Annals of the Institute of Statistical Mathematics manuscript No.
(will be inserted by the editor)

Supplementary material for "An Optimal Test for the Additive Model with Discrete or Categorical Predictors"

Abhijit Mandal

Received: date / Revised: date

Appendix B: Backfitting Estimators

Let us define $\boldsymbol{m}_p = (m_p(X_{p1}), m_p(X_{p2}), \cdots, m_p(X_{pn}))^T$ for $p \in \{1, 2, \cdots, P\}$, and $\boldsymbol{m}_{P+q} = (m_{P+q}(Z_{q1}), m_{P+q}(Z_{q2}) \ \cdots, m_{P+q}(Z_{qn}))^T$ for $q \in \{1, 2, \cdots, Q\}$. The additive components, $\boldsymbol{m}_1, \boldsymbol{m}_2, \cdots, \boldsymbol{m}_{P+Q}$, are estimated using the backfitting estimators. The first step is to select a suitable smoothing matrix $\boldsymbol{S}_d$ for $d \in \{1, 2, \cdots, P+Q\}$, where $\widehat{\boldsymbol{m}}_d = \boldsymbol{S}_d \boldsymbol{Y}_{res}$ is the estimator of $\boldsymbol{m}_d$, and $\boldsymbol{Y}_{res}$ is the residual of $\boldsymbol{Y} = (Y_1, \cdots, Y_n)^T$ given other additive components. This step is then repeated until convergence of all additive components (Hastie and Tibshirani 1990). As $\boldsymbol{X}$ contains categorical or discrete valued random variables, the bin smoother at a point mass is appropriate. Suppose, for $j = 1, 2, \cdots, k_p$, there are n_{pj} observations at $X_p = x_{pj}$, where $\sum_{j=1}^{k_p} n_{pj} = n$, $p = 1, 2, \cdots, P$. If the observations are sorted according to the values of X_p, then the smoothing matrix for m_p is given by

$$\boldsymbol{S}_p = \begin{bmatrix} n_{p1}^{-1}\boldsymbol{J}_{n_{p1}} & \boldsymbol{O}_{n_{p1},n_{p2}} & \cdots & \boldsymbol{O}_{n_{p1},n_{pk_p}} \\ \boldsymbol{O}_{n_{p2},n_{p1}} & n_{p2}^{-1}\boldsymbol{J}_{n_{p2}} & \cdots & \boldsymbol{O}_{n_{p2},n_{pk_p}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{O}_{n_{pk_p},n_{p1}} & \boldsymbol{O}_{n_{pk_p},n_{p2}} & \cdots & n_{pk_p}^{-1}\boldsymbol{J}_{n_{pk_p}} \end{bmatrix}, \text{ for } p = 1, 2, \cdots, P, \tag{A.1}$$

where $\boldsymbol{J}_n$ is a $n \times n$ matrix with elements 1, and $\boldsymbol{O}_{m,n}$ is a $m \times n$ matrix with elements 0. It essentially means that $\boldsymbol{S}_p$ is constructed such a way that $\widehat{\boldsymbol{m}}_p(x_{pj})$ is the partial mean of $\boldsymbol{Y}_{res}$, where $X_p = x_{pj}$ for $p = 1, 2, \cdots, P$ and $j = 1, 2, \cdots, k_p$.

The covariate $\boldsymbol{Z}$ may contain any type of random variable – categorical, discrete or continuous. If some components of $\boldsymbol{Z}$ are categorical or discrete, then we use bin smoother again. Otherwise, for continuous valued covariates, one may choose a smoother that uses local polynomials. For the simplicity of notation, we assume that all covariates are continuous. In fact, the situation is even simpler for categorical or discrete covariates. Let d_q be the degree of the polynomial used for smoothing of Z_q for $q = 1, 2, \cdots, Q$. Note that Nadaraya-Watson estimate (Watson 1964) is a trivial case of the polynomial smoothing where the degree of the polynomial is zero. Suppose $K(\cdot)$ is the kernel function, and denote $K_{h_q}(z) = h_q^{-1} K(\frac{z}{h_q})$, where h_q is the bandwidth parameter. Then, the smoothing matrix

A. Mandal
Department of Mathematics, Wayne State University,
656 W. Kirby, Detroit, MI 48202, U.S.A.
E-mail: abhijit.mandal@wayne.edu

of Z_q is given by

$$\boldsymbol{S}_{P+q} = \left(\boldsymbol{Z}_{z_q}^T \boldsymbol{K}_{z_q} \boldsymbol{Z}_{z_q}\right)^{-1} \boldsymbol{Z}_{z_q}^T \boldsymbol{K}_{z_q}, \text{ for } q = 1, 2, \cdots, Q, \tag{A.2}$$

where $\boldsymbol{K}_{z_q} = \text{diag}\{K_{h_q}(Z_{q1} - z_q), \cdots, K_{h_q}(Z_{qn} - z_q)\}$ is a diagonal matrix containing the kernel weight, and

$$\boldsymbol{Z}_{z_q} = \begin{bmatrix} 1 & (Z_{q1} - z_q) & \dots & (Z_{q1} - z_q)^{d_q} \\ 1 & (Z_{q2} - z_q) & \dots & (Z_{q2} - z_q)^{d_q} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (Z_{qn} - z_q) & \dots & (Z_{qn} - z_q)^{d_q} \end{bmatrix}.$$

Then, the normal equations for the backfitting estimators (Buja et al. 1989; Opsomer and Ruppert 1998) are given by

$$\begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{S}_1^* & \dots & \boldsymbol{S}_1^* \\ \boldsymbol{S}_2^* & \boldsymbol{I}_n & \dots & \boldsymbol{S}_2^* \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{P+Q}^* & \boldsymbol{S}_{P+Q}^* & \dots & \boldsymbol{I}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{m}_1 \\ \boldsymbol{m}_2 \\ \vdots \\ \boldsymbol{m}_{P+Q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{S}_1^* \\ \boldsymbol{S}_2^* \\ \vdots \\ \boldsymbol{S}_{P+Q}^* \end{bmatrix} \boldsymbol{Y}^*, \tag{A.3}$$

where $\boldsymbol{S}_d^* = (\boldsymbol{I}_n - \boldsymbol{1}_n \boldsymbol{1}_n^T/n)\boldsymbol{S}_d$ is the centered smoothing matrix for $d = 1, 2, \cdots, P+Q$, $\boldsymbol{Y}^* = \boldsymbol{Y} - \bar{Y}\boldsymbol{1}_n$ and $\boldsymbol{1}_n$ is the n-dimensional vector of elements 1. The solution to the normal equation (A.3) has the form

$$\begin{bmatrix} \widehat{\boldsymbol{m}}_1 \\ \widehat{\boldsymbol{m}}_2 \\ \vdots \\ \widehat{\boldsymbol{m}}_{P+Q} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_n & \boldsymbol{S}_1^* & \dots & \boldsymbol{S}_1^* \\ \boldsymbol{S}_2^* & \boldsymbol{I}_n & \dots & \boldsymbol{S}_2^* \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{S}_{P+Q}^* & \boldsymbol{S}_{P+Q}^* & \dots & \boldsymbol{I}_n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{S}_1^* \\ \boldsymbol{S}_2^* \\ \vdots \\ \boldsymbol{S}_{P+Q}^* \end{bmatrix} \boldsymbol{Y}^* \equiv \boldsymbol{M}^{-1}\boldsymbol{C}\boldsymbol{Y}^*,$$

provided the inverse exists. Here, $\boldsymbol{M}$ and $\boldsymbol{C}$ are the associated matrices. So, the backfitting estimator of $\boldsymbol{m}_d$ is given by

$$\widehat{\boldsymbol{m}}_d = \boldsymbol{E}_d \boldsymbol{M}^{-1}\boldsymbol{C}\boldsymbol{Y}^* \equiv \boldsymbol{W}_d \boldsymbol{Y}^*, \; d = 1, 2, \cdots, P+Q, \tag{A.4}$$

where $\boldsymbol{W}_d = \boldsymbol{E}_d \boldsymbol{M}^{-1}\boldsymbol{C}$, and $\boldsymbol{E}_d$ is a block matrix of dimension $n \times n(P+Q)$ with $n \times n$ identity matrix in the d-th block and zero elsewhere.

Let us denote $\boldsymbol{W} = \sum_{d=1}^{P+Q} \boldsymbol{W}_d$. Suppose $\boldsymbol{W}^{[-d]}$ is the smoother matrix for the additive model after dropping out the term containing $\boldsymbol{m}_d$, $d = 1, 2, \cdots, P+Q$. Then, the following lemma from Opsomer (2000) ensures the existence and uniqueness of the backfitting estimators of the additive model.

Lemma 1 *If $||\boldsymbol{S}_d^* \boldsymbol{W}^{[-d]}|| < 1$ for some $d \in (1, 2, \cdots, P+Q)$, where $||\cdot||$ denotes any matrix norm, then the backfitting estimators uniquely exist and*

$$\boldsymbol{W}_d = \boldsymbol{I}_n - \left(\boldsymbol{I}_n - \boldsymbol{S}_d^* \boldsymbol{W}^{[-d]}\right)^{-1} (\boldsymbol{I}_n - \boldsymbol{S}_d^*) = \left(\boldsymbol{I}_n - \boldsymbol{S}_d^* \boldsymbol{W}^{[-d]}\right)^{-1} \boldsymbol{S}_d^* \left(\boldsymbol{I}_n - \boldsymbol{W}^{[-d]}\right). \tag{A.5}$$

Appendix C: Proofs

Lemma 2 *Let us assume that conditions (C2)–(C5) hold, then*

$$\boldsymbol{S}_d^* = \boldsymbol{S}_d - \frac{\boldsymbol{1}_n \boldsymbol{1}_n^T}{n} + o\left(\frac{\boldsymbol{1}_n \boldsymbol{1}_n^T}{n}\right) \; a.s.,$$

for all $d = 1, 2, \cdots, P+Q$. The term $o\left(\frac{\boldsymbol{1}_n \boldsymbol{1}_n^T}{n}\right)$ means that each element is of order $o\left(\frac{1}{n}\right)$.

Proof This property is proved in Opsomer and Ruppert (1997) using a local polynomial fitting where the smoothing matrix is as defined in (A.2). This is also true for the point mass bin smoother as

$$\boldsymbol{S}_d^* = \boldsymbol{S}_d - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\boldsymbol{S}_d = \boldsymbol{S}_d - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n} \text{ for } d = 1, 2, \cdots, P. \tag{A.6}$$

Note that for $d = 1, 2, \cdots, P$ the relationship is exact, and we do not need any assumption for this.

Lemma 3 *If the predictors and covariates are pairwise independent then, under conditions (C1)–(C6), we have*

$$\boldsymbol{S}_d^*\boldsymbol{S}_{d'}^* = o\left(\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right) \text{ a.s.},$$

for all $d \neq d' \in \{1, 2, \cdots, P + Q\}$.

Proof Using equation (A.6) for p and $p' \in \{1, 2, \cdots, P\}$, we get

$$\begin{aligned}
\boldsymbol{S}_p^*\boldsymbol{S}_{p'}^* &= \left(\boldsymbol{S}_p - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right)\left(\boldsymbol{S}_{p'} - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right) \\
&= \boldsymbol{S}_p\boldsymbol{S}_{p'} - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\boldsymbol{S}_p - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\boldsymbol{S}_{p'} + \frac{\mathbf{1}_n\mathbf{1}_n^T}{n} \\
&= \boldsymbol{S}_p\boldsymbol{S}_{p'} - \frac{\mathbf{1}_n\mathbf{1}_n^T}{n}.
\end{aligned} \tag{A.7}$$

Note that $\boldsymbol{S}_p\boldsymbol{S}_p = \boldsymbol{S}_p$ for $p = 1, 2, \cdots, P$. For $p \neq p' \in \{1, 2, \cdots, P\}$, we define $\boldsymbol{U} = \boldsymbol{S}_p\boldsymbol{S}_{p'}$. Here $\boldsymbol{U}$ is a block matrix containing each element in the rs-th block equal to

$$u_{rs} = \sum_{i: X_{pi} = x_{pr}, X_{p'i} = x_{p's}} \frac{1}{n_{pr}n_{p's}},$$

where $r = 1, 2, \cdots, k_p$ and $s = 1, 2, \cdots, k_{p'}$. Using strong law of large numbers (SLLN) and assumption (C1), we get

$$\begin{aligned}
nu_{rs} &\xrightarrow{\text{a.s.}} \frac{1}{c_{pr}c_{p's}} P(X_p = x_{pr}, X_{p'} = x_{p's}) \\
&= \frac{P(X_p = x_{pr}, X_{p'} = x_{p's})}{P(X_p = x_{pr})P(X_{p'} = x_{p's})}.
\end{aligned} \tag{A.8}$$

Combining equations (A.7) and (A.8) the ij-th element of $\boldsymbol{S}_p^*\boldsymbol{S}_{p'}^*$ becomes

$$(\boldsymbol{S}_p^*\boldsymbol{S}_{p'}^*)_{ij} = \frac{1}{n}\left(\frac{P(X_p = x_{pi}, X_{p'} = x_{p'j})}{P(X_p = x_{pi})P(X_{p'} = x_{p'j})} - 1\right) \text{ a.s.}$$

So, the lemma is proved for $p \neq p' \in \{1, 2, \cdots, P\}$. For $d \neq d' \in \{P + 1, P + 2, \cdots, P + Q\}$ Opsomer and Ruppert (1997) have shown that under conditions (C2)–(C6)

$$(\boldsymbol{S}_d^*\boldsymbol{S}_{d'}^*)_{ij} = \frac{1}{n}\left(\frac{f_{dd'}(z_{di}, z_{d'j})}{f_d(z_{di})f_{d'}(z_{d'j})} - 1\right) + o\left(\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right) \text{ a.s.},$$

where $f_{dd'}(\cdot)$ is the joint distribution of Z_d and $Z_{d'}$, whereas $f_d(\cdot)$ and $f_{d'}(\cdot)$ are their marginal distributions. So, the lemma is true for $d \neq d' \in \{P + 1, P + 2, \cdots, P + Q\}$. Now, using condition (C8) and applying the same technique, we can prove this result when $d = 1, 2, \cdots, P$, and $d' = P + 1, P + 2, \cdots, P + Q$, or vise versa.

Lemma 4 *Let us denote* $\boldsymbol{W} = \sum_{d=1}^{P+Q} \boldsymbol{W}_d$, *where* $\boldsymbol{W}_d$ *is given in equation (A.4). Then, under conditions (C1)–(C6)*

$$\boldsymbol{W} \approx \boldsymbol{S}^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \quad a.s.,$$

where $\boldsymbol{S}^* = \sum_{d=1}^{P+Q} \boldsymbol{S}_d^*$.

Proof For $P + Q = 2$, we get

$$\boldsymbol{W}^{[-1]} = \boldsymbol{S}_2^* \text{ and } \boldsymbol{W}^{[-2]} = \boldsymbol{S}_1^*.$$

From equation (A.5), we have

$$\boldsymbol{W}_1 = (\boldsymbol{I}_n - \boldsymbol{S}_1^* \boldsymbol{S}_2^*)^{-1} \boldsymbol{S}_1^* (\boldsymbol{I}_n - \boldsymbol{S}_2^*). \tag{A.9}$$

Using Lemma 3, we have the following approximation

$$(\boldsymbol{I}_n - \boldsymbol{S}_1^* \boldsymbol{S}_2^*)^{-1} \approx \boldsymbol{I}_n + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.} \tag{A.10}$$

This approximation is exact when the corresponding predictors or covariates are pairwise independent. Combining equations (A.9) and (A.10), we get

$$\boldsymbol{W}_1 \approx \boldsymbol{S}_1^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.}$$

Similarly $\boldsymbol{W}_2 \approx \boldsymbol{S}_2^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right)$ a.s. Now, for all values of P and Q, we prove by recursion that

$$\boldsymbol{W}_d \approx \boldsymbol{S}_d^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.}$$

Therefore, by taking summation over $d = 1, 2, \cdots, P + Q$ the lemma is proved.

Lemma 5 *Suppose conditions (C1)–(C4) and (C8) are satisfied. Then, under* H_0^{**},

$$\boldsymbol{m}^T \boldsymbol{S}_p^* \boldsymbol{m} = o_p(1),$$

for all $p = 1, 2, \cdots, P$, *where* $\boldsymbol{m} = \sum_{p=1}^{P+Q} \boldsymbol{m}_p$.

Proof Note that

$$\boldsymbol{S}_p \boldsymbol{m}_p = \boldsymbol{m}_p \text{ for all } p = 1, 2, \cdots, P. \tag{A.11}$$

For $p = 1, 2, \cdots, P$ and $q = 1, 2, \cdots, Q$ we define $\boldsymbol{u} = \boldsymbol{S}_p \boldsymbol{m}_{P+q} = (u_1 \mathbf{1}_{n_{p1}}^T, u_2 \mathbf{1}_{n_{p2}}^T, \cdots, u_{n_{pk_p}} \mathbf{1}_{n_{pk_p}}^T)^T$. Then

$$u_j = \sum_{i: X_{pi} = x_{pj}} \frac{m_{P+q}(Z_{qi})}{n_{pj}}, \text{ for } j = 1, 2, \cdots, k_p.$$

Using strong law of large numbers (SLLN) and assumption (C8) we get

$$u_j \xrightarrow{\text{a.s.}} E[m_{P+q}(Z_q) | X_p = x_{pj}] = 0.$$

So

$$\boldsymbol{S}_p \boldsymbol{m}_{P+q} = \boldsymbol{O}_{n,1} \text{ a.s. for all } p = 1, 2, \cdots, P \text{ and } q = 1, 2, \cdots, Q. \tag{A.12}$$

Hence, under H_0^{**}, for all $p = 1, 2, \cdots, P$

$$\boldsymbol{m}^T \boldsymbol{S}_p \boldsymbol{m} = \left(\sum_{d=1}^{P} \boldsymbol{m}_p^T\right) \boldsymbol{S}_p \left(\sum_{d=1}^{P} \boldsymbol{m}_p\right) + o_p(1) = o_p(1). \tag{A.13}$$

Again using SLLN we get

$$\frac{\mathbf{1}_n^T \boldsymbol{m}_{P+q}}{n} \xrightarrow{\text{a.s.}} E[m_{P+q}(Z_q)] = 0$$

for all $q = 1, 2, \cdots, Q$. Similarly, $\mathbf{1}_n^T \boldsymbol{m}_p / n = 0$ a.s. for all $p = 1, 2, \cdots, P$. Hence using Lemma 2 the lemma is proved from equation (A.13).

Lemma 6 *Denote* $\boldsymbol{A}_{2n} = (\boldsymbol{W} - \boldsymbol{I}_n)^T (\boldsymbol{W} - \boldsymbol{I}_n)$, *then under conditions (C1)–(C6)*

$$\boldsymbol{A}_{2n} \approx \boldsymbol{S}^{*T}\boldsymbol{S}^* - \boldsymbol{S}^* - \boldsymbol{S}^{*T} + \boldsymbol{I}_n + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.,}$$

where $\boldsymbol{S}^* = \sum_{d=1}^{P+Q} \boldsymbol{S}_d^*$.

Proof Using Lemma 4, we get

$$\begin{aligned}\boldsymbol{A}_{2n} &\approx (\boldsymbol{S}^* - \boldsymbol{I}_n)^T (\boldsymbol{S}^* - \boldsymbol{I}_n) + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.} \\ &= \boldsymbol{S}^{*T}\boldsymbol{S}^* - \boldsymbol{S}^* - \boldsymbol{S}^{*T} + \boldsymbol{I}_n + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.} \end{aligned} \tag{A.14}$$

Proof (Corollary 4) Denote $\boldsymbol{A}_{1n} = (\boldsymbol{W}_{[Z]} - \boldsymbol{I}_n)^T (\boldsymbol{W}_{[Z]} - \boldsymbol{I}_n)$, where $\boldsymbol{W}_{[Z]}$ is the smoother matrix for the additive model after dropping all P predictors. Using an argument similar to that in the proof of Lemma 6, we find

$$\boldsymbol{A}_{1n} \approx \boldsymbol{S}_{[Z]}^{*T}\boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^{*T} + \boldsymbol{I}_n + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.,}$$

where $\boldsymbol{S}_{[Z]}^* = \sum_{d=1}^{Q} \boldsymbol{S}_{P+d}^*$. So

$$\boldsymbol{A}_{1n} - \boldsymbol{A}_{2n} \approx \boldsymbol{S}_{[X]}^* + \boldsymbol{S}_{[X]}^{*T} - \boldsymbol{S}_{[X]}^{*T}\boldsymbol{S}_{[X]}^* - \boldsymbol{S}_{[X]}^{*T}\boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[X]}^*\boldsymbol{S}_{[Z]}^{*T} + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.,} \tag{A.15}$$

$\boldsymbol{S}_{[X]}^* = \sum_{d=1}^{P} \boldsymbol{S}_d^*$. Now $\boldsymbol{S}_d^*\boldsymbol{S}_d^* = \boldsymbol{S}_d^*$ and $\boldsymbol{S}_d^{*T} = \boldsymbol{S}_d^*$ for $d = 1, 2, \cdots, P$. From the technique used in Lemma 3, it can be shown that $\boldsymbol{S}_{[X]}^*\boldsymbol{S}_{[Z]}^{*T}$ is a symmetric matrix. So, using Lemma 3 equation (A.15) reduces to

$$\boldsymbol{A}_{1n} - \boldsymbol{A}_{2n} \approx \boldsymbol{S}_{[X]}^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) = \sum_{d=1}^{P} \boldsymbol{S}_d^* + o\left(\frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \text{ a.s.}$$

Now

$$\begin{aligned} RSS_0^{**} - RSS_1^* &= \boldsymbol{Y}^{*T}(\boldsymbol{A}_{1n} - \boldsymbol{A}_{2n})\boldsymbol{Y}^* \\ &\approx \sum_{d=1}^{P} \boldsymbol{Y}^{*T}\boldsymbol{S}_d^*\boldsymbol{Y}^* + o_p(1) \\ &= \sum_{d=1}^{P} \boldsymbol{Y}^T \left(\boldsymbol{I}_n - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right)^T \boldsymbol{S}_d^* \left(\boldsymbol{I}_n - \frac{\mathbf{1}_n \mathbf{1}_n^T}{n}\right) \boldsymbol{Y} + o_p(1) \\ &= \sum_{d=1}^{P} \boldsymbol{Y}^T \boldsymbol{S}_d^* \boldsymbol{Y} + o_p(1) \end{aligned} \tag{A.16}$$

Let us define $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)^T$. Then

$$RSS_0^{**} - RSS_1^* = \sum_{d=1}^{P} \boldsymbol{m}^T \boldsymbol{S}_d^* \boldsymbol{m} + 2\sum_{d=1}^{P} \boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{m} + \sum_{d=1}^{P} \boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{\epsilon} + o_p(1). \tag{A.17}$$

Using Lemma 5, under H_0^{**}, we have

$$\boldsymbol{m}^T \boldsymbol{S}_d^* \boldsymbol{m} = o_p(1) \text{ for all } d = 1, 2, \cdots, P.$$

Moreover, using (A.12) it is easy to show that, under H_0^{**}, $\boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{m} = o_p(1)$ for all $d = 1, 2, \cdots, P$. Hence, from (A.17), we get

$$RSS_0^{**} - RSS_1^* = \sum_{d=1}^{P} \boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{\epsilon} + o_p(1). \tag{A.18}$$

Now, for $d = 1, 2, \cdots, P$, using the definition in (A.1), we have

$$\boldsymbol{\epsilon}^T \boldsymbol{S}_d \boldsymbol{\epsilon} = \sum_{j=1}^{k_d} \frac{1}{n_{dj}} (\boldsymbol{e}_{dj}^T \boldsymbol{\epsilon})^T \boldsymbol{e}_{dj}^T \boldsymbol{\epsilon}, \tag{A.19}$$

where $\boldsymbol{e}_{dj}$ is a vector with n_{dj} elements one and rest are zero. If $X_{di} = x_{dj}$, then the i-th element of $\boldsymbol{e}_{dj}$ is one. Note that in this definition, we did not sort X_d according to their observed values. As $E[\epsilon_i^2] < \infty$ under condition (C7), using central limit theorem (CLT), we get

$$\frac{1}{\sigma\sqrt{n_{dj}}} \boldsymbol{e}_{dj}^T \boldsymbol{\epsilon} \overset{a}{\equiv} U_{dj} \sim N(0, 1).$$

As $\boldsymbol{e}_{dj}^T \boldsymbol{\epsilon}$ and $\boldsymbol{e}_{dj'}^T \boldsymbol{\epsilon}$ are independent for all $j \neq j' \in \{1, 2, \cdots, k_d\}$, the components of $\boldsymbol{U}_d = (U_{d1}, U_{d2}, \cdots, U_{dk_d})^T$ are i.i.d. standard normal variables. Therefore, from (A.19), we have $\frac{1}{\sigma^2} \boldsymbol{\epsilon}^T \boldsymbol{S}_d \boldsymbol{\epsilon} \overset{a}{\sim} \chi^2(k_d)$. Let us define

$$\bar{U} = \sum_{j=1}^{k_d} \sqrt{\frac{n_{dj}}{n}} U_{dj}.$$

Then $\bar{U} = c_d^T \boldsymbol{U}_d$ a.s., where $c_d = (\sqrt{c_{d1}}, \sqrt{c_{d2}}, \cdots, \sqrt{c_{dk_d}})^T$. Note that $\bar{U}^2 = \frac{1}{n} \boldsymbol{\epsilon}^T \boldsymbol{\epsilon}$. Hence

$$\frac{1}{\sigma^2} \boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{\epsilon} \overset{a}{\equiv} \boldsymbol{U}_d^T \left(\boldsymbol{I}_{k_d} - c_d c_d^T \right) \boldsymbol{U}_d \sim \chi^2(k_d - 1),$$

because $(\boldsymbol{I}_{k_d} - c_d c_d^T)$ is an idempotent matrix of rank $(k_d - 1)$. If all predictors are pairwise independent, then from Lemma 3, we get

$$\boldsymbol{S}_d^* \boldsymbol{S}_{d'}^* = o\left(\frac{\boldsymbol{1}_n \boldsymbol{1}_n^T}{n} \right) \text{ a.s.}$$

for $d \neq d' \in \{1, 2, \cdots, P\}$. So, $\boldsymbol{\epsilon}^T \boldsymbol{S}_d^* \boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}^T \boldsymbol{S}_{d'}^* \boldsymbol{\epsilon}$ are asymptotically independent for all $d \neq d'$ (see p. 84 of Bapat 2012). Therefore, if the predictors are pairwise independent, then under H_0^{**}, equation (A.18) gives

$$\frac{1}{\sigma^2} (RSS_0^{**} - RSS_1^*) \overset{a}{\sim} \chi^2 \left(\sum_{d=1}^{P} (k_d - 1) \right). \tag{A.20}$$

However, in general, the above distribution comes out to be a sum of P dependent chi-square variables as

$$\frac{1}{\sigma^2} (RSS_0^{**} - RSS_1^*) \overset{a}{\equiv} \sum_{d=1}^{P} \boldsymbol{U}_d^T \left(\boldsymbol{I}_{k_d} - c_d c_d^T \right) \boldsymbol{U}_d. \tag{A.21}$$

Suppose $d \neq d' \in \{1, 2, \cdots, P\}$, $j = 1, 2, \cdots, k_d$ and $j' = 1, 2, \cdots, k_{d'}$, then the correlation between U_{dj} and $U_{d'j'}$ is given by

$$\begin{aligned}
\text{Corr}(U_{dj}, U_{d'j'}) &= \lim_{n\to\infty} \text{Corr}\left(\frac{1}{\sigma\sqrt{n_{dj}}}\boldsymbol{e}_{dj}^T\boldsymbol{\epsilon}, \frac{1}{\sigma\sqrt{n_{d'j'}}}\boldsymbol{e}_{d'j'}^T\boldsymbol{\epsilon}\right) \\
&= \lim_{n\to\infty} \sum_{i:X_{di}=x_{dj}, X_{d'i}=x_{d'j'}} \frac{1}{\sqrt{n_{dj}n_{d'j'}}} \\
&= \frac{1}{\sqrt{c_{dj}c_{d'j'}}} P(X_d = x_{dj}, X_{d'} = x_{d'j'}). \qquad \text{(A.22)}
\end{aligned}$$

The last expression is derived using the same techniques as used in equation (A.8). Note that combining equations (A.21) and (A.22), we obtain result (A.20) if the predictor variables are independent.
Now, it is easy to show that

$$\begin{aligned}
\frac{1}{n}RSS_1^* &= \frac{1}{n}\boldsymbol{Y}^{*T}\boldsymbol{A}_{2n}\boldsymbol{Y}^* \\
&\approx \frac{1}{n}\boldsymbol{Y}^{*T}\left(\boldsymbol{S}^{*T}\boldsymbol{S}^* - \boldsymbol{S}^* - \boldsymbol{S}^{*T} + \boldsymbol{I}_n + o\left(\frac{\boldsymbol{1}_n\boldsymbol{1}_n^T}{n}\right)\right)\boldsymbol{Y}^* \\
&= \sigma^2 + o_p(1). \qquad \text{(A.23)}
\end{aligned}$$

Therefore, using Slutsky's theorem, σ^2 in (A.21) may be replaced by $\frac{1}{n}RSS_1^*$, and therefore

$$\lambda_n(H_0^{**}) \overset{a}{\equiv} \sum_{p=1}^{P} \boldsymbol{U}_p^T\left(\boldsymbol{I}_{k_p} - c_p c_p^T\right)\boldsymbol{U}_p. \qquad \text{(A.24)}$$

Define $\boldsymbol{U} = (\boldsymbol{U}_1^T, \boldsymbol{U}_2^T, \cdots, \boldsymbol{U}_P^T)^T$, and $\boldsymbol{U}^* = \boldsymbol{\Sigma}_1\boldsymbol{U}$, where $\boldsymbol{\Sigma}_1$ is defined in Section 4.4. As $\boldsymbol{\Sigma}_1$ is an idempotent matrix, $\lambda_n(H_0^{**})$ in equation (A.24) is written as

$$\lambda_n(H_0^{**}) \overset{a}{\equiv} \boldsymbol{U}^T\boldsymbol{\Sigma}_1\boldsymbol{U} = \boldsymbol{U}^{*T}\boldsymbol{U}^*, \qquad \text{(A.25)}$$

Now, the covariance matrix of $\boldsymbol{U}$ is $\boldsymbol{\Sigma}_2$ (defined in Section 4.4), which is a block matrix with p-th diagonal block is an identity matrix of order k_p, and the ij-th element of the pp'-th off-diagonal block is given in (A.22). So, the covariance matrix of $\boldsymbol{U}^*$ is $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1$. Let $\lambda_1, \lambda_2, \cdots, \lambda_s$ are non-zero eigenvalues of $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1$, where s is the rank of $\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1$. Suppose $V = (V_1, V_2, \cdots, V_s)^T$ is a vector of i.i.d. standard normal variables, and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \cdots, \lambda_s)$. Then, from (A.25), the theorem is proved as

$$\lambda_n(H_0^{**}) \overset{a}{\equiv} \boldsymbol{V}^T\boldsymbol{\Lambda}\boldsymbol{V} = \sum_{i=1}^{s}\lambda_i V_i^2.$$

Theorem 3 *Let us consider the notations and assumptions of Corollary 4. Then, under* H_1*, the asymptotic distribution of the GLR test statistic coincides with* $\delta^2 + \sum_{i=1}^{s}\lambda_i V_i^2$*, where* $\delta^2 = \sum_{r,s=1}^{P} E(m_r^* m_s^*)$.

Proof If H_0^{**} is not true, then from (A.17), we get

$$RSS_0^{**} - RSS_1^* = \sum_{d=1}^{P}\boldsymbol{m}^T\boldsymbol{S}_d^*\boldsymbol{m} + \sum_{d=1}^{P}\boldsymbol{\epsilon}^T\boldsymbol{S}_d^*\boldsymbol{\epsilon} + o_p(1). \qquad \text{(A.26)}$$

Now, the result in equation (A.21) reduces to

$$\frac{1}{\sigma^2}\sum_{d=1}^{P}\boldsymbol{\epsilon}^T\boldsymbol{S}_d^*\boldsymbol{\epsilon} = \sum_{d=1}^{P}\boldsymbol{U}_d^T\left(\boldsymbol{I}_{k_d} - c_d c_d^T\right)\boldsymbol{U}_d. \qquad \text{(A.27)}$$

From the proof of Lemma 5, we have

$$\boldsymbol{m}^T \boldsymbol{S}_p \boldsymbol{m} = \left(\sum_{d=1}^{P} \boldsymbol{m}_p^T \right) \boldsymbol{S}_p \left(\sum_{d=1}^{P} \boldsymbol{m}_p \right) + o_p(1). \tag{A.28}$$

Using $\boldsymbol{S}_p \boldsymbol{m}_p = \boldsymbol{m}_p$, we get $\boldsymbol{m}_r^T \boldsymbol{S}_p \boldsymbol{m}_p = \boldsymbol{m}_r^T \boldsymbol{m}_p$ for all $p, r = 1, 2, \cdots, P$. Hence

$$\frac{1}{n} \boldsymbol{m}_r^T \boldsymbol{S}_p \boldsymbol{m}_p \xrightarrow{\text{a.s.}} E(\boldsymbol{m}_r^T \boldsymbol{m}_p).$$

Suppose u_{ij} is the (i, j)-th element of $\boldsymbol{m}_r^T \boldsymbol{S}_p \boldsymbol{m}_s$ for some $p, r, s = 1, 2, \cdots, P$. Then

$$u_{ij} = \sum_{l=1}^{k_p} \sum_{(i,j): X_{pi} = X_{pj} = x_{pl}} \frac{m_r(X_{ri}) m_s(X_{sj})}{n_{pl}}. \tag{A.29}$$

Note that

$$\sum_{(i,j): X_{pi} = X_{pj} = x_{pl}} \frac{m_r(X_{ri}) m_s(X_{sj})}{n_{pl}} \xrightarrow{\text{a.s.}} E(m_r m_s | X_p = x_{pl}).$$

So, using condition (C1), we get from equation (A.29)

$$u_{ij} \xrightarrow{\text{a.s.}} E(m_r m_s) \text{ and } \frac{1}{n} u_{ij} \xrightarrow{\text{a.s.}} 0.$$

Hence, from equation (A.28), we get

$$\frac{1}{n} \boldsymbol{m}^T \boldsymbol{S}_p \boldsymbol{m} = \sum_{r}^{P} E(m_r m_p) + o_p(1).$$

Therefore

$$\frac{1}{n} \sum_{p=1}^{P} \boldsymbol{m}^T \boldsymbol{S}_p \boldsymbol{m} = \sum_{r,s=1}^{P} E(m_r m_s) + o_p(1). \tag{A.30}$$

As $E(m_p) = 0$ for all $p = 1, 2, \cdots, P$, we get

$$\frac{1}{n} \sum_{p=1}^{P} \boldsymbol{m}^T \boldsymbol{S}_p^* \boldsymbol{m} = \sum_{r,s=1}^{P} E(m_r m_s) + o_p(1). \tag{A.31}$$

Combining (A.23), (A.26), (A.27) and (A.31) the theorem is proved.

Proof (Corollary 3) The residual sum of squares under H_0^* can be written as

$$\begin{aligned}
RSS_0^* &= \left(\boldsymbol{Y}^* - \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} - \boldsymbol{W}_{[Z]} \left(\boldsymbol{Y}^* - \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} \right) \right)^T \left(\boldsymbol{Y}^* - \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} - \boldsymbol{W}_{[Z]} \left(\boldsymbol{Y}^* - \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} \right) \right) \\
&= \left(\boldsymbol{Y}^* - \left(\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \right) \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} - \boldsymbol{W}_{[Z]} \boldsymbol{Y}^* \right)^T \left(\boldsymbol{Y}^* - \left(\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \right) \boldsymbol{X}^* \widetilde{\boldsymbol{\theta}} - \boldsymbol{W}_{[Z]} \boldsymbol{Y}^* \right) \\
&= \boldsymbol{Y}^{*T} \left(\boldsymbol{I}_n - \boldsymbol{A}_n - \boldsymbol{W}_{[Z]} \right)^T \left(\boldsymbol{I}_n - \boldsymbol{A}_n - \boldsymbol{W}_{[Z]} \right) \boldsymbol{Y}^*,
\end{aligned} \tag{A.32}$$

where

$$\boldsymbol{A}_n = \left(\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \right) \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \left(\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \right) \boldsymbol{X}^* \right)^{-1} \boldsymbol{X}^{*T} \left(\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \right). \tag{A.33}$$

Using Lemma 4 it can be shown that

$$\boldsymbol{W}_{[Z]} = \sum_{q=1}^{Q} \boldsymbol{W}_{P+q} \approx \sum_{q=1}^{Q} \boldsymbol{S}_{P+q}^* + o\left(\frac{\boldsymbol{1}_n \boldsymbol{1}_n^T}{n} \right) \text{ a.s.} \tag{A.34}$$

Hence

$$\boldsymbol{I}_n - \boldsymbol{W}_{[Z]} \approx \boldsymbol{I}_n + o\left(\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right) \text{ a.s..}$$

Therefore, equation (A.33) reduces to

$$\boldsymbol{A}_n = \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T} + o\left(\frac{\mathbf{1}_n\mathbf{1}_n^T}{n}\right) \text{ a.s.} \tag{A.35}$$

As $\boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}$ is an idempotent matrix, using (A.34) and (A.35) we get from equation (A.32)

$$\begin{aligned} RSS_0^* = \boldsymbol{Y}^{*T}\Big(&\boldsymbol{I}_n + \boldsymbol{S}_{[Z]}^{*T}\boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^{*T} - \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T} \\ &+ \boldsymbol{S}_{[Z]}^{*T}\boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T} + \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{S}_{[Z]}^*\Big)\boldsymbol{Y}^* + o_p(1). \end{aligned} \tag{A.36}$$

Suppose $\boldsymbol{\theta}_0$ is the true value of $\boldsymbol{\theta}$ under H_0^*. So, under H_0^*, the model can be written as

$$\boldsymbol{Y} = \alpha\mathbf{1}_n + \boldsymbol{X}^*\boldsymbol{\theta}_0 + \boldsymbol{m}_{[Z]} + \boldsymbol{\epsilon},$$

where $\boldsymbol{m}_{[Z]} = \sum_{q=1}^{Q}\boldsymbol{m}_{P+q}(\cdot)$. Opsomer and Ruppert (1999) have shown that $\widetilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$. Hence, under H_0^*, from equation (A.34) we get

$$\boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^* = \boldsymbol{X}^*\widetilde{\boldsymbol{\theta}} \overset{\text{a.s.}}{=} \boldsymbol{X}^*\boldsymbol{\theta}_0 = \sum_{p=1}^{P}\boldsymbol{m}_p(\cdot) = \boldsymbol{m}_{[X]}. \tag{A.37}$$

Using a similar technique of equation (A.12) it can be shown that

$$\boldsymbol{S}_q^*\boldsymbol{m}_p = \boldsymbol{O}_{n,1} \text{ a.s. for all } p = 1, 2, \cdots, P \text{ and } q = 1, 2, \cdots, Q. \tag{A.38}$$

So, combining (A.37) and (A.38) we get

$$\begin{aligned} \boldsymbol{Y}^{*T}\boldsymbol{S}_{[Z]}^{*T}\boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\boldsymbol{Y}^* &= \boldsymbol{Y}^{*T}\boldsymbol{S}_{[Z]}^*\boldsymbol{m}_{[X]} + o_p(1) \\ &= \boldsymbol{m}_{[Z]}^T\boldsymbol{m}_{[X]} + o_p(1) \\ &= o_p(1). \end{aligned} \tag{A.39}$$

Hence, equation (A.36) simplifies to

$$RSS_0^* = \boldsymbol{Y}^{*T}\left(\boldsymbol{I}_n + \boldsymbol{S}_{[Z]}^{*T}\boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^* - \boldsymbol{S}_{[Z]}^{*T} - \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\right)\boldsymbol{Y}^* + o_p(1).$$

Now, proceeding the same way as the proof of Corollary 4, we get

$$\begin{aligned} RSS_0^* - RSS_1^* = \boldsymbol{Y}^{*T}\Big(&2\boldsymbol{S}_{[X]}^* - \boldsymbol{S}_{[X]}^{*T}\boldsymbol{S}_{[X]}^* - \boldsymbol{S}_{[Z]}^{*T}\boldsymbol{S}_{[X]}^* \\ &- \boldsymbol{S}_{[X]}^{*T}\boldsymbol{S}_{[Z]}^* - \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\Big)\boldsymbol{Y}^* + o_p(1). \end{aligned}$$

Using equations (A.11) and (A.39) we get

$$RSS_0^* - RSS_1^* = \boldsymbol{Y}^{*T}\left(\boldsymbol{S}_{[X]}^* - \boldsymbol{X}^*\left(\boldsymbol{X}^{*T}\boldsymbol{X}^*\right)^{-1}\boldsymbol{X}^{*T}\right)\boldsymbol{Y}^* + o_p(1). \tag{A.40}$$

Combining equations (A.16) and (A.17) we get

$$\boldsymbol{Y}^{*T}\boldsymbol{S}_{[X]}^*\boldsymbol{Y}^* = \boldsymbol{m}_{[X]}^T\boldsymbol{S}_{[X]}^*\boldsymbol{m}_{[X]} + 2\boldsymbol{m}_{[X]}^T\boldsymbol{S}_{[X]}^*\boldsymbol{\epsilon} + \boldsymbol{\epsilon}^T\boldsymbol{S}_{[X]}^*\boldsymbol{\epsilon} + o_p(1). \tag{A.41}$$

Using CLT it is easy to establish that $\boldsymbol{m}_{[X]}^T \boldsymbol{S}_{[X]}^* \boldsymbol{\epsilon} = o_p(1)$. From equation (A.30), we have $\boldsymbol{m}_{[X]}^T \boldsymbol{S}_{[X]}^* \boldsymbol{m}_{[X]} = \boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[X]} + o_p(1)$. Then, equation (A.41) turns out to be

$$\boldsymbol{Y}^{*T} \boldsymbol{S}_{[X]}^* \boldsymbol{Y}^* = \boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[X]} + \boldsymbol{\epsilon}^T \boldsymbol{S}_{[X]}^* \boldsymbol{\epsilon} + o_p(1). \tag{A.42}$$

Note that

$$\begin{aligned} \boldsymbol{Y}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y} &= \boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[X]} + 2\boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[Z]} + 2\boldsymbol{m}_{[X]}^T \boldsymbol{\epsilon} \\ &+ \boldsymbol{m}_{[Z]}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{m}_{[Z]} + 2\boldsymbol{m}_{[Z]}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{\epsilon} \\ &+ \boldsymbol{\epsilon}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{\epsilon}. \end{aligned} \tag{A.43}$$

Using condition (C8) and equation (A.39) it can be shown that the second and the forth terms in equation (A.43) tend to zero in probability; and by CLT the third and the fifth terms is asymptotically zero. Therefore

$$\boldsymbol{Y}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y} = \boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[X]} + \boldsymbol{\epsilon}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{\epsilon} + o_p(1).$$

As $\frac{1}{n} \boldsymbol{m}_{[X]}^T \boldsymbol{1}_n \overset{\text{a.s.}}{=} E(\boldsymbol{m}_{[X]}) = 0$, we get from the above equation

$$\begin{aligned} \boldsymbol{Y}^{*T} \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y}^* &= \boldsymbol{Y}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y} \\ &\quad - \frac{\boldsymbol{1}_n^T}{n} \boldsymbol{Y}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y} \frac{\boldsymbol{1}_n}{n} \\ &= \boldsymbol{Y}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{Y} + o_p(1) \\ &= \boldsymbol{m}_{[X]}^T \boldsymbol{m}_{[X]} + \boldsymbol{\epsilon}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{\epsilon} + o_p(1) \end{aligned} \tag{A.44}$$

Combining (A.42) and (A.44), we get from (A.40)

$$\begin{aligned} RSS_0^* - RSS_1^* &= \boldsymbol{\epsilon}^T \boldsymbol{S}_{[X]}^* \boldsymbol{\epsilon} - \boldsymbol{\epsilon}^T \boldsymbol{X}^* \left(\boldsymbol{X}^{*T} \boldsymbol{X}^*\right)^{-1} \boldsymbol{X}^{*T} \boldsymbol{\epsilon} + o_p(1) \\ &\approx \sum_{p=1}^{P} \boldsymbol{\epsilon}^T \left(\boldsymbol{S}_p - \boldsymbol{R}_{p,1} \left(\boldsymbol{R}_{p,1}^T \boldsymbol{R}_{p,1}\right)^{-1} \boldsymbol{R}_{p,1}^T\right) \boldsymbol{\epsilon} + o_p(1), \end{aligned} \tag{A.45}$$

where $\boldsymbol{R}_{p,1} = {}_0^{r_p}\boldsymbol{X}_{(p)}$ and ${}_a^b\boldsymbol{X}_{(p)}$ is defined in Equation (3). It can be shown that

$$\boldsymbol{S}_p = \boldsymbol{R}_{p,2} \left(\boldsymbol{R}_{p,2}^T \boldsymbol{R}_{p,2}\right)^{-1} \boldsymbol{R}_{p,2}^T, \tag{A.46}$$

where $\boldsymbol{R}_{p,2} = {}_0^{k_p-1}\boldsymbol{X}_{(p)}$. So $\boldsymbol{S}_p$ may be regarded as the hat matrix in context of the classical regression in fitting of a k_p degree polynomial. Equation (A.46) shows that columns of the matrix $\boldsymbol{S}_p$ form an orthogonal basis for the column space of $\boldsymbol{R}_{p,2}$. Similarly, columns of $\boldsymbol{R}_{p,1} \left(\boldsymbol{R}_{p,1}^T \boldsymbol{R}_{p,1}\right)^{-1} \boldsymbol{R}_{p,1}^T$ form an orthogonal basis for the column space of $\boldsymbol{R}_{p,1}$. Using some matrix calculations it can be shown that

$$\boldsymbol{S}_p - \boldsymbol{R}_{p,1} \left(\boldsymbol{R}_{p,1}^T \boldsymbol{R}_{p,1}\right)^{-1} \boldsymbol{R}_{p,1}^T = \boldsymbol{R}_p \left(\boldsymbol{R}_p^T \boldsymbol{R}_p\right)^{-1} \boldsymbol{R}_p^T, \text{ a.s.},$$

where $\boldsymbol{R}_p = {}_{r_p+1}^{k_p-1}\boldsymbol{X}_{(p)}$. Now $\boldsymbol{R}_p \left(\boldsymbol{R}_p^T \boldsymbol{R}_p\right)^{-1} \boldsymbol{R}_p^T$ is an idempotent matrix with rank $(k_p - r_p - 1)$. Hence

$$\frac{1}{\sigma^2} \boldsymbol{\epsilon}^T \boldsymbol{R}_p \left(\boldsymbol{R}_p^T \boldsymbol{R}_p\right)^{-1} \boldsymbol{R}_p^T \boldsymbol{\epsilon} \overset{a}{\equiv} \boldsymbol{U}_p^T \boldsymbol{U}_p \sim \chi^2(k_d - r_p - 1),$$

where

$$\boldsymbol{U}_p \stackrel{a}{\equiv} \frac{1}{\sigma}\left(\boldsymbol{R}_p^T\boldsymbol{R}_p\right)^{-1/2}\boldsymbol{R}_p^T\boldsymbol{\epsilon}.$$

So $(k_d - r_p - 1)$ components of $\boldsymbol{U}_p$ are i.i.d. standard normal variables. From equation (A.45), we get

$$\frac{1}{\sigma^2}(RSS_0^* - RSS_1^*) \stackrel{a}{\equiv} \sum_{p=1}^{P}\boldsymbol{U}_p^T\boldsymbol{U}_p,$$

where

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{U}_p,\boldsymbol{U}_{p'}) &= \lim_{n\to\infty}\frac{1}{\sigma^2}\mathrm{Cov}\left(\left(\boldsymbol{R}_p^T\boldsymbol{R}_p\right)^{-1/2}\boldsymbol{R}_p^T\boldsymbol{\epsilon}, \left(\boldsymbol{R}_{p'}^T\boldsymbol{R}_{p'}\right)^{-1/2}\boldsymbol{R}_{p'}^T\boldsymbol{\epsilon}\right)\\ &= \lim_{n\to\infty}\left(\boldsymbol{R}_p^T\boldsymbol{R}_p\right)^{-1/2}\boldsymbol{R}_p^T\boldsymbol{R}_{p'}\left(\boldsymbol{R}_{p'}^T\boldsymbol{R}_{p'}\right)^{-1/2}. \end{aligned} \qquad (A.47)$$

Rest of the proof is done using the same technique as the proof of Corollary 4.

Proof (Theorem 1) In this case, we can show that

$$RSS_0 - RSS_1 \approx \sum_{p=1}^{P_1}\boldsymbol{\epsilon}^T\left(\boldsymbol{S}_p - \boldsymbol{R}_{p,1}\left(\boldsymbol{R}_{p,1}^T\boldsymbol{R}_{p,1}\right)^{-1}\boldsymbol{R}_{p,1}^T\right)\boldsymbol{\epsilon} + \sum_{p=P_1+1}^{P}\boldsymbol{\epsilon}^T\boldsymbol{S}_p^*\boldsymbol{\epsilon} + o_P(1),$$

where $\boldsymbol{R}_{p,1} = {}_{0}^{r_p}\boldsymbol{X}_{(p)}$. Hence, the proof of the theorem follows from Corollaries 3 and 4.

Proof (Theorem 2) Combining steps of Theorems 1 and 3, we get the proof of the current theorem.

References

1. Bapat, R. B. (2012). *Linear algebra and linear models.* Springer Science & Business Media.
2. Buja, A., Hastie, T., Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, 17(2), 453–555.
3. Hastie, T. J., Tibshirani, R. J. (1990). *Generalized additive models*, volume 43 of *Monographs on Statistics and Applied Probability.* Chapman and Hall, Ltd., London.
4. Opsomer, J. D. (2000). Asymptotic properties of backfitting estimators. *Journal of Multivariate Analysis*, 73(2), 166–179.
5. Opsomer, J. D., Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics*, 25(1), 186–211.
6. Opsomer, J. D., Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, 93(442), 605–619.
7. Opsomer, J. D., Ruppert, D. (1999). A root-n consistent backfitting estimator for semiparametric additive modeling. *Journal of Computational and Graphical Statistics*, 8(4), 715–732.
8. Watson, G. S. (1964). Smooth regression analysis. *Sankhyā (Statistics). The Indian Journal of Statistics. Series A*, 26, 359–372.