A Lack-of-fit Test With Screening In Sufficient Dimension Reduction

Yaowu Zhang[1], Wei Zhong[2] and Liping Zhu[3]

Shanghai University of Finance and Economics[1],

Xiamen University[2] and Renmin University of China[3]

Supplementary Material

We present the screening performance in the simulation and all technical proofs of theoretical results in the paper in this supplement.

S1 Screening Performances in Example 1

In the two-stage LOFTS procedure, the first-stage screening performance is crucial for the follow-up test according to Theorem 1. Thus, we first examine whether all the truly important covariates will be selected in $\mathcal{S}$ using the MDC-based screening based on the first half of the observations $\mathcal{D}_1$. It is obvious that the first four covariates (X_1, X_2, X_3, X_4) are truly important for Models (I) and (II) under both the null and alternative hypothesises. To evaluate the performance of the screening approach, we adopt the minimum model size that includes all active predictors as a criterion and report its 5%, 25%, 50%, 75%, and 95% quantiles of 1,000 replications in Table 1. We can clearly see that the MDC-based screening performs very well for two models since almost all the 95% quantiles of the minimum model sizes are equal to the true model size 4. In addition, we calculate the proportion of all truly active predictors included in the selected model, denoted by $\mathcal{P}_a$, when $|\mathcal{S}| = 8$ and $|\mathcal{S}| = 16$. All proportions are close to 1 under both the null and alternative hypothesises for two models. Thus, the MDC-based screening method is very effective to include the truly important covariates into the selected models. More numerical justifications of the MDC-based screening are referred to Shao and Zhang (2014).

S2 Proof of Theorem 1

It is obvious that on one hand, if there exists some $\boldsymbol{\beta}_{\mathcal{A}} \in \mathbb{R}^{|\mathcal{A}|\times d_0}$ such that $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}\mathbf{x}_{\mathcal{A}})$, for an arbitrary index set $\mathcal{S}$ such that $\mathcal{A} \subseteq \mathcal{S}$, by choosing $\boldsymbol{\beta}_{\mathcal{S}} = \left(\boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}, \mathbf{0}_{d_0\times(|\mathcal{S}|-|\mathcal{A}|)}\right)^{\mathrm{T}}$, we have

$$E(\mathbf{y} \mid \mathbf{x}_{\mathcal{S}}) = E\left\{E(\mathbf{y} \mid \mathbf{x}) \mid \mathbf{x}_{\mathcal{S}}\right\} = E\left\{E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}\mathbf{x}_{\mathcal{A}}) \mid \mathbf{x}_{\mathcal{S}}\right\} = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}).$$

On the other hand, when $E(\mathbf{y} \mid \mathbf{x}_{\mathcal{S}}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$ holds for some $\boldsymbol{\beta}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times d_0}$, with the sparsity assumption $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \mathbf{x}_{\mathcal{A}})$, it is quite

Table 1: The empirical performance of MDC-based screening.

Model	c	Minimum Model Size					$\mathcal{P}_a$	
		5%	25%	50%	75%	95%	$\lvert\mathcal{S}\rvert = 8$	$\lvert\mathcal{S}\rvert = 16$
(I)	0	4	4	4	4	4	1.000	1.000
	0.2	4	4	4	4	4	1.000	1.000
	0.4	4	4	4	4	4	1.000	1.000
	0.6	4	4	4	4	4	1.000	1.000
	0.8	4	4	4	4	4	0.999	0.999
	1	4	4	4	4	4	0.998	0.988
(II)	0	4	4	4	4	4	1.000	1.000
	0.2	4	4	4	4	4	1.000	1.000
	0.4	4	4	4	4	4	1.000	1.000
	0.6	4	4	4	4	4	0.999	1.000
	0.8	4	4	4	4	4	1.000	1.000
	1	4	4	4	4	4	0.991	0.997

straightforward that $E(\mathbf{y} \mid \mathbf{x}_{\mathcal{S}}) = E\{E(\mathbf{y} \mid \mathbf{x}) \mid \mathbf{x}_{\mathcal{S}}\} = E(\mathbf{y} \mid \mathbf{x}_{\mathcal{A}})$, which implies $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$, indicating $\mathrm{span}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}, \mathbf{0}_{d_0\times(p-|\mathcal{S}|)})^{\mathrm{T}}$ is a mean dimension-reduction subspace immediately. In addition, note that $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \mathbf{x}_{\mathcal{A}})$, $\mathrm{span}(I_{|\mathcal{A}|}, \mathbf{0}_{|\mathcal{A}|\times(p-|\mathcal{A}|)})^{\mathrm{T}}$ is also a mean dimension-reduction subspace. When both $\mathcal{S}_{E(\mathbf{y}|\mathbf{x})}$ and $\mathcal{S}_{E(\mathbf{y}|\mathbf{x}_{\mathcal{S}})}$ exist and are uniquely defined, all coefficients of $\mathbf{x}_{\mathcal{S}\cap\mathcal{A}^c}$ must be equal to zero because $\mathrm{span}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}, \mathbf{0}_{d_0\times(p-|\mathcal{S}|)})^{\mathrm{T}}$ $\bigcap \mathrm{span}(\mathbf{I}_{|\mathcal{A}|}, \mathbf{0}_{|\mathcal{A}|\times(p-|\mathcal{A}|)})^{\mathrm{T}} \subset \mathrm{span}(\mathbf{I}_{|\mathcal{A}|}, \mathbf{0}_{|\mathcal{A}|\times(p-|\mathcal{A}|)})^{\mathrm{T}}$, indicating that $\boldsymbol{\beta}_{\mathcal{S}} = (\boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}, \mathbf{0}_{d_0\times(|\mathcal{S}|-|\mathcal{A}|)})^{\mathrm{T}}$ and $\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}} = \boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}\mathbf{x}_{\mathcal{A}}$. Consequently, $E(\mathbf{y} \mid \mathbf{x}_{\mathcal{S}}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$ yields $E(\mathbf{y} \mid \mathbf{x}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{A}}^{\mathrm{T}}\mathbf{x}_{\mathcal{A}})$. $\square$

S3 Proof of Theorem 2

Define the empirical process

$$\zeta_{n_2}(\mathbf{s}) \overset{\text{def}}{=} n_2^{1/2}\xi_{n_2}(\mathbf{s}) \overset{\text{def}}{=} n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2} \widehat{\boldsymbol{\varepsilon}}_j \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) \overset{\text{def}}{=} \sum_{k=1}^{3} I_k(\mathbf{s}),$$

where $I_1(\mathbf{s})$, $I_2(\mathbf{s})$ and $I_3(\mathbf{s})$ are the relative three summations with $\widehat{\boldsymbol{\varepsilon}}_j$ split into $\mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) + \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) + \boldsymbol{\varepsilon}_j$. In the sequel, we will study the asymptotic behaviors of each $I_k(\mathbf{s})$.

We start with the first quantity I_1, which is defined as follows,

$$I_1(\mathbf{s}) \overset{\text{def}}{=} n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\left\{\mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\right\}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}).$$

Since $\widehat{\mathbf{m}}'(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \mathbf{m}'(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) = o_p(1)$, $\widehat{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}} = O_p(1/\sqrt{n_2})$, with Taylor's expansion, we have

$$I_1(\mathbf{s}) = n_2^{-1/2} \sum_{j=n_1+1}^{n_1+n_2} \{\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \widehat{\mathbf{m}}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\} \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) + o_p(1).$$

Similar to the proof for Lemma A2 of Ma and Zhu (2013), by invoking the U-process theory (Nolan and Pollard , 1987) and Zhu and Ng (2003), as $nh^{2d} \to \infty$ and $nh^{2t} \to 0$ as n grows into infinity, we obtain that

$$I_1(\mathbf{s}) = -n_2^{-1/2} \sum_{j=n_1+1}^{n_1+n_2} \boldsymbol{\varepsilon}_j E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\} + o_p(1).$$

From the definition of $I_3(\mathbf{s})$, we can easily combine $I_1(\mathbf{s})$ and $I_3(\mathbf{s})$ to obtain that

$$\begin{aligned} & I_1(\mathbf{s}) + I_3(\mathbf{s}) \\ = \; & n_2^{-1/2} \sum_{j=n_1+1}^{n_1+n_2} \boldsymbol{\varepsilon}_j \left[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\}\right] + o_p(1) \\ \stackrel{\text{def}}{=} \; & I(\mathbf{s}) + o_p(1). \end{aligned}$$

According to Zhu and Zhong (2015), $\mathrm{vecl}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}})$ can be represented as

$$n_2^{-1} \sum_{j=n_1+1}^{n_1+n_2} \alpha(\mathbf{x}_{j,\mathcal{S}}; \boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_j,$$

where $\alpha(\mathbf{x}_{\mathcal{S}}; \boldsymbol{\beta}_{\mathcal{S}}) = \left(E\left[\{\mathbf{m}'(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\}^{\mathrm{T}} \otimes \widetilde{\mathbf{x}}_{-d,\mathcal{S}}\right]\left\{\mathbf{m}'(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) \otimes \widetilde{\mathbf{x}}_{-d,\mathcal{S}}^{\mathrm{T}}\right\}\right)^{-1}$ $\cdot \left[\{\mathbf{m}'(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\}^{\mathrm{T}} \otimes \widetilde{\mathbf{x}}_{-d,\mathcal{S}}\right]$, $\widetilde{\mathbf{x}}_{-d,\mathcal{S}} = \mathbf{x}_{-d,\mathcal{S}} - E(\mathbf{x}_{-d,\mathcal{S}} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$ and $\mathbf{x}_{-d,\mathcal{S}} = (X_{d+1}, \ldots, X_{\mathcal{S}})^{\mathrm{T}}$. With Taylor's expansion and we can show without much difficulty that

$$
\begin{aligned}
I_2(\mathbf{s}) &= n_2^{-1/2} \sum_{j=n_1+1}^{n_1+n_2} \left\{ \frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{j,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})} \right\}^{\mathrm{T}} \exp(i\mathbf{s}^{\mathrm{T}} \mathbf{x}_{j,\mathcal{S}}) \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}} - \widehat{\boldsymbol{\beta}}_{\mathcal{S}}) + o_p(1) \\
&= -n_2^{-1/2} \sum_{j=n_1+1}^{n_1+n_2} E\left[\left\{ \frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})} \right\}^{\mathrm{T}} \exp(i\mathbf{s}^{\mathrm{T}} \mathbf{x}_{\mathcal{S}}) \right] \alpha(\mathbf{x}_{j,\mathcal{S}}; \boldsymbol{\beta}_{\mathcal{S}}) \boldsymbol{\varepsilon}_j + o_p(1).
\end{aligned}
$$

Recall that $\zeta_{n_2}(\mathbf{s}) = \sum_{k=1}^{3} I_k(\mathbf{s})$, when the null hypothesis holds, we can easily obtain $E\{\zeta_{n_2}(\mathbf{s})\} = o(1)$ as n_2 goes to infinity. Also, it is apparent that

$$
\zeta_{n_2}(\mathbf{s})\overline{\zeta_{n_2}^{\mathrm{T}}(\mathbf{s}_0)} = I(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)} + I_2(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)} + I(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)} + I_2(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)} + o_p(1).
$$

The first term of the above display can be easily calculated as

$$
\begin{aligned}
E\{I(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)}\} &= E\Bigg(\boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}_1^{\mathrm{T}} \left[\exp(i\mathbf{s}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}}) - E\left\{ \exp(i\mathbf{s}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}}) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}} \right\} \right] \\
&\qquad \left[\exp(-i\mathbf{s}_0^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}}) - E\left\{ \exp(-i\mathbf{s}_0^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}}) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}} \right\} \right] \Bigg).
\end{aligned}
$$

Note that $\widehat{\boldsymbol{\beta}}$ is asymptotically normal, namely, $n_2^{1/2}\mathrm{vecl}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}$. It follows that

$$
\begin{aligned}
E\{I_2(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)}\} &= E\Bigg[\left\{ \frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})} \right\}^{\mathrm{T}} \boldsymbol{\Sigma} \left\{ \frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}} \mathbf{x}_{2,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})} \right\} \\
&\qquad \exp\{i(\mathbf{s}^{\mathrm{T}} \mathbf{x}_{1,\mathcal{S}} - \mathbf{s}_0^{\mathrm{T}} \mathbf{x}_{2,\mathcal{S}})\} \Bigg] + o(1).
\end{aligned}
$$

By invoking the representation of $\mathrm{vecl}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}})$ again in Zhu and Zhong

(2015), we can obtain that

$$
\begin{aligned}
E\{I(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)}\} &= E\Bigg(\boldsymbol{\varepsilon}_2\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\alpha(\mathbf{x}_{2,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\Bigg[-\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}) \\
&\qquad +E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}\}\Bigg]\Bigg)+o(1), \\
E\{I_2(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)}\} &= E\Bigg(\boldsymbol{\varepsilon}_2\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\alpha(\mathbf{x}_{2,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\Bigg[-\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}) \\
&\qquad +E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}\}\Bigg]\Bigg)+o(1).
\end{aligned}
$$

Let $\zeta(\mathbf{s})$ denote a complex-valued Gaussian random process with mean zero and covariance matrix function $\mathrm{cov}\{\zeta(\mathbf{s}),\zeta^{\mathrm{T}}(\mathbf{s}_0)\}$ being of the form

$$
\begin{aligned}
& E\left[\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\boldsymbol{\Sigma}\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}\exp\{i(\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}-\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}})\}\right] \\
+\;& E\Bigg(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^{\mathrm{T}}\Big[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})-E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}\right\}\Big] \\
&\qquad \Big[\exp(-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})-E\left\{\exp(-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}\right\}\Big]\Bigg) \\
+\;& E\Bigg(\boldsymbol{\varepsilon}_2\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\alpha(\mathbf{x}_{2,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\Bigg[-\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}) \\
&\qquad +E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}\}\Bigg]\Bigg) \\
+\;& E\Bigg(\boldsymbol{\varepsilon}_2\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\alpha(\mathbf{x}_{2,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\Bigg[-\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}) \\
&\qquad +E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}-i\mathbf{s}_0^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}})\mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{2,\mathcal{S}}\}\Bigg]\Bigg),
\end{aligned}
\tag{S3.1}
$$

Particularly with $\mathbf{s} = \mathbf{s}_0$, we have

$$E\left\{\|\zeta_{n_2}(\mathbf{s})\|_\omega^2\right\} \;=\; \int_{\mathbf{s}} \frac{\mathrm{cov}\{\zeta(\mathbf{s}), \zeta^{\mathrm{T}}(\mathbf{s})\}}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s} + o(1).$$

According to Székely, Rizzo and Bakirov (2007) and Shao and Zhang (2014), if we can construct a sequence of random variables $Q_{n_2}(\nu)$, which satisfy

$$\text{(i)}\ \ Q_{n_2}(\nu) \xrightarrow{D} Q(\nu) \text{ for each } \nu > 0;$$

$$\text{(ii)}\ \ \limsup_{n_2\to\infty} E\Big|Q_{n_2}(\nu) - \|\zeta_{n_2}\|_\omega^2\Big| \to 0 \text{ as } \nu \to 0;$$

$$\text{(iii)}\ \ E\Big|Q(\nu) - \|\zeta\|_\omega^2\Big| \to 0 \text{ as } \nu \to 0,$$

then for any bounded, uniformly continuous function η, we have

$$\begin{aligned}
&\lim_{\nu\to 0}\Big|E\{\eta(Q(\nu))\} - E\{\eta(\|\zeta\|_\omega^2)\}\Big| \le \lim_{\nu\to 0} E\left\{\Big|\eta(Q(\nu)) - \eta(\|\zeta\|_\omega^2)\Big|\right\} \\
&\quad = \lim_{\nu\to 0} E\left\{\Big|\eta(Q(\nu)) - \eta(\|\zeta\|_\omega^2)\Big| I\big(\big|Q(\nu) - \|\zeta\|_\omega^2\big| \le \epsilon_0\big)\right\} \\
&\quad + \lim_{\nu\to 0} E\left\{\Big|\eta(Q(\nu)) - \eta(\|\zeta\|_\omega^2)\Big| I\big(\big|Q(\nu) - \|\zeta\|_\omega^2\big| > \epsilon_0\big)\right\} \\
&\quad \le c_1\epsilon_0 + c_2 \lim_{\nu\to 0} \mathrm{pr}\big(\big|Q(\nu) - \|\zeta\|_\omega^2\big| > \epsilon_0\big) \to 0,
\end{aligned}$$

as $\epsilon_0 \to 0$, where c_1 and c_2 are some positive constants. Following similar arguments, we can show that $\limsup_{n_2\to\infty}\Big|E\{\eta(Q_{n_2}(\nu))\} - E\{\eta(\|\zeta_{n_2}\|_\omega^2)\}\Big| \to 0$ as $\nu \to 0$. Theorem 8.4.1 of Resnick (1999) ensures that $\limsup_{n_2\to\infty}\Big|E\{\eta(Q_{n_2}(\nu))\} - E\{\eta(Q(\nu))\}\Big| \to 0$ as $\nu \to 0$. Combining the above results and using the triangle inequality, we obtain $\limsup_{n_2\to\infty}\Big|E\{\eta(\|\zeta_{n_2}\|_\omega^2)\} - E\{\eta(\|\zeta\|_\omega^2)\}\Big| \to 0$, indicating the weak convergence of $\|\zeta_{n_2}\|_\omega^2$ to $\|\zeta\|_\omega^2$ and therefore, we have $n_2 T_{n_2} = \|\zeta_{n_2}(\mathbf{s})\|_\omega^2 + o_p(1) \xrightarrow{d} \|\zeta(\mathbf{s})\|_\omega^2$, as $n_2 \to \infty$.

In the sequel, we construct a sequence of random variables and justify how it satisfies the above three requirements (i)-(iii). Following the construction in Shao and Zhang (2014), we define

$$Q_{n_2}(\nu) = \int_{D(\nu)} \frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s} \text{ and } Q(\nu) = \int_{D(\nu)} \frac{\|\zeta(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s},$$

where $D(\nu)$ is the region that $D(\nu) = \{\mathbf{s} : \nu \le \|\mathbf{s}\| \le 1/\nu\}$. Given $\epsilon = 1/M, M \in \mathbb{N}$, choose a partition $\{D_k\}_{k=1}^N$ of $D(\nu)$ into $N = N(\epsilon)$ measurable sets with diameter at most ϵ. Then it is clear that

$$Q_{n_2}(\nu) = \sum_{k=1}^N \int_{D_k} \frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s} \text{ and } Q(\nu) = \sum_{k=1}^N \int_{D_k} \frac{\|\zeta(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s}.$$

We define

$$Q_{n_2}^M(\nu) = \sum_{k=1}^N \int_{D_k} \frac{\|\zeta_{n_2}(\mathbf{s}_0(k))\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s} \text{ and } Q^M(\nu) = \sum_{k=1}^N \int_{D_k} \frac{\|\zeta(\mathbf{s}_0(k))\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s},$$

where $\{\mathbf{s}_0(k)\}_{k=1}^N$ are a set of distinct points such that $\mathbf{s}_0(k) \in D_k$. Following similar arguments for proving Theorem 4 in Shao and Zhang (2014), we can show that $Q_{n_2}^M(\nu) \xrightarrow{d} Q^M(\nu)$, $\limsup_{M\to\infty} E\left|Q^M(\nu) - Q(\nu)\right| = 0$ and $\limsup_{M\to\infty}\limsup_{n_2\to\infty} E\left|Q_{n_2}^M(\nu) - Q_{n_2}(\nu)\right| = 0$, which implies (i) immediately.

On the other hand,

$$
\begin{aligned}
& E\left|\int_{D(\nu)} \frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s} - \int_{\mathbf{s}} \frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s}\right| \\
= & \int_{\|\mathbf{s}\|<\nu} E\left\{\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s}\right\} + \int_{\|\mathbf{s}\|>1/\nu} E\left\{\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s}\right\}. \qquad \text{(S3.2)}
\end{aligned}
$$

According to Lemma 1 of Székely, Rizzo and Bakirov (2007), we have

$$
\int_{\mathbb{R}^{|\mathcal{S}|}} \frac{1-\cos(\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s} = c_0\|\mathbf{x}_{\mathcal{S}}\|,
$$

for a $|\mathcal{S}|$-dimensional vector $\mathbf{z} = (z_1, z_2, \ldots, z_{|\mathcal{S}|})$. Define the function

$$
G(y) = \int_{\|\mathbf{z}\|<y} \frac{1-\cos z_1}{\|\mathbf{z}\|^{1+|\mathcal{S}|}}d\mathbf{z}.
$$

Clearly $G(y)$ is bounded by c_0 and $\lim_{y\to 0} G(y) = 0$. Applying the Cauchy-Schwarz inequality, we can obtain that

$$
\|\zeta_{n_2}(\mathbf{s})\|^2 \le 4n_2^{-1} \sum_{j=n_1+1}^{n_1+n_2} \left\| \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - E\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right\|^2 \sum_{j=n_1+1}^{n_1+n_2} \widehat{\varepsilon}_j^2.
$$

For a given $\mathbf{x}_{\mathcal{S}}$, we define an orthonormal matrix $\mathbf{Q}$ with the first row being $\mathbf{x}_{\mathcal{S}}^{\mathrm{T}}/\|\mathbf{x}_{\mathcal{S}}\|$. It is easy to check that with the variable changing $\mathbf{u} = \|\mathbf{x}_{\mathcal{S}}\|\mathbf{Q}\ \mathbf{s}$. We have

$$
\int_{\|\mathbf{s}\|<\nu} \frac{1-\cos(\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s} = \int_{\|\mathbf{u}\|<\|\mathbf{x}_{\mathcal{S}}\|\nu} \frac{\|\mathbf{x}_{\mathcal{S}}\|\{1-\cos(u_1)\}}{\|\mathbf{u}\|^{1+|\mathcal{S}|}}d\mathbf{u} = \|\mathbf{x}_{\mathcal{S}}\|G(\|\mathbf{x}_{\mathcal{S}}\|\nu).
$$

Therefore, we have

$$
\int_{\|\mathbf{s}\|<\nu} \frac{\|\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\}\|^2}{\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s} \le 2E\left\{\|\mathbf{x}_{j,\mathcal{S}} - \mathbf{x}_{\mathcal{S}}\|G(\|\mathbf{x}_{j,\mathcal{S}} - \mathbf{x}_{\mathcal{S}}\|\nu) \mid \mathbf{x}_{j,\mathcal{S}}\right\},
$$

Consequently, the first summand in (S3.2) satisfies

$$
\begin{aligned}
&\int_{\|\mathbf{s}\|<\nu} E\left\{\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}\right\} d\mathbf{s} \\
\leq\ & 8c_0^{-1}n_2^{-1}E\left[\sum_{j=n_1+1}^{n_1+n_2} E\left\{\|\mathbf{x}_{j,\mathcal{S}}-\mathbf{x}_{\mathcal{S}}\|G(\|\mathbf{x}_{j,\mathcal{S}}-\mathbf{x}_{\mathcal{S}}\|\nu)\mid \mathbf{x}_{j,\mathcal{S}}\right\}\sum_{j=n_1+1}^{n_1+n_2}\widehat{\boldsymbol{\varepsilon}}_j^2\right].
\end{aligned}
$$

Without much difficulty, we can show that

$$
\sum_{j=n_1+1}^{n_1+n_2}\widehat{\boldsymbol{\varepsilon}}_j^2 = \sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_j^2 + o_p(n_2).
$$

Accordingly, $\lim_{\nu\to 0}\lim_{n_2\to\infty}\int_{\|\mathbf{s}\|<\nu} E\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s}\to 0$ for sufficiently small ν.

Now we consider the second summand in (S3.2), with the triangle inequity, we have $\|\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})-E\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\|^2\leq 4$, hence

$$
\begin{aligned}
\int_{\|\mathbf{s}\|>1/\nu} E\left\{\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}\right\} d\mathbf{s} &\leq 16\int_{\|\mathbf{s}\|>1/\nu}\frac{1}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s}E\left\{\sum_{j=n_1+1}^{n_1+n_2}\widehat{\boldsymbol{\varepsilon}}_j^2\right\} \\
&\leq 16O(1)\int_{\|\mathbf{s}\|>1/\nu}\frac{1}{c_0\|\mathbf{s}\|^{2}}d(\|\mathbf{s}\|)E\left\{\sum_{j=n_1+1}^{n_1+n_2}\widehat{\boldsymbol{\varepsilon}}_j^2\right\}.
\end{aligned}
$$

The second inequality holds due to the Jacobian of the $|\mathcal{S}|$-dimensional spherical transformation is $\|\mathbf{s}\|^{|\mathcal{S}|-1}\sin^{|\mathcal{S}|-2}(\theta_1)\cdots\sin(\theta_{p-2})$. Then, for sufficiently small ν,

$$
\lim_{\nu\to 0}\lim_{n_2\to\infty}\int_{\|\mathbf{s}\|>1/\nu} E\left\{\frac{\|\zeta_{n_2}(\mathbf{s})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}\right\} d\mathbf{s}\to 0,\ \text{for any } y.
$$

Thus, we complete the proof for (ii). A similar argument also applies to $Q(\nu)$, so (iii) holds. Therefore, $\|\zeta_{n_2}(\mathbf{s})\|_\omega^2 \xrightarrow{d} \|\zeta(\mathbf{s})\|_\omega^2$ as $n_2\to\infty$. Consequently, we have completed the proof.

$\square$

S4 Proof of Theorem 3

In the new model $\widetilde{\mathbf{y}} = \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + \widetilde{\boldsymbol{\varepsilon}}$, we can easily have $E(\widetilde{\boldsymbol{\varepsilon}} \mid \mathbf{x}_{\mathcal{S}}) = 0$ since δ is independent of $\mathbf{x}_{\mathcal{S}}$ and the null hypothesis is automatically satisfied.

Similar to the proof of Theorem 2, we define the process $\widetilde{\zeta}_{n_2}(\mathbf{s}) = \sqrt{n_2}\widetilde{\xi}_{n_2}(\mathbf{s})$. We also decompose $\widetilde{\zeta}_{n_2}(\mathbf{s})$ into $\widetilde{I}_1(\mathbf{s})+\widetilde{I}_2(\mathbf{s})+\widetilde{I}_3(\mathbf{s})$, and from the results in the proof of Theorem 2, we have $\widetilde{I}_1(\mathbf{s})+\widetilde{I}_3(\mathbf{s}) = n_2^{1/2}\sum_{j=1}^{n}\widetilde{\boldsymbol{\varepsilon}}_j\left[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) \mid \widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\}\right] + o_p(1/\sqrt{n_2})$. By denoting the dominant term of $\widetilde{I}_1(\mathbf{s}) + \widetilde{I}_3(\mathbf{s})$ as $\widetilde{I}(\mathbf{s})$, we can write $\widetilde{\zeta}_{n_2}(\mathbf{s})\overline{\widetilde{\zeta}_{n_2}^{\mathrm{T}}(\mathbf{s}_0)}$ as

$$\widetilde{\zeta}_{n_2}(\mathbf{s})\overline{\widetilde{\zeta}_{n_2}^{\mathrm{T}}(\mathbf{s}_0)} = \widetilde{I}(\mathbf{s})\overline{\widetilde{I}^{\mathrm{T}}(\mathbf{s}_0)} + \widetilde{I}_2(\mathbf{s})\overline{\widetilde{I}_2^{\mathrm{T}}(\mathbf{s}_0)} + \widetilde{I}(\mathbf{s})\overline{\widetilde{I}_2^{\mathrm{T}}(\mathbf{s}_0)} + \widetilde{I}_2(\mathbf{s})\overline{\widetilde{I}^{\mathrm{T}}(\mathbf{s}_0)} + o_p(1).$$

As the weights $\delta_i s$ are drawn identically and independently from $\{1, -1\}$ at random, $i = n_1 + 1, \dots, n$ and δ is independent of $\mathbf{x}_{\mathcal{S}}$ and $\boldsymbol{\varepsilon}$. Also, since $\widehat{\boldsymbol{\beta}}_{\mathcal{S}} - \boldsymbol{\beta}_{\mathcal{S}}$ is of order $O_p(1/\sqrt{n_2})$, $\widehat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon} = \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) - \mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + \mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) - \widehat{\mathbf{m}}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$, and by similar arguments in Lemma A1 of Ma and Zhu (2013), we have

$$n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\left\{\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) - \widehat{\mathbf{m}}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\right\}\delta_j\Big[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_j) - E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_j) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\}\Big] = o_p(1),$$

thus we can easily obtain that

$$\widetilde{I}(\mathbf{s}) = n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_j\delta_j\left[\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_j) - E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_j) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\}\right] + o_p(1),$$

which will yield

$$E\{\widetilde{I}(\mathbf{s})\overline{\widetilde{I}^{\mathrm{T}}(\mathbf{s}_0)}\} = E\{I(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)}\} + o(1).$$

Similar to the proof of Theorem 2, we have

$$\begin{aligned}\widetilde{I}_2(\mathbf{s}) &= n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\left\{\frac{\partial \mathbf{m}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})}{\partial \mathrm{vecl}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}})}\right\}^{\mathrm{T}} \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\mathrm{vecl}(\widetilde{\boldsymbol{\beta}}_{\mathcal{S}} - \widehat{\boldsymbol{\beta}}_{\mathcal{S}}) + o_p(1) \\ &= -n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2} E\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}} \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\alpha(\mathbf{x}_{j,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_j\delta_j + o_p(1),\end{aligned}$$

then it is clear that

$$E\{\widetilde{I}_2(\mathbf{s})\overline{\widetilde{I}_2^{\mathrm{T}}(\mathbf{s}_0)}\} = E\{I_2(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)}\} + o(1).$$

With similar arguments, it can be shown that $E\{\widetilde{I}(\mathbf{s})\overline{\widetilde{I}_2^{\mathrm{T}}(\mathbf{s}_0)}\} = E\{I(\mathbf{s})\overline{I_2^{\mathrm{T}}(\mathbf{s}_0)}\} + o(1)$ and $E\{\widetilde{I}_2(\mathbf{s})\overline{\widetilde{I}^{\mathrm{T}}(\mathbf{s}_0)}\} = E\{I_2(\mathbf{s})\overline{I^{\mathrm{T}}(\mathbf{s}_0)}\} + o(1)$ and accordingly we have,

$$E\left\{\widetilde{\zeta}_{n_2}(\mathbf{s})\overline{\widetilde{\zeta}_{n_2}^{\mathrm{T}}(\mathbf{s}_0)}\right\} = E\left\{\zeta_{n_2}(\mathbf{s})\overline{\zeta_{n_2}^{\mathrm{T}}(\mathbf{s}_0)}\right\} + o(1).$$

Using almost the same arguments in the proof of Theorem 2, we have $n_2\widetilde{T}_{n_2} = \|\widetilde{\zeta}_{n_2}\|_\omega^2 + o_p(1) \xrightarrow{d} \|\zeta(\mathbf{s})\|_\omega^2$, which completed the proof.

□

S5 Proof of Theorem 4

Under global alternative, in the case of $E(\mathbf{y} \mid \mathbf{x}_{\mathcal{S}}) \neq E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$ for any $\boldsymbol{\beta}_{\mathcal{S}}$ with rank d_0, $\widehat{\boldsymbol{\beta}}_{\mathcal{S}}$ converges to some $\boldsymbol{\beta}_{\mathcal{S}}$ such that $\mathbf{y} = \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + \boldsymbol{\varepsilon}$, where $\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) = E(\mathbf{y} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})$. It is straightforward that $E(\boldsymbol{\varepsilon} \mid \mathbf{x}_{\mathcal{S}}) \neq 0$ and $E(\boldsymbol{\varepsilon} \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) = 0$. In this case, by decomposing the empirical process in a similar way, we still have

$$\begin{aligned} I_1(\mathbf{s}) &= -n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_j E\left\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\right\} + o_p(1), \\ I_2(\mathbf{s}) &= -E\left[\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right] n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\alpha(\mathbf{x}_{j,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_j + o_p(1). \end{aligned}$$

However, under the alternative hypothesis, $E(\boldsymbol{\varepsilon} \mid \mathbf{x}_{\mathcal{S}}) \neq 0$. Consequently,

$$I_3(\mathbf{s}) = n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_j \exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}) = O_p(n_2^{1/2}).$$

Then it is clear that $\zeta_{n_2}(\mathbf{s}) - n_2^{1/2}E\{\boldsymbol{\varepsilon}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\}$ is of order $O_p(1)$ and

$$\begin{aligned} \|\zeta_{n_2}(\mathbf{s})\|^2 &= n_2\|E\boldsymbol{\varepsilon}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\|^2 \\ &+ n_2^{1/2}\left\{\zeta_{n_2}(\mathbf{s}) - n_2^{1/2}E\boldsymbol{\varepsilon}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right\}E\{\boldsymbol{\varepsilon}\exp(-i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\} \\ &+ n_2^{1/2}\left\{\zeta_{n_2}(-\mathbf{s}) - n_2^{1/2}E\boldsymbol{\varepsilon}\exp(-i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right\}E\boldsymbol{\varepsilon}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + o_p(1). \end{aligned}$$

Substituting $\zeta_{n_2}(\mathbf{s}) = I_1(\mathbf{s}) + I_2(\mathbf{s}) + I_3(\mathbf{s})$ into the above equation and simple algebra yields that

$$\int_{\mathbf{s}}\frac{\|n_2^{1/2}I_1(\mathbf{s})E\boldsymbol{\varepsilon}\exp(-i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}}d\mathbf{s} = \sum_{j=n_1+1}^{n_1+n_2}Z_{1,j} + o_p\left(n_2^{1/2}\right),$$

where $Z_{1,j}, j = n_1+1,\ldots,n_1+n_2$ are n_2 independent copies of

$$Z_1 = E\left\{\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2 E(\|\mathbf{x}_{1,\mathcal{S}} - \mathbf{x}_{2,\mathcal{S}}\| \mid \boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}}) \mid \mathbf{x}_{1,\mathcal{S}}, \boldsymbol{\varepsilon}_1\right\} \qquad \text{(S5.1)}$$

Similarly, we can have

$$\int_{\mathbf{s}} \frac{\|n_2^{1/2} I_2(\mathbf{s}) E\boldsymbol{\varepsilon} \exp(-i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s} \;=\; \sum_{j=n_1+1}^{n_1+n_2} Z_{2,j} + o_p\left(n_2^{1/2}\right),$$

where $Z_{2,j}, j = n_1+1, \ldots, n_1+n_2$ are n_2 independent copies of

$$Z_2 \;=\; E\left[\left\{\frac{\partial \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{1,\mathcal{S}})}{\partial \mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}} \boldsymbol{\varepsilon}_2 \|\mathbf{x}_{1,\mathcal{S}} - \mathbf{x}_{2,\mathcal{S}}\|\right] \alpha(\mathbf{x}_{1,\mathcal{S}}; \boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_1 \quad \text{(S5.2)}$$

In addition,

$$\int_{\mathbf{s}} \frac{n_2^{1/2}\left\|\left\{I_3(\mathbf{s}) - n_2^{1/2} E\boldsymbol{\varepsilon}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right\} E\boldsymbol{\varepsilon}\exp(-i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right\|^2}{c_0\|\mathbf{s}\|^{1+|\mathcal{S}|}} d\mathbf{s}$$
$$= \sum_{j=n_1+1}^{n_1+n_2} Z_{3,j} + o_p\left(n_2^{1/2}\right),$$

$Z_{3,j}, j = n_1+1, \ldots, n_1+n_2$ are n_2 independent copies of

$$Z_3 \;=\; -\boldsymbol{\varepsilon}_1 E\left\{\boldsymbol{\varepsilon}_2\|\mathbf{x}_{1,\mathcal{S}} - \mathbf{x}_{2,\mathcal{S}}\| \mid \mathbf{x}_{1,\mathcal{S}}\right\} - T \qquad \text{(S5.3)}$$

Combining the above results together allows us to write

$$T_{n_2} - T = 2n_2^{-1}\sum_{j=n_1+1}^{n_1+n_2}(Z_{1,j} + Z_{2,j} + Z_{3,j}) + o_p\left(n_2^{-1/2}\right).$$

The asymptotic expansions in (S5.1)-S5.3 are averages of n_2 independent and identically distributed variables. The first part of the proof is completed with the central limit theorem.

Under local alternative, $\mathbf{y} = \mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + C_{n_2}\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}) + \boldsymbol{\varepsilon}$, according to the Proof for Theorem 1 of Zhu and Zhong (2015) and similar to the proof for Lemma 3 of Guo, Wang and Zhu (2016), we have

$$\begin{aligned} n_2^{1/2}(\widehat{\boldsymbol{\beta}}_{\mathcal{S}}-\boldsymbol{\beta}_{\mathcal{S}}) &= n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\alpha(\mathbf{x}_{j,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_j \\ &\quad +C_{n_2}n_2^{1/2}E\alpha(\mathbf{x}_{\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})+o_p(1). \end{aligned}$$

With similar decomposition, we can easily obtain that

$$\begin{aligned} I_1(\mathbf{s}) &= -n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\{C_{n_2}\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})+\boldsymbol{\varepsilon}_j\}\,E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\mid\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\}+o_p(1) \\ &= -n_2^{1/2}C_{n_2}E\left[\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})E\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\mid\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}}\}\right] \\ &\quad -n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_jE\{\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}})\mid\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}\}+o_p(1). \end{aligned}$$

Similarly, we have

$$\begin{aligned} I_2(\mathbf{s}) &= -C_{n_2}n_2^{1/2}E\left[\left\{\frac{\partial\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\partial\mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right]E\{\alpha(\mathbf{x}_{\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\} \\ &\quad -E\left[\left\{\frac{\partial\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\partial\mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right]n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\alpha(\mathbf{x}_{j,\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\boldsymbol{\varepsilon}_j+o_p(1), \\ I_3(\mathbf{s}) &= n_2^{1/2}C_{n_2}E\{\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\}+n_2^{-1/2}\sum_{j=n_1+1}^{n_1+n_2}\boldsymbol{\varepsilon}_j\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{j,\mathcal{S}}). \end{aligned}$$

Particularly when $C_{n_2}=n_2^{-1/2}$, if we denote $\zeta_0(\mathbf{s})$ as a complex-valued Gaussian random process with mean function

$$\begin{aligned} E\zeta_0(\mathbf{s}) &= E\{\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\}-E\left[\left\{\frac{\partial\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\partial\mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right] \\ &\quad -E\left[\left\{\frac{\partial\mathbf{m}(\boldsymbol{\beta}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})}{\partial\mathrm{vecl}(\boldsymbol{\beta}_{\mathcal{S}})}\right\}^{\mathrm{T}}\exp(i\mathbf{s}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\right]E\{\alpha(\mathbf{x}_{\mathcal{S}};\boldsymbol{\beta}_{\mathcal{S}})\mathbf{g}(\mathbf{B}_{\mathcal{S}}^{\mathrm{T}}\mathbf{x}_{\mathcal{S}})\} \quad \text{(S5.4)} \end{aligned}$$

and covariance matrix function $\mathrm{cov}\{\zeta_0(\mathbf{s}),\zeta_0^{\mathrm{T}}(\mathbf{s}_0)\}$ being of the form (S3.1), with almost the same arguments, $\|\zeta_{n_2}(\mathbf{s})\|_\omega^2 \xrightarrow{d} \|\zeta_0(\mathbf{s})\|_\omega^2$ as $n_2 \to \infty$, and the proof of the second part is completed.

$\square$

S6 Proof of Theorem 5

Similar to Meinshausen, Meier and Bühlmann (2009), we omit the limes superior and the function min. Define $\pi(u) = B^{-1}\sum_{i=1}^B I(p_i \leq u)$. Recall that $Q(\gamma) = q_\gamma\left(\{p_i/\gamma; i = 1, \ldots, B\}\right)$ which indicates $\{Q(\gamma) \leq \alpha\}$ is equivalent to $\{\pi(\alpha\gamma) \geq \gamma\}$. Then we have

$$P\left\{Q(\gamma) \leq \alpha\right\} = P\{\pi(\alpha\gamma) \geq \gamma\} \leq \gamma^{-1}E\pi(\alpha\gamma).$$

By the definition of $\pi(\cdot)$,

$$\gamma^{-1}E\pi(\alpha\gamma) = (\gamma B)^{-1}\sum_{i=1}^{B} P(p_i \leq \alpha\gamma),$$

Note that $P(\mathcal{A} \subset \mathcal{S}_i) \to 1$ and B is fixed, thus $P(p_i \leq \alpha\gamma \mid \mathcal{A} \subset \mathcal{S}_i) = \alpha\gamma$ and the first assertion hold.

Since p_i is a random variable which follows the uniform distribution on $[0, 1]$ conditioning on $\mathcal{A} \subset \mathcal{S}_i$, then

$$E\left\{\sup_{\gamma\in(\gamma_{\min},1)} \gamma^{-1}I(p_i \leq \alpha\gamma)\right\} = \int_0^{\alpha\gamma_{\min}} \gamma_{\min}^{-1}dx + \int_{\alpha\gamma_{\min}}^{\alpha} \alpha x^{-1}dx = \alpha(1 - \log\gamma_{\min}),$$

subsequently we have

$$E\left[\sup_{\gamma\in(\gamma_{\min},1)} I\left\{\pi(\alpha\gamma)\geq\gamma\right\}\right]\leq\alpha(1-\log\gamma_{\min}),$$

implying that

$$P\left[\inf_{\gamma\in(\gamma_{\min},1)} Q(\gamma)(1-\log\gamma_{\min})\leq\alpha\right]\leq\alpha,$$

which completes the proof.

□

References

Guo, X., Wang, T. and Zhu, Lixing (2016) Model checking for parametric single-index models: a dimension-reduction model-adaptive approach. *Journal of the Royal Statistical Society, series B.* **78**: 1013–1035.

Ma, Y. and Zhu, L. P. (2013b) Efficiency loss and the linearity condition in dimension reduction. *Biometrika.* **100** 371–383.

Meinshausen, N., Meier L. and Bühlmann P. (2009) P-values for high-dimensional regression. *Journal of the American Statistical Association.* **104** 1671–1681.

Nolan, D. and Pollard, D. (1987). U-process: rates of convergence. *The Annals of Statistics* **15** 780–799.

Resnick, S. I. (1999) *A Probability Path.* Birkhäuser, Boston.

Shao, X. F. and Zhang, J. S. (2014) Martingale difference correlation and its use in high dimensional variable screening. *Journal of the American Statistical Association.* **109** 1302–1318.

Székely, G. J., Rizzo, M. L. and Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics* **35** 2769–2794.

Zhu, L. and Ng, K. (2003). Checking the adequacy of a partial linear Model. *Statistica Sinica*, **113** 763–781.

Zhu, L. P. and Zhong, W. (2015). Estimation and inference on central mean subspace for multivariate response data. *Computational Statistics and Data Analysis*, **92** 68–83.

Yaowu Zhang, Research Institute for Interdisciplinary Sciences, School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China.
E-mail: zhang.yaowu@mail.shufe.edu.cn

Wei Zhong, Wang Yanan Institute for Studies in Economics, Department of Statistics, School of Economics, Fujian Key Laboratory of Statistical Science, Xiamen University, Xiamen 361005, China.
E-mail: wzhong@xmu.edu.cn

Liping Zhu, Research Center for Applied Statistical Science, Institute of Statistics and Big Data, Renmin University of China, Beijing 100872, China.
E-mail: zhu.liping@ruc.edu.cn