Nonparametric Inference for Additive Model estimated via Simplified Smooth Backfitting

Suneel Babu Chatla

The University of Texas at El Paso

Supplementary Material

In this supplementary material we provide the lemmas and technical proofs for the theorems presented in the main article. We also provide additional results for the simulations presented in Section 5 of the main paper.

S1 Introduction

This article develops a hypothesis testing framework for additive models. For a random sample $\{Y_i, X_{i1}, \ldots, X_{id}\}_{i=1}^n$, we consider

$$Y_i = \alpha_0 + \sum_{j=1}^{d} m_j(X_{ij}) + \epsilon_i, \quad i = 1, \ldots, n, \tag{S1.1}$$

where $\{\epsilon_i, i = 1, \ldots, n\}$ is a sequence of i.i.d. random variables with mean zero and finite variance σ^2. Each additive component function $m_j(\cdot)$, $j = 1, \ldots, d$, is assumed to be an unknown smooth function and identifiable subject to the constraint, $E[m_j(\cdot)] = 0$. For simplicity in presentation, the following hypothesis testing problem is considered

$$H_0 : m_d(x_d) = 0 \quad \text{vs.} \qquad H_1 : m_d(x_d) \neq 0, \tag{S1.2}$$

which tests whether the dth covariate is significant or not.

For readability, we repeat some notations and definitions that are provided in the main document. Let $\boldsymbol{m}_j = (m_j(X_{1j}), \ldots, m_j(X_{nj}))^T$ and $\boldsymbol{x}_j = (X_{1j}, \ldots, X_{nj})^T$ for $j = 1, \ldots, d$. Let $\mathbb{X}_j = [\mathbf{1}\ \boldsymbol{x}_j\ \cdots\ \boldsymbol{x}_j^{p_j}]$ for $j = 1, \ldots, d$, and $\mathbb{X} = [\mathbf{1}\ \boldsymbol{x}_1\ \cdots\ \boldsymbol{x}_d\ \cdots\ \boldsymbol{x}_1^{p_1}\ \cdots\ \boldsymbol{x}_d^{p_d}]$, where $\mathbf{1}$ is the vector of ones. Let $\mathbb{X}^{[-0]} = [\boldsymbol{x}_1\ \cdots\ \boldsymbol{x}_d\ \cdots\ \boldsymbol{x}_1^{p_1}\ \cdots\ \boldsymbol{x}_d^{p_d}]$ which is same as $\mathbb{X}$ without the column of ones. For any matrix $\boldsymbol{A}$, define $\boldsymbol{A}^\perp = \boldsymbol{I} - \boldsymbol{A}$ and $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$.

The following definitions are needed for the theoretical results. Let $\mathcal{M}_1(\boldsymbol{H}^*_{p_j,j})$ be a space spanned by the eigenvectors of $\boldsymbol{H}^*_{p_j,j}$ with eigenvalue 1. It includes polynomials of $\boldsymbol{x}_j$ up to p_jth order because $\boldsymbol{H}^*_{p_j,j}\boldsymbol{x}_j^k = \boldsymbol{x}_j^k$, $k = 0, 1 \ldots, p_j$, and $j = 1, \ldots, d$. Suppose $\boldsymbol{G}$ is an orthogonal projection onto the space $\mathcal{M}_1(\boldsymbol{H}^*_{p_1,1}) + \cdots + \mathcal{M}_1(\boldsymbol{H}^*_{p_d,d})$ and $\boldsymbol{G}_j$ is an orthogonal projection onto the space $\mathcal{M}_1(\boldsymbol{H}^*_{p_j,j})$, $j = 1, \ldots, d$. Then,

$$\boldsymbol{G} = \boldsymbol{P}_{\mathbb{X}} = \boldsymbol{P}_{\mathbf{1}} + \boldsymbol{P}_{\boldsymbol{P}_{\mathbf{1}}^\perp \mathbb{X}^{[-0]}}, \qquad \boldsymbol{G}_j = \boldsymbol{P}_{\mathbb{X}_j}, \tag{S1.3}$$

where $\boldsymbol{P}_{\mathbb{X}} = \mathbb{X}(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}$ and $\boldsymbol{P}_{\mathbf{1}}$, $\boldsymbol{P}_{\mathbb{X}_j}$ and $\boldsymbol{P}_{\boldsymbol{P}_{\mathbf{1}}^\perp \mathbb{X}^{[-0]}}$ are defined similarly. Let $\boldsymbol{G}_{[-d]} = \boldsymbol{P}_{\mathbb{X}^{[-d]}}$ where $\mathbb{X}^{[-d]} = [\mathbf{1}\ \boldsymbol{x}_1\ \cdots\ \boldsymbol{x}_{d-1}\ \cdots\ \boldsymbol{x}_1^{p_1}\ \cdots\ \boldsymbol{x}_{d-1}^{p_{d-1}}]$ and $\boldsymbol{x}_j^k = (X_{1j}^k, \ldots, X_{nj}^k)^T$ for $k = 0, 1, \ldots, p_j$, as in (S1.3).

Define

$$\boldsymbol{C} = \boldsymbol{P}_{\boldsymbol{G}^\perp_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{G}^\perp \boldsymbol{H}^*_{p_d,d} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^\perp - \boldsymbol{G}^\perp \boldsymbol{H}^*_{p_d,d}\boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^\perp + O(n^{-1}h_d^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J}), \tag{S1.4}$$

$$\boldsymbol{D} = \boldsymbol{G}^\perp - \sum_{j=1}^{d} \left\{ \boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^\perp + O(n^{-1}h_j^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J}) \right\}, \tag{S1.5}$$

$$\boldsymbol{E} = \boldsymbol{P}_{\boldsymbol{G}^\perp_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^\perp + O(n^{-1}h_d^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J}), \tag{S1.6}$$

where $\boldsymbol{P}_{\boldsymbol{G}^\perp_{[-d]}\mathbb{X}_d^{[-0]}} = \boldsymbol{G}^\perp_{[-d]}\mathbb{X}_d^{[-0]} \left(\mathbb{X}_d^{[-0]^T} \boldsymbol{G}^\perp_{[-d]} \mathbb{X}_d^{[-0]} \right)^{-1} \mathbb{X}_d^{[-0]^T} \boldsymbol{G}^\perp_{[-d]}$, and $\mathbb{X}_d^{[-0]} = [\boldsymbol{x}_d \cdots \boldsymbol{x}_d^{p_d}]$, $\boldsymbol{J}$ is the matrix of ones, and $\boldsymbol{I}$ is an identity matrix of size n.

S1.1 Generalized Likelihood Ratio test:

The GLR test statistic is defined as

$$\lambda_n(H_0) = [\ell(H_1) - \ell(H_0)] \approxeq \frac{n}{2}\log\frac{RSS_0}{RSS_1} \approx \frac{n}{2}\frac{RSS_0 - RSS_1}{RSS_1}, \tag{S1.7}$$

where RSS_0 and RSS_1 are the residual sum of squares under the null and alternative, respectively, and reject the null hypothesis when $\lambda_n(H_0)$ is large. Analogously, the following F-type of test (Huang & Davidson 2010) is also developed

$$F_\lambda = \frac{\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}\frac{\mathrm{tr}(\boldsymbol{D})}{\mathrm{tr}(\boldsymbol{C})}, \tag{S1.8}$$

where $\mathrm{tr}(\cdot)$ denotes the trace.

S1.2 Loss Function test:

The LF test statistic is defined as

$$q_n(H_0) = \frac{Q_n}{n^{-1}RSS_1} \approx \frac{d''(0)/2\sum_{i=1}^n(\widehat{m}_+(X_{i1},\dots,X_{id}) - \widetilde{m}_+^{(-d)}(X_{i1},\dots,X_{i(d-1)}))^2 + R}{n^{-1}RSS_1}, \tag{S1.9}$$

where $\widehat{m}_+$ and $\widetilde{m}_+^{(-d)}$ are the fitted values of the models under null and alternative, respectively, RSS_1 is the residual sum of squares under alternative and R is the remainder term in the Taylor expansion of $d(\cdot)$. We reject the null hypothesis when $q_n(H_0)$ is large. The corresponding F-type of test statistic is define as

$$F_q = \frac{\boldsymbol{y}^T\boldsymbol{E}^T\boldsymbol{E}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}\frac{\mathrm{tr}(\boldsymbol{D})}{\mathrm{tr}(\boldsymbol{E}^T\boldsymbol{E})}, \tag{S1.10}$$

for $\boldsymbol{E}$ defined in (S1.6).

S2 Assumptions

We repeat the assumptions that are outlined in the main document.

(A.1). *The densities $f_j(\cdot)$ of X_j are Lipschitz-continuous and bounded away from 0 and have bounded support Ω_j for $j = 1,\dots,d$. The joint density of X_j and $X_{j'}$, $f_{j,j'}(\cdot,\cdot)$, for $1 \le j \ne j' \le d$, is also Lipschitz continuous and have bounded support.*

(A.2). *The kernel $K(\cdot)$ is a bounded symmetric density function with bounded support and satisfies Lipschitz condition. The bandwidth $h_j \to 0$ and $nh_j^2/(\ln n)^2 \to \infty$, $j = 1,\dots,d$, as $n \to \infty$.*

(A.3). *The $(2p_j + 2)-th$ derivative of $m_j(\cdot)$, $j = 1,\dots,d$, exists.*

(A.4). *The error ϵ has mean 0, variance σ^2, and finite fourth moment.*

(A.5). *The loss function $d : \mathbb{R} \to \mathbb{R}^+$ has a unique minimum at 0, and $d(z)$ is monotonically nondecreasing as $|z| \to \infty$. Furthermore, $d(z)$ is twice continuously differentiable at 0 with $d(0) = 0$, $d'(0) = 0$, $M = \frac{1}{2}d''(0) \in (0,\infty)$, and $|d''(z) - d''(0)| \le C|z|$ for any z near 0.*

S3 Required Lemmas and Proofs

The explicit expressions for the estimators $\widehat{\boldsymbol{m}}_j^*$, $j = 1,\dots,d$ are provided as follows. Let

$$\boldsymbol{A}_j = (\boldsymbol{I} - (\boldsymbol{H}^*_{p_j,j} - \boldsymbol{G}_j))^{-1}(\boldsymbol{H}^*_{p_j,j} - \boldsymbol{G}_j) \;\text{ and }\; \boldsymbol{A} = \sum_{j=1}^d \boldsymbol{A}_j. \tag{S3.11}$$

By Proposition 3 in Buja et al. (1989), we obtain $\widehat{\boldsymbol{m}}_j^* = \boldsymbol{A}_j(\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{G}^\perp\boldsymbol{y}$. Therefore, the fitted response under the alternative can be written as

$$\widehat{\boldsymbol{y}}(H_1) = \sum_{j=1}^d \widehat{\boldsymbol{m}}_j^* + \boldsymbol{G}\boldsymbol{y} = \left(\sum_{j=1}^d \boldsymbol{A}_j(\boldsymbol{I} + \boldsymbol{A})^{-1}\boldsymbol{G}^\perp + \boldsymbol{G}\right)\boldsymbol{y} := \boldsymbol{W}\boldsymbol{y}. \tag{S3.12}$$

Analogously, the fitted response under the null can be written as

$$\widehat{\boldsymbol{y}}(H_0) = \left(\sum_{j=1}^{d-1} \boldsymbol{A}_j \left(\boldsymbol{I} + \boldsymbol{A}\right)^{-1} \boldsymbol{G}_{[-d]}^{\perp} + \boldsymbol{G}_{[-d]} \right) \boldsymbol{y} := \boldsymbol{W}^{[-d]} \boldsymbol{y}. \tag{S3.13}$$

The following lemma simplifies the expressions for the $RSS_0 - RSS_1$.

Lemma 1 *Denote $A_{n1} = (\boldsymbol{I} - \boldsymbol{W}^{[-d]})^T (\boldsymbol{I} - \boldsymbol{W}^{[-d]})$ and $A_{n2} = (\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W})$ where $\boldsymbol{W}$ and $\boldsymbol{W}^{[-d]}$ are defined in (S3.12) and (S3.13), respectively. If assumptions (A.1)–(A.3) hold, then*

$$RSS_0 - RSS_1 = \boldsymbol{y}^T (A_{n1} - A_{n2}) \boldsymbol{y} \tag{S3.14}$$

and

$$A_{n1} - A_{n2} = \boldsymbol{P}_{\boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]}} + \boldsymbol{G}^{\perp} \boldsymbol{H}_{p_d,d}^* + \boldsymbol{H}_{p_d,d}^* \boldsymbol{G}^{\perp} - \boldsymbol{G}^{\perp} \boldsymbol{H}_{p_d,d}^* \boldsymbol{H}_{p_d,d}^* \boldsymbol{G}^{\perp} + O(n^{-1} h_d^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}), \tag{S3.15}$$

where $\boldsymbol{P}_{\boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]}} = \boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]} \left(\mathbb{X}_d^{[-0]T} \boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]} \right)^{-1} \mathbb{X}_d^{[-0]T} \boldsymbol{G}_{[-d]}^{\perp}$, and $\mathbb{X}_d^{[-0]} = [\boldsymbol{x}_d \cdots \boldsymbol{x}_d^{p_d}]$, $\boldsymbol{J}$ is the matrix of ones, and $\boldsymbol{I}$ is an identity matrix of size n.

Proof. As shown in Huang & Yu (2019), the diagonal elements of $\boldsymbol{H}_{p_j,j}^*$ and $\boldsymbol{H}_{p_j,j}^* \boldsymbol{H}_{p_j,j}^*$, $j = 1, \ldots, d$, are of order $O(n^{-1} h_j^{-1})$ and the off-diagonal elements are of order $O(n^{-1})$. Similarly, the elements of $\boldsymbol{H}_{p_j,j}^* \boldsymbol{H}_{p_l,l}^*$ are of order $O(n^{-1})$ for $j \neq l$. Since the elements of $\boldsymbol{H}_{p_j,j}^*$ are of smaller order, we can write

$$\begin{aligned} \boldsymbol{A}_j &= \left[\boldsymbol{I} - (\boldsymbol{H}_{p_j,j}^* - \boldsymbol{G}_j) \right]^{-1} (\boldsymbol{H}_{p_j,j}^* - \boldsymbol{G}_j) = \boldsymbol{H}_{p_j,j}^* - \boldsymbol{G}_j + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}), \\ \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A})^{-1} &= \left[\sum \boldsymbol{A}_j \right] \left[\boldsymbol{I} + \sum \boldsymbol{A}_j \right]^{-1} = \sum_{j=1}^{d} \left\{ \boldsymbol{H}_{p_j,j}^* - \boldsymbol{G}_j + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right\}, \end{aligned} \tag{S3.16}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{J}$ is the matrix of 1's of size n. Since $\boldsymbol{G}_j \in \mathcal{M}_1(\boldsymbol{H}_{p_j,j}^*)$, it follows that $\boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j = \boldsymbol{G}_j$. Therefore,

$$\boldsymbol{H}_{p_j,j}^* - \boldsymbol{G}_j = \boldsymbol{H}_{p_j,j}^* (\boldsymbol{I} - \boldsymbol{G}_j) = \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j^{\perp}.$$

Consequently, we write (S3.12) as

$$\begin{aligned} \boldsymbol{W} &= \boldsymbol{A}(\boldsymbol{I} + \boldsymbol{A})^{-1} \boldsymbol{G}^{\perp} + \boldsymbol{G} \\ &= \sum_{j=1}^{d} \left\{ \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j^{\perp} + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right\} \boldsymbol{G}^{\perp} + \boldsymbol{G} \\ &= \sum_{j=1}^{d} \left\{ \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}^{\perp} + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right\} + \boldsymbol{G}, \end{aligned} \tag{S3.17}$$

where the last step uses the fact that $\boldsymbol{G}_j^{\perp} \boldsymbol{G}^{\perp} = \boldsymbol{G}^{\perp}$. Let $\boldsymbol{G}_{[-d]}$ be the parametric projection matrix of the first $d-1$ components defined similar to (S1.3). Based on the properties of the projection matrices, we obtain

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{G}_{[-d]} + \boldsymbol{P}_{\boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]}}, \\ \boldsymbol{G}^{\perp} &= \boldsymbol{G}_{[-d]}^{\perp} - \boldsymbol{P}_{\boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]}}, \end{aligned} \tag{S3.18}$$

where $\boldsymbol{P}_{\boldsymbol{G}_{[-d]}^{\perp} \mathbb{X}_d^{[-0]}}$ defined in (S1.6). Combination of (S3.17) and (S3.18) and some rearrangement of terms yields

$$\boldsymbol{I} - \boldsymbol{W} = \boldsymbol{G}^{\perp} - \left[\sum_{j=1}^{d-1} \left\{ \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_{[-d]}^{\perp} + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right\} + \boldsymbol{H}_{p_d,d}^* \boldsymbol{G}^{\perp} + O(n^{-1} h_d^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right].$$

Observe that $\boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_{[-d]}^{\perp} = \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j^{\perp} \boldsymbol{G}_{[-d]}^{\perp}$, for $j = 1, \ldots, d-1$. The elements of $\boldsymbol{G}_j^{\perp} \boldsymbol{H}_{p_j,j}^* \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j^{\perp}$ are of smaller order than the elements of $\boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_j^{\perp}$ since the latter has eigenvalues in $[0, 1)$. Therefore,

$$\begin{aligned} &(\boldsymbol{I} - \boldsymbol{W})^T (\boldsymbol{I} - \boldsymbol{W}) \\ &= \boldsymbol{G}^{\perp} - \left[\sum_{j=1}^{d-1} \left\{ \boldsymbol{G}_{[-d]}^{\perp} \boldsymbol{H}_{p_j,j}^* + \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_{[-d]}^{\perp} - \boldsymbol{G}_{[-d]}^{\perp} \boldsymbol{H}_{p_j,j}^* \boldsymbol{H}_{p_j,j}^* \boldsymbol{G}_{[-d]}^{\perp} + O(n^{-1} h_j^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right\} \right. \\ &\qquad \left. + \boldsymbol{G}^{\perp} \boldsymbol{H}_{p_d,d}^* + \boldsymbol{H}_{p_d,d}^* \boldsymbol{G}^{\perp} - \boldsymbol{G}^{\perp} \boldsymbol{H}_{p_d,d}^* \boldsymbol{H}_{p_d,d}^* \boldsymbol{G}^{\perp} + O(n^{-1} h_d^{-1} \boldsymbol{I} + n^{-1} \boldsymbol{J}) \right]. \end{aligned}$$

Similar computations yield

$$(\boldsymbol{I}-\boldsymbol{W}^{[-d]})^T(\boldsymbol{I}-\boldsymbol{W}^{[-d]})$$
$$=\boldsymbol{G}^{\perp}_{[-d]}-\left[\sum_{j=1}^{d-1}\left\{\boldsymbol{G}^{\perp}_{[-d]}\boldsymbol{H}^{*}_{p_j,j}+\boldsymbol{H}^{*}_{p_j,j}\boldsymbol{G}^{\perp}_{[-d]}-\boldsymbol{G}^{\perp}_{[-d]}\boldsymbol{H}^{*}_{p_j,j}\boldsymbol{H}^{*}_{p_j,j}\boldsymbol{G}^{\perp}_{[-d]}+O(n^{-1}h_j^{-1}\boldsymbol{I}+n^{-1}\boldsymbol{J})\right\}\right].$$

Hence,

$$A_{n1}-A_{n2}=\boldsymbol{P}_{\boldsymbol{G}^{\perp}_{[-d]}\mathbb{X}_d^{[-0]}}+\boldsymbol{G}^{\perp}\boldsymbol{H}^{*}_{p_d,d}+\boldsymbol{H}^{*}_{p_d,d}\boldsymbol{G}^{\perp}-\boldsymbol{G}^{\perp}\boldsymbol{H}^{*}_{p_d,d}\boldsymbol{H}^{*}_{p_d,d}\boldsymbol{G}^{\perp}+O(n^{-1}h_d^{-1}\boldsymbol{I}+n^{-1}\boldsymbol{J}).$$

■

Lemma 2 *If assumptions (A.1)–(A.4) hold, then under* $H_0:\boldsymbol{m}_d=0$

$$\begin{aligned}d_{1n}&\equiv\boldsymbol{m}_+^T(A_{n1}-A_{n2})\boldsymbol{m}_++2\epsilon^T(A_{n1}-A_{n2})\boldsymbol{m}_+\\&=O_p\left(1+\sum_{j=1}^{d}nh_j^{4(p_j+1)}+\sum_{j=1}^{d}\sqrt{n}h_j^{2(p_j+1)}\right),\end{aligned}\tag{S3.19}$$

where A_{n1} *and* A_{n2} *are defined in Lemma 1 and* $\boldsymbol{m}_+=\boldsymbol{m}_1+\ldots+\boldsymbol{m}_d$.

Proof. From Huang & Chan (2014), we have $\boldsymbol{H}^{*}_{p_j,j}\boldsymbol{m}_j=\boldsymbol{m}_j+\boldsymbol{1}\cdot O_P(h_j^{2(p_j+1)})$ for $p_j=0,1,2,3$, and $\boldsymbol{G}_j\boldsymbol{m}_j=\boldsymbol{1}\cdot O_P(1/\sqrt{n})$ for $j=1,\ldots,d$, where $\boldsymbol{1}$ is the vector of ones. The calculations analogous to Lemma 1 yield, under $H_0:\boldsymbol{m}_d=0$, that

$$\begin{aligned}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{m}_+&=\left(\boldsymbol{I}-\boldsymbol{G}-\sum_{j=1}^{d}\left\{\boldsymbol{H}^{*}_{p_j,j}\boldsymbol{G}^{\perp}+O(n^{-1}h_j^{-1}\boldsymbol{I}+n^{-1}\boldsymbol{J})\right\}\right)\boldsymbol{m}_+\\&=\boldsymbol{m}_+-\boldsymbol{m}_++\boldsymbol{1}\cdot O_p\left(\sum_{j=1}^{d}h_j^{2(p_j+1)}\right)+\boldsymbol{1}\cdot O_p\left(1/\sqrt{n}\right)\\&=\boldsymbol{1}\cdot O_p\left(\sum_{j=1}^{d}h_j^{2(p_j+1)}\right)+\boldsymbol{1}\cdot O_p\left(1/\sqrt{n}\right).\end{aligned}\tag{S3.20}$$

Consequently,

$$\boldsymbol{m}_+^T(\boldsymbol{I}-\boldsymbol{W})^T(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{m}_+=O_p\left(1+\sum_{j=1}^{d}nh_j^{4(p_j+1)}\right),\tag{S3.21}$$

$$\boldsymbol{m}_+^T(\boldsymbol{I}-\boldsymbol{W}^{[-d]})^T(\boldsymbol{I}-\boldsymbol{W}^{[-d]})\boldsymbol{m}_+=O_p\left(1+\sum_{j=1}^{d-1}nh_j^{4(p_j+1)}\right).$$

Moreover,

$$(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{\epsilon}=\boldsymbol{\epsilon}+\boldsymbol{1}.o_P\left(1\right)$$

which implies that under assumption (A.4)

$$\boldsymbol{\epsilon}^T(\boldsymbol{I}-\boldsymbol{W})^T(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{m}_+=O_p(1+\sum_{j=1}^{d}\sqrt{n}h_j^{2(p_j+1)}).\tag{S3.22}$$

Hence, the stated result (S3.19) follows from (S3.20) and (S3.22). ■

Theorem 1 *(GLR test) Suppose that conditions (A.1)–(A.4) hold and* $0\le p_j\le 3$, $j=1,\ldots,d$. *Then, under* H_0 *for the testing problem (11)*

$$P\left\{\sigma_n^{-1}\left(\lambda_n(H_0)-\mu_n-\frac{1}{2\sigma^2}d_{1n}\right)<t|\mathcal{X}\right\}\xrightarrow{d}\boldsymbol{\Phi}(t),\tag{S3.23}$$

where $d_{1n}=O_p\left(1+\sum_{j=1}^{d}nh_j^{4(p_j+1)}+\sum_{j=1}^{d}\sqrt{n}h_j^{2(p_j+1)}\right)$ *and* $\boldsymbol{\Phi}(\cdot)$ *is the standard normal distribution. Furthermore, if* $nh_j^{4(p_j+1)}h_d\to 0$ *for* $j=1,\ldots,d$, *conditional on the sample space* $\mathcal{X}$, $r_k\lambda_n(H_0)\to\chi^2_{r_k\mu_n}$ *as* $n\to\infty$. *Similarly,*

$$F_\lambda=\frac{2\lambda_n(H_0)tr(\boldsymbol{D})}{ntr(\boldsymbol{C})}\to F_{tr(C),tr(D)},\tag{S3.24}$$

as $n\to\infty$, *where* $tr(C)$ *and* $tr(D)$ *are the degrees of freedom.*

Proof.: Recall that

$$\lambda_n(H_0) \approx \frac{n}{2} \frac{RSS_0 - RSS_1}{RSS_1}. \tag{S3.25}$$

Proof of S3.23:

(i) Asymptotic Expression for $RSS_0 - RSS_1$:
Using the notation from Lemma 1, we write

$$\begin{aligned} RSS_0 - RSS_1 &= \boldsymbol{y}^T(A_{n1} - A_{n2})\boldsymbol{y} \\ &= \boldsymbol{\epsilon}^T(A_{n1} - A_{n2})\boldsymbol{\epsilon} + \Big[\boldsymbol{m}_+^T(A_{n1} - A_{n2})\boldsymbol{m}_+ + 2\boldsymbol{\epsilon}^T(A_{n1} - A_{n2})\boldsymbol{m}_+\Big] \\ &= \boldsymbol{\epsilon}^T\boldsymbol{C}\boldsymbol{\epsilon} + d_{1n}, \end{aligned} \tag{S3.26}$$

where

$$\begin{aligned} \boldsymbol{C} &= A_{n1} - A_{n2} \\ &= \boldsymbol{P}_{\boldsymbol{G}^{\perp}_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{G}^{\perp}\boldsymbol{H}^*_{p_d,d} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^{\perp} - \boldsymbol{G}^{\perp}\boldsymbol{H}^*_{p_d,d}\boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^{\perp} + O(n^{-1}h_d^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J}) \\ &= (c_{ij})_{1\le i,j\le n}, \end{aligned} \tag{S3.27}$$

and $d_{1n} = \boldsymbol{m}_+^T(A_{n1} - A_{n2})\boldsymbol{m}_+ + 2\boldsymbol{\epsilon}^T(A_{n1} - A_{n2})\boldsymbol{m}_+$. With the help of Lemma 2, we can bound the bias term d_{1n} by $O_p\left(1 + \sum_{k=1}^d nh_k^{4(p_k+1)} + \sum_{k=1}^d \sqrt{n}h_k^{2(p_k+1)}\right)$. We write

$$\boldsymbol{\epsilon}^T\boldsymbol{C}\boldsymbol{\epsilon} = \sum_{i=1}^n \epsilon_i^2 c_{ii} + \sum_{i\neq j}^n \epsilon_i\epsilon_j c_{ij} = L_1 + L_2. \tag{S3.28}$$

Since the leading terms of $\boldsymbol{C}$ in (S3.27) come from $\boldsymbol{H}^*_{p_d,d}$, we obtain $c_{ii} = O(n^{-1}h_d^{-1} + n^{-1})$. Combination of Assumption (A.4) and Chebyshev inequality yields $L_1 = \sigma^2 E(\sum_{i=1}^n c_{ii}) + O_p(1/\sqrt{n}h_d)$. After some algebra,

$$\begin{aligned} E(\sum_{i=1}^n c_{ii}) = \frac{2|\Omega_d|}{h_d}\Bigg(&\sum_{l=0}^{p_d}\sum_{m=0}^{p_d} v_{l+m}s^{(m+1),(l+1)} \\ &- \frac{1}{2}\int\Big\{\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(u)(-1)^m s^{(m+1),(l+1)}\Big\}^2 du\Bigg) + o_p(h_d^{-1}), \end{aligned}$$

where $|\Omega_d|$ is the length of the support of the density $f_d(x_d)$ of X_d. It remains to show that L_2 converges to normal in distribution. Note that $E[L_2] = 0$ and

$$\mathrm{Var}(L_2|\mathcal{X}) = \mathrm{var}\left(\sum_{i\neq j}^n \epsilon_i\epsilon_j c_{ij}\right) = 4\sigma^4\sigma_n^2,$$

where

$$\begin{aligned} \sigma_n^2 = \sum_{i<j} c_{ij}^2 = \frac{|\Omega_d|}{h_d}\int\Bigg\{&\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(u)(-1)^m s^{(m+1),(l+1)} \\ &- \frac{1}{2}\int\Big[\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(u+v)(-1)^m s^{(m+1),(l+1)}\Big] \\ &\times\Big[\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(v)(-1)^m s^{(m+1),(l+1)}\Big]dv\Bigg\}^2 du + o_p(h_d^{-1}). \end{aligned}$$

Application of Proposition 3.2 of de Jong (1987) yields

$$\frac{1}{2\sigma^2}\sigma_n^{-1}L_2|\mathcal{X} \xrightarrow{d} N(0,1). \tag{S3.29}$$

(ii) Asymptotic Expression for RSS_1/n: By the definition of $RSS_!$,

$$\begin{aligned} RSS_1 &= \boldsymbol{\epsilon}^T A_{n2}\boldsymbol{\epsilon} + \boldsymbol{m}_+^T A_{n2}\boldsymbol{m}_+ + 2\boldsymbol{\epsilon}^T A_{n2}\boldsymbol{m}_+ \\ &= \boldsymbol{\epsilon}^T A_{n2}\boldsymbol{\epsilon} + d_{0n}. \end{aligned}$$

The arguments analogous to Lemma 2 yields

$$\begin{aligned} d_{0n} &= \boldsymbol{m}_+^T A_{n2} \boldsymbol{m}_+ + 2\boldsymbol{\epsilon}^T A_{n2} \boldsymbol{m}_+ \\ &= O_p\left(1 + \sum_{k=1}^d n h_k^{4(p_k+1)} + \sum_{k=1}^d \sqrt{n} h_k^{2(p_k+1)}\right). \end{aligned}$$

Note that, under the condition (A.2), $d_{0n}/n = o_p(1)$. Thus, it remains to show that $n^{-1}\boldsymbol{\epsilon}^T A_{n2}\boldsymbol{\epsilon} = \sigma^2 + o_p(1)$. From the proof of Lemma 1, we obtain

$$\begin{aligned} n^{-1}\boldsymbol{\epsilon}^T A_{n2}\boldsymbol{\epsilon} &= n^{-1}\boldsymbol{\epsilon}^T \left(\boldsymbol{I} - \boldsymbol{G} - \left(\sum_{j=1}^d \left\{\boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^\perp + O(n^{-1}h_j^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J})\right\}\right)\right)\boldsymbol{\epsilon} + o_p(1) \\ &= n^{-1}\sum_{i=1}^n \epsilon_i^2 + o_p(1) \\ &= \sigma^2 + o_p(1), \end{aligned}$$

which follows from the Chebyshev inequality and using the arguments analogous to the derivation of variance for (S3.28).

(iii) Conclusion : By part(i), part(ii) and definition of $\lambda_n(H_0)$, we have

$$\begin{aligned} \lambda_n(H_0) &\approxeq \frac{RSS_0 - RSS_1}{2RSS_1/n} \\ &= \frac{d_{1n} + L_1 + L_2}{2\sigma^2} \\ &= \frac{d_{1n} + \sigma^2 E(\sum_{i=1}^n c_{ii}) + L_2}{2\sigma^2} + o_p(h_d^{-1}) \\ &\approxeq \frac{d_{1n}}{2\sigma^2} + \mu_n + \frac{L_2}{2\sigma^2}, \end{aligned}$$

where $\mu_n = E(\sum_{i=1}^n c_{ii})/2$. Therefore, (S3.29) implies

$$P\left\{\sigma_n^{-1}\left(\lambda_n(H_0) - \mu_n - \frac{1}{2\sigma^2}d_{1n}\right) < t|\mathcal{X}\right\} \xrightarrow{d} \Phi(t).$$

If $nh_k^{4(p_k+1)}h_d \to 0$ for $k = 1, \ldots, d$, then $d_{1n} = o_p(h_d^{-1})$ which is dominated by $\mu_n = O(h_d^{-1})$. Then $r_k\lambda_n(H_0)|\mathcal{X} \to \chi^2_{r_k\mu_n}$ as $n \to \infty$.

Proof of (S3.24):

By virtue of Lemma 1, the GLR test statistic is defined as

$$\begin{aligned} \lambda_n(H_0) &\approxeq \frac{n\boldsymbol{y}^T(A_{n1} - A_{n2})\boldsymbol{y}}{2\boldsymbol{y}^T A_{n2}\boldsymbol{y}} \\ &= \frac{n\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}}{2\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}}, \end{aligned}$$

for $\boldsymbol{D} = \boldsymbol{G}^\perp - \left(\sum_{j=1}^d \left\{\boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^\perp + O(n^{-1}h_j^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J})\right\}\right)$ and $\boldsymbol{C}$ defined in (S3.27). As discussed in Huang & Davidson (2010), for $F-$type statistics,

$$F = \frac{\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}/\mathrm{tr}(\boldsymbol{C})}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}/\mathrm{tr}(\boldsymbol{D})}, \tag{S3.30}$$

the $F-$ distribution is warranted if $\boldsymbol{C}$ and $\boldsymbol{D}$ are both projection matrices (symmetric and idempotent) and they are orthogonal to each other. Clearly, both $\boldsymbol{C}$ and $\boldsymbol{D}$ are not projection matrices and not orthogonal to each other. However, following Huang & Chen (2008), we show these properties hold asymptotically. It is straightforward to show both $\boldsymbol{C}$ and $\boldsymbol{D}$ are asymptotically idempotent. Now it remains to show that they are asymptotically orthogonal. Observe

$$E\{\boldsymbol{C}\boldsymbol{D}\boldsymbol{y}|\mathcal{X}\} = \boldsymbol{C}\left(\sum_{j=1}^d \left[O(h_j^{2(p_j+1)}) + O_p(1/\sqrt{nh_j} + 1/\sqrt{n})\right]\right) = o(1).$$

Based on the definition of asymptotic orthogonality in Huang & Chen (2008), we claim $\boldsymbol{C}$ and $\boldsymbol{D}$ are asymptotically orthogonal. Therefore,

$$F = \frac{\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}/\mathrm{tr}(\boldsymbol{C})}{\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}/\mathrm{tr}(\boldsymbol{D})} = \frac{2\lambda_n(H_0)}{n}\frac{\mathrm{tr}(\boldsymbol{D})}{\mathrm{tr}(\boldsymbol{C})} = \frac{2\lambda_n(H_0)}{\mathrm{tr}(\boldsymbol{C})},$$

because $\mathrm{tr}(\boldsymbol{D})/n = \frac{1}{n}\mathrm{tr}\left(\boldsymbol{G}^{\perp} - \left(\sum_{j=1}^{d}\left\{\boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^{\perp} + O(n^{-1}h_j^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J})\right\}\right)\right) \to 1$ as $nh_d \to \infty$ and $n \to \infty$. ■

Theorem 2 *(LF test) Suppose that conditions (A.1)–(A.5) hold and $0 \le p_j \le 3$, $j = 1,\ldots,d$. Then, under H_0 for the testing problem (11)*

$$P\left\{\delta_n^{-1}\left(\frac{q_n(H_0)}{M} - \nu_n - \frac{1}{\sigma^2}b_{1n}\right) < t|\mathcal{X}\right\} \xrightarrow{d} \boldsymbol{\Phi}(t), \tag{S3.31}$$

where $b_{1n} = O_P\left(1 + \sum_{j=1}^{d} nh_j^{4(p_j+1)}\right)$. Furthermore, if $nh_j^{4(p_j+1)}h_d \to 0$ for $j = 1,\ldots,d$, conditional on $\mathcal{X}$, $s_kM^{-1}q_n(H_0) \to \chi^2_{s_k\nu_n}$ as $n \to \infty$. Similarly,

$$F_q = \frac{q_n(H_0)tr(\boldsymbol{D})}{Mntr(\boldsymbol{E}^T\boldsymbol{E})} \to F_{tr(\boldsymbol{E}^T\boldsymbol{E}),tr(\boldsymbol{D})}, \tag{S3.32}$$

as $n \to \infty$.

Proof. Proof of (S3.31):
Consider the LF test statistic in (S1.9)

$$q_n(H_0) = \frac{Q_n}{n^{-1}SSR1} = \frac{\sum_{i=1}^{n} d\left\{\sum_{j=1}^{n} e_{ij}Y_j\right\}}{n^{-1}RSS_1},$$

where $d(\cdot)$ is the loss function defined in Assumption (A.5) and e_{ij} is the (i,j)th, $1 \le i,j \le n$, element of $\boldsymbol{P}_{\boldsymbol{G}^{\perp}_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^{\perp} + O(n^{-1}h_d^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J})$. The arguments analogous to Lemma 1 yield that , under H_0,

$$\sum_{j=1}^{n} e_{ij}m_{+j} = O\left(\sum_{k=1}^{d} h_k^{2(p_k+1)}\right) + O_P\left(1/\sqrt{n}\right).$$

Note that the dominant orders for the elements e_{ij}'s come from $\boldsymbol{H}^*_{p_d,d}$. Therefore, diagonal elements e_{ii}'s are of order $O(n^{-1}h_d^{-1}+n^{-1})$ and the off-diagonal elements $e_{ii'}$, $i \ne i'$, are of order $O(n^{-1})$. By Taylor series expansion of loss function $d(\cdot)$ in the neighborhood of 0, we obtain

$$d(z) \approx d(0) + d'(0)z + Mz^2 + 1/2(d''(\bar{z}) - d''(0))z^2 = Mz^2 + R,$$

where $d(0) = 0$, $d'(0) = 0$, $M = d''(0)/2! \in (0,\infty)$ and $\bar{z}$ lies between 0 and z. Assumption (A.5) implies $R \le Cz^3$. Therefore

$$\sum_{i=1}^{n} d\left\{\sum_{j=1}^{n} e_{ij}Y_j\right\} = M\sum_{i=1}^{n}\left(\sum_{j=1}^{n} e_{ij}\epsilon_j\right)^2 + O_P\left(1 + \sum_{k=1}^{d} nh_k^{4(p_k+1)}\right) + \sum_{i=1}^{n} R_i, \tag{S3.33}$$

where each $R_i \le C|\sum_{j=1}^{n} e_{ij}\epsilon_j|^3$. The idea is to show that the first term in (S3.33) converges to normal distribution and the third term is of smaller order. Using the relation $E|x|^3 \le [E|x|^4]^{3/4}$, we obtain

$$\begin{aligned}\sum_{i=1}^{n} R_i \le C\sum_{i=1}^{n} E|\sum_{j=1}^{n} e_{ij}\epsilon_j|^3 \le C\sum_{i=1}^{n}\left\{E\left|\sum_{j=1}^{n} e_{ij}\epsilon_j\right|^4\right\}^{3/4} \\ \le CC^*\sum_{i=1}^{n}\left\{\sum_{j=1}^{n} e_{ij}^4 E[\epsilon_j]^4\right\}^{3/4} + CC^*\sum_{i=1}^{n}\left\{\sum_{j\ne j'}^{n} e_{ij}^2e_{ij'}^2 E[\epsilon_j\epsilon_{j'}]^2\right\}^{3/4} \\ = O_p\left(n(1/n^3h_d^3) + n(n(n-1)/n^4h_d^2)^{3/4}\right) \\ = O_p(n^{-2}h_d^{-3}) + O_p(1/h_d\sqrt{nh_d}) = O_p(1/h_d\sqrt{nh_d}),\end{aligned} \tag{S3.34}$$

where C^* is some positive constant and the exact value of it can be calculated using the expression in page 101 of Lin & Bai (2010). Note that, the first term in (S3.33) can be written as

$$M\sum_{i=1}^{n}\left(\sum_{j=1}^{n}e_{ij}\epsilon_j\right)^2 = M\sum_{i=1}^{n}\sum_{j=1}^{n}e_{ij}^2\epsilon_j^2 + M\sum_{i=1}^{n}\sum_{j\neq j'}^{n}e_{ij}e_{ij'}\epsilon_j\epsilon_{j'} = T_{n1}+T_{n2}. \tag{S3.35}$$

After some algebra, we obtain

$$\sum_{i,j}e_{ij}^2 = \frac{|\Omega_d|}{h_d}\int\left\{\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(u)(-1)^m s^{(m+1),(l+1)}\right\}^2 du + o(h_d^{-1}).$$

Application of Chebyshev inequality yields that $T_{n1} = M\sigma^2\nu_n + O_p(1/\sqrt{n}h_d)$ where $\nu_n = E(\sum_{i,j}e_{ij}^2)$. Now it remains to show that T_{n2} converges to normal in distribution. Observe that $E(T_{n2}) = 0$ and

$$var(T_{n2}|\mathcal{X}) = M^2\sigma^4\sum_{j\neq j'}\left(\sum_{i=1}^{n}e_{ij}e_{ij'}\right)^2 = M^2\sigma^4\sum_{j\neq j'}(\boldsymbol{e}_j^T\boldsymbol{e}_{j'})^2 = M^2\sigma^4\delta_n^2,$$

where $\boldsymbol{e}_k = (e_{1k},\ldots,e_{nk})^T$ and $\delta_n^2 = \sum_{j\neq j'}(\boldsymbol{e}_j^T\boldsymbol{e}_{j'})^2$. We note that the leading terms of

$$\delta_n^2 = \frac{|\Omega_d|}{h_d}\int\left\{\int\left[\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(u+v)(-1)^m s^{(m+1),(l+1)}\right]\right.$$
$$\left.\times\left[\sum_{l=0}^{p_d}\sum_{m=0}^{p_d}(K_l * K_m)(v)(-1)^m s^{(m+1),(l+1)}\right]dv\right\}^2 du + o_p(h_d^{-1}).$$

Therefore, application of Proposition 3.2 of de Jong (1987) yields

$$\frac{1}{M\sigma^2}\delta_n^{-1}T_{n2}|\mathcal{X} \xrightarrow{d} N(0,1). \tag{S3.36}$$

By plugging (S3.34) and (S3.35) in (S3.33), we obtain

$$Q_n = T_{n1} + T_{n2} + \sum_{i=1}^{n}R_i + O_p\left(1+\sum_{k=1}^{d}nh_k^{4(p_k+1)}\right).$$

Since $n^{-1}RSS_1 = \sigma^2 + o_p(1)$ and $q_n = Q_n/n^{-1}RSS_1$, we have,

$$\frac{q_n(H_0)}{M} - \nu_n - \frac{b_{1n}}{\sigma^2} + o_p(h_d^{-1}) \approxeq \frac{T_{n2}}{M\sigma^2}, \tag{S3.37}$$

where $b_{1n} = O_p\left(1+\sum_{k=1}^{d}nh_k^{4(p_k+1)}\right)$. Therefore, combination of (S3.36) and (S3.37) yields

$$P\left\{\delta_n^{-1}\left(\frac{q_n(H_0)}{M} - \nu_n - \frac{1}{\sigma^2}b_{1n}\right) < t|\mathcal{X}\right\} \xrightarrow{d} \Phi(t).$$

If $nh_k^{4(p_k+1)}h_d \to 0$ for $k = 1,\ldots,d$, then $b_{1n} = o_p(h_d^{-1})$ which is dominated by ν_n. Then $s_kM^{-1}q_n(H_0)|\mathcal{X} \to \chi^2_{s_k\nu_n}$ as $n\to\infty$.

Proof of (S3.32): Recall

$$q_n(H_0) = \frac{Q_n}{n^{-1}SSR1} = \frac{\sum_{i=1}^{n}d\left\{\sum_{j=1}^{n}e_{ij}Y_j\right\}}{n^{-1}RSS_1}.$$

By Taylor's expansion, as in part(a), the numerator can be written as $M\boldsymbol{y}^T\boldsymbol{E}^T\boldsymbol{E}\boldsymbol{y}+R_n$ where R_n is the remainder term which is of order $o_p(h_d^{-1})$. As in part (b) of Theorem 1, we show that $\boldsymbol{E}^T\boldsymbol{E}$ is asymptotically an idempotent matrix and $\boldsymbol{E}^T\boldsymbol{E}$ and $\boldsymbol{D}$ are asymptotically orthogonal. Hence, the LFT statistic is

$$q_n(H_0) \approxeq \frac{M\boldsymbol{y}^T\boldsymbol{E}^T\boldsymbol{E}\boldsymbol{y}}{n^{-1}\boldsymbol{y}^T\boldsymbol{D}\boldsymbol{y}} = F\frac{Mtr(\boldsymbol{E}^T\boldsymbol{E})}{n^{-1}tr(\boldsymbol{D})}, \tag{S3.38}$$

which implies that, as in part(b) of Theorem 1, we have

$$\frac{q_n(H_0)}{M\mathrm{tr}(\boldsymbol{E}^T\boldsymbol{E})} \to F_{\mathrm{tr}(\boldsymbol{E}^T\boldsymbol{E}),\mathrm{tr}(\boldsymbol{D})},$$

as $nh_d \to \infty$ and $n \to \infty$.

■

For the following theorem, we consider the contiguous alternative of the form

$$H_{1n} : m_d(X_d) = M_n(X_d), \tag{S3.39}$$

where $M_n(X_d) \to 0$ and $M_n \in \mathcal{M}_n(\rho;\eta)$.

Theorem 3 *Suppose* $E\{M_n(X_d)|X_1,\ldots,X_{d-1}\} = 0$ *and* $h_d \cdot \sum_{i=1}^n M_n^2(X_{id}) \xrightarrow{P} C_M$ *for some constant* C_M*. Suppose* $0 \le p_j \le 3$*, for* $j = 1,\ldots,d$.

(i) [GLR test] Suppose that conditions (A.1)-(A.4) hold. Under H_{1n} *for the testing problem (11),*

$$P\left\{\sigma_n^{-1}\left(\lambda_n(H_0) - \mu_n - \frac{d_{1n}+d_{2n}}{2\sigma^2}\right) < t|\mathcal{X}\right\} \xrightarrow{d} \boldsymbol{\Phi}(t),$$

where μ_n*,* d_{1n} *and* σ_n *are same as those in Theorem 1 and*

$$d_{2n} = \sum_{i=1}^n M_n^2(X_{id})(1+o_p(1)).$$

(ii) [LF test] Suppose that conditions (A.1)-(A.5) hold. Under H_{1n} *for the testing problem (11),*

$$P\left\{\delta_n^{-1}\left(\frac{q_n(H_0)}{M} - \nu_n - \frac{b_{1n}+b_{2n}}{\sigma^2}\right) < t|\mathcal{X}\right\} \xrightarrow{d} \boldsymbol{\Phi}(t),$$

where ν_n*,* b_{1n} *and* δ_n *are same as those in Theorem 2 and*

$$b_{2n} = \sum_{i=1}^n M_n^2(X_{id})(1+o_p(1)).$$

Proof. Part (i): Under H_{1n} (S3.39), the arguments analogous to Lemma 1 yields,

$$\begin{aligned}(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{m}_+ &= (\boldsymbol{G}^\perp - \sum_{j=1}^d \boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^\perp)(\boldsymbol{m}_1+\ldots+\boldsymbol{m}_{d-1}+\boldsymbol{M}_n) + o_p(1)\\ &= \boldsymbol{1}\cdot O\left(\sum_{k=1}^d h_k^{2(p_k+1)}\right) + \boldsymbol{1}\cdot O_p\left(1/\sqrt{n}\right),\end{aligned}$$

where $\boldsymbol{M}_n(\boldsymbol{x}_d) = (M_n(X_{1d}),\ldots,M_n(X_{nd}))^T$, $\boldsymbol{M}_n \in \mathcal{M}_n(\rho;\eta)$ defined in (26) and $\boldsymbol{1}$ is the vector of ones of size n. Similarly,

$$(\boldsymbol{I}-\boldsymbol{W}^{[-d]})\boldsymbol{m}_+ = \left(\boldsymbol{G}^\perp_{[-d]} - \sum_{j=1}^{d-1}\boldsymbol{H}^*_{p_j,j}\boldsymbol{G}^\perp_{[-d]}\right)\boldsymbol{M}_n + \boldsymbol{1}\cdot O\left(\sum_{k=1}^{d-1} h_k^{2(p_k+1)}\right) + \boldsymbol{1}\cdot O_p\left(1/\sqrt{n}\right).$$

Observe that, same set of arguments yield $(\boldsymbol{I}-\boldsymbol{W})\boldsymbol{\epsilon} = \boldsymbol{\epsilon} + o_p(1)$ and $(\boldsymbol{I}-\boldsymbol{W}^{[-d]})\boldsymbol{\epsilon} = \boldsymbol{\epsilon} + o_p(1)$. Consider,

$$\begin{aligned}RSS_0 - RSS_1 &= \boldsymbol{y}^T(A_{n1}-A_{n2})\boldsymbol{y}\\ &= \boldsymbol{\epsilon}^T(A_{n1}-A_{n2})\boldsymbol{\epsilon} + \boldsymbol{m}_+^T(A_{n1}-A_{n2})\boldsymbol{m}_+ + 2\boldsymbol{\epsilon}^T(A_{n1}-A_{n2})\boldsymbol{m}_+\\ &= I_{n1}+I_{n2}+I_{n3}.\end{aligned} \tag{S3.40}$$

straightforward computations yield

$$I_{n2} = \boldsymbol{M}_n^T\boldsymbol{M}_n + O_p\left(1+\sum_{k=1}^d nh_k^{4(p_k+1)}\right) \qquad \text{and}$$

$$I_{n3} = \boldsymbol{\epsilon}^T\boldsymbol{M}_n + O_p\left(1+\sum_{k=1}^d \sqrt{n}h_k^{2(p_k+1)}\right).$$

Plugging the above results and the I_{n1} value from Theorem 1 in (S3.40), we obtain

$$RSS_0 - RSS_1 = L_1 + L_2 + C_n + d_{2n} + d_{1n}, \tag{S3.41}$$

where L_1, L_2, d_{1n} are same as defined in Theorem 1,

$$d_{2n} = \sum_{i=1}^{n} M_n^2(X_{id}) + o_p(h_d^{-1}) \qquad \text{and}$$
$$C_n = \sum_{i=1}^{n} \epsilon_i M_n(X_{id}).$$

The proof follows by proceeding similar to Theorem 1.

Part (ii): The arguments analogous to Lemma 1 yield that, under H_1,

$$\begin{aligned}(\boldsymbol{W} - \boldsymbol{W}^{[-d]})\boldsymbol{m}_+ &= (\boldsymbol{P}_{\boldsymbol{G}^{\perp}_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^{\perp})(\boldsymbol{m}_1 + \ldots + \boldsymbol{m}_{d-1} + \boldsymbol{M}_n) + o_p(1)\\ &= \boldsymbol{M}_n + \boldsymbol{1}\cdot O\left(\sum_{k=1}^{d} h_k^{2(p_k+1)}\right) + \boldsymbol{1}\cdot O_p\left(1/\sqrt{n}\right).\end{aligned}$$

Similarly, $(\boldsymbol{W} - \boldsymbol{W}^{[-d]})\boldsymbol{\epsilon} = (\sum_{j=1}^{n} e_{1j}\epsilon_j, \ldots, \sum_{j=1}^{n} e_{nj}\epsilon_j)^T$ where e_{ij} is the (i,j)th $1 \le i,j \le n$, element in the matrix $\left\{\boldsymbol{P}_{\boldsymbol{G}^{\perp}_{[-d]}\mathbb{X}_d^{[-0]}} + \boldsymbol{H}^*_{p_d,d}\boldsymbol{G}^{\perp} + O(n^{-1}h_d^{-1}\boldsymbol{I} + n^{-1}\boldsymbol{J})\right\}$. Note that the leading terms of e_{ij}'s are of the same order as the elements in $\boldsymbol{H}^*_{p_d,d}$. By Taylor expansion,

$$\begin{aligned}\sum_{i=1}^{n} d\left\{\sum_{j=1}^{n} e_{ij}Y_j\right\} &= M\sum_{i=1}^{n}\left(\sum_{j=1}^{n} e_{ij}\epsilon_j\right)^2 + M\sum_{i=1}^{n} M_n^2(X_{id})\\ &\quad + O_p\left(1 + \sum_{k=1}^{d} nh_k^{4(p_k+1)}\right) + \sum_{i=1}^{n} R_i,\end{aligned} \tag{S3.42}$$

where each $R_i \le C|\sum_{j=1}^{n} e_{ij}\epsilon_j|^3$. The proof follows by proceeding similar to Theorem 2. ■

Theorem 4 *Under conditions (A.1)-(A.5), if* $h_k^{2(p_k+1)} = O(h_d^{2(p_d+1)})$ *and* $0 \le p_k \le 3$*, for* $k = 1, \ldots, d-1$*, then for the testing problem (11), both GLR and LF tests can detect alternatives with rate* $\rho_n = n^{-\frac{4(p_d+1)}{8p_d+9}}$ *when* $h_d = c_* n^{-\frac{2}{8p_d+9}}$ *for some constant* c_**.*

Proof. The proof uses arguments analogous to Theorem 5 in Fan & Jiang (2005). We provide proof only for the GLR test and similar arguments can be used to prove the LF test. Under the contiguous alternative $H_{1n} : m_d(X_d) = M_n(X_d)$, it follows from (i) of Theorem 3,

$$\lambda_n(H_0) = \mu_n + \frac{L_2}{2\sigma^2} + \frac{d_{2n} + C_n}{2\sigma^2} + O_p\left(1 + \sum_{k=1}^{d} nh_k^{4(p_k+1)} + \sum_{k=1}^{d}\sqrt{n}h_k^{2(p_k+1)}\right), \tag{S3.43}$$

where $d_{2n} = \sum_{i=1}^{n} M_n^2(X_{id})$ and $C_n = \sum_{i=1}^{n}\epsilon_i M_n(X_{id})$. Since the probability of the type II error at H_{1n} is defined as $\beta(\alpha, M_n) = P(\phi_h = 0|m_d = M_n)$, it implies that

$$\begin{aligned}\beta(\alpha, M_n) &= P\{\sigma_n^{-1}(-\lambda_n(H_0) + \mu_n) \ge z_\alpha|\mathcal{X}\}\\ &= P\left\{\sigma_n^{-1}\left(-\frac{L_2}{2\sigma^2} - \frac{d_{2n}+C_n}{2\sigma^2} + O_p\left(1 + \sum_{k=1}^{d} nh_k^{4(p_k+1)} + \sum_{k=1}^{d}\sqrt{n}h_k^{2(p_k+1)}\right)\right) \ge z_\alpha|\mathcal{X}\right\}\\ &= P_{1n} + P_{2n},\end{aligned}$$

with

$$P_{1n} = P\left\{\sigma_n^{-1}(-\frac{L_2}{2\sigma^2}) + \sqrt{n}h_d^{(4p_d+5)/2}t_{1n} + nh_d^{(8p_d+9)/2}t_{2n} - \sqrt{h_d}t_{3n} \ge z_\alpha, |t_{1n}| \le M, |t_{2n}| \le M|\mathcal{X}\right\},$$
$$P_{2n} = P\left\{\sigma_n^{-1}(-\frac{L_2}{2\sigma^2}) + \sqrt{n}h_d^{(4p_d+5)/2}t_{1n} + nh_d^{(8p_d+9)/2}t_{2n} - \sqrt{h_d}t_{3n} \ge z_\alpha, |t_{1n}| \ge M, |t_{2n}| \ge M|\mathcal{X}\right\},$$

and

$$t_{1n} = \left(\sqrt{n}h_d^{(4p_d+5)/2}\sigma_n\right)^{-1} O_p\left(1 + \sum_{k=1}^{d} \sqrt{n}h_k^{2(p_k+1)}\right) = O_p(1),$$

$$t_{2n} = \left(nh_d^{(8p_d+9)/2}\sigma_n\right)^{-1} O_p(\sum_{k=1}^{d} nh_k^{4(p_k+1)}) = O_p(1),$$

$$t_{3n} = (\sqrt{h_d}\sigma^2\sigma_n)^{-1}\frac{1}{2}[d_{2n} + C_n].$$

Note that $E[C_n|\mathcal{X}] = 0$ and $var(C_n|\mathcal{X}) = O(\sum_{i=1}^n M_n^2(X_{id}))$ and hence $C_n = O_p(\sqrt{d_{2n}})$. Analogous arguments to Lemma B.7 of Fan & Jiang (2005) lead to

$$\sqrt{h_d}t_{3n} \to \infty \qquad \text{only when} \qquad n\sqrt{h_d}\rho^2 \to \infty.$$

We choose $h_d \le c_0^{-\frac{1}{2(p_d+1)}} n^{-\frac{1}{4(p_d+1)}}$. This implies, $\sqrt{n}h_d^{(4p_d+5)/2} \ge c_0 nh_d^{(8p_d+9)/2}$, $\sqrt{n}h_d^{(4p_d+5)/2} \to 0$, and $nh_d^{(8p_d+9)/2} \to 0$. Hence, for $h_d \to 0$ and $nh_d \to \infty$, it follows that $\beta(\alpha,\rho) \to 0$ only when $n\sqrt{h_d}\rho^2 \to +\infty$. This implies $\rho_n^2 = n^{-1}h_d^{-1/2}$ and the possible minimum value of ρ_n in this setting is $n^{\frac{-(8p_d+7)}{16(p_d+1)}}$. When $nh_d^{4(p_d+1)} \to \infty$, for any $\delta > 0$, there exists a constant $M > 0$ such that $P_{2n} < \frac{\delta}{2}$ uniformly in $M_n \in \mathcal{M}_n(\rho;\eta)$. Then

$$\beta(\alpha,\rho) \le \frac{\delta}{2} + P_{1n}.$$

We note that $\sup_{\mathcal{M}_n(\rho;\eta)} P_{1n} \to 0$ only when $B(h_d) \equiv nh_d^{(8p_d+9)/2}M - nh_d^{1/2}\rho^2 \to -\infty$. The function $B(h_d)$ attains the minimum value

$$-\frac{8(p_d+1)}{8p_d+9}[M(8p_d+9)]^{-\frac{1}{8(p_d+1)}} n\rho^{\frac{8p_d+9}{4(p_d+1)}}$$

at $h_d = \left[\frac{\rho^2}{M(8p_d+9)}\right]^{\frac{1}{4(p_d+1)}}$. With simple algebra, in this setting, we obtain the corresponding minimum value of $\rho_n = n^{-\frac{4(p_d+1)}{8p_d+9}}$ at $h_d = c_* n^{-\frac{2}{8p_d+9}}$ for some constant c_*. ■

Theorem 5 *[Relative efficiency] Suppose Conditions (A.1)–(A.5) hold, $h \propto n^{-\omega}$ for $\omega \in (0, 1/(4p_d+5))$ and $p_j = 0$ for $j = 1, \ldots, d$. Then Pitman's relative efficiency of the LF test over the GLR test under H_n in (29) is given by*

$$ARE(q_n, \lambda_n) = \left[\frac{\int\left\{2(K_0 * K_0)(u) - \int(K_0 * K_0)(u+v)(K_0 * K_0)(v)dv\right\}^2 du}{\int\left\{\int(K_0 * K_0)(u+v)(K_0 * K_0)(v)dv\right\}^2 du}\right]^{1/(2-3\omega)}.$$

The asymptotic relative efficiency $ARE(q_n, \lambda_n)$ is larger than 1 for any kernel satisfying Condition (A.2) and $K(\cdot) \le 1$.

Proof. Pitman's asymptotic relative efficiency of the LF test over the GLR test is the limit of the ratio of the sample sizes required by the two tests to have the same asymptotic power at the same significance level, under the same local alternative [Pitman (2018), Chapter 7]. Suppose n_1 and n_2 are the sample sizes required for the LF test and the GLR test, respectively. The Pitman's asymptotic relative efficiency of q_n to λ_n is defined as

$$\text{ARE}(q_n, \lambda_n) = \lim_{n_1,n_2\to\infty} \frac{n_1}{n_2},$$

under the condition that λ_n and q_n have the same asymptotic power under the same local alternatives $n_1^{-1/2}h_{d_1}^{-1/2}g_1(x_d) \sim n_2^{-1/2}h_{d_2}^{-1/2}g_2(x_d)$ in the sense that

$$\lim_{n_1,n_2\to\infty} \frac{n_1^{-1/2}h_{d_1}^{-1/2}g_1(x_d)}{n_2^{-1/2}h_{d_2}^{-1/2}g_2(x_d)} = 1.$$

Given $h_{d_i} = cn_i^{-\omega}$, $i = 1, 2$, we have $n_1^{-2\gamma}\sum_{i=1}^n g_1^2(X_{di}) \sim n_2^{-2\gamma}\sum_{i=1}^n g_2^2(X_{di})$, where $\gamma = (1-\omega)/2$. Hence,

$$\lim_{n_1,n_2\to\infty}\left(\frac{n_1}{n_2}\right)^{2\gamma} = \frac{\sum_{i=1}^n g_1^2(X_{di})}{\sum_{i=1}^n g_2^2(X_{di})}. \tag{S3.44}$$

From Theorem 3(i), we have

$$\frac{\lambda_{n_1}(H_0)-\mu_{n_1}}{\sigma_{n_1}} \xrightarrow{d} N(\xi,1),$$

under $H_{n_1}: m_d(x_d) = n_1^{-1/2}h_{d_1}^{-1/2}g_1(x_d)$, where $\xi = [\sum_{i=1}^n g_1^2(X_{di})]/(2\sigma^2\sigma_{n_1})$ with σ_{n_1} is defined in Theorem 1. Also, from Theorem 3(ii), we have

$$\frac{M^{-1}q_{n_2}(H_0)-\nu_{n_2}}{\delta_{n_2}} \xrightarrow{d} N(\psi,1),$$

under $H_{n_2}: m_d(x_d) = n_2^{-1/2}h_{d_2}^{-1/2}g_2(x_d)$, where $\psi = [\sum_{i=1}^n g_2^2(X_{di})]/(\sigma^2\delta_{n_2})$ with δ_{n_2} is defined in Theorem 2. To have the same asymptotic power, the noncentrality parameters must be equal which means $\xi = \psi$ or

$$\frac{\sum_{i=1}^n g_1^2(X_{di})}{\sum_{i=1}^n g_2^2(X_{di})} = \frac{2\sigma_{n_1}}{\delta_{n_2}}. \tag{S3.45}$$

Combination of (S3.44) and (S3.45) yields, for $p_j = 0$, $j = 1,\dots,d$,

$$\begin{aligned}
\mathrm{ARE}(q_n,\lambda_n) &= \left[\frac{2h_{d_1}^{1/2}\sigma_{n_1}}{h_{d_2}^{1/2}\delta_{n_2}}\right]^{2/(2-3\omega)} = \left[\frac{4h_{d_1}\sigma_{n_1}^2}{h_{d_2}\delta_{n_2}^2}\right]^{1/(2-3\omega)} \\
&= \left[\frac{\int\left\{2(K_0*K_0)(u)-\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du}{\int\left\{\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du}\right]^{1/(2-3\omega)}.
\end{aligned}$$

Now, we show $\mathrm{ARE}(q_n,\lambda_n) \geq 1$ for any positive kernels with $K(\cdot) \leq 1$. It is sufficient to show that

$$\begin{aligned}
&\int\left\{2(K_0*K_0)(u)-\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du \\
&\qquad \geq \int\left\{\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du.
\end{aligned}$$

From Jensen's inequality and Fubini's theorem we obtain

$$\begin{aligned}
&\int\left\{\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du \\
&\qquad \leq \int\int(K_0*K_0)^2(u+v)(K_0*K_0)(v)dvdu \\
&\qquad = \int(K_0*K_0)^2(u)du.
\end{aligned} \tag{S3.46}$$

Triangle inequality and (S3.46) yields that

$$\begin{aligned}
&\left\{\int\left\{2(K_0*K_0)(u)-\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du\right\}^{1/2} \\
&\geq 2\left\{\int(K_0*K_0)^2(u)du\right\}^{1/2} - \left\{\int\left\{\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\}^2du\right\}^{1/2} \\
&\geq 2\left\{\int(K_0*K_0)^2(u)du\right\}^{1/2} - \left\{\int(K_0*K_0)^2(u)du\right\}^{1/2} \\
&= \left\{\int(K_0*K_0)^2(u)du\right\}^{1/2}.
\end{aligned} \tag{S3.47}$$

Combination of (S3.47) and (S3.46) yields

$$\begin{aligned}
&\left\|2(K_0*K_0)(u)-\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\|_2 \\
&\qquad \geq \left\|(K_0*K_0)(u)\right\|_2 \geq \left\|\int(K_0*K_0)(u+v)(K_0*K_0)(v)dv\right\|_2.
\end{aligned}$$

Hence, the LF test is asymptotically more efficient than the GLR test.

■

S4 Numerical Comparison- Extra results

S4.1 Conditional Bootstrap

(a) Fix the bandwidths at their estimated values $(\widehat{h}_1, \widehat{h}_2, \widehat{h}_3, \widehat{h}_4)$ and then obtain the estimates of additive functions under both null and unrestricted additive models.
(b) Compute λ_n, q_n, $\lambda_n(FJ)$, F_λ, F_q, S_n and the residuals $\widehat{\epsilon}_i$, $i = 1, \ldots, n$, from the unrestricted model.
(c) For each $(X_{1i}, X_{2i}, X_{3i}, X_{4i})$, draw a bootstrap residual $\widehat{\epsilon}_i^*$ from the centered empirical distribution of $\widehat{\epsilon}_i$ and compute $Y_i^* = \widehat{m}_0 + \widehat{m}_1(X_{1i}) + \widehat{m}_3(X_{3i}) + \widehat{m}_4(X_{4i}) + \widehat{\epsilon}_i^*$, where $\widehat{m}_1$, $\widehat{m}_3$ and $\widehat{m}_4$ are the estimated additive functions under the restricted model in step (a). This forms a conditional bootstrap sample $(Y_i^*, X_{1i}, X_{2i}, X_{3i}, X_{4i})_{i=1}^n$.
(d) Using the bootstrap sample in step (c) and bandwidths in step (a), obtain λ_n^*, q_n^*, $\lambda_n^*(FJ)$, F_λ^*, F_q^*, S_n^*.
(e) Repeat steps (c) and (d) for a total of B times, where B is large number. We then obtain a sample of statistics.
(f) Compute the bootstrap P values $P_\lambda^* = B^{-1} \sum_{l=1}^{B} \mathbf{1}(\lambda_n < \lambda_{nl}^*)$ for all the statistics. Reject H_0 at a prespecified significance level α if and only if $P_\lambda^* < \alpha$. Repeat this process for the all the above statistics.

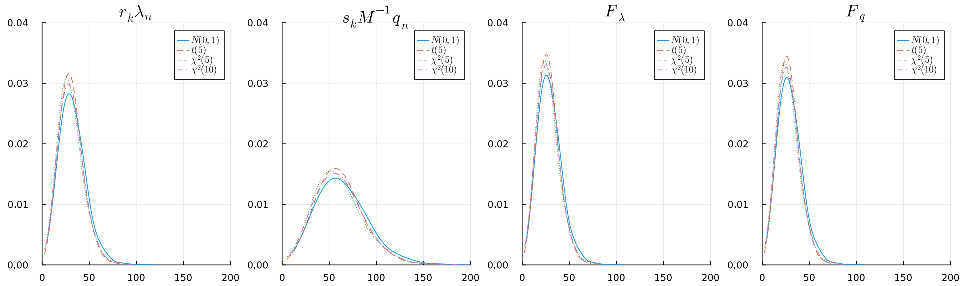

Fig. S4.1 Estimated densities for scaled GLR and LF test statistics, and F statistics, among 1000 simulations under different errors (— normal; $- - -t(5)$; $\cdots \chi^2(5)$; $- \cdot - \chi^2(10)$). Here, the errors except normal are scaled to have mean 0 and variance 1.

References

1. Buja, A., Hastie, T. & Tibshirani, R. (1989), 'Linear smoothers and additive models', *The Annals of Statistics* **17**(2), 453–510.

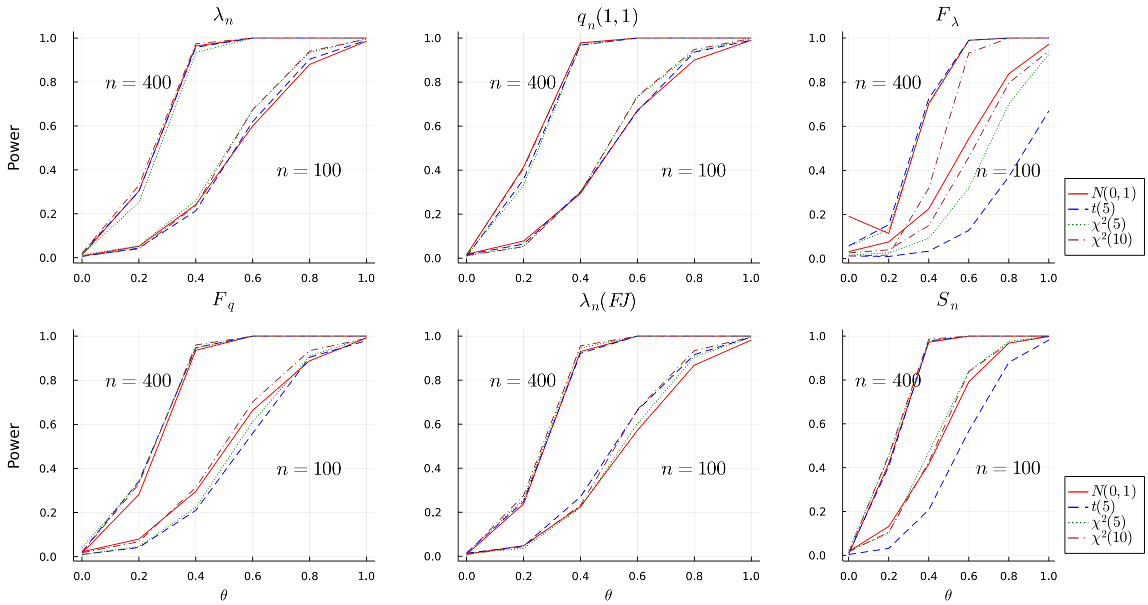

Fig. S4.2 Power of the tests under alternative model sequence (32) using bandwidths $S_X n^{-2/17}$ at 1% level of significance. Only the LF test with LINEX loss function (31) for $s = 1, t = 1$ is reported. The power values are similar for other choices of s and t.

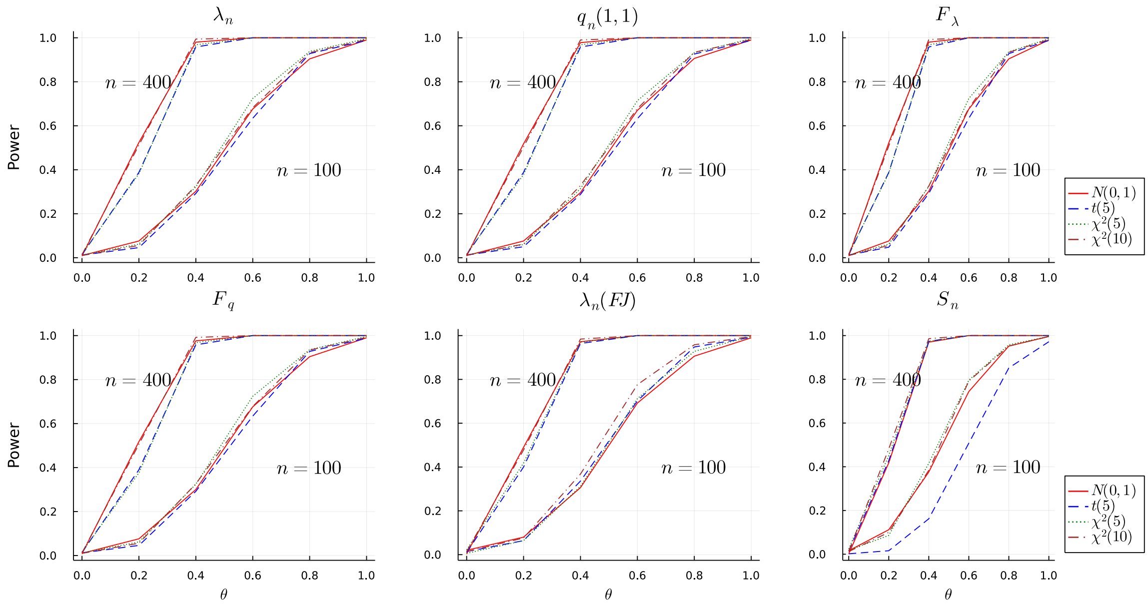

Fig. S4.3 Power of the tests under alternative model sequence (32) using optimal bandwidths (cross-validation) at 1% level of significance. Only the LF test with LINEX loss function (31) for $s = 1, t = 1$ is reported. The power values are similar for other choices of s and t.

2. de Jong, P. (1987), 'A central limit theorem for generalized quadratic forms', *Probability Theory and Related Fields* **75**(2), 261–277.
3. Fan, J. & Jiang, J. (2005), 'Nonparametric inferences for additive models', *Journal of the American Statistical Association* **100**(471), 890–907.
4. Huang, L.-S. & Chan, K.-S. (2014), 'Local polynomial and penalized trigonometric series regression', *Statistica Sinica* **24**, 1215–1238.
5. Huang, L.-S. & Chen, J. (2008), 'Analysis of variance, coefficient of determination and f − test for local polynomial regression', *The Annals of Statistics* **36**(5), 2085–2109.
6. Huang, L.-S. & Davidson, P. W. (2010), 'Analysis of variance and f-tests for partial linear models with applications to environmental health data', *Journal of the American Statistical Association* **105**(491), 991–1004.
7. Huang, L.-S. & Yu, C.-H. (2019), 'Classical backfitting for smooth-backfitting additive models', *Journal of Computational and Graphical Statistics* **28**(2), 386–400.
8. Lin, Z. & Bai, Z. (2010), Probability inequalities of random variables, *in* 'Probability Inequalities', Springer, pp. 37–50.
9. Pitman, E. J. (2018), *Some basic theory for statistical inference: Monographs on applied probability and statistics*, CRC Press.