Singular Additive Models for Function to Function Regression

Byeong U. Park, Chun-Jui Chen, Wenwen Tao, and Hans-Georg Müller

Seoul National University and University of California at Davis

Supplementary Material

A1 Dependency of Functional Singular Components

Under mild conditions the eigenfunctions φ_{Xk}, $k \geq 1$, which appear in the Karhunen-Loève representation for predictor processes X as in (1.4), $X^c(s) = \sum_{k=1}^{\infty} \xi_{Xk}\varphi_{Xk}(s)$, form an orthonormal basis of L^2 and can be used to represent the singular functions of X, i.e., $\phi_m(t) = \sum_{k\geq 1} \alpha_{mk}\varphi_{Xk}(t)$, with $\alpha_{mk} = \int \phi_m(s)\varphi_{Xk}(s)ds$. We use properties (3.6) and the representation of the covariance function $\text{cov}(X(s), X(t)) = \sum_{k\geq 1} \lambda_k\varphi_{Xk}(s)\varphi_{Xk}(t)$ with eigenvalues $\lambda_k \geq 0$, which is associated with the Karhunen-Loève representation of X, to compute the covariance of the functional singular components of X, obtaining

$$E\zeta_m\zeta_p = \sum_{k=1}^{\infty} \lambda_k\alpha_{mk}\alpha_{pk}.$$

Since the α_{mk} depend on both the cross-variance between X and response processes Y through the ϕ_m, as well as on the auto-covariance of X through the φ_{Xk}, it is clear that only under very special circumstances one can have $E\zeta_m\zeta_p \neq 0$ when $m \neq p$. One can bear this out further in special cases, for example when the processes X are random straight lines on a finite domain, with random intercepts and slopes. One then finds that the singular functions of X have to satisfy a specific algebraic constraint if they do not coincide with the eigenfunctions φ_{Xk} (possibly in a permuted order) in order to produce $E\zeta_m\zeta_p = 0$. Only very specific models would satisfy such constraints. Functional singular components will therefore in general be correlated unless very restrictive additional assumptions are made.

A2 Proof of Equation 5.2

Note that

$$\begin{aligned}\|\hat{\mathcal{A}}_{XYX} - \mathcal{A}_{XYX}\|_{\text{op}} \;\leq\; & \|\hat{\mathcal{C}}_{XY} - \mathcal{C}_{XY}\|_{\text{op}}\|\hat{\mathcal{C}}_{YX} - \mathcal{C}_{YX}\|_{\text{op}} + \|\mathcal{C}_{XY}\|_{\text{op}}\|\hat{\mathcal{C}}_{YX} - \mathcal{C}_{YX}\|_{\text{op}} \\ & + \|\mathcal{C}_{YX}\|_{\text{op}}\|\hat{\mathcal{C}}_{XY} - \mathcal{C}_{XY}\|_{\text{op}}.\end{aligned}$$

The facts that $E\|\hat{\mathcal{C}}_{XY} - \mathcal{C}_{XY}\|_{\text{op}}^2$ and $E\|\hat{\mathcal{C}}_{YX} - \mathcal{C}_{YX}\|_{\text{op}}^2$ are bounded by $n^{-1}E\|X\|^2\|Y\|^2$, and that $\|\mathcal{C}_{XY}\|_{\text{op}}^2$ and $\|\mathcal{C}_{YX}\|_{\text{op}}^2$ are bounded by $E\|X\|^2\|Y\|^2$, imply

$$E\|\hat{\mathcal{A}}_{XYX} - \mathcal{A}_{XYX}\|_{\mathrm{op}} \le (n^{-1} + 2\,n^{-1/2})E\|X\|^2\|Y\|^2.$$

Due to Lemma 4.3 in Bosq (2000), for fixed $j > 1$,

$$\|\hat{\phi}_j - \phi_j\| \le \frac{2\sqrt{2}}{\delta_j}\|\hat{\mathcal{A}}_{XYX} - \mathcal{A}_{XYX}\|_{\mathrm{op}},$$

$$\|\hat{\psi}_j - \psi_j\| \le \frac{2\sqrt{2}}{\delta_j}\|\hat{\mathcal{A}}_{YXY} - \mathcal{A}_{YXY}\|_{\mathrm{op}},$$

where $\delta_j = \min_{1\le k\le j}(\sigma_k^2 - \sigma_{k+1}^2)$, implying

$$\|\hat{\phi}_j - \phi_j\| = O_p(n^{-1/2}), \quad \|\hat{\psi}_j - \psi_j\| = O_p(n^{-1/2}), \qquad \text{(A2.1)}$$

provided that $E\|X\|^2\|Y\|^2 < \infty$ and the eigenvalues up to σ_{j+1}^2 are separated.

The approximation errors of the estimated singular components $\hat{\zeta}_{ij}$ and $\hat{\xi}_{ij}$ may be obtained from (A2.1). Suppose that $E\|X\|^{2\alpha} < \infty$ and $E\|Y\|^{2\alpha} < \infty$ for some $\alpha \ge 2$. Then it follows from a simple application of Markov inequality that

$$\max_{1\le i\le n}\|X_i - \mu_X\|^2 = O_p(n^{1/\alpha}), \quad \max_{1\le i\le n}\|Y_i - \mu_Y\|^2 = O_p(n^{1/\alpha}). \qquad \text{(A2.2)}$$

The fact that $\|\hat{\mu}_X - \mu_X\| = O_p(n^{-1/2})$ and $\|\hat{\mu}_Y - \mu_Y\| = O_p(n^{-1/2})$, together with (A2.2), leads to the results (5.2).

A3 Proof of Theorem 1

Before providing details of the proof, some perspective is in order. Let $\hat{p}_j^{*I}$ and $\hat{p}_{jl}^{*I}$ denote the versions of $\hat{p}_j^I$ and $\hat{p}_{jl}^I$, respectively, defined at (4.4) with $\hat{\boldsymbol{\zeta}}_i$ being replaced by the true $\boldsymbol{\zeta}_i$. Likewise, let $\tilde{f}_{kj}^*$ be the corresponding versions of $\tilde{f}_{kj}$ defined at (4.4). Define $\hat{f}_{kj}^*$ to be the solution of the version of the equation (4.5) subject to the versions of the constraints (4.6) where $\tilde{f}_{kj}$, $\hat{p}_j^I$ and $\hat{p}_{jl}^I$ in (4.5) and (4.6) are replaced by $\tilde{f}_{kj}^*$, $\hat{p}_j^{*I}$ and $\hat{p}_{jl}^{*I}$, respectively. Let $\hat{f}_{kj}^{*[r]}$ be the updates in the corresponding backfitting iteration. Then, an analogue of Theorem 1 for the versions $\hat{f}_{kj}^*$ and $\hat{f}_{kj}^{*[r]}$ where the arguments are fully available may be proved along the lines of the proofs in Mammen et al. (1999). Thus, in the proof we provide here we focus on the extra complications that arise due to the fact that the true singular scores are not available but must be estimated from the data, i.e., from having to use the estimated singular components $\hat{\zeta}_{ij}$ and $\hat{\xi}_{ik}$ in the estimation of the f_{kj}.

For simplicity of notation, let $\beta = 1/\alpha < 1/5$, where α is the number in the condition (A6). In terms of β, the results (5.2) can be rewritten as

$$\max_{1\le i\le n} |\hat{\zeta}_{ij} - \zeta_{ij}| = O_p(n^{-(1-\beta)/2}), \quad \max_{1\le i\le n} |\hat{\xi}_{ik} - \xi_{ik}| = O_p(n^{-(1-\beta)/2}), \quad 1 \le j \le M. \tag{A3.1}$$

We assume $\mu_X \equiv 0 \equiv \mu_Y$ and take $\hat{\mu}_X \equiv 0 \equiv \hat{\mu}_Y$, without loss of generality, since the estimation errors of $\hat{\mu}_X$ and $\hat{\mu}_Y$ have the parametric rate $n^{-1/2}$ and thus do not affect the first-order accuracy of the estimation of various nonparametric functions in the estimating equations (4.5).

Below we give two lemmas based on (A3.1) for the approximations of some relevant terms in the analysis of the backfitting equations. We write $K_{ij}(u) = K_{h_j}(u, \hat{\zeta}_{ij})$ and $K^*_{ij}(u) = K_{h_j}(u, \zeta_{ij})$ for brevity, and $\mathbb{I}_i = \mathbb{I}(\hat{\boldsymbol{\zeta}}_i \in I)$, $\mathbb{I}^*_i = \mathbb{I}(\boldsymbol{\zeta}_i \in I)$. Let $\hat{\vartheta}_{ik} = \hat{\xi}_{ik} - \sum_{j=1}^M f_{kj}(\zeta_{ij})$ and $\vartheta_{ik} = \xi_{ik} - \sum_{j=1}^M f_{kj}(\zeta_{ij})$. Define

$$\tilde{f}^A_{kj}(u) = \frac{1}{\hat{p}^I_0 \hat{p}^I_j(u)} n^{-1} \sum_{i=1}^n K_{ij}(u) \mathbb{I}_i \hat{\vartheta}_{ik},$$

$$\tilde{f}^B_{kj}(u) = \frac{1}{\hat{p}^I_0 \hat{p}^I_j(u)} n^{-1} \sum_{i=1}^n K_{ij}(u) \mathbb{I}_i \left[f_{kj}(\zeta_{ij}) - f_{kj}(u) \right],$$

$$\tilde{f}^C_{kjl}(u) = n^{-1} \sum_{i=1}^n K_{ij}(u) \mathbb{I}_i \int_0^1 K_{il}(v) \left[f_{kl}(\zeta_{il}) - f_{kl}(v) \right] dv.$$

Likewise, define $\tilde{f}^{*A}_{kj}$, $\tilde{f}^{*B}_{kj}$ and $\tilde{f}^{*C}_{kjl}$, replacing $\hat{\vartheta}_{ik}$, $\hat{p}^I_0$, $\hat{p}^I_j$, $\mathbb{I}_i$, K_{ij}, K_{il} in the definitions of $\tilde{f}^A_{kj}$, $\tilde{f}^B_{kj}$ and $\tilde{f}^C_{kjl}$ by ϑ_{ik}, $\hat{p}^{*I}_0$, $\hat{p}^{*I}_j$, $\mathbb{I}^*_i$, K^*_{ij}, K^*_{il}, respectively.

Lemma 1. Under the conditions of Theorem 1, we have

$$\hat{p}^I_0 - \hat{p}^{*I}_0 = O_p(n^{-(1-\beta)/2}),$$

$$\sup_{u \in [0,1]} |\hat{p}^I_j(u) - \hat{p}^{*I}_j(u)| = O_p(n^{-(3-5\beta)/10}) \quad \text{for all } j,$$

$$\sup_{u,v \in [0,1]} |\hat{p}^I_{jk}(u,v) - \hat{p}^{*I}_{jk}(u,v)| = O_p(n^{-(3-5\beta)/10}) \quad \text{for all } j \neq k.$$

Proof. For the proof of the first claim, we may assume that $\max_{i,j} |\hat{\zeta}_{ij} - \zeta_{ij}| \le C_0 n^{-(1-\beta)/2}$ for some positive constant C_0, due to (A3.1). Define

$I_n = I_n^L / I_n^S$, where

$$I_n^L = \{\mathbf{u} : -C_0 n^{-(1-\beta)/2} \le u_j \le 1 + C_0 n^{-(1-\beta)/2}\, 1 \le j \le M\},$$

$$I_n^S = \{\mathbf{u} : C_0 n^{-(1-\beta)/2} \le u_j \le 1 - C_0 n^{-(1-\beta)/2}\, 1 \le j \le M\}.$$

The volume of I_n in $\mathbb{R}^M$ is of order $n^{-(1-\beta)/2}$. Thus, we have

$$|\hat{p}_0^I - \hat{p}_0^{*I}| \le n^{-1} \sum_{i=1}^n \mathbb{I}(\boldsymbol{\zeta}_i \in I_n) = O_p(n^{-(1-\beta)/2}).$$

Among the last two claims, we only prove the third one. The second one follows by similar arguments. From (A1) and (A3.1),

$$\max_{1 \le i \le n} \sup_{u \in [0,1]} |K_{ij}(u) - K_{ij}^*(u) - (\hat{\zeta}_{ij} - \zeta_{ij}) K'_{h_j}(u, \zeta_{ij})| \le L(\hat{\zeta}_{ij} - \zeta_{ij})^2 h_j^{-3}, \quad \text{(A3.2)}$$

for some constant $L > 0$, where $K'_g(u, v) = \partial K_g(u, v)/\partial v$, and

$$\begin{aligned} \sup_{u \in [0,1]} n^{-1} \sum_{i=1}^n |K_{ij}^*(u)| &= O_p(1), \\ \sup_{u \in [0,1]} n^{-1} \sum_{i=1}^n |K'_{h_j}(u, \zeta_{ij})| &= O_p(h_j^{-1}), \\ \sup_{u,v \in [0,1]} n^{-1} \sum_{i=1}^n |K'_{h_j}(u, \zeta_{ij}) K_{ik}^*(v)| &= O_p(h_j^{-1}), \\ \sup_{u,v \in [0,1]} n^{-1} \sum_{i=1}^n |K'_{h_j}(u, \zeta_{ij}) K'_{h_k}(v, \zeta_{ik})| &= O_p(h_j^{-1} h_k^{-1}). \end{aligned} \quad \text{(A3.3)}$$

From (A3.2) and (A3.3) we may deduce

$$n^{-1} \sum_{i=1}^n K_{ij}(u) K_{ik}(v) \mathbb{I}_i = n^{-1} \sum_{i=1}^n K_{ij}^*(u) K_{ik}^*(v) \mathbb{I}_i + O_p(n^{-(3-5\beta)/10})$$

uniformly for $u, v \in [0,1]$. Now, assuming $\max_{i,j} |\hat{\zeta}_{ij} - \zeta_{ij}| \le C_0 n^{-(1-\beta)/2}$ as in the proof of the first claim, we may prove

$$n^{-1}\sum_{i=1}^{n} K^*_{ij}(u)K^*_{ik}(v)|\mathbb{I}_i - \mathbb{I}^*_i| \le n^{-1}\sum_{i=1}^{n} K^*_{ij}(u)K^*_{ik}(v)\mathbb{I}(\boldsymbol{\zeta}_i \in I_n) \qquad \text{(A3.4)}$$
$$= O_p(n^{-(1-\beta)/2+(1/5)}) = O_p(n^{-(3-5\beta)/10}),$$

uniformly for $u, v \in [0,1]$. This completes the proof of the third part of the lemma. □

Lemma 2. Under the conditions of Theorem 1,

$$\sup_{u\in[0,1]} |\tilde{f}^A_{kj}(u) - \tilde{f}^{*A}_{kj}(u)| = O_p(n^{-(1-\beta)/2}) \quad \text{for all } j,$$

$$\sup_{u\in[0,1]} |\tilde{f}^B_{kj}(u) - \tilde{f}^{*B}_{kj}(u)| = O_p(n^{-(1-\beta)/2}) \quad \text{for all } j,$$

$$\sup_{u\in[0,1]} |\tilde{f}^C_{kjl}(u) - \tilde{f}^{*C}_{kjl}(u)| = O_p(n^{-(1-\beta)/2}) \quad \text{for all } j \ne l.$$

Proof. We prove the first and the third parts only. The second part follows from the arguments used in the proof of the third part. For the first part we note that from the second part of (A3.1) and the inequality (A3.2),

$$n^{-1}\sum_{i=1}^{n} K_{ij}(u)\mathbb{I}_i\hat{\vartheta}_{ik} = n^{-1}\sum_{i=1}^{n} K_{ij}(u)\mathbb{I}_i\vartheta_{ik} + O_p(n^{-(1-\beta)/2})$$
$$= n^{-1}\sum_{i=1}^{n} K^*_{ij}(u)\mathbb{I}_i\vartheta_{ik} + n^{-1}\sum_{i=1}^{n}(\hat{\zeta}_{ij} - \zeta_{ij})K'_{h_j}(u, \zeta_{ij})\mathbb{I}_i\vartheta_{ik} \qquad \text{(A3.5)}$$
$$+ O_p(n^{-(1-\beta)/2})$$

uniformly for $u \in [0,1]$. By an application of an exponential inequality, conditioning on $(X_i : 1 \le i \le n)$ and the use of (A3.1), we may show that the second term on the right hand side of the second equation of (A3.5) is of order $O_p(n^{-1/2}h_j^{-3/2}(\log n)^{1/2}n^{-(1-\beta)/2}) = O_p(n^{-(7-5\beta)/10}(\log n)^{1/2})$, uniformly for $u \in [0,1]$. Similarly,

$$n^{-1}\sum_{i=1}^{n} K_{ij}^*(u)(\mathbb{I}_i - \mathbb{I}_i^*)\vartheta_{ik} = O_p(n^{-1/2}h_j^{-1}n^{-(1-\beta)/4}(\log n)^{1/2}),$$

uniformly for $u \in [0,1]$, where we used

$$n^{-1}\sum_{i=1}^{n} K_{ij}^*(u)^2(\mathbb{I}_i - \mathbb{I}_i^*)^2 = O_p(h_j^{-2}n^{-(1-\beta)/2}),$$

uniformly for $u \in [0,1]$. This completes the proof of the first part.

For the proof of the third part, we replace $K_{ij}(u)$ in $\tilde{f}_{kjl}^C(u)$, as defined right before Lemma 1, by $K_{ij}^*(u) + (\hat{\zeta}_{ij} - \zeta_{ij})K'_{h_j}(u,\zeta_{ij}) + (\text{remainder})$, with the remainder being of order $n^{-(2-5\beta)/5}$, uniformly for $u \in [0,1]$. Likewise, we replace $K_{il}(v)$ in $\tilde{f}_{kjl}^C(u)$ by similar terms. This gives a decomposition of $\tilde{f}_{kjl}^C(u) - \tilde{f}_{kjl}^{*C}(u)$ into several terms. The three leading terms are $\mathbf{I}+\mathbf{II}+\mathbf{III}$, where

$$\mathbf{I} = \frac{1}{\hat{p}_0^I\hat{p}_j^I(u)}n^{-1}\sum_{i=1}^{n} K_{ij}^*(u)\mathbb{I}_i(\hat{\zeta}_{il} - \zeta_{il})\int_0^1 K'_{h_l}(v,\zeta_{il})\,[f_{kl}(\zeta_{il}) - f_{kl}(v)]\,dv$$

$$\mathbf{II} = \frac{1}{\hat{p}_0^I\hat{p}_j^I(u)}n^{-1}\sum_{i=1}^{n} K'_{h_j}(u,\zeta_{ij})\mathbb{I}_i(\hat{\zeta}_{ij} - \zeta_{ij})\int_0^1 K_{il}^*(v)\,[f_{kl}(\zeta_{il}) - f_{kl}(v)]\,dv$$

$$\mathbf{III} = \frac{1}{\hat{p}_0^I\hat{p}_j^I(u)}n^{-1}\sum_{i=1}^{n} K_{ij}^*(u)(\mathbb{I}_i - \mathbb{I}_i^*)\int_0^1 K_{il}^*(v)\,[f_{kl}(\zeta_{il}) - f_{kl}(v)]\,dv,$$

while the other terms are of smaller order. Using the third property of (A3.3) and the fact that

$$|K'_{h_l}(v,\zeta_{il})(f_{kl}(\zeta_{il}) - f_{kl}(v))| \le Ch_lK'_{h_l}(v,\zeta_{il}) \qquad \text{(A3.6)}$$

for some constant $C > 0$, we get that $\mathbf{I}$ is of order $O_p(n^{-(1-\beta)/2})$, uniformly for $u \in [0,1]$. Note that (A3.6) also holds with $K^*_{il}(v)$ replacing $K'_{h_l}(v, \zeta_{il})$ on both sides of the inequality. This gives $\mathbf{II} = O_p(n^{-(1-\beta)/2})$ uniformly for $u \in [0,1]$. Furthermore, together with (A3.4) it gives $\mathbf{III} = O_p(n^{-(1-\beta)/2})$ uniformly for $u \in [0,1]$. This completes the proof of the lemma. $\square$

We now come to the proof of Theorem 1. We assume $f_{k0} = 0$ and ignore $\hat{f}_{k0}$ and $\hat{f}^*_{k0}$ in the backfitting equation (4.8) and its version with true ζ_{ij}, respectively. This is justified because $\hat{f}^*_{k0} - f_{k0}$ is of order $n^{-1/2}$ and $\hat{f}_{k0} - \hat{f}^*_{k0}$ is of order $n^{-(1-\beta)/2} = o(n^{-2/5})$. Define linear operators

$$\pi_j(g) = \int_{I_{-j}} g(\mathbf{u}) \frac{p^I(\mathbf{u})}{p^I_j(u_j)}\, d\mathbf{u}_{-j}, \quad 1 \le j \le M,$$

and likewise $\hat{\pi}_j$ and $\hat{\pi}^*_j$, respectively, replacing (p^I, p^I_j) by $(\hat{p}^I, \hat{p}^I_j)$ and $(\hat{p}^{*I}, \hat{p}^{*I}_j)$, where $p^I(\mathbf{u}) = p(\mathbf{u})/p^I_0$, $\hat{p}^I(\mathbf{u}) = \hat{p}(\mathbf{u})/\hat{p}^I_0$ and $\hat{p}^{*I}(\mathbf{u}) = \hat{p}^*(\mathbf{u})/\hat{p}^{*I}_0$. Define a linear operator

$$T = (I - \pi_M)(I - \pi_{M-1}) \cdots (I - \pi_2)(I - \pi_1),$$

and likewise $\hat{T}$ and $\hat{T}^*$ with π_j being replaced by $\hat{\pi}_j$ and $\hat{\pi}^*_j$, respectively. For a linear operator F that maps the space of additive functions to itself, we define its norm $\|F\|$ by

$$\|F\|^2 = \sup\left\{ \int F(g)(\mathbf{u})^2 p^I(\mathbf{u})\, d\mathbf{u} : g \text{ is additive and } \int g(\mathbf{u})^2 p^I(\mathbf{u})\, d\mathbf{u} = 1 \right\}$$

Lemma 3. Under the conditions of Theorem 1, it holds that $\|\hat{T}^* - T\| = o_p(1)$ and $\|T\| < \gamma$ for some constant $0 < \gamma < 1$.

Proof. The lemma follows along the lines of the proof of Theorem 1 in Mammen et al. (1999). Let $\mathcal{H}(p^I)$ denote the space of additive functions g such that $g(\mathbf{u}) = \sum_{j=1}^M g_j(u_j)$ for some univariate functions g_j with $\int_0^1 g_j(u_j)^2 p_j^I(u_j)\, du_j < \infty$. Also, let $\mathcal{H}_k(p^I)$ denote its subspaces consisting of functions such that $g(\mathbf{u}) = g_k(u_k)$ for some univariate function g_k. The second result of the lemma follows from an application of Proposition A.4.2 of Bickel et al. (1993) to the projection operators π_j. The key argument is that the projection π_j restricted to $\mathcal{H}_k(p)$ for $k \neq j$ is Hilbert-Schmidt, that is

$$\int_{[0,1]^2} \left[\frac{p_{jk}^I(u_j, u_k)}{p_j^I(u_j)p_k^I(u_k)}\right]^2 p_j^I(u_j)p_k^I(u_k)\, du_j\, du_k < \infty.$$

For the proof of the first part, it suffices to show $\|\hat{\pi}_j^* - \pi_j\| = o_p(1)$ for $1 \leq j \leq M$ since $\|\pi_j\| = 1$. Another application of Proposition A.4.2 of Bickel et al. (1993) entails that there exists a constant $0 < c < \infty$ such that for any $g \in \mathcal{H}(p^I)$ there exists a decomposition $g(\mathbf{u}) = \sum_{j=1}^M g_j(u_j)$ such that

$$\max\{\|g_1\|_2, \ldots, \|g_M\|_2\} \leq c\|g\|_2. \tag{A3.7}$$

Let g be an arbitrary element of $\mathcal{H}(p^I)$. Due to (A3.7), we have

$$\begin{aligned}
\|(\hat{\pi}_j^* - \pi_j)g\|_2 &= \left[\int_0^1 \left(\sum_{k\neq j}^d \int_0^1 g_k(u_k)\left[\frac{\hat{p}_{jk}^{*I}(u_j,u_k)}{\hat{p}_j^{*I}(u_j)} - \frac{p_{jk}^I(u_j,u_k)}{p_j^I(u_j)}\right] du_k\right)^2 p_j^I(u_j)\, du_j\right]^{1/2} \\
&\leq \sum_{k\neq j}^d \left[\int_0^1 \left(\int_0^1 g_k(u_k)\left[\frac{\hat{p}_{jk}^{*I}(u_j,u_k)}{\hat{p}_j^{*I}(u_j)p_k^I(u_k)} - \frac{p_{jk}^I(u_j,u_k)}{p_j^I(u_j)p_k^I(u_k)}\right] p_k^I(u_k)\, du_k\right)^2 \right. \\
&\qquad\qquad \left. \times\, p_j^I(u_j)\, du_j\right]^{1/2} \\
&\leq \sum_{k\neq j}^d \|g_k\|_2 \left(\int_{[0,1]^2}\left[\frac{\hat{p}_{jk}^{*I}(u_j,u_k)}{\hat{p}_j^{*I}(u_j)p_k^I(u_k)} - \frac{p_{jk}^I(u_j,u_k)}{p_j^I(u_j)p_k^I(u_k)}\right]^2 p_j^I(u_j)p_k^I(u_k)\, du_j\, du_k\right)^{1/2} \\
&\leq c\cdot \hat{r}_{nj}\cdot \|g\|_2,
\end{aligned}$$

where the constant c is as given at (A3.7) which does not depend on g and $\hat{r}_{nj} = o_p(1)$. This proves $\|\hat{\pi}_j^* - \pi_j\| = o_p(1)$. □

Proof of (i) *and* (ii). Let

$$\begin{aligned}
\tilde{f}_{k\oplus} &= \tilde{f}_{kM} + (I-\hat{\pi}_M)\tilde{f}_{k,M-1} + (I-\hat{\pi}_M)(I-\hat{\pi}_{M-1})\tilde{f}_{k,M-2} \\
&\qquad + \cdots + (I-\hat{\pi}_M)\cdots(I-\hat{\pi}_2)\tilde{f}_{k1},
\end{aligned} \tag{A3.8}$$

and define likewise $\tilde{f}_{k\oplus}^*$ with $\tilde{f}_{kj}$ and $\hat{\pi}_j$ being replaced by $\tilde{f}_{kj}^*$ and $\hat{\pi}_j^*$, respectively. Also, define the following additive functions,

$$\begin{aligned}
\hat{f}_{k+}(\mathbf{u}) &= \sum_{j=1}^M \hat{f}_{kj}(u_j), \quad \hat{f}_{k+}^*(\mathbf{u}) = \sum_{j=1}^M \hat{f}_{kj}^*(u_j), \\
\hat{f}_{k+}^{[r]}(\mathbf{u}) &= \sum_{j=1}^M \hat{f}_{kj}^{[r]}(u_j), \quad \hat{f}_{k+}^{*[r]}(\mathbf{u}) = \sum_{j=1}^M \hat{f}_{kj}^{*[r]}(t,u_j).
\end{aligned}$$

Then, the backfitting equation (4.8) and its version with true ζ_{ij}, respectively, can be written as

$$\hat{f}_{k+} = \tilde{f}_{k\oplus} + \hat{T}\hat{f}_{k+}, \quad \hat{f}^*_{k+} = \tilde{f}^*_{k\oplus} + \hat{T}^*\hat{f}^*_{k+},$$

and the updating algorithms can be also written as

$$\hat{f}^{[r]}_{k+} = \tilde{f}_{k\oplus} + \hat{T}\hat{f}^{[r-1]}_{k+}, \quad \hat{f}^{*[r]}_{k+} = \tilde{f}^*_{k\oplus} + \hat{T}^*\hat{f}^{*[r-1]}_{k+}.$$

As a consequence of Lemma 3, (i) and (ii) of the theorem follow if we prove

$$\|\hat{T} - \hat{T}^*\| = o_p(1). \tag{A3.9}$$

The property (A3.9) is a direct consequence of Lemma 1 since the second and third parts of the lemma imply $\|\hat{\pi}_j - \hat{\pi}^*_j\| = O_p(n^{-(3-5\beta)/10}) = o_p(1)$.

Proof of (iii). From the backfitting equation (4.8),

$$\begin{aligned}\hat{f}_{kj}(u) = f_{kj}(u) + \tilde{f}^A_{kj}(u) + \tilde{f}^B_{kj}(u) - \frac{1}{\hat{p}^I_0\hat{p}^I_j(u)} n^{-1}\sum_{i=1}^n \mathbb{I}_i K_{ij}(u) \\ \times \sum_{l\neq j}^M \int_0^1 [\hat{f}_{kl}(v) - f_{kl}(\zeta_{il})] K_{il}(v)\, dv.\end{aligned} \tag{A3.10}$$

For (A3.10) we have used $\int K_{ij}(u)\, du = 1$. By Lemma 2,

$$\tilde{f}^C_{kjl}(u) = \tilde{f}^{*C}_{kjl}(u) + o_p(n^{-2/5}), \tag{A3.11}$$

uniformly for $u \in [0, 1]$. To further approximate $\tilde{f}^{*C}_{kjl}(u)$, define

$$\delta_{ij} = \int [f_{kj}(\zeta_{ij}) - f_{kj}(z)]\, K_{ij}(z)\, dz.$$

Then, $\hat{p}_0^I \hat{p}_j^I(u) \cdot \tilde{f}_{kjl}^{*C}(u) = n^{-1} \sum_{i=1}^n \delta_{il} \mathbb{I}_i^* K_{ij}^*(u)$. From standard results on kernel smoothing,

$$\sup_{u \in [0,1]} \left| n^{-1} \sum_{i=1}^n \left[\delta_{il} - E(\delta_{il} | \zeta_{ij}, \mathbb{I}_i^*)\right] K_{ij}^*(u) \mathbb{I}_i^* \right| = O_p(n^{-3/5} \sqrt{\log n}). \quad \text{(A3.12)}$$

We next compute $E(\delta_{il} | \zeta_{ij} = v, \boldsymbol{\zeta}_i \in I)$. Define

$$\mu_l = \int u^l K(u)\, du, \quad \mu_{l,j}(z) = h_j^{-l} \int (w - z)^l K_{h_j}(z, w)\, dw,$$

where we note that $\mu_{l,j}(z) = 0$ for $z \in [2h_j, 1 - 2h_j]$, if l is an odd positive integer and the baseline kernel K is symmetric. Also, let $a_j(z) = \mu_{1,j}(z) f'_{kj}(z)$, $b_j(z) = \mu_2\, f''_{kj}(z)/2$ and

$$c_{jl}(v, z) = \mu_2\, f'_{kl}(z) p_{jl}^I(v, z)^{-1} \partial p_{jl}^I(v, z)/\partial z.$$

Note that a_j, b_j and c_{jl} depend on the index k, but we suppress k for simplicity of notation. By expanding of $f_{kj}(w) - f_{kj}(z)$ and the conditional density $p_{jl}^I(v, w)/p_j^I(v)$ for w near z, we get

$$E(\delta_{il} | \zeta_{ij} = v, \boldsymbol{\zeta}_i \in I) = \int_0^1 \frac{p_{jl}^I(v, z)}{p_j^I(v)} [h_l a_l(z) + h_l^2 b_l(z) + h_l^2 c_{jl}(v, z)]\, dz + o_p(n^{-2/5}),$$

uniformly for $v \in [0, 1]$. For this we have used the formula

$$E[f(\boldsymbol{\zeta})\, | \zeta_j = u_j, \boldsymbol{\zeta} \in I] = \left(\int_{I_{-j}} p(\mathbf{u})\, d\mathbf{u}_{-j} \right)^{-1} \int_{I_{-j}} f(\mathbf{u}) p(\mathbf{u})\, d\mathbf{u}_{-j}, \quad \text{(A3.13)}$$

for $u_j \in [0,1]$. This together with (A3.12) gives that, uniformly for $u \in [0,1]$,

$$\begin{aligned}\hat{p}_0^I \hat{p}_j^I(u) \cdot \tilde{f}_{kjl}^{*C}(u) &= n^{-1} \sum_{i=1}^{n} E(\delta_{il} | \zeta_{ij}, \mathbb{I}_i^*) \mathbb{I}_i^* K_{ij}^*(u) + o_p(n^{-2/5}) \\ &= n^{-1} \sum_{i=1}^{n} \mathbb{I}_i^* \int_0^1 \frac{p_{jl}^I(\zeta_{ij}, z)}{p_j^I(\zeta_{ij})} \Big[h_l a_l(z) + h_l^2 b_l(z) \\ &\qquad + h_l^2 c_{jl}(u,z) \Big] K_{ij}^*(u)\, dz + o_p(n^{-2/5}).\end{aligned} \tag{A3.14}$$

Noting that $a_j(z) = O_p(1)$, $b_j(z) = O_p(1)$, $c_{jl}(u,z) = O_p(1)$ uniformly for $u, z \in [0,1]$, and $a_j(z) = 0$ for $z \in [2h_j, 1-2h_j]$, we infer that, uniformly for $u \in [0,1]$,

$$\begin{aligned}o_p(n^{-2/5}) &= n^{-1} \sum_{i=1}^{n} \int_0^1 \left[h_l \frac{a_l(z)}{\mu_{0,l}(z)} + h_l^2 b_l(z) + h_l^2 c_{jl}(u,z) \right] \\ &\qquad \times \left[K_{il}^*(z) - E\left(K_{il}^*(z) \middle| \zeta_{ij}, \mathbb{I}_i^* \right) \right] K_{ij}^*(u)\, dz \\ &= \hat{p}_0^{*I} \int_0^1 \left[h_l \frac{a_l(z)}{\mu_{0,l}(z)} + h_l^2 b_l(z) + h_l^2 c_{jl}(u,z) \right] \\ &\qquad \times \hat{p}_{jl}^{*I}(u,z)\, dz - n^{-1} \sum_{i=1}^{n} \mathbb{I}_i^* \int_0^1 \frac{p_{jl}^I(\zeta_{ij}, z)}{p_j^I(\zeta_{ij})} \Big[h_l a_l(z) \\ &\qquad + h_l^2 b_l(z) + h_l^2 c_{jl}(u,z) \Big] K_{ij}^*(u)\, dz.\end{aligned} \tag{A3.15}$$

By Lemma 1, we may replace $\hat{p}_0^{*I}$ and $\hat{p}_{jk}^{*I}$ by $\hat{p}_0^I$ and $\hat{p}_{jk}^I$, respectively, on the right hand side of the second equality at (A3.15), with an approximation error $o_p(n^{-2/5})$ uniformly for $u \in [0,1]$. This together with (A3.11), (A3.14) and (A3.15) leads to

$$\tilde{f}^{C}_{kjl}(u) = \int_0^1 \left[h_l \frac{a_l(z)}{\mu_{0,l}(z)} + h_l^2 b_l(z) + h_l^2 c_{jl}(u,z)\right] \frac{\hat{p}^I_{jl}(u,z)}{\hat{p}^I_j(u)}\, dz + o_p(n^{-2/5}), \tag{A3.16}$$

uniformly for $u \in [0.1]$. Furthermore, $\tilde{f}^{B}_{kj}(u) = \tilde{f}^{*B}_{kj}(u) + o_p(n^{-2/5})$, uniformly for $u \in [0,1]$ by Lemma 2, and

$$\tilde{f}^{*B}_{kj}(u) = h_j \frac{a_j(u)}{\mu_{0,j}(u)} + h_j^2 b_j(u) + h_j^2 c_j(u) + r_j(u), \tag{A3.17}$$

where $c_j(u) = \mu_2 f'_{kj}(u) p^I_j(u)^{-1} \partial p^I_j(u)/\partial u$ and r_j denotes a generic stochastic term such that

$$\sup_{u\in[2h_j,1-2h_j]} |r_j(u)| = o_p(n^{-2/5}), \quad \sup_{u\in[0,1]} |r_j(u)| = O_p(n^{-2/5}).$$

Now, (A3.10), (A3.16), (A3.17) and Lemma 2 give

$$\begin{aligned}\hat{f}_{kj}(u) =& f_{kj}(u) + \tilde{f}^{*A}_{kj}(u) + h_j \frac{a_j(u)}{\mu_{0,j}(u)} + h_j^2 b_j(u) + \tilde{\Delta}_{kj}(u) \\ & - \sum_{l\neq j}^{M} \int_0^1 \Big[\hat{f}_{kl}(v) - f_{kl}(v) - \tilde{f}^{*A}_{kl}(v) - h_l \frac{a_l(v)}{\mu_{0,l}(v)} \\ & - h_l^2 b_l(v)\Big] \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)}\, dv + r_j(u),\end{aligned}$$

with

$$\tilde{\Delta}_{kj}(u) = \mu_2 \sum_{l=1}^{M} h_l^2 E\Big[f'_{kl}(\zeta_l) \frac{p^{(1)}_l(\boldsymbol{\zeta})}{p(\boldsymbol{\zeta})} \big| \zeta_j = u, \boldsymbol{\zeta} \in I\Big].$$

In the above equation, we have used

$$\sup_{u\in[0,1]} \Big| \int_0^1 \tilde{f}^{*A}_{kl}(v) \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)}\, dv \Big| = o_p(n^{-2/5}),$$

$$\sup_{u\in[0,1]} \Big| \int_0^1 c_{jl}(u,v) \Big(\frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)} - \frac{p^I_{jl}(u,v)}{p^I_j(u)} \Big)\, dv \Big| = o_p(1).$$

The above approximations follow from Lemma 1 and standard results for kernel smoothing. With

$$\hat{\Delta}_{kj}(u) = \hat{f}_{kj}(u) - f_{kj}(u) - \tilde{f}^{*A}_{kj}(u) - h_j \frac{a_j(u)}{\mu_{0,j}(u)} - h_j^2 b_j(u) - r_j(u),$$

(A3.18) implies that, up to a remainder that is uniformly of order $o_p(n^{-2/5})$, the tuple $(\hat{\Delta}_{kj} : 1 \le j \le M)$ satisfies the system of equations

$$\hat{\Delta}_{kj}(u) = \tilde{\Delta}_{kj}(u) - \sum_{l \neq j}^{M} \int_0^1 \hat{\Delta}_{kl}(v) \frac{\hat{p}^I_{jl}(u,v)}{\hat{p}^I_j(u)} \, dv. \tag{A3.18}$$

Let the tuple $(\Delta_{kj} : 1 \le j \le M)$ be the solution of the system of equations

$$\Delta_{kj}(u) = \tilde{\Delta}_{kj}(u) - \sum_{l \neq j}^{M} \int_0^1 \Delta_{kl}(v) \frac{p^I_{jl}(u,v)}{p^I_j(u)} \, dv \tag{A3.19}$$

subject to

$$\int_0^1 \Delta_{kj}(u) p^I_j(u) \, du = \mu_2 \, h_j^2 \int_0^1 f'_{kj}(u) \frac{\partial}{\partial u} p^I_j(u) \, du. \tag{A3.20}$$

Note that the tuple that satisfies the system of equations (A3.19) is unique up to an additive constant vector. This can be seen from the fact that replacing $\Delta_{kj}(u)$ by $\Delta_{kj}(u) + c$ for a constant c on the left hand side and $\Delta_{kl}(v)$ by $\Delta_{kl}(v) - c$ for a particular l on the right hand side gives another solution. With the constraints at (A3.20), however, the tuple is uniquely determined. We claim

$$\hat{\Delta}_{kj}(u) = \Delta_{kj}(u) + r_j(u), \tag{A3.21}$$

so that

$$\hat{f}_{kj}(u) = f_{kj}(u) + \tilde{f}^{*A}_{kj}(u) + h_j \frac{a_j(u)}{\mu_{0,j}(u)} + \frac{1}{2}\, h_j^2\, \mu_2\, f''_{kj}(u) + \Delta_{kj}(u) + r_j(u).$$

This completes the proof of (iii) since $\tilde{f}^{*A}_{kj}$ for $1 \le j \le M$ are asymptotically independent and $n^{2/5}\tilde{f}^{*A}_{kj}(u)$ converges in distribution to $N(0, \tau_j^2(u))$.

It remains to prove (A3.21). Define $\Delta_{k\oplus}$ as $\tilde{f}_{k\oplus}$ at (A3.8) with $\tilde{f}_{kj}$ and $\hat{\pi}_j$, respectively, being replaced by $\tilde{\Delta}_{kj}$ and π_j. Also, define $\hat{\Delta}_{k\oplus}$ with only $\tilde{f}_{kj}$ being replaced by $\tilde{\Delta}_{kj}$. For

$$\hat{\Delta}_{k+}(\mathbf{u}) = \sum_{j=1}^{M} \hat{\Delta}_{kj}(u_j), \quad \Delta_{k+}(\mathbf{u}) = \sum_{j=1}^{M} \Delta_{kj}(u_j),$$

the backfitting equations (A3.18) and (A3.19) can be written as $\hat{\Delta}_{k+} = \hat{\Delta}_{k\oplus} + \hat{T}\hat{\Delta}_{k+}$ and $\Delta_{k+} = \Delta_{k\oplus} + T\Delta_{k+}$, respectively. Since $\sup_{u\in[0,1]} |\tilde{\Delta}_{kj}(u)| = O(n^{-2/5})$, and $\hat{\Delta}_{k\oplus}$ differs from $\Delta_{k\oplus}$ only in that it uses $\hat{\pi}_j$ instead of π_j, it follows from Lemma 1 that

$$\sup_{\mathbf{u}\in I_0} |\hat{\Delta}_{k\oplus}(\mathbf{u}) - \Delta_{k\oplus}(\mathbf{u})| = o_p(n^{-2/5}), \tag{A3.22}$$

where $I_0 = \{\mathbf{u} : 2h_j < u_j < 1 - 2h_j,\ 1 \le j \le M\}$. From Lemma 3 and (A3.9), we also have $\|\hat{T} - T\| = o_p(1)$ and $\|T\| < 1$. Together with (A3.22), this entails

$$\sup_{\mathbf{u}\in I_0} |\hat{\Delta}_{k+}(\mathbf{u}) - \Delta_{k+}(\mathbf{u})| = o_p(n^{-2/5}), \quad \sup_{\mathbf{u}\in I} |\hat{\Delta}_{k+}(\mathbf{u}) - \Delta_{k+}(\mathbf{u})| = O_p(n^{-2/5}),$$

so that

$$\hat{\Delta}_{kj}(u) = \Delta_{kj}(u) + n^{-2/5} Z_j + r_j(u) \tag{A3.23}$$

for some random variables Z_j such that $\sum_{j=1}^{M} Z_j = o_p(1)$. We prove $Z_j = o_p(1)$ for all j, which establishes (A3.21).

From the definition of $\hat{\Delta}_{kj}$, its expansion at (A3.23) and the constraints for $\hat{f}_{kj}$ at (4.9), we have

$$\begin{aligned} 0 = &\int_0^1 f_{kj}(u)\hat{p}_j^I(u)\,du + h_j \int_0^1 \frac{a_j(u)}{\mu_{0,j}(u)}\hat{p}_j^I(u)\,du \\ &+ \frac{1}{2}\,h_j^2\,\mu_2 \int_0^1 f_{kj}''(u)p_j^I(u)\,du + \int_0^1 \Delta_{kj}(u)p_j^I(u)\,du + n^{-2/5}Z_j + o_p(n^{-2/5}). \end{aligned} \tag{A3.24}$$

Here, we also have used Lemma 1 and the fact that $\sup_{u\in[0,1]} |\Delta_{kj}(u)| = O(n^{-2/5})$. Using $\int K_{h_j}(u,v)\,du = 1$ for all $v \in [0,1]$, we get

$$\begin{aligned} \int_0^1 f_{kj}(u)\hat{p}_j^I(u)\,du = n^{-1}\sum_{i=1}^{n} \mathbb{I}_i \int_0^1 [f_{kj}(u) - f_{kj}(\zeta_{ij})]K_{h_j}(u,\hat{\zeta}_{ij})\,du/\hat{p}_0^I \\ + n^{-1}\sum_{i=1}^{n} \mathbb{I}_i\, f_{kj}(\zeta_{ij})/\hat{p}_0^I. \end{aligned}$$

The second term on the right hand side of the above equation is of order $n^{-(1-\beta)/2}$. This is due to the constraint (4.3), $n^{-1}\sum_{i=1}^{n} |\mathbb{I}_i - \mathbb{I}_i^*| = O_p(n^{-(1-\beta)/2})$ and

$$n^{-1}\sum_{i=1}^{n} \mathbb{I}_i^*[f_{kj}(\zeta_{ij}) - E(f_{kj}(\zeta_{ij})|\boldsymbol{\zeta}_i \in I)] = O_p(n^{-1/2}).$$

For the first term, denoted by **IV**, we get

$$\begin{aligned}
\mathbf{IV} &= n^{-1}\sum_{i=1}^{n}\mathbb{I}_i^*\int_0^1[f_{kj}(u)-f_{kj}(\zeta_{ij})]K_{h_j}(u,\zeta_{ij})\,du/\hat{p}_0^{*I}+o_p(n^{-2/5})\\
&= \int_0^1[f_{kj}(u)-f_{kj}(v)]K_{h_j}(u,v)p_j^I(v)\,dv\,du+o_p(n^{-2/5})\\
&= -h_j\int_0^1 a_j(u)p_j^I(u)\,du-\frac{1}{2}\,h_j^2\,\mu_2\int_0^1 f_{kj}''(u)p_j^I(u)\,du\\
&\qquad\qquad -h_j^2\,\mu_2\int_0^1 f_{kj}'(u)\frac{\partial}{\partial u}p_j^I(u)\,du+o_p(n^{-2/5}).
\end{aligned}$$

In the first approximation of **IV**, we have used (A3.1) and Lemma 1. We also have

$$h_j\int_0^1\frac{a_j(u)}{\mu_{0,j}(u)}\hat{p}_j^I(u)\,du=h_j\int_0^1 a_j(u)p_j^I(u)\,du+o_p(n^{-2/5}).$$

These approximations of the terms in (A3.24) and the constraint of Δ_{kj} at (A3.20) give $Z_j=o_p(1)$. □