{(j)}$, where $\mathcal{F}_{1\alpha}^{(j)} = \{\psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}) - \psi_1^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}_0) : \boldsymbol{\alpha} \in \mathcal{U}_{\alpha_0}, \boldsymbol{\theta} \in \Theta_\theta\}$. We can show that, with probability $1 - o(1)$, $\sup_{f\in\mathcal{F}_{1\alpha}} \|\mathbb{G}_n f\| \lesssim 2\sqrt{C_0 p}L_{1n}^{1/2}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| + 4C_0 p M_n n^{-1/2} = o(1)$, which implies that (S1.1) holds.

It follows from the central limit theorem

$$\mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}) = O_p(n^{-1/2}), \tag{S1.3}$$

which combining (S1.1) indicates that $\mathbf{P}[\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)] = O(n^{-1/2})$ with probability $1 - o(1)$, and consequently, we have, by Assumption (A1), $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = o(1)$ with probability $1 - o(1)$.

For any $\boldsymbol{\alpha} \in \mathcal{U}_{\alpha_0}$, we have, by Assumption (A2) and Taylor's theorem,

$$\mathbf{P}[\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)] = \frac{\partial \mathbf{P}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + O(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|^2),$$

which indicates that, with probability $1 - o(1)$, $\mathbf{P}[\psi_1(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}) - \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)] = J^{-1}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0)(1 + o(1))$. Thus, combining (S1.1) and (S1.3) we have $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0 = O_p(n^{-1/2})$, and that (S1.2) holds.

Lemma 3. If Assumptions (A1)–(A5) hold, then for any $\boldsymbol{\beta} \in \Theta_\beta$ and $\boldsymbol{\theta} \in \Theta_\theta$, we have

$$\mathbb{P}_n\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbb{P}_n\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}) - K(\boldsymbol{\beta}, \boldsymbol{\theta})J\mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0) + o_p(n^{-1/2}),$$

where $\boldsymbol{\alpha}_0$ is given in Assumption (A1), $K(\boldsymbol{\beta}, \boldsymbol{\theta}) = \partial \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})/\partial \boldsymbol{\alpha} \in \mathbb{R}^{p \times r}$, and $J = \partial \mathbf{P}\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)/\partial \boldsymbol{\alpha} \in \mathbb{R}^{r \times r}$.

Proof of Lemma 3. Note that $\mathbf{P}[\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})] = \mathbb{P}_n\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \mathbb{P}_n\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}) - (\mathbb{P}_n - \mathbf{P})\{\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})\}$. It suffices to prove that

$$(\mathbb{P}_n - \mathbf{P})\{\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})\} = o_p(n^{-1/2}), \tag{S1.4}$$

and

$$\mathbf{P}[\psi(\boldsymbol{V},\hat{\boldsymbol{\alpha}},\boldsymbol{\beta},\boldsymbol{\theta})-\psi(\boldsymbol{V},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta})]=-K(\boldsymbol{\theta},\boldsymbol{\beta})J\mathbb{P}_n\psi_1(\boldsymbol{V},\boldsymbol{\alpha}_0)+o(n^{-1/2}). \tag{S1.5}$$

We first show that (S1.4) holds. Let $\mathcal{U}_{\alpha_0}=\{\boldsymbol{\alpha}\in\Theta_\alpha:\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|=O(n^{-1/2})\}$ and $\mathcal{F}_\alpha=\cup_{1\le j\le p}\mathcal{F}_\alpha^{(j)}$, where $\mathcal{F}_\alpha^{(j)}=\{\psi^{(j)}(\boldsymbol{V},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta})-\psi^{(j)}(\boldsymbol{V},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta}):\boldsymbol{\alpha}\in\mathcal{U}_{\alpha_0},\boldsymbol{\beta}\in\Theta_\beta,\boldsymbol{\theta}\in\Theta_\theta\}$. We define the envelope function of $\mathcal{F}_\alpha$ as $F_\alpha=\max_{j=1,\cdots,p}F_\alpha^{(j)}$, where $F_\alpha^{(j)}=\sup_{\boldsymbol{\alpha}\in\mathcal{U}_{\alpha_0},\boldsymbol{\theta}\in\Theta_\theta}|\psi^{(j)}(\boldsymbol{V},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\theta})-\psi^{(j)}(\boldsymbol{V},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta})|$. Assumption (A3) implies that $\|F_\alpha\|_{\mathbf{P},2}\le\|F\|_{\mathbf{P},2}\le L_n^{1/2}\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$, and by the fact that $\mathcal{F}_\alpha^{(j)}\subseteq\mathcal{F}^{(j)}-\mathcal{F}^{(j)}$,

$$\begin{aligned}\log\sup_Q N(\epsilon\|F_\alpha\|_{Q,2},\mathcal{F}_\alpha,\|\cdot\|_{Q,2})\le&2p\log\sup_Q N\left(\frac{\epsilon}{2}\|F_j\|_{Q,2},\mathcal{F}^{(j)},\|\cdot\|_{Q,2}\right)\\ \le&2C_0p\log(2e/\epsilon),\end{aligned}$$

which results in that, for $\delta\in(0,1)$,

$$\begin{aligned}\boldsymbol{J}(\delta,\mathcal{F}_\alpha,F_\alpha)=&\int_0^\delta\sup_Q\sqrt{1+\log N(\epsilon\|F_\alpha\|_{Q,2},\mathcal{F}_\alpha,\|\cdot\|_{Q,2})}d\epsilon\\ \le&\int_0^\delta\sqrt{1+2C_0p\log(2e/\epsilon)}d\epsilon\\ =&2e\int_{2e/\delta}^\infty\frac{1}{\epsilon^2}\sqrt{1+2C_0p\log(\epsilon)}d\epsilon.\end{aligned}$$

According to the fact that $1+\log(\epsilon)\ge1$ for all $\epsilon\in[1,\infty)$ and the integration

by parts, we have

$$\int_{2e/\delta}^{\infty} \frac{1}{\epsilon^2}\sqrt{1+\log(\epsilon)}d\epsilon = -\frac{1}{\epsilon}\sqrt{1+\log(\epsilon)}\Big|_{\epsilon=2e/\delta}^{\infty} + \frac{1}{2}\int_{2e/\delta}^{\infty} \frac{1}{\epsilon^2\sqrt{1+\log(\epsilon)}}d\epsilon$$
$$\leq \frac{\delta}{2e}\sqrt{1+\log(2e/\delta)} + \frac{1}{2}\int_{2e/\delta}^{\infty} \frac{1}{\epsilon^2}\sqrt{1+\log(\epsilon)}d\epsilon,$$

which implies that $\int_{2e/\delta}^{\infty} \frac{1}{\epsilon^2}\sqrt{1+\log(\epsilon)}d\epsilon \leq \frac{\delta}{e}\sqrt{1+\log(2e/\delta)}$, and consequently $\boldsymbol{J}(\delta, \mathcal{F}_\alpha, F_\alpha) \leq \delta\sqrt{2C_0 p(1+\log(2e/\delta))}$. Therefore, by Assumption (A4) and Theorem 5.1 and 5.2 of Chernozhukov et al. (2014) with $\sigma^2 = \|F_\alpha\|_{\mathbf{P},2}^2$, we have that, with probability $1-o(1)$, $\sup_{f\in\mathcal{F}_\alpha} \|\mathbb{G}_n f\| \lesssim 2\sqrt{C_0 p}L_n^{1/2}\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| + 4C_0 p M_n n^{-1/2} = o(1)$, which implies that, with probability $1-o(1)$, $\sup_{f\in\mathcal{F}_\alpha} \|(\mathbb{P}_n - \mathbf{P})f\| = o(n^{-1/2})$ because of the fact that $\hat{\boldsymbol{\alpha}} \in \mathcal{U}_{\alpha_0}$ with probability $1-o(1)$ by Lemma 2, and consequently (S1.4) holds.

We, now show (S1.5) holds. For any $\boldsymbol{\alpha} \in \mathcal{U}_{\alpha_0}$, we have, by Assumption (A2) and Taylor's theorem, $\mathbf{P}[\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})] = \frac{\partial \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}^{\mathrm{T}}}(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + O(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|^2)$, which combining Lemma 2 imply that, with probability $1-o(1)$, $\mathbf{P}[\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})] = -K(\boldsymbol{\theta}, \boldsymbol{\beta})J\mathbb{P}_n\psi_1(\boldsymbol{V}, \alpha_0) + O(n^{-1})$, and consequently (S1.5) holds.

Proof of Lemma 1 in the main paper. By applying the central limit theorem to i.i.d. sample $\{\boldsymbol{V}_i\}_{i=1}^n$, we have that under H_0

$$n^{1/2}\left\{\mathbb{P}_n\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - K(\boldsymbol{\theta})J\mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\right\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Gamma(\boldsymbol{\theta})),$$

for any fixed $\boldsymbol{\theta} \in \Theta_\theta$, where $\Gamma(\boldsymbol{\theta}) = \mathbf{P}\{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - K(\boldsymbol{\theta})J\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\}^{\otimes 2}$.

By Theorem 2.1, it is straightforward that for any fixed $\boldsymbol{\theta} \in \Theta_\theta$,

$$\tilde{T}_n(\boldsymbol{\theta}) = \|\mathbb{P}_n\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - K(\boldsymbol{\theta})J\mathbb{P}_n\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\|^2_{\Gamma(\boldsymbol{\theta})^{-1}} + o_p(1) \xrightarrow{\mathcal{L}} \chi^2_p.$$

To complete the proof, it suffices to show that $\tilde{V}(\boldsymbol{\theta}) \xrightarrow{p} \Gamma(\boldsymbol{\theta})$ for any fixed $\boldsymbol{\theta} \in \Theta_\theta$. In fact, it is easy to show that $\tilde{V}(\boldsymbol{\theta}) \xrightarrow{p} \Gamma(\boldsymbol{\theta})$ for any fixed $\boldsymbol{\theta} \in \Theta_\theta$ by the law of large numbers.

Lemma 4. If Assumptions (A1)–(A5) hold, then $\tilde{T}_n$ converges in distribution to $\sup_{\boldsymbol{\theta}\in\Theta_\theta} \mathcal{G}^2(\boldsymbol{\theta})$ under H_0 as $n \to \infty$, where $\{\mathcal{G}(\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta_\theta\}$ is a zero-mean Gaussian process with the covariance function

$$\Sigma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{P}\psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_1)\psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_2)^{\mathrm{T}},$$

indexed by $\boldsymbol{\theta}_1 \in \Theta_\theta$ and $\boldsymbol{\theta}_2 \in \Theta_\theta$, where $\psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta})^{-1/2}\{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - K(\boldsymbol{\theta})J\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\}$ and $\Gamma(\boldsymbol{\theta}) = \mathbf{P}\{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - K(\boldsymbol{\theta})J\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\}^{\otimes 2}$, with $K(\boldsymbol{\theta})$ and J defined in Assumption (A5).

Proof of Lemma 4. Define a class of measurable functions $\mathcal{F}_2 = \cup_{1\le j\le p}\mathcal{F}_2^{(j)}$ and its entropy function $F_2 = \max_{j=1,\cdots,p} \sup_{f\in\mathcal{F}_2^{(j)}} |f|$, where $\mathcal{F}_2^{(j)} = \{\boldsymbol{V} \mapsto \psi_*^{(j)}(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) : \boldsymbol{\beta} \in \Theta_\beta, \boldsymbol{\theta} \in \Theta_\theta\}$. Assumption (A2) and (A4) imply that $\mathbf{P}F_2^2(\boldsymbol{V}) \le C + C_K^2 C_c^2 C_1$. and by the fact that $\mathcal{F}_2^{(j)} \subseteq \mathcal{F}^{(j)} +$

$\underbrace{\mathcal{F}_1^{(j)} + \cdots + \mathcal{F}_1^{(j)}}_{r}$, and

$$\begin{aligned}\log \sup_Q N(\epsilon \|F_2\|_{Q,2}, \mathcal{F}_2, \|\cdot\|_{Q,2}) \leq & p \log \sup_Q N\left(\frac{\epsilon}{2}\|F_j\|_{Q,2}, \mathcal{F}^{(j)}, \|\cdot\|_{Q,2}\right) \\ & + rp \log \sup_Q N\left(\frac{\epsilon}{2}\|F_{1j}\|_{Q,2}, \mathcal{F}_1^{(j)}, \|\cdot\|_{Q,2}\right) \\ \leq & 2C_0 rp \log(2e/\epsilon),\end{aligned}$$

which results in that, along the line of Lemma 2.1, for $\delta \in (0,1)$,

$$\begin{aligned}\boldsymbol{J}(\delta, \mathcal{F}_2, F_2) = & \int_0^\delta \sup_Q \sqrt{1 + \log N(\epsilon \|F_2\|_{Q,2}, \mathcal{F}_2, \|\cdot\|_{Q,2})} d\epsilon \\ \leq & \delta \sqrt{2C_0 rp\{1 + \log(2e/\delta)\}} < \infty.\end{aligned}$$

By Theorem 2.5.2 of van der Vaart and Wellner (1996), $\mathcal{F}_2$ is $\mathbf{P}$-Donsker, which means that $\mathbb{G}_n \psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})$ converges weekly to a mean zero Gaussian process with the covariance $\Sigma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{P}\psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_1)\psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_2)^{\mathrm{T}}$. By Theorem 2.1, it remains to show that $\tilde{V}(\boldsymbol{\theta}) \xrightarrow{p} \Gamma(\boldsymbol{\theta})$ for any fixed $\boldsymbol{\theta} \in \Theta_\theta$. In fact, it is seen that $\tilde{V}(\boldsymbol{\theta}) \xrightarrow{p} \Gamma(\boldsymbol{\theta})$ for any fixed $\boldsymbol{\theta} \in \Theta_\theta$ by the law of large numbers.

Lemma 5. If Assumptions (A1)–(A5) hold, then $\tilde{T}_n$ converges in distribution to $\sup_{\boldsymbol{\theta} \in \Theta_\theta} \mathcal{G}_{r_n}^2(\boldsymbol{\theta})$ under H_{1n} as $n \to \infty$, where $\{\mathcal{G}_{r_n}(\boldsymbol{\theta}); \boldsymbol{\theta} \in \Theta_\theta\}$ is a Gaussian process with the mean function $\mu(\boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta})^{-1/2} \frac{\partial \mathbf{P}_0 \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \boldsymbol{\xi}$ and the covariance function $\Sigma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ as defined in Lemma 4. Here, $\mathbf{P}_0$ is the probability measure under the null hypothesis.

Proof of Lemma 5. Under the local alternative, as in Lemma 4, we can show that$(\mathbb{P}_n - \mathbf{P})\{\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})\} = o_p(n^{-1/2})$, and consequently,

$$\begin{aligned}\mathbb{P}_n\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) =& \mathbb{P}_n\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) \\ &+ \mathbf{P}[\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})] + o_p(n^{-1/2}).\end{aligned} \tag{S1.6}$$

By the differentiability of $\mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ in Assumption (A2), we have, for any $\boldsymbol{\alpha} \in \Theta_\alpha$ and $\boldsymbol{\theta} \in \Theta_\theta$, $\mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta}) + \boldsymbol{\beta}^{\mathrm{T}} \frac{\partial \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} + O(\|\boldsymbol{\beta}\|^2)$. Thus, for any $\boldsymbol{\alpha}$ satisfying $\boldsymbol{\alpha} - \boldsymbol{\alpha}_0 = O(n^{-1/2})$, we have, by Assumption (A2) and Taylor's theorem,

$$\begin{aligned}\mathbf{P}[\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})] =& \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) - \boldsymbol{\beta}^{\mathrm{T}} \frac{\partial \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} \\ &- \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}) + \boldsymbol{\beta}^{\mathrm{T}} \frac{\partial \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}} + O(\|\boldsymbol{\beta}\|^2) \\ =& K(\boldsymbol{\beta}, \boldsymbol{\theta})(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) + \boldsymbol{\beta}^{\mathrm{T}} \frac{\partial^2 \mathbf{P}\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\alpha}^{\mathrm{T}}} (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \\ &+ O(\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|^2) + O(\|\boldsymbol{\beta}\|^2),\end{aligned}$$

which combining Lemma 2 implies that, with probability $1-o(1)$, $\mathbf{P}[\psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) -$

$\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})] = K(\boldsymbol{\beta}, \boldsymbol{\theta}) J \mathbb{P}_n \psi_1(\boldsymbol{V}, \alpha_0) + O(n^{-1})$, which implies that

$$\begin{aligned}\mathbb{P}_n \psi(\boldsymbol{V}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) =& \mathbb{P}_n \{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) + K(\boldsymbol{\beta}, \boldsymbol{\theta}) J \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\} + o_p(n^{-1/2}) \\ =& \mathbb{P}_n \{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}) + K(\boldsymbol{\beta}, \boldsymbol{\theta}) J \psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)\} \\ &+ \mathbb{P}_n \{\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) - \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})\} + o_p(n^{-1/2}).\end{aligned}$$

Let $\mathbf{P}_0$ denote the probability measure under the null. Since under the local alternative, $\mu(\boldsymbol{\theta}) = \Gamma(\boldsymbol{\theta})^{-1/2} \frac{\partial \mathbf{P}_0 \psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \boldsymbol{\xi}$, we have, along the line of Theorem 2.1, $\mathbb{G}_n \psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})$ converges weekly to a Gaussian process with mean function $\mu(\boldsymbol{\theta})$ and the covariance $\Sigma(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \mathbf{P} \psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_1) \psi_*(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_2)^{\mathrm{T}}$. Therefore, we complete the proof by (S1.6).

Lemma 6. Let $H = \int_{\boldsymbol{\theta} \in \Theta_\theta} J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} K(\boldsymbol{\theta}) J w(\boldsymbol{\theta}) d\boldsymbol{\theta}$, where J and $K(\boldsymbol{\theta})$ are defined in Assumption (A5), and $w(\boldsymbol{\theta})$ is a weight satisfying $w(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta} \in \Theta_\theta$ and $\int_{\boldsymbol{\theta} \in \Theta_\theta} w(\boldsymbol{\theta}) d\boldsymbol{\theta} = 1$. If Assumption (A5) holds, we have

$$\lambda_{\max}(H) \leq \lambda_{\max}^2(J) \lambda_{\max} \left(\int_{\boldsymbol{\theta} \in \Theta_\theta} \Sigma_K(\boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta} \right),$$

and

$$\operatorname{trace}(H) \leq \lambda_{\max}^2(J) \operatorname{trace} \left(\int_{\boldsymbol{\theta} \in \Theta_\theta} \Sigma_K(\boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta} \right),$$

where $\Sigma_K(\boldsymbol{\theta}) = K(\boldsymbol{\theta})^{\mathrm{T}} K(\boldsymbol{\theta})$. Furthermore, by Assumption (A5), $\lambda_{\max}(H)$ and $\operatorname{trace}(H)$ are bounded.

Proof of Lemma 6. It is straightforward to complete the proof by utilizing Assumption (A5).

Lemma 7. If Assumption (A4) holds, we have for any fixed $\boldsymbol{\xi} \in \Theta_\beta$,

$$\begin{aligned}\mathbf{P}[h(\boldsymbol{V}_1, \boldsymbol{V}_2)^2] \leq & 4 \int_{\boldsymbol{\theta} \in \Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta})) \mathrm{trace}(\Sigma_\psi(\boldsymbol{\theta})) w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ & + 4\lambda_{\max}^2(\Sigma_{\psi_1}) \mathrm{trace}(H^2) \\ & + 8\lambda_{\max}(\Sigma_{\psi_1}) \lambda_{\max}^2(J) \int_{\boldsymbol{\theta} \in \Theta_\theta} \lambda_{\max}(\Sigma_\psi) \mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right) w(\boldsymbol{\theta}) d\boldsymbol{\theta},\end{aligned}$$

where $h(\boldsymbol{V}_1, \boldsymbol{V}_2)$ is defined in (10) in the main paper, and $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are generated from the null distribution with density $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$ and are independent of each other. Furthermore, if Assumption (A5) holds, $\mathrm{E}[h(\boldsymbol{V}_1, \boldsymbol{V}_2)^2]$ is bounded.

Here we defined matrices as following,

$$\Sigma_\psi(\boldsymbol{\theta}) = \mathbf{P}\left[\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\otimes 2}\right], \Sigma_{\psi_1}(\boldsymbol{\theta}) = \mathbf{P}\left[\psi_1(\boldsymbol{V}, \boldsymbol{\alpha}_0)^{\otimes 2}\right], \Sigma_K(\boldsymbol{\theta}) = K(\boldsymbol{\theta})^{\mathrm{T}} K(\boldsymbol{\theta}).$$

Proof of Lemma 7. We deal with each term in $h(\boldsymbol{V}_1, \boldsymbol{V}_2)$ separately. It is seen that

$$\begin{aligned}& \mathbf{P}\left[\left\{\int_{\boldsymbol{\theta} \in \Theta_\theta} \psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_j, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}\right\}^2\right] \\ \leq & \int_{\boldsymbol{\theta} \in \Theta_\theta} \mathbf{P}\left[\left(\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_j, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})\right)^2\right] w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ \leq & \int_{\boldsymbol{\theta} \in \Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta})) \mathbf{P}\left[\|\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})\|^2\right] w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ \leq & \int_{\boldsymbol{\theta} \in \Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta})) \mathrm{trace}(\Sigma_\psi(\boldsymbol{\theta})) w(\boldsymbol{\theta}) d\boldsymbol{\theta}.\end{aligned}$$

By the definition of K_i, we have

$$\begin{aligned}\mathbf{P}\|K_i\|^2 \leq & \int_{\boldsymbol{\theta}\in\Theta_\theta} \text{trace}\left\{\mathbf{P}\left[J^{\mathrm{T}}K(\boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})^{\mathrm{T}}K(\boldsymbol{\theta})J\right]\right\} w(\boldsymbol{\theta})d\boldsymbol{\theta} \\ \leq & \int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\text{trace}\left(J^{\mathrm{T}}K(\boldsymbol{\theta})^{\mathrm{T}}K(\boldsymbol{\theta})J\right) w(\boldsymbol{\theta})d\boldsymbol{\theta} \\ = & \lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi)\text{trace}\left(\Sigma_K(\boldsymbol{\theta})\right) w(\boldsymbol{\theta})d\boldsymbol{\theta},\end{aligned}$$

and consequently,

$$\mathbf{P}\left[(\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j)^2\right] = \mathbf{P}\left[\left(K_j^{\mathrm{T}}\Sigma_{\psi_1}K_j\right)\right] \leq \lambda_{\max}(\Sigma_{\psi_1})\lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi)\text{trace}\left(\Sigma_K(\boldsymbol{\theta})\right) w(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

And similarly, it can be obtained that

$$\mathbf{P}\left[\left(\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right)^2\right] \leq \lambda^2_{\max}(\Sigma_{\psi_1})\text{trace}(H^2).$$

From the foregoing considerations, it is seen that

$$\begin{aligned}\mathbf{P}[h(\boldsymbol{V}_1,\boldsymbol{V}_2)^2] \leq & 4\mathbf{P}\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0) + 4\mathbf{P}\left[S_0(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}HS_0(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right] + 8\mathbf{P}\left[S_0(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j\right] \\ \leq & 4\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\text{trace}(\Sigma_\psi(\boldsymbol{\theta}))w(\boldsymbol{\theta})d\boldsymbol{\theta} \\ & + 4\lambda^2_{\max}(\Sigma_{\psi_1})\text{trace}(H^2) \\ & + 8\lambda_{\max}(\Sigma_{\psi_1})\lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi)\text{trace}\left(\Sigma_K(\boldsymbol{\theta})\right) w(\boldsymbol{\theta})d\boldsymbol{\theta}.\end{aligned}$$

By Assumption (A5) and Lemma 6, it is straightforeward to see that $\mathrm{E}[h(\boldsymbol{V}_1,\boldsymbol{V}_2)^2]$ is bounded.

Lemma 8. If Assumption (A4) – (A6) holds, we have for any fixed $\boldsymbol{\xi}\in\Theta_\beta$,

$$\mathbf{P}[h(\boldsymbol{V}_1,\boldsymbol{V}_2)^2\{1+b(\boldsymbol{V}_1,\boldsymbol{\xi})\}\{1+b(\boldsymbol{V}_2,\boldsymbol{\xi})\}]$$

is bounded, where $b(\boldsymbol{v}, \boldsymbol{\xi})$ is defined in Assumption (A6), $h(\boldsymbol{V}_1, \boldsymbol{V}_2)$ is defined in (10) in the main paper, and $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are generated from the null distribution with density $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$ and are independent of each other.

Proof of Lemma 8. We defined $\Sigma_{\psi,b}$ and $\Sigma_{\psi_1,b}$ as following,

$$\Sigma_{\psi,b} = \mathbf{P}\left[b(\boldsymbol{V}_i, \boldsymbol{\xi})\psi(\boldsymbol{V}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\otimes 2}\right],$$

$$\Sigma_{\psi_1,b} = \mathbf{P}\left[b(\boldsymbol{V}_i, \boldsymbol{\xi})\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\otimes 2}\right].$$

Along the lines of Lemma 7, we have by Assumption (A6),

$$\begin{aligned}
&\mathbf{P}\left[\left\{\int_{\boldsymbol{\theta}\in\Theta_\theta} \psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta}\right\}^2 b(\boldsymbol{V}_i, \boldsymbol{\xi})\right] \\
\leq &\int_{\boldsymbol{\theta}\in\Theta_\theta} \mathbf{P}\left[\left\{\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})\right\}^2 b(\boldsymbol{V}_i, \boldsymbol{\xi})\right] w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
\leq &\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi)\mathbf{P}\left[\|\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})\|^2 b(\boldsymbol{V}_i, \boldsymbol{\xi})\right] w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
\leq &\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\mathrm{trace}(\Sigma_{\psi,b}(\boldsymbol{\theta}))w(\boldsymbol{\theta})d\boldsymbol{\theta},
\end{aligned}$$

By the definition of K_i, we have

$$\begin{aligned}
&\mathbf{P}\left[\|K_i\|^2 b(\boldsymbol{V}_i, \boldsymbol{\xi})\right] \\
\leq &\int_{\boldsymbol{\theta}\in\Theta_\theta} \mathrm{trace}\left\{\mathbf{P}\left[J^{\mathrm{T}}K(\boldsymbol{\theta})^{\mathrm{T}}S(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})S(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}}K(\boldsymbol{\theta})Jb(\boldsymbol{V}_i, \boldsymbol{\xi})\right]\right\} w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
\leq &\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(J^{\mathrm{T}}K(\boldsymbol{\theta})^{\mathrm{T}}K(\boldsymbol{\theta})J\right) w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
= &\lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right) w(\boldsymbol{\theta})d\boldsymbol{\theta},
\end{aligned}$$

and consequently,

$$\begin{aligned}\mathbf{P}\left[\{\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})\right] =&\mathbf{P}\left[K_j^{\mathrm{T}}\Sigma_{\psi_1,b}K_j\right]\\ \leq&\lambda_{\max}(\Sigma_{\psi_1,b})\lambda_{\max}^2(J)\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta},\end{aligned}$$

and

$$\begin{aligned}\mathbf{P}\left[\{\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})\right] =&\mathbf{P}\left[K_i^{\mathrm{T}}\Sigma_{\psi_1}K_i b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]\\ \leq&\lambda_{\max}(\Sigma_{\psi_1})\lambda_{\max}^2(J)\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta}.\end{aligned}$$

And similarly, it can be obtained that

$$\mathbf{P}\left[\left\{\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]\leq\lambda_{\max}(\Sigma_{\psi_1,b})\lambda_{\max}(\Sigma_{\psi_1})\mathrm{trace}(H^2).$$

From the foregoing considerations, it is seen that by Assumption (A7),

$$\begin{aligned}&\mathbf{P}[h(\boldsymbol{V}_i,\boldsymbol{V}_j)^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})]\\ \leq&4\mathbf{P}\left[\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]\\ &+4\mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]\\ &+4\mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]+4\mathbf{P}\left[\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]\\ \leq&4\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\mathrm{trace}(\Sigma_{\psi,b}(\boldsymbol{\theta}))w(\boldsymbol{\theta})d\boldsymbol{\theta}\\ &+4\lambda_{\max}(\Sigma_{\psi_1,b})\lambda_{\max}^2(J)\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_\psi(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta}\\ &+4\lambda_{\max}(\Sigma_{\psi_1})\lambda_{\max}^2(J)\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta}\\ &+4\lambda_{\max}(\Sigma_{\psi_1,b})\lambda_{\max}(\Sigma_{\psi_1})\mathrm{trace}(H^2).\end{aligned}\tag{S1.7}$$

Furthermore, we have

$$
\begin{aligned}
&\mathbf{P}\left[\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]\\
=&\mathbf{P}\left[\left\{\int_{\boldsymbol{\theta}\in\Theta_\theta}\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta}\right\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]\\
\leq&\int_{\boldsymbol{\theta}\in\Theta_\theta}\mathbf{P}\left[\left\{\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})\right\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]w(\boldsymbol{\theta})d\boldsymbol{\theta}\\
\leq&\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathbf{P}\left[\left\|\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})\right\|^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})\right]w(\boldsymbol{\theta})d\boldsymbol{\theta}\\
\leq&\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}(\Sigma_{\psi,b}(\boldsymbol{\theta}))w(\boldsymbol{\theta})d\boldsymbol{\theta},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{P}\left[\{\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]=&\mathbf{P}\left[K_j^{\mathrm{T}}\Sigma_{\psi_1,b}K_j b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]\\
\leq&\lambda_{\max}(\Sigma_{\psi_1,b})\lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta}\lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta},
\end{aligned}
$$

and $\mathbf{P}\left[\left\{\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right\}^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right]\leq\lambda^2_{\max}(\Sigma_{\psi_1,b})\mathrm{trace}(H^2)$.

Therefore, we have

$$
\begin{aligned}
&\mathbf{P}[h(\boldsymbol{V}_i,\boldsymbol{V}_j)^2 b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})] \\
\leq& 4\mathbf{P}\left[\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right] \\
&+ 4\mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}} H \psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right] \\
&+ 4\mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}} K_j b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right] + 4\mathbf{P}\left[\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)^{\mathrm{T}} K_i b(\boldsymbol{V}_i,\boldsymbol{\xi})b(\boldsymbol{V}_j,\boldsymbol{\xi})\right] \\
\leq& 4\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}(\Sigma_{\psi,b}(\boldsymbol{\theta}))w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&+ 8\lambda_{\max}(\Sigma_{\psi_1,b})\lambda^2_{\max}(J)\int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}(\Sigma_{\psi,b}(\boldsymbol{\theta}))\mathrm{trace}\left(\Sigma_K(\boldsymbol{\theta})\right)w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&+ 4\lambda^2_{\max}(\Sigma_{\psi_1,b})\mathrm{trace}(H^2),
\end{aligned}
$$

which combining with (S1.7) and Lemmas 6 and 7 implies that $\mathbf{P}[h(\boldsymbol{V}_1,\boldsymbol{V}_2)^2\{1+b(\boldsymbol{V}_i,\boldsymbol{\xi})\}\{1+b(\boldsymbol{V}_j,\boldsymbol{\xi})\}]$ is bounded, and consequently completes the proof.

Lemma 9. If Assumption (A4) – (A5) and (A7) holds, we have for any fixed $\boldsymbol{\xi}_1 \in \Theta_\alpha$,

$$\mathbf{P}[h(\boldsymbol{V}_1,\boldsymbol{V}_2)^2\{1+b_1(\boldsymbol{V}_1,\boldsymbol{\xi}_1)\}\{1+b_1(\boldsymbol{V}_2,\boldsymbol{\xi}_1)\}]$$

is bounded, where $b_1(\boldsymbol{v},\boldsymbol{\xi}_1)$ is defined in Assumption (A7), $h(\boldsymbol{V}_1,\boldsymbol{V}_2)$ is defined in (10) in the main paper, and $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are generated from the null distribution with density $f(\boldsymbol{v},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)$ and are independent of each other.

Proof of Lemma 9. We defined Σ_{ψ,b_1} and Σ_{ψ_1,b_1} as following,

$$\begin{aligned}\Sigma_{\psi,b_1} =&\mathbf{P}\left[b_1(\boldsymbol{V}_i,\boldsymbol{\xi}_1)\psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})^{\otimes 2}\right],\\ \Sigma_{\psi_1,b_1} =&\mathbf{P}\left[b_1(\boldsymbol{V}_i,\boldsymbol{\xi}_1)\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\otimes 2}\right].\end{aligned}$$

It is straightforward to complete the proof by the same arguments in Lemma 8.

Proof of Theorem 1 in the main paper. For easy to express, we give some additional notations as follows.

$$\begin{aligned}K =&\int_{\boldsymbol{\theta}\in\Theta_\theta} K(\boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta},\\ \tilde{\psi}_i =&\int_{\boldsymbol{\theta}\in\Theta_\theta} \psi(\boldsymbol{V}_i,\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta},\\ \tilde{\psi}_{ij}(\boldsymbol{\alpha}) =&\int_{\boldsymbol{\theta}\in\Theta_\theta} \psi(\boldsymbol{V}_i,\boldsymbol{\alpha},\mathbf{0},\boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j,\boldsymbol{\alpha},\mathbf{0},\boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta}.\end{aligned}$$

By Lemma 3, we have $\mathbb{P}_n\psi(\boldsymbol{V},\hat{\boldsymbol{\alpha}},\mathbf{0},\boldsymbol{\theta}) = \mathbb{P}_n\psi(\boldsymbol{V},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta})+K(\boldsymbol{\theta})J\Psi_{1n}(\boldsymbol{\alpha}_0)+o_p(n^{-1/2})$, and

$$\begin{aligned}\frac{1}{n}\sum_{i\neq j}\tilde{\psi}_{ij}(\hat{\boldsymbol{\alpha}}) =&\frac{1}{n}\sum_{i\neq j}\int_{\boldsymbol{\theta}\in\Theta_\theta}\psi(\boldsymbol{V}_i,\hat{\boldsymbol{\alpha}},\mathbf{0},\boldsymbol{\theta})^{\mathrm{T}}\psi(\boldsymbol{V}_j,\hat{\boldsymbol{\alpha}},\mathbf{0},\boldsymbol{\theta})w(\boldsymbol{\theta})d\boldsymbol{\theta}\\ =&\frac{1}{n}\sum_{i\neq j}\left\{\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)+\frac{1}{n}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i+\frac{1}{n}K_j^{\mathrm{T}}\Psi_{1n}(\boldsymbol{\alpha}_0)+\frac{1}{n^2}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}H\Psi_{1n}(\boldsymbol{\alpha}_0)\right\}\\ &+\frac{1}{n}\sum_{i\neq j}\left\{\tilde{\psi}_i+\frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\}o_p(n^{-1/2})\\ &+\frac{1}{n}\sum_{i\neq j}\left\{\tilde{\psi}_j+\frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\}o_p(n^{-1/2})+o_p(n^{-1}).\end{aligned}$$

By notations, we have, for any $i \neq j$, $\mathrm{E}[K_i] = \mathbf{0}$ and $\mathrm{E}[\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)] = 0$. It can be shown by the central limit theorem (CLT) that

$$\frac{1}{n}\sum_{i\neq j}\left\{\tilde{\psi}_i + \frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\} = \frac{n-1}{n}\sum_{i=1}^{n}\left\{\tilde{\psi}_i + \frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\} = O_p(n^{1/2}),$$

which implies that

$$\begin{aligned}
T_n =& \frac{1}{n(n-1)}\sum_{i\neq j}\tilde{\psi}_{ij}(\hat{\boldsymbol{\alpha}}) \\
=& \frac{1}{n(n-1)}\sum_{i\neq j}\left\{\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0) + \frac{1}{n}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i + \frac{1}{n}K_j^{\mathrm{T}}\Psi_{1n}(\boldsymbol{\alpha}_0) + \frac{1}{n^2}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}H\Psi_{1n}(\boldsymbol{\alpha}_0)\right\} \\
& + o_p(n^{-1}) \\
\equiv& T_{n1} + T_{n2} + T_{n3} + o_p(n^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
T_{n1} =& \frac{1}{n(n-1)}\sum_{i\neq j}\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0), \\
T_{n2} =& \frac{2}{n^2}\sum_{i=1}^{n}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i, \\
T_{n3} =& \frac{1}{n^2}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}H\Psi_{1n}(\boldsymbol{\alpha}_0).
\end{aligned}$$

By some tedious calculations we have that $\mathbf{P}T_{n2} = \frac{2}{n^2}\sum_{i=1}^{n}\mathbf{P}[\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\mathrm{T}}K_i] = \frac{2}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1, \boldsymbol{\alpha}_0)^{\mathrm{T}}K_1]$, and $\mathbf{P}T_{n3} = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{P}[\psi_1(\boldsymbol{V}_k, \boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_k, \boldsymbol{\alpha}_0)] =$

$\frac{1}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)]$, which implies that

$$\begin{aligned}\mathbf{P}\left[T_{n1}+T_{n2}+T_{n3}\right] =& \frac{2}{n}\mathbf{P}\left[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_1\right] + \frac{1}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)]\Big] \\ &+ O(n^{-2}) \\ \equiv& \frac{\mu_0}{n} + O(n^{-2}).\end{aligned}$$

Thus, we can rewrite T_n as $T_n - \frac{\mu_0}{n} = \tilde{T}_n + R_n + o_p(n^{-1})$, where $\tilde{T}_n = \frac{1}{n(n-1)}\sum_{i\neq j} h(\boldsymbol{V}_i,\boldsymbol{V}_j)$ with the kernel given in (10) in the main paper

$$h(\boldsymbol{V}_i,\boldsymbol{V}_j) = \tilde{\psi}_{ij}(\boldsymbol{\alpha}_0) + \psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j + K_i^{\mathrm{T}}\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0) + \psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)$$

and

$$\begin{aligned}R_n =& -\frac{\mu_0}{n} + \frac{2}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i + \frac{1}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0) \\ &+ \frac{2}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j - \frac{1}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0).\end{aligned}$$

We are now going to verify that $R_n = o_p(n^{-1})$. In fact, we have, by the law of large numbers,

$$\frac{2}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i + \frac{1}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0) - \frac{\mu_0}{n} = O_p(n^{-2}),$$

and

$$\frac{2}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j - \frac{1}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0) = o_p(n^{-1}).$$

Therefor, to complete the proof, it remains to show that $n\tilde{T}_n \xrightarrow{\mathcal{L}} \zeta_0$, where ζ_0 is a random variable of the form $\zeta_0 = \sum_{j=1}^{\infty}\lambda_j(\chi^2_{1j}-1)$, and

$\chi^2_{11}, \chi^2_{12}, ...$ are independent χ^2_1 variables, that is, ζ_0 has characteristic function $E\left[e^{it\zeta_0}\right] = \prod_{j=1}^{\infty}(1-2it\lambda_j)^{-\frac{1}{2}}e^{-it\lambda_j}$. In fact, this can be shown by Theorem in 5.5.2 of Serfling (1980) by checking its conditions. It can be shown that $g_1(v_1) = \mathrm{E}[h(\boldsymbol{V}_1, \boldsymbol{V}_2)|\boldsymbol{V}_1 = v_1] = 0$, and by Lemma 7, $\mathrm{Var}(h(\boldsymbol{V}_1, \boldsymbol{V}_2)) = \mathrm{E}[h(\boldsymbol{V}_1, \boldsymbol{V}_2)^2] = \nu^2 < \infty$.

Proof of Theorem 2 in the main paper. For easy to express, we use the same notations K, $\tilde{\psi}_i$, K_i, H and $\tilde{\psi}_{ij}(\boldsymbol{\alpha})$ in the proof of Theorem 1. Under the alternative hypothesis and by Assumption (A2), we have, for any $i \neq j$,

$$\begin{aligned}
\mathbf{P}[\tilde{\psi}_i] =& \int_{\boldsymbol{\theta}\in\Theta_\theta} \mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta})]w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
=& \int_{\boldsymbol{\theta}\in\Theta_\theta} \mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})]w(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&+ \boldsymbol{\beta}^{\mathrm{T}} \int_{\boldsymbol{\theta}\in\Theta_\theta} \partial\mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta})/\partial\boldsymbol{\beta}]w(\boldsymbol{\theta})d\boldsymbol{\theta}(1+o(1)) \\
=& O(n^{-1/2}),
\end{aligned}$$

and $\mathbf{P}[\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)] = \boldsymbol{\beta}^{\mathrm{T}}\partial\mathbf{P}[\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)]/\partial\boldsymbol{\beta}(1+o(1)) = O(n^{-1/2})$. It can be shown by the law of large numbers that

$$\frac{1}{n}\sum_{i\neq j}\left\{\tilde{\psi}_i + \frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\} = \frac{n-1}{n}\sum_{i=1}^{n}\left\{\tilde{\psi}_i + \frac{1}{n}KJ\Psi_{1n}(\boldsymbol{\alpha}_0)\right\} = O_p(n^{1/2}),$$

which, by a similar arguments as in the proof of Theorem 1, implies that $T_n \equiv T_{n1} + T_{n2} + T_{n3} + o_p(n^{-1})$, where $T_{n1} = \frac{1}{n(n-1)}\sum_{i\neq j}\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0), T_{n2} =$

$\frac{2}{n^2}\sum_{i=1}^{n}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i$, and $T_{n3} = \frac{1}{n^2}\Psi_{1n}(\boldsymbol{\alpha}_0)^{\mathrm{T}}H\Psi_{1n}(\boldsymbol{\alpha}_0)$. By some tedious calculations we have that $\mathbf{P}T_{n2} = \frac{2}{n^2}\sum_{i=1}^{n}\mathbf{P}[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i] = \frac{2}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_1]$, and $\mathbf{P}T_{n3} = \frac{1}{n^2}\sum_{k=1}^{n}\mathbf{P}[\psi_1(\boldsymbol{V}_k,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_k,\boldsymbol{\alpha}_0)] = \frac{1}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)]$, which implies that

$$\begin{aligned}\mathbf{P}\left[T_{n2}+T_{n3}\right] =& \frac{2}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_1] + \frac{1}{n}\mathbf{P}[\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_1,\boldsymbol{\alpha}_0)] + O(n^{-2})\\ \equiv& \frac{\mu_0}{n} + O(n^{-2}).\end{aligned}$$

Thus, we can rewrite T_n as $T_n - \frac{\mu_0}{n} = \tilde{T}_n + R_n + o_p(n^{-1})$, where $\tilde{T}_n = \frac{1}{n(n-1)}\sum_{i\neq j}h(\boldsymbol{V}_i,\boldsymbol{V}_j)$ with the kernel $h(\boldsymbol{V}_i,\boldsymbol{V}_j)$ defined in (10) in the main paper and

$$\begin{aligned}R_n =& -\frac{\mu_0}{n} + \frac{2}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_i + \frac{1}{n^2}\sum_{i=1}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)\\ &+ \frac{2}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j - \frac{1}{n^2(n-1)}\sum_{i\neq j}^{n}\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0).\end{aligned}$$

Similarly to the proof of Theorem 1, we can verify that $R_n = o_p(n^{-1})$.

Therefore, we complete the proof if we can show that $n\tilde{T}_n \xrightarrow{\mathcal{L}} \zeta_1$, and consequently by Theorem 9.5.2 of Resnick (2019), it suffices to show that for each $t \in \mathbb{R}$, $\left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n}\right] - \mathbf{P}\left[e^{it\zeta_1}\right]\right| \to 0$, as n goes to infinity.

Let $\{\boldsymbol{V}_{0i}, i = 1,\cdots,n\}$ be independent and identically distributed sample for each $\boldsymbol{V}_{0i}$ from the null distribution with density $f(\boldsymbol{v},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)$. We show this Lemma along the lines of Theorem in Section 5.5.2 of Serfling (1980). As shown in Theorem in Section 5.5.2 of Serfling (1980), we rewrite,

by the representation of $h(\boldsymbol{v}_1, \boldsymbol{v}_2)$ (Page 1087 in Dunford and Schwartz (1963)), $h(\boldsymbol{v}_1, \boldsymbol{v}_2) = \sum_{k=1}^{\infty} \lambda_k \phi_k(\boldsymbol{v}_1)\phi_k(\boldsymbol{v}_2)$, where $\{\phi_k(\cdot)\}$ and $\{\lambda_k\}$ have following properties that

$$\mathbf{P}[\phi_j(\boldsymbol{V}_{01})\phi_k(\boldsymbol{V}_{01})] = \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise,} \end{cases} \qquad \mathbf{P}[\phi_k(\boldsymbol{V}_{01})] = 0, \quad \text{for all } k, \tag{S1.8}$$

and $\mathbf{P}[h(\boldsymbol{V}_{01}, \boldsymbol{V}_{02})^2] = \sum_{k=1}^{\infty} \lambda_k^2 < \infty$.

Let $T_{nK} = \frac{1}{n}\sum_{i\neq j}\sum_{k=1}^{K} \lambda_k \phi_k(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_j)$. According to the fact that $|e^{it} - 1| \leq |t|$, we have

$$\begin{aligned} \left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n}\right] - \mathbf{P}\left[e^{itT_{nK}}\right]\right| \leq & \mathbf{P}\left[\left|e^{it(n-1)\tilde{T}_n} - e^{itT_{nK}}\right|\right] \\ \leq & |t|\mathbf{P}\left[\left|(n-1)\tilde{T}_n - T_{nK}\right|\right] \\ \leq & |t|\left\{\mathbf{P}\left[(n-1)\tilde{T}_n - T_{nK}\right]^2\right\}^{1/2}. \end{aligned}$$

Next we claim that

$$\mathbf{P}\left[(n-1)\tilde{T}_n - T_{nK}\right]^2 \leq 4(2C_\xi^2 + 1)\sum_{k=K+1}^{\infty}\lambda_k^2 + \left(\sum_{k=K+1}^{\infty}\lambda_k\mu_{ak}^2\right)^2. \tag{S1.9}$$

In fact, $(n-1)\tilde{T}_n - T_{nK}$ has basically the form of a U-statistic, that is, $(n-1)\tilde{T}_n - T_{nK} = \frac{2}{n}\binom{n}{2}U_{nK}$, where $U_{nK} = \frac{1}{n(n-1)}\sum_{i\neq j} g_K(\boldsymbol{V}_i, \boldsymbol{V}_j)$ with

$$g_K(\boldsymbol{V}_i, \boldsymbol{V}_j) = h(\boldsymbol{V}_i, \boldsymbol{V}_j) - \sum_{k=1}^{K}\lambda_k\phi_k(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_j) = \sum_{k=K+1}^{\infty}\lambda_k\phi_k(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_j).$$

Before showing the inequality (S1.9), we derive following useful inequalities. It is seen that $\mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})] = \mathbf{P}\left[\psi(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\frac{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0)}{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right]$,

and consequently, by Assumption (A6) and Cauchy-Schwarz inequality, there is a constant $C_{\boldsymbol{\xi}}$ such that $\|\mathbf{P}\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\| \leq \frac{1}{\sqrt{n}} C_{\boldsymbol{\xi}} \left\{\mathbf{P}\|\psi(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\|^2\right\}^{1/2}$. By Assumption (A5), it is followed from the above that there is a constant $C_1 > 0$ such that

$$\begin{aligned}
\mathbf{P}[\tilde{\psi}_{ij}(\boldsymbol{\alpha}_0)] =& \int_{\boldsymbol{\theta}\in\Theta_\theta} \|\mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})]\|^2 \, w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
\leq& \frac{1}{n} C_{\boldsymbol{\xi}}^2 \int_{\boldsymbol{\theta}\in\Theta_\theta} \mathbf{P}\|\psi(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\|^2 w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
=& \frac{1}{n} C_1 C_{\boldsymbol{\xi}}^2,
\end{aligned}$$

and $\mathbf{P}[K_i] = \int_{\boldsymbol{\theta}\in\Theta_\theta} J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} \mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})] w(\boldsymbol{\theta}) d\boldsymbol{\theta}$, and by Assumption (A3), there is a constant $C_3 > 0$ such that

$$\begin{aligned}
\|\mathbf{P}[K_i]\|^2 \leq& \int_{\boldsymbol{\theta}\in\Theta_\theta} \|J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} \mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})]\|^2 w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
\leq& \lambda_{\max}^2(J) \int_{\boldsymbol{\theta}\in\Theta_\theta} \lambda_{\max}\left(K(\boldsymbol{\theta}) K(\boldsymbol{\theta})^{\mathrm{T}}\right) \|\mathbf{P}\left[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\right]\|^2 w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
\leq& \frac{1}{n} C_{\boldsymbol{\xi}}^2 \lambda_{\max}^2(J) \int_{\boldsymbol{\theta}\in\Theta_\gamma} \lambda_{\max}\left(K(\boldsymbol{\theta}) K(\boldsymbol{\theta})^{\mathrm{T}}\right) \mathbf{P}\|\psi(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, 0, \boldsymbol{\theta})\|^2 w(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
\leq& \frac{1}{n} C_3 C_{\boldsymbol{\xi}}^2.
\end{aligned}$$

As the same arguments, we have $\mathbf{P}[\psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)] = \mathbf{P}\left[\psi_1(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0) \frac{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0)}{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right]$, and there are constant $C_2 > 0$,

$$\|\mathbf{P}[\psi(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)]\| \leq \frac{1}{\sqrt{n}} C_{\boldsymbol{\xi}} \left\{\mathbf{P}\left[\|\psi_1(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0)\|^2\right]\right\}^{1/2} \leq \frac{1}{\sqrt{n}} C_2 C_{\boldsymbol{\xi}}.$$

From the foregoing considerations, we have

$$\begin{aligned}|\mathbf{P}[h(\boldsymbol{V}_i,\boldsymbol{V}_j)]| =& \int_{\boldsymbol{\theta}\in\Theta_\gamma} \|\mathbf{P}[\psi(\boldsymbol{V}_i,0,\boldsymbol{\theta})]\|^2 w(\boldsymbol{\theta})d\boldsymbol{\theta} \\ &+ \mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}K_j + K_i^{\mathrm{T}}\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right] \\ &+ \mathbf{P}\left[\psi_1(\boldsymbol{V}_i,\boldsymbol{\alpha}_0)^{\mathrm{T}}H\psi_1(\boldsymbol{V}_j,\boldsymbol{\alpha}_0)\right] \\ \leq& \frac{1}{n}C_1C_{\boldsymbol{\xi}}^2 + 2\frac{1}{n}C_3C_{\boldsymbol{\xi}}^2 + \frac{1}{n}C_2C_{\boldsymbol{\xi}}^2 \\ =& O(n^{-1}),\end{aligned} \tag{S1.10}$$

and by the mean value theorem and combining the dominated convergence theorem and Assumption (A6),

$$\begin{aligned}\mathbf{P}[\phi_k(\boldsymbol{V}_i)] =& \mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})\frac{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta}_0)}{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)}\right] \\ =& \frac{1}{n^{1/2}}\mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})\frac{1}{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)}\frac{\partial f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\tilde{\boldsymbol{\beta}},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\boldsymbol{\xi}\right] \\ =& \frac{1}{n^{1/2}}\mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})\frac{1}{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)}\frac{\partial f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\boldsymbol{\xi}\right](1+o(1)) \\ =& \frac{1}{n^{1/2}}\mu_{ak} + o(n^{-1/2}),\end{aligned} \tag{S1.11}$$

where $\tilde{\boldsymbol{\beta}}$ is between $\mathbf{0}$ and $\boldsymbol{\beta}$, $|\mathbf{P}[\phi_k(\boldsymbol{V}_i)]| \leq \frac{1}{\sqrt{n}}C_{\boldsymbol{\xi}}\left\{\mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})^2\right]\right\}^{1/2} = \frac{1}{\sqrt{n}}C_{\boldsymbol{\xi}}$, and

$$\begin{aligned}\mathbf{P}[\phi_k(\boldsymbol{V}_i)^2] =& \mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})^2\frac{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta}_0)}{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\mathbf{0},\boldsymbol{\theta}_0)}\right] \\ \leq& 1 + \frac{1}{\sqrt{n}}\mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})^2 b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\right] \\ =& 1 + O(n^{-1/2}).\end{aligned}$$

Since by Lemma 8 combining Assumption (A6),

$$\begin{aligned}
\mathbf{P}[g_K(\boldsymbol{V}_i, \boldsymbol{V}_j)^2] =& \int\int g_K(\boldsymbol{v}_1, \boldsymbol{v}_2)^2 f(\boldsymbol{v}_1, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0) f(\boldsymbol{v}_2, \boldsymbol{\alpha}_0, \boldsymbol{\beta}, \boldsymbol{\theta}_0) d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\
\leq& \mathbf{P}[g_K(\boldsymbol{V}_{0i}, \boldsymbol{V}_{0j})^2 \{1 + n^{-1/2} b(\boldsymbol{V}_{0i}, \boldsymbol{\xi})\}\{1 + n^{-1/2} b(\boldsymbol{V}_{0j}, \boldsymbol{\xi})\}] \\
\leq& \mathbf{P}[g_K(\boldsymbol{V}_{0i}, \boldsymbol{V}_{0j})^2 \{1 + n^{-1/2} b(\boldsymbol{V}_{0i}, \boldsymbol{\xi})\}^2] \\
=& \sum_{k=K+1}^{\infty} \lambda_k^2 \mathbf{P}[\phi_k(\boldsymbol{V}_{0i})^2 \{1 + n^{-1/2} b(\boldsymbol{V}_{0i}, \boldsymbol{\xi})\}^2] \mathbf{P}[\phi_k(\boldsymbol{V}_{0j})^2] \\
=& \sum_{k=K+1}^{\infty} \lambda_k^2 \mathbf{P}[\phi_k(\boldsymbol{V}_{0i})^2 (1 + n^{-1/2} b(\boldsymbol{V}_{0i}, \boldsymbol{\xi}))^2] \\
\leq& (1 + 3Cn^{-1/2}) \sum_{k=K+1}^{\infty} \lambda_k^2 \\
<& \infty,
\end{aligned} \tag{S1.12}$$

where C is a constant such that $\mathbf{P}[\phi_k(\boldsymbol{V}_i)^2 b(\boldsymbol{V}_{0i}, \boldsymbol{\xi})^2] \leq C$, thus, it is straightforward to obtain that by the dominated theorem and for large n,

$$\mathbf{P}[g_K(\boldsymbol{V}_i, \boldsymbol{V}_j)] = \sum_{k=K+1}^{\infty} \lambda_k (\mathbf{P}[\phi_k(\boldsymbol{V}_i)])^2 = \frac{1}{n} \left(\sum_{k=K+1}^{\infty} \lambda_k \mu_{ak}^2 \right) (1 + o(1)),$$

and

$$\begin{aligned}\mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)^2] =& \int\int g_K(\boldsymbol{v}_1,\boldsymbol{v}_2)^2 f(\boldsymbol{v}_1,\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta}_0) f(\boldsymbol{v}_2,\boldsymbol{\alpha}_0,\boldsymbol{\beta},\boldsymbol{\theta}_0) d\boldsymbol{v}_1 d\boldsymbol{v}_2 \\ \leq& \mathbf{P}[g_K(\boldsymbol{V}_{0i},\boldsymbol{V}_{0j})^2](1+o(1)) \\ =& (1+o(1)) \sum_{k=K+1}^{\infty} \lambda_k^2 \\ \leq& 2 \sum_{k=K+1}^{\infty} \lambda_k^2,\end{aligned}$$

which implies that $\mathrm{Var}(g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)) \leq \mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)^2] \leq 2\sum_{k=K+1}^{\infty}\lambda_k^2$. Let $b(\boldsymbol{v}) = \mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)|\boldsymbol{V}_j = \boldsymbol{v}]$. Noting that $b(\boldsymbol{v}) = \mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)|\boldsymbol{V}_j = \boldsymbol{v}] = \sum_{k=K+1}^{\infty}\lambda_k\phi_k(\boldsymbol{v})\mathbf{P}[\phi_k(\boldsymbol{V}_i)]$, we have, along same lines as (S1.12), for n large enough,

$$\begin{aligned}\mathbf{P}\left[b(\boldsymbol{V}_i)^2\right] =& \mathbf{P}\left[b(\boldsymbol{V}_{0i})^2 \frac{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\theta_0)}{f(\boldsymbol{V}_{0i},\boldsymbol{\alpha}_0,\boldsymbol{0},\theta_0)}\right] \\ \leq& \mathbf{P}\left[b(\boldsymbol{V}_{0i})^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}\right] \\ =& \mathbf{P}\left[(\mathrm{E}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)|\boldsymbol{V}_i = \boldsymbol{V}_{0i}])^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}\right] \\ \leq& \mathbf{P}\left[\mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}|\boldsymbol{V}_i = \boldsymbol{V}_{0i}]\right] \\ =& \mathbf{P}\left[g_K(\boldsymbol{V}_{0i},\boldsymbol{V}_j)^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}\right] \\ =& \mathbf{P}\left[g_K(\boldsymbol{V}_{0i},\boldsymbol{V}_{0j})^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}\frac{f(\boldsymbol{V}_{0j},\boldsymbol{\alpha}_0,\boldsymbol{\beta},\theta_0)}{f(\boldsymbol{V}_{0j},\boldsymbol{\alpha}_0,\boldsymbol{0},\theta_0)}\right] \\ \leq& \mathbf{P}\left[g_K(\boldsymbol{V}_{0i},\boldsymbol{V}_{0j})^2\{1+n^{-1/2}b(\boldsymbol{V}_{0i},\boldsymbol{\xi})\}\{1+n^{-1/2}b(\boldsymbol{V}_{0j},\boldsymbol{\xi})\}\right] \\ <& \infty,\end{aligned}$$

and consequently by the dominated theorem,

$$\begin{aligned}\mathrm{Var}(\mathbf{P}[g_K(\boldsymbol{V}_i,\boldsymbol{V}_j)|\boldsymbol{V}_i]) \leq & \mathbf{P}\left[b(\boldsymbol{V}_i)^2\right] \\ = & \mathbf{P}\left[b(\boldsymbol{V}_{0i})^2\right](1+o(1)) \\ = & (1+o(1))\sum_{k=K+1}^{\infty}\lambda_k^2(\mathbf{P}[\phi_k(\boldsymbol{V}_j)])^2 \\ \leq & \frac{2}{n}C_\xi^2\sum_{k=K+1}^{\infty}\lambda_k^2.\end{aligned}$$

Lemma A in Section 5.2.1 of Serfling (1980) states that

$$\begin{aligned}\mathrm{Var}(U_{nK}) \leq & \binom{n}{2}^{-1}\left(4C_\xi^2\sum_{k=K+1}^{\infty}\lambda_k^2+2\sum_{k=K+1}^{\infty}\lambda_k^2\right) \\ \leq & 2(2C_\xi^2+1)\binom{n}{2}^{-1}\sum_{k=K+1}^{\infty}\lambda_k^2,\end{aligned}$$

which implies that

$$\begin{aligned}\mathbf{P}\left[(n-1)\tilde{T}_n - T_{nK}\right]^2 = & (n-1)^2\left\{\mathrm{Var}(U_{nK})+(\mathbf{P}[U_{nK}])^2\right\} \\ \leq & 4(2C_\xi^2+1)\sum_{k=K+1}^{\infty}\lambda_k^2+(\mathrm{E}[U_{nK}])^2,\end{aligned}$$

and consequently (S1.9) holds.

It is immediately obtained that $n\mathbf{P}[h(\boldsymbol{V}_i,\boldsymbol{V}_j)] < +\infty$ by (S1.10), which implies that $\sum_{k=1}^{\infty}\lambda_k\mu_{ak}^2 = n\mathbf{P}[h(\boldsymbol{V}_i,\boldsymbol{V}_j)] < \infty$, and consequently that $\lim_{K\to\infty}\sum_{k=K+1}^{\infty}\lambda_k\mu_{ak}^2 = 0$. For fixed t and given $\epsilon > 0$, choose and fix K large enough such that $2|t|\left(\sum_{k=K+1}^{\infty}\lambda_k^2\right)^{1/2} < \epsilon$, and $|t|\sum_{k=K+1}^{\infty}\lambda_k\mu_{ak}^2 < \epsilon$. Then it has been shown that for all n sufficiently large, $|t|^2(\mathbf{P}[U_{nK}])^2 =$

$|t|^2 \left(\sum_{k=K+1}^{\infty} \lambda_k \mu_{ak}^2\right)^2 < \epsilon^2$, and immediately,

$$\left| \mathbf{P}\left[e^{it(n-1)\tilde{T}_n}\right] - \mathbf{P}\left[e^{itT_{nK}}\right] \right| < (2C_{\boldsymbol{\xi}}^2 + 2)^{1/2}\epsilon. \tag{S1.13}$$

We now are going to show that

$$T_{nK} \xrightarrow{\mathcal{L}} Y_K = \sum_{k=1}^{K} \lambda_k (W_k^2 - 1), \tag{S1.14}$$

where $\{W_k, k = 1, \cdots, K\}$ are independent and each W_k is normally distributed with mean μ_{ak} and unit variance. We can rewrite $T_{nK} = \sum_{k=1}^{K} \lambda_k (W_{nk}^2 - Z_{nk})$, where $W_{nk} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_k(\boldsymbol{V}_i)$ and $Z_{nk} = \frac{1}{n} \sum_{i=1}^{n} \phi_k^2(\boldsymbol{V}_i)$. Recalling the previous arguments, it can be obtained that

$$\begin{aligned}
\mathbf{P}[W_{nk}] =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{\sqrt{n}} \mu_{ak}(1 + o(1)) = \mu_{ak}(1 + o(1)), \\
\mathrm{Var}(W_{nk}) =& \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(\phi_k(\boldsymbol{V}_i)) = 1 + O(n^{-1/2}) - \frac{1}{n}\mu_{ak}^2 \to 1,
\end{aligned}$$

and if $j \neq k$,

$$\begin{aligned}
\mathrm{Cov}(W_{nj}, W_{nk}) =& \mathbf{P}[W_{nj}W_{nk}] - \mathbf{P}[W_{nj}]\mathbf{P}[W_{nk}] \\
=& \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}[\phi_j(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_i)] + \frac{1}{n} \sum_{i \neq l}^{n} \mathbf{P}[\phi_j(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_l)] - \mu_{aj}\mu_{ak}(1 + o(1)) \\
=& \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}[\phi_j(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_i)] + \frac{1}{n} \sum_{i \neq l}^{n} \frac{1}{n}\mu_{aj}\mu_{ak}(1 + o(1)) - \mu_{aj}\mu_{ak}(1 + o(1)) \\
=& \mathbf{P}[\phi_j(\boldsymbol{V}_i)\phi_k(\boldsymbol{V}_i)] - \frac{1}{n}\mu_{aj}\mu_{ak}(1 + o(1)) \\
=& \mathbf{P}\left[\phi_j(\boldsymbol{V}_{0i})\phi_k(\boldsymbol{V}_{0i})\right](1 + n^{-1/2}C_{\boldsymbol{\xi}}) - \frac{1}{n}\mu_{aj}\mu_{ak}\{1 + o(1)\} \\
\to& 0,
\end{aligned}$$

which implies that if $j \neq k$, $|\mathrm{Cov}(W_{nj}, W_{nk})| \to 0$, and consequently,

$$\mathrm{Cov}(W_{nj}, W_{nk}) \to \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, by Lindeberg-Levy central limit theorem, we have $(W_{n1}, \cdots, W_{nK})^{\mathrm{T}} \xrightarrow{\mathcal{L}} N(\boldsymbol{\mu}_a, \mathbf{1}_{K\times K})$, where $\boldsymbol{\mu}_a = (\mu_{a1}, \cdots, \mu_{aK})^{\mathrm{T}}$. Since $\mathbf{P}[\phi_k^2(\boldsymbol{V}_i)] \to 1$, it is seen that by the strong law of large numbers, $Z_{nk} \xrightarrow{a.s.} 1$, which implies that $(Z_{n1}, \cdots, Z_{nK})^{\mathrm{T}} \xrightarrow{a.s.} (1, \cdots, 1)^{\mathrm{T}}$. Thus, (S1.14) holds and consequently,

$$\left|\mathbf{P}\left[e^{itT_{nK}}\right] - \mathbf{P}\left[e^{itY_K}\right]\right| < \epsilon, \tag{S1.15}$$

for all n sufficiently large.

Noting that by (S1.13) and (S1.15), for all n large enough,

$$\begin{aligned}\left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n}\right] - \mathbf{P}\left[e^{it\zeta_1}\right]\right| \leq & \left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n}\right] - \mathbf{P}\left[e^{itT_{nK}}\right]\right| + \left|\mathbf{P}\left[e^{itT_{nK}}\right] - \mathbf{P}\left[e^{itY_K}\right]\right| \\ & + \left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_1}\right]\right| \\ < & (2C_{\boldsymbol{\xi}}^2 + 2)^{1/2}\epsilon + \left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_1}\right]\right|.\end{aligned}$$

Therefore, to complete the proof, it suffices to establish that for all K sufficiently large,

$$\left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_1}\right]\right| < \epsilon\left\{\mathbf{P}[(W_1^2 - 1)^2]\right\}^{1/2}. \tag{S1.16}$$

Let the random variable W_1, W_2, $\cdots$ be defined on a common probability space and denote by ζ_1 the limit in mean square of Y_K as $K \to \infty$. Then,

by the definition of Y_K, we have

$$\begin{aligned}\left|\mathbf{P}\left[e^{itY_K}\right]-\mathbf{P}\left[e^{it\zeta_1}\right]\right| \leq & |t|\left(\mathbf{P}\left[Y_K-\zeta_1\right]^2\right)^{1/2} \\ \leq & |t|\left(\mathbf{P}\left[W_1^2-1\right]^2\right)^{1/2}\left(\sum_{k=K+1}^{\infty}\lambda_k^2\right)^{1/2} \\ < & \epsilon\left(\mathbf{P}\left[W_1^2-1\right]^2\right)^{1/2},\end{aligned}$$

which arrives at (S1.16), and then completes the proof.

Proof of Theorem 3 in the main paper. It is straightforward to complete the proof if we have $nT_n^*-\mu_0 \xrightarrow{\mathcal{L}} \zeta_0$, where ζ_0 is defined in Theorem 1. Thus, it suffices to show that

$$nT_n^*-\mu_0 \xrightarrow{\mathcal{L}} \zeta_0. \tag{S1.17}$$

Since the above asymptotic distribution is derived given the observed sample $\{\boldsymbol{V}_i, i=1,\cdots,n\}$, we can treat $\hat{\boldsymbol{\alpha}}$ as a deterministic sequence satisfying $\hat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0=\frac{1}{\sqrt{n}}\boldsymbol{\xi}_1$ for fixed $\boldsymbol{\xi}_1 \in \Theta_\alpha$, where we use the fact that $\hat{\boldsymbol{\alpha}}$ is $\sqrt{n}$-consistent by Lemma 2.

Denote by E^* the expectation generated by the resampling from $\boldsymbol{V}$. We

give additional notations as follows.

$$
\begin{aligned}
K^*(\boldsymbol{\theta}) =& \frac{\partial \mathrm{E}^* \psi(\boldsymbol{V}^*, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta})}{\partial \boldsymbol{\alpha}}, \\
K^* =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} K^*(\boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
H^* =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} J^{\mathrm{T}} K^*(\boldsymbol{\theta})^{\mathrm{T}} K^*(\boldsymbol{\theta}) J w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
\tilde{\psi}_i^* =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} \psi(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}}^*, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
\tilde{K}_i^* =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} J^{\mathrm{T}} K^*(\boldsymbol{\theta})^{\mathrm{T}} S(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}}^*, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
K_i^* =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} J^{\mathrm{T}} K(\boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}}^*, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}, \\
\tilde{\psi}_{ij}^*(\boldsymbol{\alpha}) =& \int_{\boldsymbol{\theta} \in \Theta_\gamma} \psi(\boldsymbol{V}_i^*, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})^{\mathrm{T}} \psi(\boldsymbol{V}_j^*, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta}) w(\boldsymbol{\theta}) d\boldsymbol{\theta}.
\end{aligned}
$$

Similar to the proof of Theorem 1, we can obtain that $T_n^* \equiv T_{n1}^* + T_{n2}^* + T_{n3}^* + o_{p^*}(n^{-1})$, where $T_{n1}^* = \frac{1}{n(n-1)} \sum_{i \neq j} \tilde{\psi}_{ij}^*(\hat{\boldsymbol{\alpha}}), T_{n2}^* = \frac{2}{n^2} \sum_{i=1}^n \Psi_{1n}(\hat{\boldsymbol{\alpha}})^{\mathrm{T}} \tilde{K}_i^*$, and $T_{n3}^* = \frac{1}{n^2} \Psi_{1n}(\hat{\boldsymbol{\alpha}})^{\mathrm{T}} H^* \Psi_{1n}(\hat{\boldsymbol{\alpha}})$. Furthermore, we have that

$$
\begin{aligned}
\mathrm{E}^* \left[T_{n1}^* + T_{n2}^* + T_{n3}^* \right] =& \frac{2}{n} \mathrm{E}^* [\psi_1(\boldsymbol{V}^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \tilde{K}_1^*] + \frac{1}{n} \mathrm{E}^* [\psi_1(\boldsymbol{V}^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} H^* \psi_1(\boldsymbol{V}^*, \hat{\boldsymbol{\alpha}})] + O(n^{-2}) \\
\equiv& \frac{\mu_0(\hat{\boldsymbol{\alpha}})}{n} + O(n^{-2}).
\end{aligned}
$$

By the continuity of $\mu_0(\hat{\boldsymbol{\alpha}})$, we have that $\mathrm{E}^* \left[T_{n1}^* + T_{n2}^* + T_{n3}^* \right] = \frac{\mu_0(\hat{\boldsymbol{\alpha}})}{n} (1 + o_p(1))$.

By the Assumption (A7) and the dominated convergence theorem, it is

seen that for all $\boldsymbol{\theta} \in \Theta_\gamma$,

$$\begin{aligned} K^*(\boldsymbol{\theta}) =& \partial \int \psi(\boldsymbol{v}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) f(\boldsymbol{v}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}) d\boldsymbol{v} / \partial \boldsymbol{\alpha} \\ =& \{1 + O(n^{-1/2})\} \partial \int \psi(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}) d\boldsymbol{v} / \partial \boldsymbol{\alpha} \\ =& K(\boldsymbol{\theta}) \{1 + O(n^{-1/2})\}, \end{aligned}$$

and consequently, $K^* = K\{1 + o(1)\}$, and $H^* = H\{1 + o(1)\}$. By the same arguments, we have that $\mu_0(\hat{\boldsymbol{\alpha}}) = \mu_0(1 + o(1))$. Thus, we can rewrite T_n^* as $T_n^* - \frac{\mu_0}{n}\{1 + o(1)\} = \tilde{T}_n^*\{1 + o_{p^*}(1)\} + R_n^* + o_{p^*}(n^{-1})$, where

$$\begin{aligned} \tilde{T}_n^* =& \frac{1}{n(n-1)} \sum_{i \neq j} h(\boldsymbol{V}_i^*, \boldsymbol{V}_j^*, \hat{\boldsymbol{\alpha}}), \\ R_n^* =& -\frac{\mu_0(\hat{\boldsymbol{\alpha}})}{n} + \frac{2}{n^2} \sum_{i=1}^n \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \tilde{K}_i^* + \frac{1}{n^2} \sum_{i=1}^n \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} H^* \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}}) \\ & - \frac{2}{n^2(n-1)} \sum_{i \neq j}^n \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} \tilde{K}_j^* - \frac{1}{n^2(n-1)} \sum_{i \neq j}^n \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} H^* \psi_1(\boldsymbol{V}_j^*, \hat{\boldsymbol{\alpha}}), \end{aligned}$$

and

$$h(\boldsymbol{V}_i^*, \boldsymbol{V}_j^*, \hat{\boldsymbol{\alpha}}) = \tilde{\psi}_{ij}^*(\hat{\boldsymbol{\alpha}}) + \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} K_j^* + (K_i^*)^{\mathrm{T}} \psi_1(\boldsymbol{V}_j^*, \hat{\boldsymbol{\alpha}}) + \psi_1(\boldsymbol{V}_i^*, \hat{\boldsymbol{\alpha}})^{\mathrm{T}} H \psi_1(\boldsymbol{V}_j^*, \hat{\boldsymbol{\alpha}}).$$

We can verify $R_n^* = o_{p^*}(n^{-1})$ in a similar way as in the proof of Theorem 1.

Next, we are going to show that $n\tilde{T}_n^* \xrightarrow{\mathcal{L}} \zeta_0$. Thus, by Theorem 9.5.2 of Resnick (2019), it remains to show that for each $t \in \mathbb{R}$, as n goes to infinity,

$$\left| \mathbf{P}\left[e^{it(n-1)\tilde{T}_n^*} \right] - \mathbf{P}\left[e^{it\zeta_0} \right] \right| \to 0. \tag{S1.18}$$

Let $T_{nK} = \frac{1}{n}\sum_{i\neq j}\sum_{k=1}^{K}\lambda_k\left(\phi_k(\boldsymbol{V}_i^*) - \mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]\right)\left(\phi_k(\boldsymbol{V}_j^*) - \mathbf{P}[\phi_k(\boldsymbol{V}_j^*)]\right)$.

First, we are going to show that for any fixed K,

$$T_{nK} \xrightarrow{\mathcal{L}} Y_K = \sum_{k=1}^{K}\lambda_k(W_k^2 - 1), \tag{S1.19}$$

where $\{W_k, k = 1, \cdots, K\}$ are independent and each W_k is normally distributed with mean μ_{ak} and variance 1. We can rewrite $T_{nK} = \sum_{k=1}^{K}\lambda_k(W_{nk}^2 - Z_{nk})$, where $W_{nk} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\phi_k(\boldsymbol{V}_i^*) - \mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]\}$, and $Z_{nk} = \frac{1}{n}\sum_{i=1}^{n}\{\phi_k(\boldsymbol{V}_i^*) - \mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]\}^2$. By the mean value theorem and combing the dominated convergence theorem and Assumption (A7), it is seen from a similar way with (S1.11) that

$$\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)] = \frac{1}{n^{1/2}}\mu_{ak} + o(n^{-1/2}), \tag{S1.20}$$

and then, we have that $|\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]| \leq \frac{1}{\sqrt{n}}C_f(\boldsymbol{\xi}_1)$, and

$$\begin{aligned}\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)^2] =& \mathbf{P}\left[\phi_k(\boldsymbol{V}_{0i})^2\frac{f(\boldsymbol{V}_{0i}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}_0)}{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right] \\ \leq& 1 + \frac{1}{\sqrt{n}}\mathrm{E}\left[\phi_k(\boldsymbol{V}_{0i})^2 b_1(\boldsymbol{V}_{0i}, \boldsymbol{\xi}_1)\right] \\ =& 1 + \frac{1}{\sqrt{n}}C_f(\boldsymbol{\xi}_1).\end{aligned} \tag{S1.21}$$

Thus, it can be obtained that $\mathbf{P}[W_{nk}] = 0$, and $\mathrm{Var}(W_{nk}) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{Var}(\phi_k(\boldsymbol{V}_i^*)) =$

$1 + O(n^{-1/2}) - \frac{1}{n}\mu_{ak}^2 \to 1$, and if $j \neq k$, by assumption (A7),

$$\begin{aligned}
\mathrm{Cov}(W_{nj}, W_{nk}) =& \mathbf{P}[W_{nj}W_{nk}] - \mathbf{P}[W_{nj}]\mathbf{P}[W_{nk}] \\
=& \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{P}[\phi_j(\boldsymbol{V}_i^*)\phi_k(\boldsymbol{V}_i^*)] - \mathbf{P}[\phi_j(\boldsymbol{V}_i^*)]\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]\} \\
=& \mathbf{P}[\phi_j(\boldsymbol{V}_i^*)\phi_k(\boldsymbol{V}_i^*)] - \mathbf{P}[\phi_j(\boldsymbol{V}_i^*)]\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)] \\
=& \frac{1}{n^{1/2}}\mathbf{P}\left[\phi_j(\boldsymbol{V}_{0i})\phi_k(\boldsymbol{V}_{0i})\boldsymbol{\xi}_1^{\mathrm{T}}\frac{\partial f(\boldsymbol{V}_{0i}, \hat{\boldsymbol{\alpha}}, \mathbf{0}, \boldsymbol{\theta}_0)/\partial\boldsymbol{\alpha}}{f(\boldsymbol{V}_{0i}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)}\right] \\
& - \mathbf{P}[\phi_j(\boldsymbol{V}_i^*)]\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)] \\
\leq& \frac{1}{n^{1/2}}\mathbf{P}\left[|\phi_j(\boldsymbol{V}_{0i})\phi_k(\boldsymbol{V}_{0i})|b_1(\boldsymbol{V}_{0i}, \boldsymbol{\xi}_1)\right] + \frac{1}{n}C_f(\boldsymbol{\xi}_1) \\
=& O(n^{-1/2}) + O(n^{-1}) \\
\to& 0,
\end{aligned}$$

which implies that if $j \neq k$, $|\mathrm{Cov}(W_{nj}, W_{nk})| \to 0$, and consequently,

$$\mathrm{Cov}(W_{nj}, W_{nk}) \to \begin{cases} 1, & \text{if } j = k, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, by Lindeberg-Levy central limit theorem, we have $(W_{n1}, \cdots, W_{nK})^{\mathrm{T}} \xrightarrow{\mathcal{L}} N(\mathbf{0}_K, \mathbf{1}_{K\times K})$. Since $\mathbf{P}[\phi_k^2(\boldsymbol{V}_i^*)] \to 1$ and $(\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)])^2 \to 0$, it is seen that by the strong law of large numbers, $Z_{nk} \xrightarrow{a.s.} 1$, which implies that $(Z_{n1}, \cdots, Z_{nK})^{\mathrm{T}} \xrightarrow{a.s.} (1, \cdots, 1)^{\mathrm{T}}$. Thus, (S1.19) holds and consequently,

$$\left|\mathbf{P}\left[e^{itT_{nK}}\right] - \mathbf{P}\left[e^{itY_K}\right]\right| < \epsilon, \tag{S1.22}$$

for all n sufficiently large.

If we have (we will show this later) that for fixed t and given $\epsilon > 0$, choose and fix K large enough such that for all n sufficiently large,

$$\left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n^*}\right] - \mathbf{P}\left[e^{itT_{nK}}\right]\right| < 6\{C_f(\boldsymbol{\xi}_1) + 1\}\epsilon, \tag{S1.23}$$

which combing with (S1.22) implies that, for all n large enough,

$$\begin{aligned}\left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n^*}\right] - \mathbf{P}\left[e^{it\zeta_0}\right]\right| \leq & \left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n^*}\right] - \mathbf{P}\left[e^{itT_{nK}}\right]\right| + \left|\mathbf{P}\left[e^{itT_{nK}}\right] - \mathbf{P}\left[e^{itY_K}\right]\right| \\ & + \left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_0}\right]\right| \\ < & \left\{2C_f^2(\boldsymbol{\xi}_1) + 2\right\}^{1/2}\epsilon + \left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_0}\right]\right|.\end{aligned}$$

Therefore, to obtain the inequality (S1.18), it suffices to establish that for all K sufficiently large,

$$\left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_0}\right]\right| < \epsilon\left\{\mathbf{P}[(W_1^2 - 1)^2]\right\}^{1/2}. \tag{S1.24}$$

Let the random variable W_1, W_2, $\cdots$ be defined on a common probability space and denote by ζ_0 the limit in mean square of Y_K as $K \to \infty$. Then according to the fact that $|e^{it} - 1| \leq |t|$, we have by the definition of Y_K,

$$\begin{aligned}\left|\mathbf{P}\left[e^{itY_K}\right] - \mathbf{P}\left[e^{it\zeta_0}\right]\right| \leq & |t|\left(\mathbf{P}\left[Y_K - \zeta_0\right]^2\right)^{1/2} \\ \leq & |t|\left(\mathbf{P}\left[W_1^2 - 1\right]^2\right)^{1/2}\left(\sum_{k=K+1}^{\infty}\lambda_k^2\right)^{1/2} \\ < & \epsilon\left(\mathbf{P}\left[W_1^2 - 1\right]^2\right)^{1/2},\end{aligned}$$

which arrives at (S1.24). Therefore, the inequality (S1.18) holds.

The proof will be completed if we claim that (S1.23) holds. To facilitate the expression, we give some additional notations below. Let

$$\begin{aligned}
\tilde{T}_{nK} =& \frac{1}{n}\sum_{i\neq j}\sum_{k=1}^{K}\lambda_k\phi_k(\boldsymbol{V}_i^*)\phi_k(\boldsymbol{V}_j^*),\\
R_{1nK} =& (n-1)\sum_{k=1}^{K}\lambda_k(\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)])^2,\\
R_{2nK} =& \frac{n-1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K}\lambda_k\mathbf{P}[\phi_k(\boldsymbol{V}_j^*)]\phi_k(\boldsymbol{V}_i^*),\\
T_{0n}^* =& \frac{1}{n}\sum_{i\neq j}h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*),\\
\tilde{T}_{0n}^* =& \frac{1}{n}\sum_{i\neq j}\mathbf{P}[h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*)|\boldsymbol{V}_i^*).
\end{aligned}$$

Thus, $(n-1)\tilde{T}_n^* - T_{nK}$ can be decomposed as follows,

$$\begin{aligned}
(n-1)\tilde{T}_n^* - T_{nK} =& (n-1)\tilde{T}_n^* - T_{0n}^* + 2\tilde{T}_{0n}^* - \mathbf{P}[T_{0n}^*] + T_{0n}^* - \tilde{T}_{nK}\\
&+ \mathbf{P}[T_{0n}^*] - R_{1nK} + 2R_{2nK} - 2\tilde{T}_{0n}^*\\
\equiv& \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4,
\end{aligned}$$

where

$$\begin{aligned}
\Pi_1 =& \frac{1}{n}\sum_{i\neq j}\big\{h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*,\hat{\boldsymbol{\alpha}}) - h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*)\\
&+ \mathbf{P}[h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*)|\boldsymbol{V}_j^*) + \mathbf{P}[h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*)|\boldsymbol{V}_i^*) - \mathbf{P}[h(\boldsymbol{V}_i^*,\boldsymbol{V}_j^*)]\big\},
\end{aligned}$$

and $\Pi_2 = T_{0n}^* - \tilde{T}_{nK}$, $\quad \Pi_3 = \mathbf{P}[T_{0n}^*] - R_{1nK}$, $\quad \Pi_4 = 2R_{2nK} - 2\tilde{T}_{0n}^*$.

We can obtain the upper bounds for all Π_j, $j = 1, 2, 3, 4$ through similar techniques used in above proofs. For fixed t and given $\epsilon > 0$, choose and fix

K large enough such that $|t|\left(\sum_{k=K+1}^{\infty}\lambda_k^2\right)^{1/2}<\epsilon$, and $|t|\sum_{k=K+1}^{\infty}\lambda_k\mu_{ak}^2<\epsilon$. According to (S1.20), it is seen that

$$\mathbf{P}[g_{1K}(\boldsymbol{V}_i^*)]=\sum_{k=K+1}^{\infty}\lambda_k\left(\mathbf{P}[\phi_k(\boldsymbol{V}_i^*)]\right)^2=\frac{1}{n}\sum_{k=K+1}^{\infty}\lambda_k\mu_{ak}^2+o(n^{-1}).$$

Thus, we can obtain that

$$\begin{aligned}
|t|\mathbf{P}\left[|\Pi_2|\right]\leq&|t|\left(\mathbf{P}\left[\Pi_2^2\right]\right)^{1/2}\leq 2\{2C_f^2(\boldsymbol{\xi}_1)+1\}^{1/2}\epsilon,\\
|t|\mathbf{P}\left[|\Pi_3|\right]\leq&\epsilon,\\
|t|\mathbf{P}\left[|\Pi_4|\right]\leq&|t|\left\{\mathbf{P}\left[\Pi_4^2\right]\right\}^{1/2}\leq 2\{C_f(\boldsymbol{\xi}_1)+1\}\epsilon.
\end{aligned}$$

According to the fact that $|e^{it}-1|\leq|t|$, we have, by the above arguments and for all n sufficiently large,

$$\begin{aligned}
\left|\mathbf{P}\left[e^{it(n-1)\tilde{T}_n^*}\right]-\mathbf{P}\left[e^{itT_{nK}}\right]\right|\leq&\mathbf{P}\left[\left|e^{it(n-1)\tilde{T}_n^*}-e^{itT_{nK}}\right|\right]\\
\leq&|t|\mathbf{P}\left[\left|(n-1)\tilde{T}_n^*-T_{nK}\right|\right]\\
\leq&|t|\left\{\mathbf{P}[|\Pi_1|]+\mathbf{P}[|\Pi_2|]+\mathbf{P}[|\Pi_3|]+\mathbf{P}[|\Pi_4|]\right\}\\
<&6\{C_f(\boldsymbol{\xi}_1)+1\}\epsilon,
\end{aligned}$$

which establishes (S1.23) and then completes the whole proof.

S2 Simulation Studies for large q

S2.1 Change plane analysis with sparse $\boldsymbol{Z}$ for GLMs

Consider the GLM in the main paper with the canonical parameter

$$\mu = \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} \geq 0).$$

For Gaussian and Poisson families, we set $\alpha_1 = 0.5$, and for binomial family α_1 is chosen so that the proportion of cases in the data set is $1/3$ on average under the null hypothesis. $\theta_2, \cdots, \theta_q$ are generated randomly from uniform distribution $U(1, 3)$, and θ_1 is chosen as the negative of the 0.65 percentile of $Z_2\theta_2 + \cdots + Z_q\theta_q$, which means that $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta}$ divides the population into two groups with 0.35 and 0.65 observations, respectively. For the predictor, we generate $(Z_1, Z_2, \cdots, Z_q)^{\mathrm{T}}$ from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{I}_q$. We set $\widetilde{X}_j = \mathbf{1}(v_j > 0)$ with $j = 2, \cdots, r$, $X_k = \mathbf{1}(v_k > 0)$ with $k = 2, \cdots, p$, and $\widetilde{X}_1 = 1$, $X_1 = 1$, and $Z_1 = 1$.

For all models with change plane analysis, we evaluated the power under a sequence of alternative models indexed by κ, that is $H_1^{\kappa} : \boldsymbol{\beta}^{\kappa} = \kappa\boldsymbol{\beta}^*$ with $\kappa = i/10$ for semiparmatric model and $\kappa = i/20$ for others, $i = 1, \cdots, 10$, where $\boldsymbol{\beta}^* = (1, \cdots, 1)^{\mathrm{T}}$. We set sample size $n = (300, 600)$ for GLMs, 1000 repetitions and 1000 bootstrap samples, and report in Figure 1-3 the performance for the WAST, SST and SLRT. We calculate the SST and SLRT

over $\{\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}} : k = 1, \cdots, K\}$ with the number of threshold values $K = 1000$. Let $\boldsymbol{\theta}_{-1}^{(k)} = \tilde{\boldsymbol{\theta}}_{-1}^{(k)} / \|\tilde{\boldsymbol{\theta}}_{-1}^{(k)}\|$, where $\boldsymbol{\theta}_{-1}^{(k)} = (\theta_2^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)} = (\tilde{\theta}_2^{(k)}, \cdots, \tilde{\theta}_q^{(k)})^{\mathrm{T}}$, and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)}$ is drawn independently form $(r-1)$-variate standard normal distribution. For each $\theta_1^{(k)}$, $k = 1, \cdots, K$, we selected it by equal grid search in the range from the lower 10th percentile to upper 10th percentile of the data points of $\{\theta_2^{(k)} Z_{2i} + \cdots + \theta_q^{(k)} Z_{qi}\}_{i=1}^n$, which is same as that in Yu and Fan (2020). Here we consider four combinations of $(p, q) = (2, 100), (2, 500), (6, 100), (6, 500), (11, 100), (11, 500)$.

Type I errors ($\kappa = 0$) for GLMs are list in Table 1. We can see from Table 1 that the sizes of the proposed WAST and the SLRT are close to the nominal significance level 0.05, but for most scenarios the size of the SST are much smaller than 0.05. Figure 1-3 indicate that powers become greater as sample size n increases, which are verified by the asymptotic theory. The proposed WAST has comparable power with the SST for the semiparametric model, but the size of the SST is much less than the nominal level 0.05 in Figure 1-3 show that the powers of the proposed WAST are uniformly greater than the competitors SST and SLRT.

Table 1: Type I errors of the proposed WAST, SST and SLRT based on resampling for GLMs with large numbers of sparse $\boldsymbol{Z}$. The nominal significant level is 0.05.

Model	(p, q)	$n = 300$			$n = 600$		
		WAST	SST	SLRT	WAST	SST	SLRT
Gaussian	$(2, 100)$	0.043	0.028	0.066	0.054	0.028	0.044
	$(2, 500)$	0.045	0.024	0.048	0.048	0.035	0.040
	$(6, 100)$	0.052	0.014	0.054	0.055	0.024	0.056
	$(6, 500)$	0.055	0.017	0.052	0.055	0.026	0.044
	$(11, 100)$	0.055	0.005	0.042	0.042	0.010	0.048
	$(11, 500)$	0.058	0.005	0.052	0.043	0.012	0.053
binomial	$(2, 100)$	0.051	0.067	0.038	0.040	0.078	0.048
	$(2, 500)$	0.050	0.081	0.033	0.044	0.070	0.044
	$(6, 100)$	0.046	0.216	0.084	0.056	0.352	0.079
	$(6, 500)$	0.043	0.191	0.088	0.054	0.341	0.070
	$(11, 100)$	0.056	0.051	0.036	0.050	0.077	0.038
	$(11, 500)$	0.058	0.050	0.026	0.054	0.085	0.048
Poisson	$(2, 100)$	0.049	0.030	0.060	0.047	0.049	0.049
	$(2, 500)$	0.054	0.042	0.055	0.046	0.051	0.057
	$(6, 100)$	0.061	0.024	0.064	0.047	0.020	0.060
	$(6, 500)$	0.050	0.014	0.059	0.047	0.018	0.053
	$(11, 100)$	0.044	0.008	0.066	0.049	0.015	0.076
	$(11, 500)$	0.060	0.004	0.078	0.044	0.015	0.060

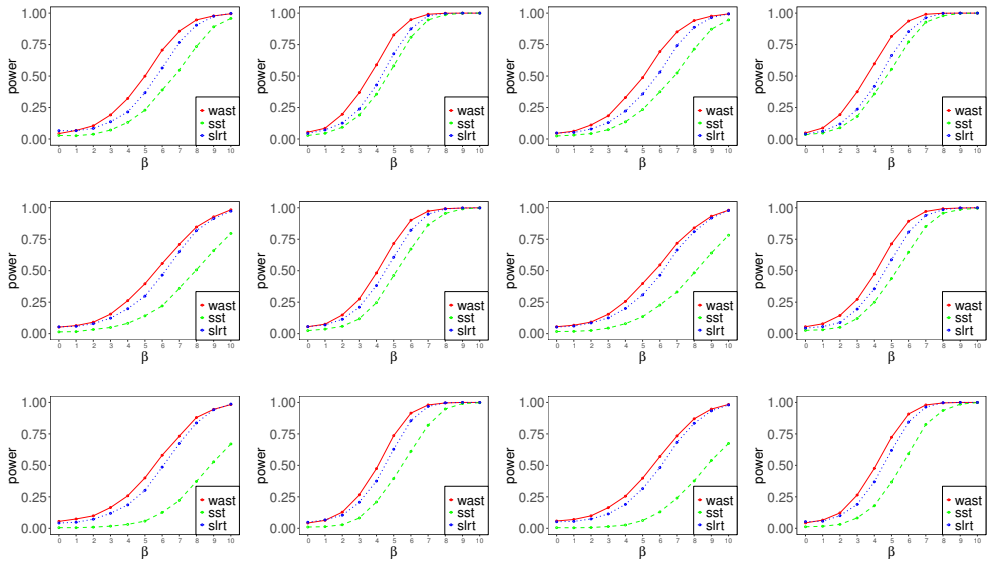

Figure 1: *Powers of test statistic for GLM with Gaussian and with large numbers of sparse* $\boldsymbol{Z}$ *by the proposed WAST (solid line), SST (dashed line) and SLRT (dotted line). From top to bottom, each row depicts the powers for* $p = 2, 6, 11$. *From left to right, each column depicts the powers for* $(n, q) = (300, 100), (600, 100), (300, 500), (600, 500)$.

S2.2 Comparison of the WAST

In this section, we compare the WAST with the closed form (3.12) and with approximation (3.11) in the main text. The simulation settings are same as that in Section S2.1, in which we set $(r, p, q) = (2, 11, 100)$ and apply sparse scenario with the number of nonzero parameters being 10. We set $N = 500, 1000, 5000, 10000$ in the formula (12) of the main paper. Figure 4 indicate that the WAST has slightly greater power than the WAST with

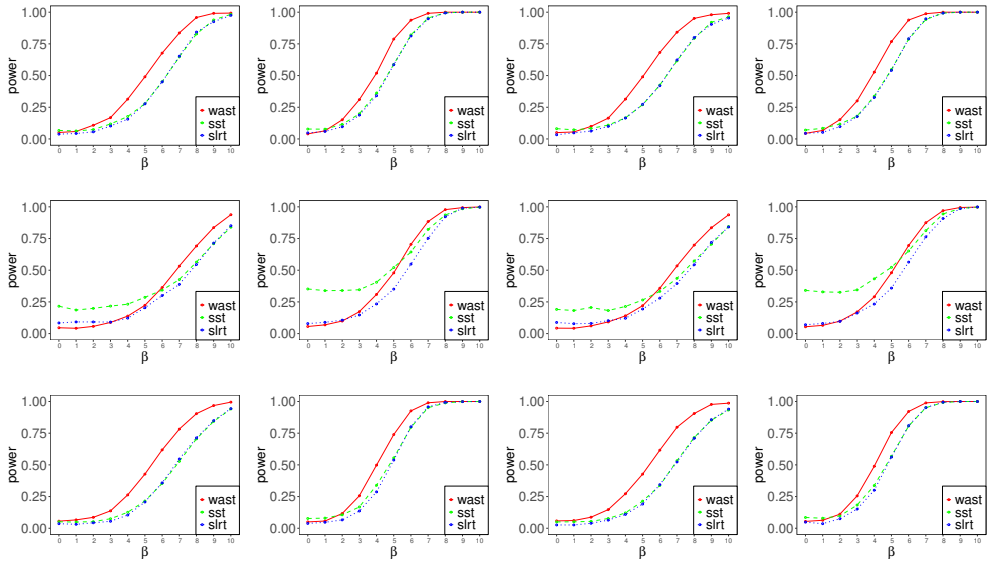

Figure 2: *Powers of test statistic for GLM with binomial and with large numbers of sparse* $\boldsymbol{Z}$ *by the proposed WAST (solid line), SST (dashed line) and SLRT (dotted line). From top to bottom, each row depicts the powers for* $p = 2, 6, 11$. *From left to right, each colum depicts the powers for* $(n, q) = (300, 100), (600, 100), (300, 500), (600, 500)$.

approximation for each distribution and sample size. The number N of the generated zs is not senstive for the WAST with approximation when $N \geq 500$.

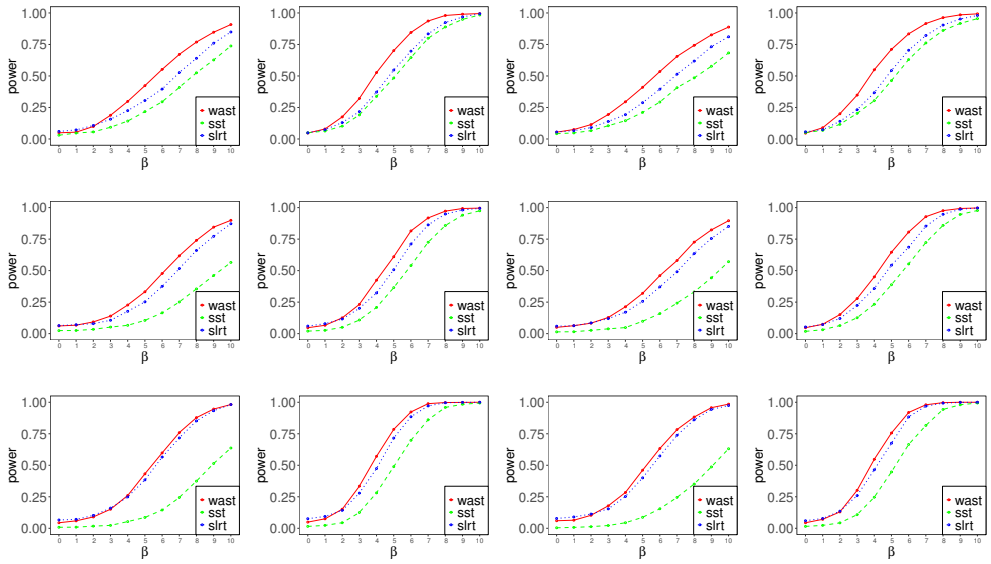

Figure 3: *Powers of test statistic for GLM with Poisson and with large numbers of sparse* $\boldsymbol{Z}$ *by the proposed WAST (solid line), SST (dashed line) and SLRT (dotted line). From top to bottom, each row depicts the powers for* $p = 2, 6, 11$. *From left to right, each colum depicts the powers for* $(n, q) = (300, 100), (600, 100), (300, 500), (600, 500)$.

S3 Simulation Studies for small q

S3.1 Single change plane for GLMs

For fair comparison, we use the same settings as considered by Huang et al. (2021). Consider the GLMs with the canonical parameter

$$\mu = \tilde{\boldsymbol{X}}^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{\beta}\mathbf{1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta} \geq 0),$$

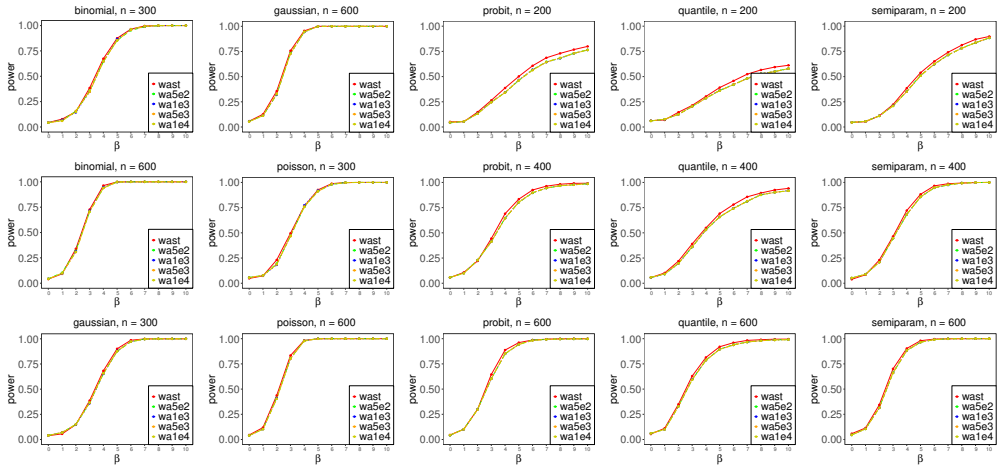

Figure 4: Powers of test statistics by WAST (solid line) and WAST with approximation ("wa5e2" denotes $N = 500$, "wa1e3" denotes $N = 1000$, "wa5e3" denotes $N = 5000$, "wa1e4" denotes $N = 10000$).

where $\boldsymbol{\beta} = \kappa \mathbf{1}_p$, $\alpha_1 = 0.5$ and $\alpha_2 = \cdots = \alpha_r = \log(1.4)$. $\theta_2, \cdots, \theta_q$ are equally spaced numbers from -1 to 1. For the predictor, we generate $(v_2, \cdots, v_{\max\{r,p\}}, Z_2, \cdots, Z_q)^{\mathrm{T}}$ from a multivariate normal distribution with mean $\mathbf{0}$ and covariance $\Sigma = (\tilde{\rho}_{ij})$, where $\tilde{\rho}_{ij} = 1$ if $i = j$ and $\tilde{\rho}_{ij} = \rho$ otherwise. Here, we consider both $\rho = 0$ and $\rho = 0.3$. We set $\widetilde{X}_j = \mathbf{1}(v_j > 0)$ with $j = 2, \cdots, r$, $X_k = \mathbf{1}(v_k > 0)$ with $k = 2, \cdots, p$, and $\widetilde{X}_1 = 1$, $X_1 = 1$, and $Z_1 = 1$.

We evaluate the power under a sequence of alternative models indexed by κ, i.e., $H_1^{\kappa} : \boldsymbol{\beta} = \kappa \mathbf{1}_p$ with κ equally spaced in the range $(0, \kappa_{\max}]$. We set the sample size as $n = (300, 600)$, and we set 1000 repetitions

and 1000 bootstrap samples. For comparison, we consider three methods here: (i) the proposed WAST, (ii) the SST, and (iii) the approximated supremum of likelihood-ratio test statistic (SLRT) given by Huang et al. (2021). We calculate both SLRT and SST over $\{\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}} : k = 1, \cdots, K\}$. Let $\boldsymbol{\theta}_{-1}^{(k)} = \tilde{\boldsymbol{\theta}}_{-1}^{(k)} / \|\tilde{\boldsymbol{\theta}}_{-1}^{(k)}\|$, where $\boldsymbol{\theta}_{-1}^{(k)} = (\theta_2^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)} = (\tilde{\theta}_2^{(k)}, \cdots, \tilde{\theta}_q^{(k)})^{\mathrm{T}}$, and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)}$ is drawn independently from the multivariate normal distribution $\mathcal{N}(\mathbf{0}_{q-1}, \mathbf{1}_{q-1})$. For each $\theta_1^{(k)}$, $k = 1, \cdots, K$, we select it by an equal grid search in the range from the lower tenth percentile to the upper tenth percentile of the data points of $\{\theta_2^{(k)} Z_{i2} + \cdots + \theta_q^{(k)} Z_{iq}\}_{i=1}^n$, which is the same as that in Huang et al. (2021). Here, we set $K = 1000$.

Table 2 lists the type-I errors ($\kappa = 0$) for poisson families with the nominal significance level of 0.05. As can be seen, the sizes of the proposed method are close to 0.05. For the SST, its sizes are far from 0.05 in most of the scenarios, and it performs the worst of these three methods. For the SLRT, its sizes are close to 0.05 except for $(r, p, q) = (2, 51, 11)$ and $(r, p, q) = (6, 51, 11)$.

Figure 5 shows the powers for all the scenarios considered in Table 2. As can be seen, the powers increase as the sample size n increases, which verifies the asymptotic theory as expected. For the Poisson families, the power of the proposed WAST increases faster than those of the SST and

Table 2: Type-I errors of the WAST, SST, and SLRT for GLMs. The nominal significance level is 0.05.

		$\rho = 0$						$\rho = 0.3$					
Family	(r, p, q)	$n = 300$			$n = 600$			$n = 300$			$n = 600$		
		WAST	SST	SLRT	WAST	SST	SLRT	WAST	SST	SLRT	WAST	SST	SLRT
Poisson	$(2, 2, 3)$	0.048	0.038	0.043	0.055	0.055	0.048	0.046	0.039	0.050	0.052	0.046	0.055
	$(6, 6, 3)$	0.061	0.023	0.056	0.057	0.042	0.054	0.065	0.014	0.052	0.056	0.029	0.048
	$(2, 2, 11)$	0.047	0.031	0.059	0.049	0.044	0.042	0.050	0.039	0.057	0.050	0.041	0.038
	$(6, 6, 11)$	0.061	0.020	0.051	0.051	0.023	0.066	0.052	0.014	0.052	0.055	0.022	0.056
	$(2, 51, 11)$	0.048	0.030	0.036	0.042	0.002	0.098	0.050	0.036	0.033	0.044	0.000	0.093
	$(6, 51, 11)$	0.047	0.033	0.045	0.040	0.001	0.063	0.049	0.021	0.032	0.048	0.000	0.060

SLRT, and this superiority is particularly evident when n is small or p and q increase. In summary, the proposed test statistic shows very competitive performance.

S3.2 Multiple change plane analysis

It is straightforward to extend the proposed test procedure for models with a change plane to the case of multiple change planes, as considered by Li et al. (2021). For instance, we consider the quantile regression model with multiple change planes:

$$Q_{Y_i}(\tau|\tilde{\boldsymbol{X}}_i, \boldsymbol{X}_i, \boldsymbol{Z}_i) = \tilde{\boldsymbol{X}}_i^{\mathrm{T}}\boldsymbol{\alpha}(\tau) + \sum_{t=1}^{T} \boldsymbol{X}_{ti}^{\mathrm{T}}\boldsymbol{\beta}_t(\tau)\mathbf{1}(\boldsymbol{Z}_{ti}^{\mathrm{T}}\boldsymbol{\theta}_t(\tau) \geq 0),$$

where $\boldsymbol{\alpha}(\tau) = (\alpha_1(\tau), \cdots, \alpha_r(\tau))^{\mathrm{T}} \in \Theta_\alpha \subseteq \mathbb{R}^r$, $\boldsymbol{\beta}_t(\tau) = (\beta_{t1}(\tau), \cdots, \beta_{tp_t}(\tau))^{\mathrm{T}} \in \Theta_{\beta_t} \subseteq \mathbb{R}^{p_t}$, and $\boldsymbol{\theta}_t(\tau) = (\theta_{t1}(\tau), \cdots, \theta_{tq_t}(\tau))^{\mathrm{T}} \in \Theta_{\boldsymbol{\theta}_t} \subseteq \mathbb{R}^{q_t}$ are unknown

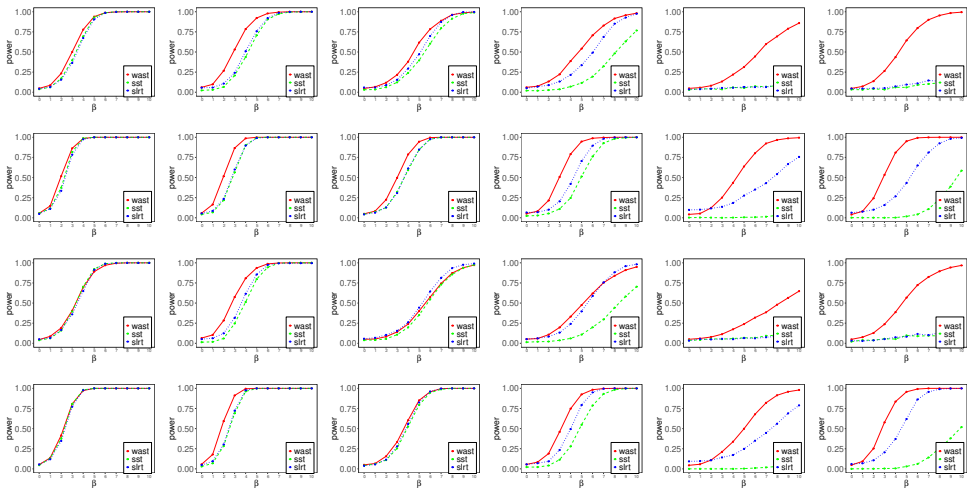

Figure 5: *Powers of testing Poisson regression by the proposed WAST (solid line), SST (dashed line), and SLRT (dotted line). From top to bottom, each row depicts the powers for* $(n, \rho) = (300, 0), (600, 0), (300, 0.3)$, *and* $(600, 0.3)$. *From left to right, each column depicts the powers for* $(r, p, q) = (2, 2, 3)$, $(6, 6, 3)$, $(2, 2, 11)$, $(6, 6, 11)$, $(2, 51, 11)$, *and* $(6, 51, 11)$ *respectively.*

parameters, and the number of change planes T is a preset integer, $\boldsymbol{X}_i = (\boldsymbol{X}_{1i}^{\mathrm{T}}, \cdots, \boldsymbol{X}_{Ti}^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (\boldsymbol{Z}_{1i}^{\mathrm{T}}, \cdots, \boldsymbol{Z}_{Ti}^{\mathrm{T}})^{\mathrm{T}}$. We still denote $\boldsymbol{\beta}(\tau) = (\boldsymbol{\beta}_1(\tau)^{\mathrm{T}}, \cdots, \boldsymbol{\beta}_T(\tau)^{\mathrm{T}})^{\mathrm{T}}$ and $\boldsymbol{\theta}(\tau) = (\boldsymbol{\theta}_1(\tau)^{\mathrm{T}}, \cdots, \boldsymbol{\theta}_T(\tau)^{\mathrm{T}})^{\mathrm{T}}$, and $\Theta_\beta = \Theta_{\beta_1} \times \cdots \times \Theta_{\beta_T}$ and $\Theta_\theta = \Theta_{\theta_1} \times \cdots \times \Theta_{\theta_T}$.

Testing whether $\boldsymbol{\beta}_t(\tau) = \mathbf{0}$ for all $t \in \{1, \cdots, T\}$ is of interest, i.e.,

$$H_0 : \boldsymbol{\beta}(\tau) = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\beta}(\tau) \neq \mathbf{0}. \tag{S3.1}$$

Similar to the WAST in the main paper, we propose the integrated statistic

$$T_n^{(m)} = \frac{1}{n(n-1)} \sum_{i \neq j} \tilde{\omega}_{ij} \{\mathbf{1}(Y_i - \tilde{\boldsymbol{X}}_i^{\mathrm{T}} \hat{\boldsymbol{\alpha}}(\tau) \leq 0) - \tau\} \{\mathbf{1}(Y_j - \tilde{\boldsymbol{X}}_j^{\mathrm{T}} \hat{\boldsymbol{\alpha}}(\tau) \leq 0) - \tau\}$$

with

$$\tilde{\omega}_{ij} = \sum_{t=1}^{T} \boldsymbol{X}_{ti}^{\mathrm{T}} \boldsymbol{X}_{tj} \int_{\boldsymbol{\theta}_t \in \Theta_{\theta_t}} \mathbf{1}(\boldsymbol{Z}_{ti}^{\mathrm{T}} \boldsymbol{\theta}_t \geq 0) \mathbf{1}(\boldsymbol{Z}_{tj}^{\mathrm{T}} \boldsymbol{\theta}_t \geq 0) w_t(\boldsymbol{\theta}_t) d\boldsymbol{\theta}_t. \tag{S3.2}$$

Taking the weight $w(\boldsymbol{\theta})$ to be the standard multivariate normal density leads to, for $i \neq j$,

$$\tilde{\omega}_{ij} = \frac{1}{4} \sum_{t=1}^{T} \boldsymbol{X}_{ti}^{\mathrm{T}} \boldsymbol{X}_{tj} + \frac{1}{2\pi} \sum_{t=1}^{T} \boldsymbol{X}_{ti}^{\mathrm{T}} \boldsymbol{X}_{tj} \arctan \left(\frac{\varrho_{t,ij}}{\sqrt{1 - \varrho_{t,ij}^2}} \right),$$

where $\varrho_{t,ij} = \boldsymbol{Z}_{ti}^{\mathrm{T}} \boldsymbol{Z}_{tj} (\|\boldsymbol{Z}_{ti}\| \|\boldsymbol{Z}_{tj}\|)^{-1}$.

As in model (7) in the main paper, we consider the empirical estimating functions

$$\begin{aligned} \Psi_n(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) &= \sum_{i=1}^{n} \boldsymbol{W}_i(\boldsymbol{\theta}) \psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}), \\ \Psi_{1n}(\boldsymbol{\alpha}) &= \sum_{i=1}^{n} \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}), \end{aligned} \tag{S3.3}$$

where $\boldsymbol{W}_i(\boldsymbol{\theta}) = (\boldsymbol{X}_{1i}^{\mathrm{T}} \mathbf{1}(\boldsymbol{Z}_{1i}^{\mathrm{T}} \boldsymbol{\theta}_1 \geq 0), \cdots, \boldsymbol{X}_{Ti}^{\mathrm{T}} \mathbf{1}(\boldsymbol{Z}_{Ti}^{\mathrm{T}} \boldsymbol{\theta}_T \geq 0))^{\mathrm{T}}$. For ease of expression, we denote $\psi^{(m)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \boldsymbol{W}(\boldsymbol{\theta}) \psi_0^{(m)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$, and $\psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha}) = \psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha}, \mathbf{0}, \boldsymbol{\theta})$, where $\psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha})$ is free of $\boldsymbol{\theta}$, and $\psi^{(m)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta})$ satisfies $\mathrm{E}\psi^{(m)}(\boldsymbol{V}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}) = 0$ with true parameters $\boldsymbol{\alpha}_0$, $\boldsymbol{\beta}_0$, and $\boldsymbol{\theta}_0$. We propose the integrated statistic with $\tilde{\omega}_{ij}$ defined in (S3.2), i.e.,

$$T_n^{(m)} = \frac{1}{n(n-1)} \sum_{i \neq j} \tilde{\omega}_{ij} \psi_0^{(m)}(\boldsymbol{V}_i, \hat{\boldsymbol{\alpha}}) \psi_0^{(m)}(\boldsymbol{V}_j, \hat{\boldsymbol{\alpha}}).$$

Let $\tilde{K}^t(\boldsymbol{\theta}) = \partial \mathrm{E}[\boldsymbol{X}_{ti}\mathbf{1}(\boldsymbol{Z}_{ti}^{\mathrm{T}}\boldsymbol{\theta}_1 \geq 0)\psi_0^{(m)}(\boldsymbol{V}, \boldsymbol{\alpha}_0)]/\partial \boldsymbol{\alpha}^{\mathrm{T}}$, $\tilde{K}(\boldsymbol{\theta}) = (\tilde{K}^1(\boldsymbol{\theta})^{\mathrm{T}}, \cdots, \tilde{K}^T(\boldsymbol{\theta})^{\mathrm{T}})^{\mathrm{T}}$, and the kernel of a U-statistic under the null hypothesis be

$$\begin{aligned} h^{(m)}(\boldsymbol{V}_i, \boldsymbol{V}_j) =& \tilde{\omega}_{ij}\psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)\psi_0^{(m)}(\boldsymbol{V}_j, \boldsymbol{\alpha}_0) - \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\mathrm{T}}\tilde{K}_j \\ & - \tilde{K}_i^{\mathrm{T}}\psi_1(\boldsymbol{V}_j, \boldsymbol{\alpha}_0) + \psi_1(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)^{\mathrm{T}}\tilde{\boldsymbol{H}}\psi_1(\boldsymbol{V}_j, \boldsymbol{\alpha}_0), \end{aligned} \tag{S3.4}$$

where $\tilde{K}_i = \sum_{t=1}^{T} \int_{\boldsymbol{\theta}_t \in \Theta_{\theta_t}} J^{\mathrm{T}}\tilde{K}^t(\boldsymbol{\theta})^{\mathrm{T}}\boldsymbol{X}_{ti}\mathbf{1}(\boldsymbol{Z}_{ti}^{\mathrm{T}}\boldsymbol{\theta}_t \geq 0)\psi_0^{(m)}(\boldsymbol{V}_i, \boldsymbol{\alpha}_0)w_t(\boldsymbol{\theta}_t)d\boldsymbol{\theta}_t$ and $\tilde{\boldsymbol{H}} = \sum_{t=1}^{T} \int_{\boldsymbol{\theta}_t \in \Theta_{\theta_t}} J^{\mathrm{T}}\tilde{K}^t(\boldsymbol{\theta})^{\mathrm{T}}\tilde{K}^t(\boldsymbol{\theta})Jw_t(\boldsymbol{\theta}_t)d\boldsymbol{\theta}_t$.

We are now ready to derive the asymptotic distributions of $T_n^{(m)}$ under the null and local alternative hypotheses.

Theorem 1. If Assumptions (A1)–(A5) hold, then under the null hypothesis, we have

$$nT_n^{(m)} - \mu_0 \xrightarrow{\mathcal{L}} \zeta_0,$$

where $\mu_0 = -2\mathrm{E}[\psi_1(\boldsymbol{V}_1, \boldsymbol{\alpha}_0)^{\mathrm{T}}\tilde{K}_1] + \mathrm{E}[\psi_1(\boldsymbol{V}_1, \boldsymbol{\alpha}_0)^{\mathrm{T}}\tilde{\boldsymbol{H}}\psi_1(\boldsymbol{V}_1, \boldsymbol{\alpha}_0)]$, ζ_0 is a random variable of the form $\zeta_0 = \sum_{j=1}^{\infty}\lambda_j(\chi_{1j}^2 - 1)$, and $\chi_{11}^2, \chi_{12}^2, ...$ are independent χ_1^2 variables. Here, $\{\lambda_j\}$ are the eigenvalues of the kernel $h^{(m)}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ under $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$, i.e., they are the solutions of $\lambda_j g_j(\boldsymbol{v}_2) = \int_0^{\infty} h^{(m)}(\boldsymbol{v}_1, \boldsymbol{v}_2)g_j(\boldsymbol{v}_1)f(\boldsymbol{v}_1, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)d\boldsymbol{v}_1$ for nonzero g_j.

Proof. This theorem can be proved in the same way as Theorem 1.

Theorem 2. If Assumptions (A1)–(A6) hold, then under the local alternative hypothesis that is $\boldsymbol{\beta} = n^{-1/2}\boldsymbol{\xi}$ with a deterministic vector $\boldsymbol{\xi} \in \Theta_\beta$,

we have

$$nT_n^{(m)} - \mu_0 \xrightarrow{\mathcal{L}} \zeta_1,$$

where μ_0 is defined in Theorem 1, ζ_1 is a random variable of the form $\zeta_1 = \sum_{j=1}^{\infty} \lambda_j(\chi_{1j}^2(\mu_{aj}) - 1)$, and $\chi_{11}^2(\mu_{a1}), \chi_{12}^2(\mu_{a2}), \cdots$ are independent noncentral χ_1^2 variables. Here, $\{\lambda_j\}$ are the eigenvalues of the kernel $h^{(m)}(\boldsymbol{v}_1, \boldsymbol{v}_2)$ defined in (S3.4) under $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$, i.e., they are the solutions of $\lambda_j g_j(\boldsymbol{v}_2) = \int_0^{\infty} h^{(m)}(\boldsymbol{v}_1, \boldsymbol{v}_2) g_j(\boldsymbol{v}_1) f(\boldsymbol{v}_1, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0) d\boldsymbol{v}_1$ for nonzero g_j, and each noncentrality parameter of $\chi_{1j}^2(\mu_{aj})$ is

$$\mu_{aj} = \mathrm{E}\left[\phi_j(\boldsymbol{V}_0)\boldsymbol{\xi}^{\mathrm{T}} \partial \log(f(\boldsymbol{V}_0, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0))/\partial \boldsymbol{\beta}\right], \quad j = 1, 2, \cdots,$$

where $\{\phi_j(\boldsymbol{v})\}$ denote orthonormal eigenfunctions corresponding to the eigenvalues $\{\lambda_j\}$, and $\boldsymbol{V}_0$ is generated from the null distribution $f(\boldsymbol{v}, \boldsymbol{\alpha}_0, \mathbf{0}, \boldsymbol{\theta}_0)$.

The proof of this theorem can be completed along the same lines of Theorem 2 in the main paper and the above Theorem 1.

Next, we demonstrate the finite-sample performance of the proposed WAST in multiple change-plane models. Consider the GLM in example 2 with the canonical parameter

$$\mu = \tilde{\boldsymbol{X}}^{\mathrm{T}} \boldsymbol{\alpha} + \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_1 \mathbf{1}(\boldsymbol{Z}_1^{\mathrm{T}} \boldsymbol{\theta}_1 \geq 0) + \boldsymbol{X}^{\mathrm{T}} \boldsymbol{\beta}_2 \mathbf{1}(\boldsymbol{Z}_2^{\mathrm{T}} \boldsymbol{\theta}_2 \geq 0),$$

where $\boldsymbol{\alpha}$ is the same as that in Section 5 in the main paper, $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \tau \mathbf{1}_q$, and $\tilde{\boldsymbol{X}}$, $\boldsymbol{X}$, and $\boldsymbol{Z}_1 = \boldsymbol{Z}_2$ are generated as in Section 5 in the main paper.

Let $\boldsymbol{Z}_{-1,s}$ be $\boldsymbol{Z}_s$ excluding the first component, and let $\boldsymbol{\theta}_{-1,s}$ be $\boldsymbol{\theta}_s$ excluding the first component, $s = 1, 2$. θ_{11} and θ_{21} are chosen as the negatives of the 0.3 and 0.6 percentiles of $\boldsymbol{Z}_{-1,1}^{\mathrm{T}}\boldsymbol{\theta}_{-1,1}$ and $\boldsymbol{Z}_{-1,1}^{\mathrm{T}}\boldsymbol{\theta}_{-1,1}$, respectively, which means that $\boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{\theta}_1$ divides the population into two groups with 0.7 and 0.3 observations, and $\boldsymbol{Z}_1^{\mathrm{T}}\boldsymbol{\theta}_1$ divides the population into two groups with 0.4 and 0.6 observations. When $(r, p, q) = (2, 2, 3)$ or $(r, p, q) = (6, 6, 3)$, the maximal value of the range $\kappa_{\max}$ is $\log(0.7)$, $\log(0.35)$ and $\log(0.7)$ for Gaussian, binomial and Poisson regression, respectively. While $(r, p, q) = (2, 2, 11)$ or $(r, p, q) = (6, 6, 11)$, the maximal value of the range $\kappa_{\max}$ is $\log(0.9)$, $\log(0.75)$, and $\log(0.9)$ for Gaussian, binomial, and Poisson regression, respectively. The other settings are the same as in Section 5 in the main paper.

Table 3 lists the sizes of the proposed WAST with multiple change planes; as can be seen, the sizes are close to the nominal level of 0.05, as expected. Figure 6 shows the power of the proposed WAST with multiple change planes; as can be seen, all the powers (i) increase as the sample size increases and (ii) increase rapidly as $\boldsymbol{\beta}$ departs from the null hypothesis.

Table 3: Type-I errors of proposed WAST for multiple change planes. The nominal significant level is 0.05.

(r, p, q)	n	$\rho = 0$			$\rho = 0.3$		
		Gaussian	Binomial	Poisson	Gaussian	Binomial	Poisson
$(2, 2, 3)$	300	0.043	0.048	0.053	0.055	0.048	0.046
	600	0.051	0.058	0.050	0.051	0.055	0.052
$(6, 6, 3)$	300	0.046	0.053	0.056	0.059	0.061	0.065
	600	0.060	0.064	0.045	0.046	0.057	0.056
$(2, 2, 11)$	300	0.040	0.049	0.046	0.052	0.047	0.050
	600	0.044	0.039	0.057	0.050	0.049	0.050
$(6, 6, 11)$	300	0.041	0.040	0.052	0.043	0.061	0.052
	600	0.039	0.042	0.036	0.043	0.051	0.055

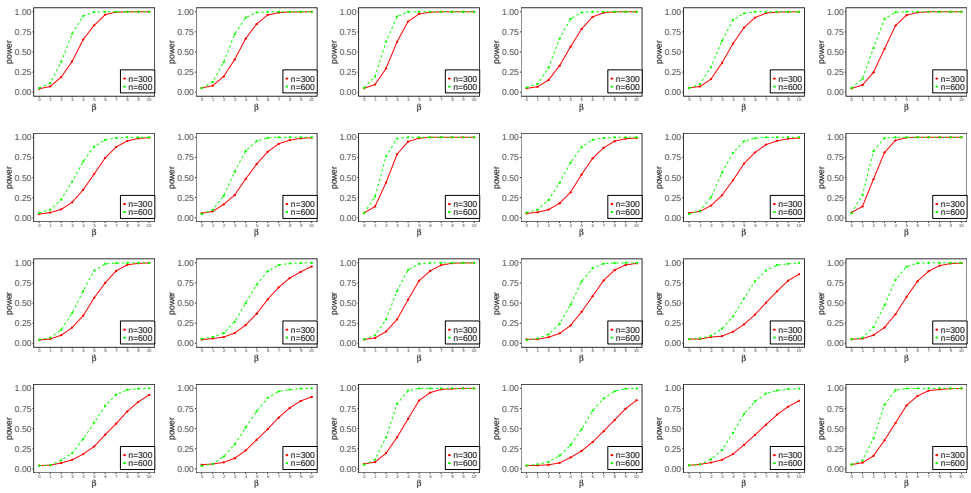

Figure 6: Powers of proposed WAST with $n = 300$ (solid line) and $n = 600$ (dashed line) for multiple change planes. From top to bottom, each row depicts the powers for the cases $(r, p, q) = (2, 2, 3)$, $(6, 6, 3)$, $(2, 2, 11)$, and $(6, 6, 11)$. From left to right, each column depicts the powers for Gaussian, binomial, and poisson families with $\rho = 0$ and for these three families with $\rho = 0.3$.

S3.3 Change plane analysis for quantile and semiparametric models

For fair comparison, in this section, we consider the settings as similar to the existing literature Lee et al. (2011); Fan et al. (2017), which are different from the GLMs in Section 4 in the main paper.

Consider the quantile regression model in the main paper,

$$Y_i = 0.5 + \widetilde{X}_i + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}(\tau)\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta}(\tau) \geq 0) + \epsilon_i,$$

where $\widetilde{X}_i$ is generated from standard normal distribution, $\boldsymbol{X}_i$ from p-variate normal distribution with mean $\mathbf{0}$ and variance $\sqrt{2}I$, and $Z_{1i} = 1$ and $(Z_{2i}, \cdots, Z_{qi})^{\mathrm{T}}$ from $(q-1)$-variate standard normal distribution. We set $\boldsymbol{\beta} = (1, \cdots, 1)^{\mathrm{T}}$ under H_0, and $(\theta_2, \cdots, \theta_q)^{\mathrm{T}} = (1, 2, \cdots, 2)^{\mathrm{T}}$ under H_1, where θ_1 is chosen as the negative of the 0.65 percentile of $Z_2\theta_2 + \cdots + Z_q\theta_q$, which means that $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\theta}$ divides the population into two groups with 0.35 and 0.65 observations, respectively. The error ϵ_i is generated from standard normal distribution.

Considered the semiparamteric model in the main paper,

$$Y_i = \gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2) + \boldsymbol{X}_i^{\mathrm{T}}\boldsymbol{\beta}A_i\mathbf{1}(\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\theta} \geq 0) + \epsilon_i.$$

As the same settings provided by Fan et al. (2017), we consider two settings for the propensity score model $\pi(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_1)$, denoted by P1 and P2:

(P1) $\pi(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_1) = 0.5$;

(P2) $\pi(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_1) = \frac{\exp(0.5\widetilde{X}_{2i} + 0.5\widetilde{X}_{3i})}{1+\exp(0.5\widetilde{X}_{2i} + 0.5\widetilde{X}_{3i})}$,

and two baseline mean functions for $\gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2)$, denoted by B1 and B2:

(B1) $\gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2) = 1 + 0.5\widetilde{X}_{2i} + \widetilde{X}_{3i}^2$;

(B2) $\gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2) = 1 + \sin(\widetilde{X}_{2i} + \widetilde{X}_{3i})$.

We calculate the WAST and SST in the main paper by fitting a linear model

for baseline function $\gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2)$ and a logistic regression for the propensity score model $\pi(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_1)$. Therefore, two baseline models $\gamma(\tilde{\boldsymbol{X}}_i, \boldsymbol{\alpha}_2)$ are misspecified. We generate v_{1i} independently from binomial distribution $b(1, 0.5)$, v_{2i} independently from uniform distribution $U(-1, 1)$, and X_{ji} independently from uniform distribution $U(-1, 1)$, where $i = 1, \cdots, n$ and $j = 1, \cdots, p$. We set $\tilde{\boldsymbol{X}}_i = (1, v_{1i}, v_{2i})^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (1, v_{1i}, v_{2i}, X_{1i}, \cdots, X_{(p-3)i})^{\mathrm{T}}$. The parameters $\boldsymbol{\beta}$ under H_0 and $\boldsymbol{\theta}$ under H_1 are same as these in quantile model.

For all models with change plane analysis, we evaluated the power under a sequence of alternative models indexed by κ, that is $H_1^{\kappa} : \boldsymbol{\beta}^{\kappa} = \kappa \boldsymbol{\beta}^*$ with $\kappa = i/10$ for semiparmatric model and $\kappa = i/20$ for others, $i = 1, \cdots, 10$, where $\boldsymbol{\beta}^* = (1, \cdots, 1)^{\mathrm{T}}$. We set sample size $n = (200, 400, 600)$, 1000 repetitions and 1000 bootstrap samples, and report in Figure 7-10 the performance for both the WAST and SST. We calculate SST over $\{\boldsymbol{\theta}^{(k)} = (\theta_1^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}} : k = 1, \cdots, K\}$ with the number of threshold values $K = 2000$. Let $\boldsymbol{\theta}_{-1}^{(k)} = \tilde{\boldsymbol{\theta}}_{-1}^{(k)} / \|\tilde{\boldsymbol{\theta}}_{-1}^{(k)}\|$, where $\boldsymbol{\theta}_{-1}^{(k)} = (\theta_2^{(k)}, \cdots, \theta_q^{(k)})^{\mathrm{T}}$ and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)} = (\tilde{\theta}_2^{(k)}, \cdots, \tilde{\theta}_q^{(k)})^{\mathrm{T}}$, and $\tilde{\boldsymbol{\theta}}_{-1}^{(k)}$ is drawn independently form $(r-1)$-variate standard normal distribution. For each $\theta_1^{(k)}$, $k = 1, \cdots, K$, we selected it by equal grid search in the range from the lower 10th percentile to upper 10th percentile of the data points of $\{\theta_2^{(k)} Z_{2i} + \cdots + \theta_q^{(k)} Z_{qi}\}_{i=1}^n$, which is

same as that in Yu and Fan (2020). Here we consider four combinations of $(p, q) = (1, 3), (5, 5), (10, 10), (50, 20)$.

Type I errors ($\kappa = 0$) with sample size $n = (200, 400, 600)$ are list in Table 4. We can see from Table 4 that the size of the proposed WAST are close to the nominal significance level 0.05, but for most scenarios the size of the SST are much smaller than 0.05. Figure 7-8 indicate that powers become greater as sample size n increases, which verified the asymptotic theory. The proposed WAST has comparable power with the SST for the semiparametric model, but the size of the SST is much less than the nominal level 0.05 in Table 4. SST has wrong size, and is less powerful than the WAST for quantile regression models in four combinations of (p, q). Since the size is much less than the nominal level 0.05 when dimension of $\boldsymbol{X}$ not smaller than 5 in Table 4, we only depict in Figure 9 the power curves for the WAST. Figure 9 shows that the power increases fast as the sample size becomes large as expected from the theory, and the power increases rapidly in all models and two combinations $(p, q) = (10, 10)$ and $(50, 20)$ as the alternative is farther away from the null.

Table 4: Type I errors of the proposed WAST and SST based on resampling for quantile regression (QuantRE) and semiparametric model (SPModel). The nominal significant level is 0.05.

Model	(p, q)	$n = 200$		$n = 400$		$n = 600$	
		WAST	SST	WAST	SST	WAST	SST
SPModel	$(1, 3)$	0.053	0.051	0.063	0.045	0.046	0.048
(B1+P1)	$(5, 5)$	0.056	0.008	0.049	0.022	0.046	0.012
	$(10, 10)$	0.056	0.001	0.047	0.007	0.052	0.018
	$(50, 20)$	0.058	0.033	0.046	0.035	0.052	0.000
SPModel	$(1, 3)$	0.057	0.044	0.054	0.048	0.046	0.053
(B2+P2)	$(5, 5)$	0.050	0.007	0.059	0.013	0.032	0.013
	$(10, 10)$	0.053	0.002	0.048	0.006	0.038	0.008
	$(50, 20)$	0.058	0.031	0.058	0.032	0.042	0.000
QuantRE	$(1, 3)$	0.044	0.015	0.057	0.027	0.043	0.030
($\tau = 0.2$)	$(5, 5)$	0.051	0.000	0.059	0.016	0.058	0.020
	$(10, 10)$	0.041	0.000	0.048	0.000	0.055	0.002
	$(50, 20)$	0.056	0.022	0.043	0.031	0.046	0.000
QuantRE	$(1, 3)$	0.060	0.039	0.058	0.051	0.053	0.034
($\tau = 0.5$)	$(5, 5)$	0.049	0.005	0.046	0.024	0.047	0.033
	$(10, 10)$	0.052	0.009	0.046	0.014	0.058	0.026
	$(50, 20)$	0.057	0.050	0.054	0.043	0.051	0.005
QuantRE	$(1, 3)$	0.044	0.020	0.050	0.025	0.053	0.037
($\tau = 0.7$)	$(5, 5)$	0.048	0.005	0.055	0.021	0.045	0.024
	$(10, 10)$	0.045	0.000	0.052	0.009	0.055	0.017
	$(50, 20)$	0.036	0.022	0.035	0.042	0.039	0.001

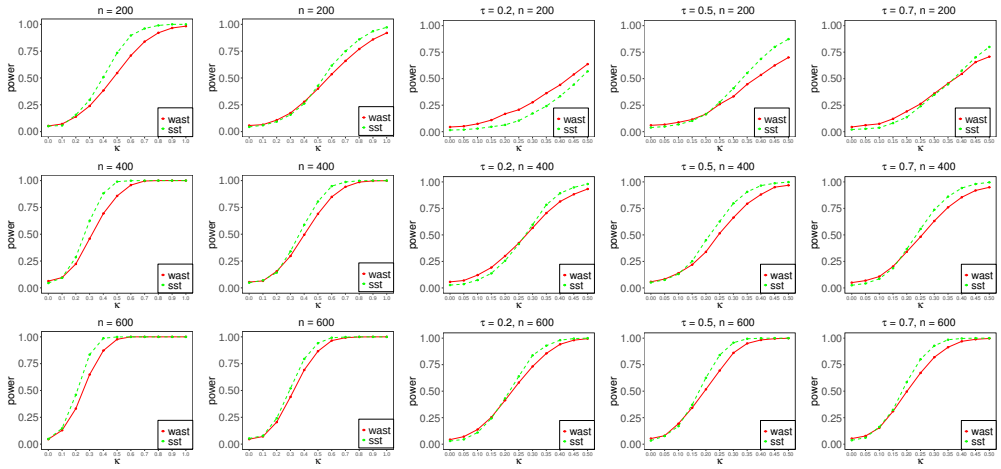

Figure 7: *Powers of test statistic by the proposed WAST (solid line) and SST (dashed line) for* $(p, q) = (1, 3)$. *From left to right, each column depicts the powers for semiparametric model, quantile regression with* $\tau = 0.2$, $\tau = 0.5$ *and* $\tau = 0.7$, *respectively.*

S4 Case study

We show the proposed test procedure in application to cancer type BRCA downloaded from the Cancer Genome Atlas https://tcga-data.nci.nih.gov/.

It is well known that around the world, breast cancer is the most common cancer among women. *BRCA1*, *BRCA2*, and *PALB2* are breast-cancer susceptibility genes, and the *PALB2* gene makes a protein that works with the *BRCA2* gene protein to repair damaged DNA and stop tumour growth. Studies have shown that for women with *BRCA* mutations, their chance of developing breast cancer during their lifetime can be increased by up

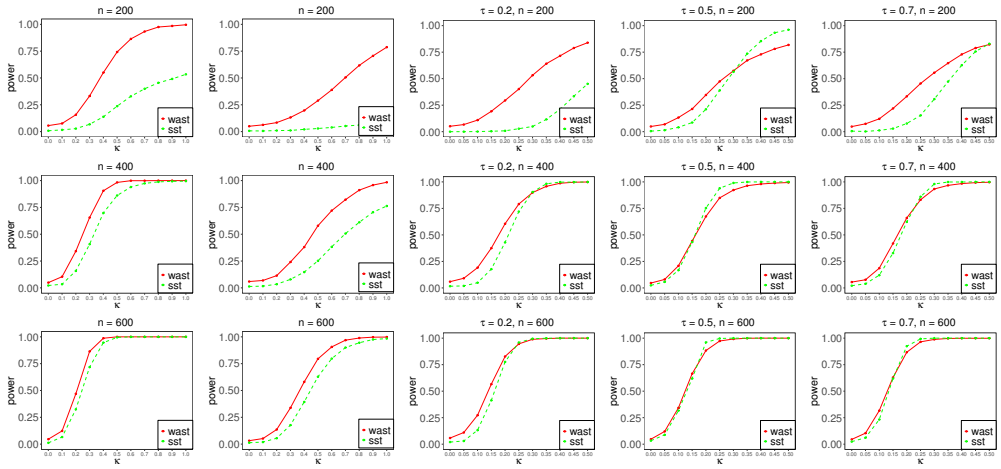

Figure 8: *Powers of test statistic by the proposed WAST (solid line) and SST (dashed line) for* $(p, q) = (5, 5)$. *From left to right, each column depicts the powers for semiparametric model, quantile regression with* $\tau = 0.2$, $\tau = 0.5$, *and* $\tau = 0.7$, *respectively.*

to 72%; also women with *PALB2* mutations have a 14% chance of breast cancer by age 50 and a 35% chance by age 70. Our aim here is to detect whether there is a subgroup of the BRCA tumour stage based on *PALB2* gene mutation (Antoniou et al., 2014).

We treat tumour stages I and II as one type coded by 0, and the rest (stages III–V) as another type coded by 1, which is binomial and treated as response variable y. The covariate $\boldsymbol{X}$ includes age, race, and clinical calls for estrogen receptor (ER) and progesterone receptor (PR). After removing missing values, there are 423 subjects in the age range from 26 to 90. For race, we code 0 for white people (350 subjects) and 1 for others (73 subjects).

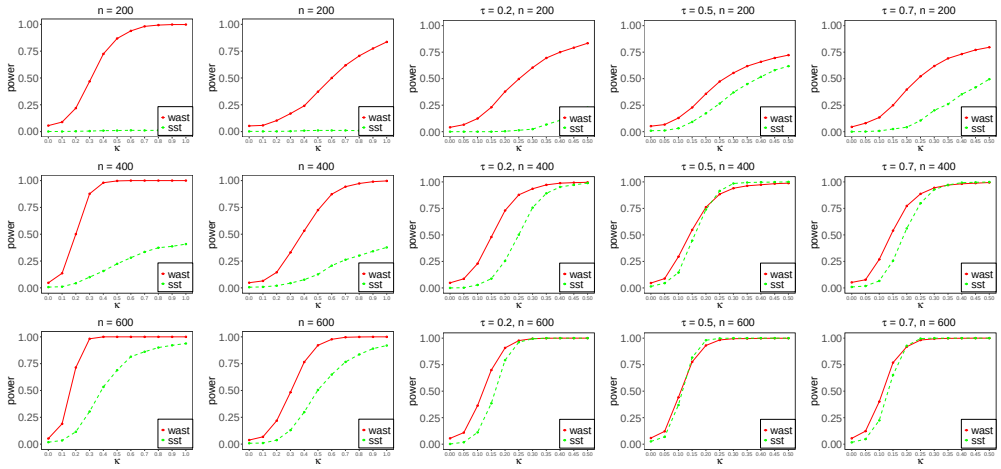

Figure 9: *Powers of test statistic by the proposed WAST (solid line) and SST (dashed line) for* $(p, q) = (10, 10)$. *From left to right, each column depicts the powers for semi-parametric model, quantile regression with* $\tau = 0.2$, $\tau = 0.5$, *and* $\tau = 0.7$, *respectively.*

There are 320 subjects with positive ER (coded 1) and 272 subjects with positive PR (coded 1). We consider the logistic regression in example 2 with

$$\tilde{\boldsymbol{X}}_i = \boldsymbol{X}_i = (1, \mathrm{AGE}_i, \mathrm{RACE}_i, \mathrm{ER}_i, \mathrm{PR}_i)^{\mathrm{T}},$$

$$\boldsymbol{Z}_i = (1, GENE_i^{\mathrm{T}})^{\mathrm{T}}, \quad i = 1, \cdots, 423,$$

where the $GENE_i$ consists of 500 genes, including the eight important genes *BRCA1*, *BRCA2*, *PALB2*, *PIK3CA*, *PTEN*, *TP53*, *ATM*, and *CDH1*, whhich are highly correlated with breast cancer (Antoniou et al., 2014).

Calculated based on 5000 bootstrap samples, the p-value of the WAST

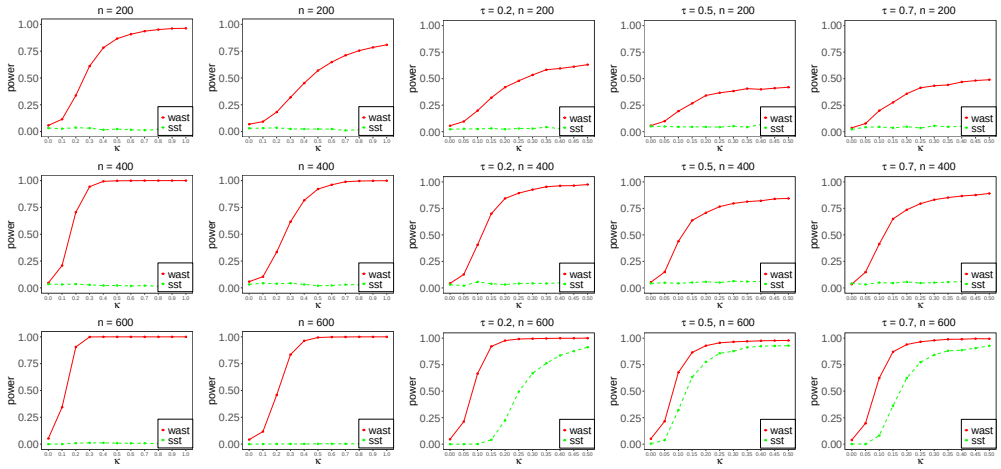

Figure 10: *Powers of test statistic by the proposed WAST (solid line) and SST (dashed line) for* $(p, q) = (50, 20)$. *From left to right, each column depicts the powers for semi-parametric model, quantile regression with* $\tau = 0.2$, $\tau = 0.5$, *and* $\tau = 0.7$, *respectively.*

is 0.0142 (using SST and SLRT with 2000 candidate grouping parameters, the p-values are 0.095 and 0.0062, respectively). This indicates that based on the WAST, we have strong evidence for rejecting H_0, i.e., there is a subgroup in which these 500 genes mutation are sensitive.

Bibliography

Antoniou, A. C., S. Casadei, and et al. (2014). Breast-cancer risk in families with mutations in palb2. *New England Journal of Medicine 371*(6), 497–506.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.

Dunford, N. and J. T. Schwartz (1963). *Linear operators, part 2: spectral theory*. John Wiley & Sons.

Fan, A., S. Rui, and W. Lu (2017). Change-plane analysis for subgroup detection and sample size calculation. *Journal of the American Statistical Association 112*(518), 769–778.

Huang, Y., J. Cho, and Y. Fong (2021). Threshold-based subgroup testing in logistic regression models in two phase sampling designs. *Journal of the Royal Statistical Society Series C: Applied Statistics 70*(2), 291–311.

Lee, S., M. H. Seo, and Y. Shin (2011). Testing for threshold effects in regression models. *Journal of the American Statistical Association 106*(493), 220–231.

Li, J., Y. Li, B. Jin, and M. R. Kosorok (2021). Multithreshold change plane model: Estimation theory and applications in subgroup identification. *Statistics in Medicine 40*(15), 3440–3459.

Resnick, S. (2019). *A probability path*. Springer.

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

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

Yu, P. and X. Fan (2020). Threshold regression with a threshold boundary. *Journal of Business and Economic Statistics 39*(4), 953–971.