Submitted to the Annals of Statistics

ON THE OPTIMAL RECONSTRUCTION OF PARTIALLY OBSERVED FUNCTIONAL DATA

By Alois Kneip and Dominik Liebl

University of Bonn

We propose a new reconstruction operator that aims to recover the missing parts of a function given the observed parts. This new operator belongs to a new, very large class of functional operators which includes the classical regression operators as a special case. We show the optimality of our reconstruction operator and demonstrate that the usually considered regression operators generally cannot be optimal reconstruction operators. Our estimation theory allows for autocorrelated functional data and considers the practically relevant situation in which each of the n functions is observed at m discretization points. We derive rates of consistency for our nonparametric estimation procedures using a double asymptotic ($n \to \infty, m \to \infty$). For data situations, as in our real data application where m is considerably smaller than n, we show that our functional principal components based estimator can provide better rates of convergence than any conventional nonparametric smoothing method.

1. Introduction. Our work is motivated by a data set from energy economics which is shown in Figure 1. The data consist of partially observed price functions. Practitioners use these functions, for instance, to do comparative statics, i.e., a ceteris-paribus analysis of price effects with respect to changes in electricity demand (cf. Weigt, 2009; Hirth, 2013). The possibilities of such an analysis, however, are limited by the extent to which we can observe the price functions. This motivates the goal of our work, which is to develop a reconstruction procedure that allows us to recover the total functions from their partial observations.

Let $X_1, \ldots, X_n$ be an identically distributed, possibly weakly dependent sample of continuous random functions, where each function X_i is an element of the separable Hilbert space $\mathbb{L}^2([a,b])$ with $[a,b] \subset \mathbb{R}$ and $\mathbb{E}(||X_i||_2^4) < \infty$, where $||X_i||_2^2 = \int_a^b X_i^2 dx$.

We denote the observed and missing parts of X_i by $X_i^{O_i}$ and $X_i^{M_i}$, where

$$X_i^{O_i}(u) := X_i(u) \quad \text{for} \quad u \in O_i \subseteq [a,b]$$
$$X_i^{M_i}(u) := X_i(u) \quad \text{for} \quad u \in M_i = [a,b] \setminus O_i,$$

AMS 2000 subject classifications: Primary62M20, 62H25, 62G05, 62G08

Keywords and phrases: functional data analysis, functional principal components, incomplete functions

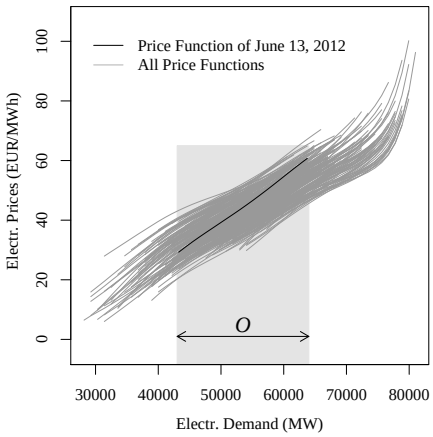

FIG 1. *Partially observed electricity price functions* $X_i^O(u)$ *with* $u \in O_i \subseteq [a,b]$.

and where $O_i = [A_i, B_i] \subseteq [a,b]$ is a random subinterval, independent from X_i, and with $B_i - A_i > 0$ almost surely. In what follows we use "O" and "M" to denote a given realization of O_i and M_i. In addition, we use the following shorthand notation for conditioning on O_i and M_i:

$$X_i^O(u) := X_i^{O_i}(u) \,|(O_i = O)$$
$$X_i^M(u) := X_i^{M_i}(u)|(M_i = M);$$

typical realizations of X_i^O and O are shown in Figure 1. In order to denote the inner product and norm of $\mathbb{L}^2(O)$, we use generic notations $\langle ., . \rangle_2$ and $||.||_2$; their dependency on O will be made obvious by writing, for instance, $\langle x^O, y^O \rangle_2$ and $||x^O||_2^2 = \langle x^O, x^O \rangle_2$ for all $x^O, y^O \in \mathbb{L}^2(O)$, where $\langle x^O, y^O \rangle_2 = \int_O x(u)y(u)du$.

For a start, we consider centered random functions, i.e., $\mathbb{E}(X_i(u)) = \mu(u)$ with $\mu(u) = 0$ for all $u \in [a,b]$. Our object of interest is the following linear reconstruction problem:

$$X_i^M = L(X_i^O) + Z_i, \quad u \in M, \tag{1}$$

which aims to reconstruct the unobserved missing parts $X_i^M \in \mathbb{L}^2(M)$ given the partial observation $X_i^O \in \mathbb{L}^2(O)$. Our objective is to identify the optimal linear reconstruction operator $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ which minimizes the mean squared error loss $\mathbb{E}\left[\left(X_i^M(u) - [L(X_i^O)](u)\right)^2\right]$ at any $u \in M$.

The case of partially observed functional data was first considered in Goldberg, Ritov and Mandelbaum (2014) and Kraus (2015). The work of

Gromenko et al. (2017) is also related as it proposes an inferential framework for incomplete spatially and temporally correlated functional data. Goldberg, Ritov and Mandelbaum (2014) consider the case of finite dimensional functional data and their results have well-known counterparts in multivariate statistics. Kraus (2015) starts by deriving his "optimal" reconstruction operator as a solution to the Frécht-type normal equation, where he *assumes* the existence of a bounded solution. The theoretical results in our paper imply that this assumption generally holds only under the very restrictive case of linear regression operators, i.e., Hilbert-Schmidt operators. For showing consistency of his empirical reconstruction operator, Kraus (2015) restricts his work to this case of Hilbert-Schmidt operators. We demonstrate, however, that a Hilbert-Schmidt operator generally cannot be the optimal reconstruction operator.

In order to see the latter, we need some conceptional work. Hilbert-Schmidt operators on $\mathbb{L}^2$ spaces correspond to linear regression operators,

$$L(X_i^O) = \int_O \beta(u,v) X_i^O(v) dv, \quad \text{with} \quad \beta \in \mathbb{L}^2(M \times O). \tag{2}$$

However, such a regression operator generally does *not* provide the optimal solution of the reconstruction problem in (1). For instance, let us consider the "last observed"(="first missing") points, namely, the boundary points[1] $\vartheta \in \partial M$ of M. For any optimal reconstruction operator L, it must hold that the "first reconstructed" value, $[L(X_i^O)](\vartheta)$, connects with the "last observed" value, $X_i^O(\vartheta)$, i.e., that

$$X_i^O(\vartheta) = [L(X_i^O)](\vartheta) \quad \text{for all} \quad \vartheta \in \partial M.$$

There is no hope, though, of finding a slope function $\beta(\vartheta, .) \in \mathbb{L}^2(O)$ that fulfills the equation $X_i^O(\vartheta) = \int_O \beta(\vartheta, v) X_i^O(v) dv$, except in a degenerated manner if $\beta(\vartheta, .)$ equals the Dirac-δ function $\delta_\vartheta(v) = 1_{(v=\vartheta)}$. Dirac-$\delta$ functions in $\mathbb{L}^2(O)$ are, however, equivalent to the zero function $z(v) \equiv 0$, since $||\delta_\vartheta - z||_2 = 0$, which makes it impossible to identify the optimal reconstruction operator L within the class of linear regression operators defined by (2).

Best possible linear reconstruction operators depend, of course, on the structure of the random function X_i, and possible candidates have only to be well-defined for any function in the support of X_i. We therefore consider the class of all linear operators L with $\mathbb{V}([L(X_i^O)](u)) < \infty$ and thus

[1]The boundary ∂M of a subset M is defined as $\partial M := \overline{M} \cap \overline{O}$, where $\overline{M}$ and $\overline{O}$ denote the closures of the subsets M and O.

$\mathbb{P}(|[L(X_i^O)](u)| < \infty) = 1$ for any $u \in M$. This class of reconstruction operators is much larger than the class of regression operators and contains the latter as a special case. A theoretical characterization is given in Section 2. We then show that the optimal linear reconstruction operator, minimizing $\mathbb{E}[(X_i^M(u) - [L(X_i^O)](u))^2]$ for all $u \in M$, is given by

$$
\begin{aligned}
[\mathcal{L}(X_i^O)](u) &= \sum_{k=1}^{\infty} \frac{\langle \phi_k^O, X_i^O \rangle_2 \; \mathbb{E}[X_i^M(u) \langle \phi_k^O, X_i^O \rangle_2]}{\lambda_k^O} \\
&= \sum_{k=1}^{\infty} \frac{\langle \phi_k^O, X_i^O \rangle_2 \, \langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O}, \quad u \in M, \qquad (3)
\end{aligned}
$$

where $(\phi_k^O, \lambda_k^O)_{k \geq 1}$ denote the pairs of eigenfunctions and nonzero eigenvalues of the covariance operator $\Gamma^O(x)(u) = \int \gamma^O(u,v)x(v)dv$ with $x \in \mathbb{L}^2(O)$. Here $\gamma^O(u,v) = \mathrm{Cov}(X_i^O(u), X_i^O(v))$ denotes the covariance function of X_i^O, and $\gamma_u(v) = \gamma(u,v)$ the covariance function $\gamma(u,v) = \mathrm{Cov}(X_i^M(u), X_i^O(v))$.

The general structure of $\mathcal{L}$ in (3) coincides with the structure of the operators considered in the literature on functional linear regression, which, however, additionally postulates that $\mathcal{L}$ has an (restrictive) integral-representation as in (2); see, for instance, Cardot, Mas and Sarda (2007), Cai and Hall (2006), Hall and Horowitz (2007) in the context of functional linear regression, or Kraus (2015) in a setup similar to ours. There is, though, no reason to expect that the optimal reconstruction operator $\mathcal{L}$ satisfies (2).

To see the point, consider again the reconstruction at a boundary point $\vartheta \in \partial M$, where $\langle \phi_k^O, \gamma_\vartheta \rangle_2$ simplifies to $\langle \phi_k^O, \gamma_\vartheta \rangle_2 = \lambda_k^O \phi_k^O(\vartheta)$, since for boundary points ϑ we have $\gamma_\vartheta = \gamma_\vartheta^O$ and $\gamma_\vartheta^O(.) = \gamma^O(\vartheta, .) = \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(\vartheta)\phi_k^O(.)$. Plugging this simplification into (3) and using the Karhunen-Loéve decomposition of X_i^O implies that $[\mathcal{L}(X_i^O)](\vartheta) = \sum_{k=1}^{\infty} \langle \phi_k^O, X_i^O \rangle_2 \phi_k^O(\vartheta) = X_i^O(\vartheta)$. That is, as required for optimal reconstructions, our reconstruction operator $\mathcal{L}$ connects the "last observed" value $X_i^O(\vartheta)$ with the "first reconstructed" value $[\mathcal{L}(X_i^O)](\vartheta)$. As we argue above, though, this property generally cannot be achieved by a regression operator as defined in (2); therefore, we abstain from imposing such a restrictive assumption.

So far we have considered only boundary points ϑ, though we will show in Section 2.3 that in most practically important situations there is also no reason to expect that regression operators are the optimal reconstruction operators for any non-boundary point $u \in M$ either.

The problem of estimating $\mathcal{L}(X_i^O)$ from real data is considered in Section 3. Motivated by our real data application, our estimation theory allows for an autocorrelated time series of functional data and considers the practically relevant case where the functions X_i^O are only observed at m many

discretization points $(Y_{i1}, U_{i1}), \ldots, (Y_{im}, U_{im})$ with $Y_{ij} = X_i^O(U_{ij}) + \varepsilon_{ij}$, $i = 1, \ldots, n$, and $j = 1, \ldots, m$.

We basically follow the standard approach to estimate $\mathcal{L}(X_i^O)$ through approximating the infinite series (3) by a truncated sequence relying only on the K largest eigenvalues of the covariance operator. But note that our data structure implies that we are faced with two simultaneous estimation problems. One is efficient estimation of $[\mathcal{L}(X_i^O)](u)$ for $u \in M$, the other one is a best possible estimation of the function $X_i(u)$ for $u \in O$ from the observations $(Y_{i1}, U_{i1}), \ldots, (Y_{im}, U_{im})$. We consider two different estimation strategies; both allow us to accomplish these two estimation problems.

The first consists in using a classical functional principal components based approximation of X_i on O, which is simply given by extending the operator $\mathcal{L}$ in (3) by extending $\gamma_u(v) = \mathrm{Cov}(X_i^M(u), X_i^O(u))$ to $\gamma_u(v) = \mathrm{Cov}(X_i(u), X_i(u))$. This way the empirical counterpart of the truncated sum

$$[\mathcal{L}_K(X_i^O)](u) = \sum_{k=1}^{K} \frac{\langle \phi_k^O, X_i^O \rangle_2 \langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O}, \quad \text{for} \quad u \in O \cup M$$

will simultaneously provide estimates of the true function $X_i^O(u)$ on the observed interval O and of the optimal reconstruction $[\mathcal{L}(X_i^O)](u)$ on the unobserved interval M.

The second consists in estimating the true function $X_i^O(u)$ on the observed interval O directly from the observations $(Y_{i1}, U_{i1}), \ldots, (Y_{im}, U_{im})$ using, for instance, a local linear smoother and to estimate $[\mathcal{L}(X_i^O)](u)$ for $u \in M$ through approximating the infinite series (3) by its truncated version. But a simple truncation would result in a jump at a boundary point ϑ_u, with ϑ_u denoting the closest boundary point to the considered $u \in M$, i.e., $\vartheta_u = A_i$ if $|A_i - u| < |B_i - u|$ and $\vartheta_u = B_i$ otherwise. We know, however, that for any $u \approx \vartheta_u$ we must have $\langle \phi_k^O, \gamma_u \rangle_2 \approx \lambda_k^O \phi_k^O(\vartheta)$ for all $k \geq 1$, since $\langle \phi_k^O, \gamma_\vartheta \rangle_2 = \lambda_k^O \phi_k^O(\vartheta)$ for all boundary points $\vartheta_u \in \partial M$. Therefore, we explicitly incorporate boundary points and estimate $\mathcal{L}(X_i^O)$ by the empirical counterpart of the truncated sum

$$[\mathcal{L}_K^*(X_i^O)](u) = X_i^O(\vartheta) + \sum_{k=1}^{K} \langle \phi_k^O, X_i^O \rangle_2 \left(\frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O} - \phi_k^O(\vartheta) \right), \quad u \in M.$$

The above truncation does not lead to an artificial jump at a boundary point ϑ_u, since $(\langle \phi_k^O, \gamma_u \rangle_2 / \lambda_k^O - \phi_k^O(\vartheta_u)) \to 0$ continuously as $u \to \vartheta_u$ for all $k = 1, \ldots, K$.

For estimating the mean and covariance functions—the basic ingredients of our reconstruction operator—we suggest using Local Linear Kernel (LLK)

estimators. These LLK estimators are commonly used in the context of sparse functional data (see, e.g., Yao, Müller and Wang, 2005a), though, we do *not* consider the case of sparse functional data. In the context of partially observed functional data, it is advisable to use LLK estimators, since these will guarantee smooth estimation results, which is not the case when using the empirical moment estimators for partially observed functions as proposed in Kraus (2015).

We derive consistency as well as uniform rates of convergence under a double asymptotic ($n \to \infty, m \to \infty$ and $K \equiv K_{m,n} \to \infty$) which allows us to investigate all data scenarios from almost sparse to dense functional data. This leads to different convergence rates depending on the relative order of m and n. For data situations, as in our real data application where m is considerably smaller than n and the sample curves are very smooth, we show that our functional principal components based estimator achieves almost parametric convergence rates and can provide better rates of convergence than any conventional nonparametric smoothing method.

Our development focuses on the regular situation where (with probability tending to 1) there exist functions that are observed over the total interval $[a, b]$. Only then is it possible to consistently estimate the covariance function $\gamma(u, v)$ for all possible pairs (u, v). In our application this is not completely fulfilled, and there is no information on $\gamma(u, v)$ for very large values $|u - v|$. Consequently, for some intervals O and M the optimal reconstruction operator cannot be identified. This situation corresponds to the case of so-called fragmentary observations, as considered by Delaigle and Hall (2013), Delaigle and Hall (2016), and Descary and Panaretos (2017). To solve this problem we suggest an iterative reconstruction algorithm. Optimal reconstruction operators are determined for a number of smaller subintervals, and a final operator for a larger interval is obtained by successively plugging in the reconstructions computed for the subintervals. We also provide some inequality bounding the accumulating reconstruction error.

The rest of this paper is structured as follows: Section 2 introduces our reconstruction operator and contains the optimality result. Section 3 comprises our estimation procedure. The asymptotic results are presented in Section 4. Section 5 describes the iterative reconstruction algorithm. Section 6 contains the simulation study and Section 7 the real data application. All proofs can be found in the online supplement supporting this article.

2. Optimal reconstruction of partially observed functions. Let our basic setup be as described in Section 1. Any (centered) random function

X_i^O then adopts the well-known Karhunen-Loéve (KL) representation

$$X_i^O(u) = \sum_{k=1}^{\infty} \xi_{ik}^O \phi_k^O(u), \quad u \in O, \tag{4}$$

with the pc-scores $\xi_{ik}^O = \langle X_i^O, \phi_k^O \rangle_2$, where $\mathbb{E}(\xi_{ik}^O) = 0$ and $\mathbb{E}(\xi_{ik}^O \, \xi_{il}^O) = \lambda_k^O$ for all $k = l$ and zero else and $\lambda_1^O > \lambda_2^O > \cdots > 0$.

By the classical eigen-equations we have that

$$\phi_k^O(u) = \frac{\langle \phi_k^O, \gamma_u^O \rangle_2}{\lambda_k^O}, \quad u \in O, \tag{5}$$

where $\gamma_u^O(v) = \gamma^O(u, v) = \mathbb{E}(X_i^O(u) X_i^O(v))$. Equation (5) can obviously be generalized for all $u \in M$ which leads to the following "extrapolated" kth basis function:

$$\tilde{\phi}_k^O(u) = \frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O}, \quad u \in M, \tag{6}$$

where $\gamma_u(v) = \mathbb{E}(X_i^M(u) X_i^O(v))$. Equation (6) leads to the definition of our reconstruction operator $\mathcal{L}_u$ as a generalized version of the KL representation in (4):

$$[\mathcal{L}(X_i^O)](u) = \sum_{k=1}^{\infty} \xi_{ik}^O \, \tilde{\phi}_k^O(u), \quad u \in M. \tag{7}$$

Remark. Note that the KL representation provides the very basis of a majority of the works in functional data analysis (cf. Ramsay and Silverman, 2005; Horváth and Kokoszka, 2012). Functional Principal Component Analysis (FPCA) relies on approximating X_i by its first K principal components. This is justified by the **best basis property**, i.e., for any $K \geq 1$

$$\begin{aligned} \sum_{k=K+1}^{\infty} \lambda_k^O &= \mathbb{E}\left(\|X_i^O(u) - \sum_{k=1}^{K} \xi_{ik}^O \phi_k^O(u)\|_2^2 \right) \\ &= \min_{v_1, \ldots, v_K \in \mathbb{L}^2(O)} \mathbb{E}\left(\min_{a_{i1}, \ldots, a_{ik} \in \mathbb{R}} \|X_i^O(u) - \sum_{k=1}^{K} a_{ik} v_k(u)\|_2^2 \right). \end{aligned} \tag{8}$$

Remark. For later use it is important to note that the definitions of $\tilde{\phi}_k^O(u)$ and $[\mathcal{L}(X_i^O)](u)$ in (6) and (7) can be extended for all $u \in O \cup M$ by setting $\gamma_u = \mathbb{E}(X_i(u) X_i(v))$. Then by construction $\tilde{\phi}_k^O(u) = \phi_k^O(u)$ for all $u \in O$ and, therefore, $[\mathcal{L}(X_i^O)](u) = X_i^O(u)$ for all $u \in O$.

2.1. *A theoretical framework for reconstruction operators.* Before we consider the optimality properties of $\mathcal{L}$, we need to define a sensible class of operators against which to compare our reconstruction operator. We cannot simply choose the usual class of regression operators, since $\mathcal{L}$ does generally not belong to this class, as pointed out in Section 1. Therefore, we introduce the following (very large) class of "reconstruction operators":

DEFINITION 2.1 (Reconstruction operators). *Let the (centered) random function X_i^O have a KL representation as in (4). We call every linear operator $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ a "reconstruction operator with respect to X_i^O" if $\mathbb{V}([L(X_i^O)](u)) < \infty$ for all $u \in M$.*

It is important to note that this definition of "reconstruction operators" is specific to the considered process X_i. This should not be surprising, since a best possible linear reconstitution will of course depend on the structure of the relevant random function X_i. The following theorem provides a useful representation of this class of linear operators:

THEOREM 2.1 (Representation of reconstruction operators). *Let $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ be a "reconstruction operator with respect to X_i^O" according to Definition 2.1. Then there exists a unique (deterministic) parameter function $\alpha_u \in H$ such that almost surely*

$$[L(X_i^O)](u) = \langle \alpha_u, X_i^O \rangle_H, \quad u \in M,$$

where $H := \left\{ f \in \mathbb{L}^2(O) : \ ||f||_H^2 < \infty \right\}$ is a Hilbert space with inner product $\langle f, g \rangle_H := \sum_{k=1}^{\infty} \langle f, \phi_k^O \rangle_2 \, \langle g, \phi_k^O \rangle_2 / \lambda_k^O$ for all $f, g \in \mathbb{L}^2(O)$ and induced norm $||f||_H = \sqrt{\langle f, f \rangle_H}$.

The space H is the Reproducing Kernel Hilbert Space (RKHS) that takes the covariance kernel $\gamma^O(u, v) = \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(u) \phi_k^O(v)$ as its reproducing kernel. By construction, we obtain that the variance of $[L(X_i^O)](u)$ equals the H-norm of the parameter function α_u, i.e., $\mathbb{V}([L(X_i^O)](u)) = ||\alpha_u||_H$.

Let us consider two examples of possible reconstruction operators:

Example 1 - Point of impact: Consider $[L(X_i^O)](u) = X_i^O(\tau)$, i.e., a model with one "impact point" $\tau \in O$ for all missing points $u \in M$. With $\gamma_\tau(v) := \gamma(\tau, v) = \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(\tau) \phi_k^O(v)$ we have $\lambda_k^O \phi_k^O(\tau) = \langle \gamma_\tau, \phi_k^O \rangle_2$, and

hence

$$[L(X_i^O)](u) = X_i^O(\tau) = \sum_{k=1}^{\infty} \xi_{ik}^O \phi_k^O(\tau) = \sum_{k=1}^{\infty} \frac{\langle X_i^O, \phi_k^O \rangle_2 \lambda_k^O \phi_k^O(\tau)}{\lambda_k^O} =$$

$$= \sum_{k=1}^{\infty} \frac{\langle X_i^O, \phi_k^O \rangle_2 \langle \gamma_\tau, \phi_k^O \rangle_2}{\lambda_k^O} = \langle \gamma_\tau, X_i^O \rangle_H, \tag{9}$$

where $\gamma_\tau(\cdot) := \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(\tau) \phi_k^O(\cdot) \in H$ with $||\gamma_\tau||_H^2 = \sum_{k=1}^{\infty} \frac{(\lambda_k^O)^2 \phi_k^O(\tau)^2}{\lambda_k^O} = \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(\tau)^2 = \mathbb{V}(X_i(\tau)) < \infty$.

Example 2 - Regression operator: Let L be a regression operator (see (2)). Then there exists a $\beta_u \in \mathbb{L}^2(O)$ such that $[L(X_i^O)](u) = \langle \beta_u, X_i^O \rangle_2$. Since eigenfunctions can be completed to an orthonormal basis of $\mathbb{L}^2(O)$, we necessarily have that $\sum_{k=1}^{\infty} \beta_{u,k}^2 < \infty$ for $\beta_{u,k} := \langle \beta_u, \phi_k^O \rangle_2$. Then

$$[L(X_i^O)](u) = \langle \beta_u, X_i^O \rangle_2 = \sum_{k=1}^{\infty} \xi_{ik}^O \beta_{u,k} = \sum_{k=1}^{\infty} \frac{\langle X_i^O, \phi_k^O \rangle \lambda_k^O \beta_{u,k}}{\lambda_k^O}$$

$$= \sum_{k=1}^{\infty} \frac{\langle X_i^O, \phi_k^O \rangle \langle \alpha_u, \phi_k^O \rangle}{\lambda_k^O} = \langle \alpha_u, X_i^O \rangle_H, \tag{10}$$

where $\alpha_u(\cdot) := \sum_{j=1}^{\infty} \lambda_k^O \beta_{u,k} \phi_k^O(\cdot) \in H$ with $||\alpha||_H^2 = \beta^2 \sum_{k=1}^{\infty} \frac{\lambda_k^O \beta_{u,k}^2}{\lambda_k^O} = \sum_{k=1}^{\infty} \lambda_k^O \beta_{u,k}^2 < \infty$. Also note that for any k we have $\langle \alpha_u, \phi_k^O \rangle_2 = \lambda_k^O \beta_{u,k}$. This means that for $\alpha_u \in H$ the operator $\langle \alpha_u, X_i^O \rangle_H$ constitutes a regression operator if and only if in addition to $\|\alpha_u\|_H = \sum_{k=1}^{\infty} \langle \alpha_u, \phi_k^O \rangle_2^2 / \lambda_k^O < \infty$ we also have that $\sum_{k=1}^{\infty} \langle \alpha_u, \phi_k^O \rangle_2^2 / (\lambda_k^O)^2 < \infty$ (the latter is not satisfied in Example 1).

These examples show that Definition 2.1 leads to a very large class of linear operators which contains the usually considered class of regression operators as a special case. Of course, the class of reconstruction operators as defined by Definition 2.1 also contains much more complex operators than those illustrated in the examples.

Using Theorem 2.1, our reconstruction problem in (3) of finding a "best linear" reconstruction operator minimizing the squared error loss can now be restated in a theoretically precise manner: Find the linear operator $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ which for all $u \in M$ minimizes

$$\mathbb{E}\left[\left(X_i^M(u) - [L(X_i^O)](u)\right)^2\right]$$

with respect to all reconstruction operators L satisfying $[L(X_i^O)](u) = \langle \alpha_u, X_i^O \rangle_H$ for some $\alpha_u \in H$. In the next subsection we show that the solution is given by the operator $\mathcal{L}$ defined in (7) which can now be rewritten in the form

$$[\mathcal{L}(X_i^O)](u) = \langle \gamma_u, X_i^O \rangle_H, \quad u \in M, \tag{11}$$

where $\gamma_u(v) = \gamma(u, v)$ for $v \in O$ and $u \in M$. In particular, Theorem 2.2 below shows that $\mathbb{V}([\mathcal{L}(X_i^O)](u)) = ||\gamma_u||_H < \infty$ for any $u \in M$, i.e., that $\mathcal{L}$ is indeed a reconstruction operator according to Definition 2.1.

Remark. In the context of reconstructing functions, problems with the use of regression operators are clearly visible. But the above arguments remain valid for standard functional linear regression, where for some real-valued (centered) response variable Y_i with $\mathbb{V}(Y_i) < \infty$ one aims to determine the best linear functional $\tilde{L} : \mathbb{L}^2(O) \to \mathbb{R}$ according to the model $Y_i = \tilde{L}(X_i^O) + \varepsilon_i$. Straightforward generalization of Theorems 2.2 and 2.3 below then shows that the optimal functional $\tilde{\mathcal{L}}(X_i^O)$ is given by

$$\tilde{\mathcal{L}}(X_i^O) = \langle \sigma, X_i^O \rangle_H,$$

where $\sigma(u) := \mathbb{E}(Y_i X_i^O(u))$ for $u \in O$. Following the arguments of Example 2) it is immediately seen that it constitutes a restrictive, additional condition, to assume that $\tilde{\mathcal{L}}(X_i^O)$ can be rewritten in the form $[L(X_i^O)](u) = \langle \beta, X_i^O \rangle_2$ for some $\beta_u \in \mathbb{L}^2(O)$.

2.2. *Theoretical properties.* Result *(a)* of the following theorem assures that $\mathcal{L}$ is a reconstruction operator according to Definition 2.1, and result *(b)* assures unbiasedness.

THEOREM 2.2. *Let the (centered) random function X_i^O have a KL representation as in (4).*

(a) $[\mathcal{L}(X_i^O)](u)$ *in (7) has a continuous and finite variance function, i.e.,* $\mathbb{V}([\mathcal{L}(X_i^O)](u)) < \infty$ *for all* $u \in M$.
(b) $[\mathcal{L}(X_i^O)](u)$ *is unbiased in the sense that* $\mathbb{E}([\mathcal{L}(X_i^O)](u)) = 0$ *for all* $u \in M$.

The following theorem describes the fundamental properties of the reconstruction error

$$\mathcal{Z}_i := X_i^M - \mathcal{L}(X_i^O), \quad \mathcal{Z}_i \in \mathbb{L}^2(M),$$

and contains the optimality result for our reconstruction operator $\mathcal{L}$. Result *(a)* shows that the reconstruction error $\mathcal{Z}_i$ is orthogonal to X_i^O. This result

serves as an auxiliary result for result *(b)* which shows that $\mathcal{L}(X_i^O)$ is the optimal linear reconstruction of X_i^M. Finally, result *(c)* allows us to identify cases where X_i^M can be reconstructed without any reconstruction error.

THEOREM 2.3 (Optimal linear reconstruction). *Under our setup it holds that:*

(a) For every $v \in O$ and $u \in M$,

$$\mathbb{E}\left(X_i^O(v)\mathcal{Z}_i(u)\right) = 0 \quad \textit{and} \tag{12}$$

$$\mathbb{V}(\mathcal{Z}_i(u)) = \mathbb{E}\left((\mathcal{Z}_i(u))^2\right) = \gamma(u,u) - \sum_{k=1}^{\infty} \lambda_k^O \left(\tilde{\phi}_k^O(u)\right)^2. \tag{13}$$

(b) For any linear operator $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ that is a reconstruction operator with respect to X_i^O, according to Definition 2.1,

$$\mathbb{E}\left(\left(X_i^M(u) - [L(X_i^O)](u)\right)^2\right) \geq \mathbb{V}(\mathcal{Z}_i(u)), \quad \textit{for all} \quad u \in M.$$

(c) Assume that the underlying process is Gaussian, and let X_i and X_j, $i \neq j$, be two independent copies of the random variable X_i. Then for all $u \in M$ the variance of the reconstruction error can be written as

$$\mathbb{V}(\mathcal{Z}_i(u)) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left(\left(X_{i,1}(u) - X_{i,2}(u)\right)^2 \middle| X_i^O = X_j^O\right)\right) \tag{14}$$

where $X_i^O = X_j^O$ means equality on O, i.e., $X_i(v) = X_j(v)$ for all $v \in O$.

Whether or not a sensible reconstruction of partially observed functions is possible, of course, depends on the character of the underlying process. For very rough and unstructured processes no satisfactory results can be expected. An example is the standard Brownian motion on $[0,1]$ which is a pure random process with independent increments. If Brownian motions X_i are only observed on an interval $O := [0,\vartheta]$, it is well known that the "best" (i.e., the only unbiased) prediction of $X_i(u)$ for $u \in M := (\vartheta, 1]$ is the last observed value $X_i^O(\vartheta)$. This result is consistent with our definition of an "optimal" operator $\mathcal{L}$: The covariance function of the Brownian motion is given by $\gamma_u(v) = \min(u,v)$, and hence for all $v \in [0,\vartheta]$ and $u \geq \vartheta$ one obtains $\gamma_u(v) = \min(u,v) = \min(\vartheta,v) = \gamma_\vartheta(v) = v$. Therefore, by (11) and (9) we have $[\mathcal{L}(X_i^O)](u) = \langle \gamma_u, X_i^O\rangle_H = \langle \gamma_\vartheta, X_i^O\rangle_H = X_i^O(\vartheta)$ for all $u \in [\vartheta, 1]$. This obviously does not constitute a regression operator. In this paper we focus on processes that lead to smooth, regularly shaped sample curves which may allow for a sensible reconstruction.

Result *(c)* of Theorem 2.3 may be useful to identify cases that allow for a perfect reconstruction. By (14) there is no reconstruction error, i.e., $\mathbb{V}(\mathcal{Z}_i(u)) = 0$ for $u \in M$ if the event $X_i^O = X_j^O$ implies that also $X_i^M = X_j^M$. This might be fulfilled for very simply structured processes. It is necessarily satisfied for finite dimensional random functions $X_i^K(u) = \sum_{k=1}^K \xi_{ik}\phi_k(u)$, $\lambda_{K+1} = \lambda_{K+2} = \cdots = 0$, as long as the basis functions $\phi_1, \ldots, \phi_K$ are linear independent over O.

2.3. *A deeper look at the structure of $\mathcal{L}$.* Remember that the definition of $\mathcal{L}$ can be extended to an operator $\mathcal{L} : \mathbb{L}^2(O) \to \mathbb{L}^2(O \cup M)$. For elements $u \in O$ of the observed part O the best "reconstruction" of $X_i(u)$ is obviously the observed value $X_i(u)$ itself, and indeed for any $u \in O$ (11) yields $[\mathcal{L}(X_i^O)](u) = \langle \gamma_u, X_i^O \rangle_H = X_i(u)$. Equation (7) then holds with

$$\tilde{\phi}_k^O(u) := \frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O} = \phi_k^O(u), \quad u \in O.$$

Since $\gamma_u(v) = \gamma(u,v) = \mathbb{E}(X_i(u)X_i(v))$ is a continuous function on $O \cup M$ it follows that the resulting "reconstructed" function $[\mathcal{L}(X_i^O)]$ is continuous on $O \cup M$. In particular, $[\mathcal{L}(X_i^O)]$ is continuous at any boundary point $\vartheta_u \in \partial M$, and

$$\lim_{u \in M, u \to \vartheta_u} [\mathcal{L}(X_i^O)](u) = X_i(\vartheta_u), \text{ as well as}$$
$$\lim_{u \in M, u \to \vartheta_u} \tilde{\phi}_k^O(u) = \phi_k^O(\vartheta_u), \ k = 1, 2, \ldots$$

Equation (7) together with our definition of $\mathcal{Z}_i$ imply that the complete function X_i on $O \cup M$ can be represented in the form

(15)
$$X_i(v) = \sum_{k=1}^{\infty} \xi_{ik}^O \phi_k^O(v), v \in O, \text{ and } X_i(u) = \sum_{k=1}^{\infty} \xi_{ik}^O \tilde{\phi}_k^O(u) + \mathcal{Z}_i(u), u \in M.$$

This sheds some additional light on result (14). We will have $\mathcal{Z}_i(u) \approx 0$ and $X_i(u) \approx \sum_{k=1}^{\infty} \xi_{ik}^O \tilde{\phi}_k^O(u)$ if on the segment M the process is *essentially* driven by the same random components ξ_{ik}^O as those determining its structure on O. Additional random components $\mathcal{Z}_i(u)$, not present on O, and uncorrelated with ξ_{ik}^O, then have to be of minor importance. If the observed interval is sufficiently long, then this may be approximately true for processes with smooth, similarly shaped trajectories. Note that even if $X_i(u) = \sum_{k=1}^{\infty} \xi_{ik}^O \tilde{\phi}_k^O(u)$, the eigenfunctions of X_i^M will usually not coincide with $\tilde{\phi}_k^O$ since there is no reason to expect that these functions be mutually orthogonal.

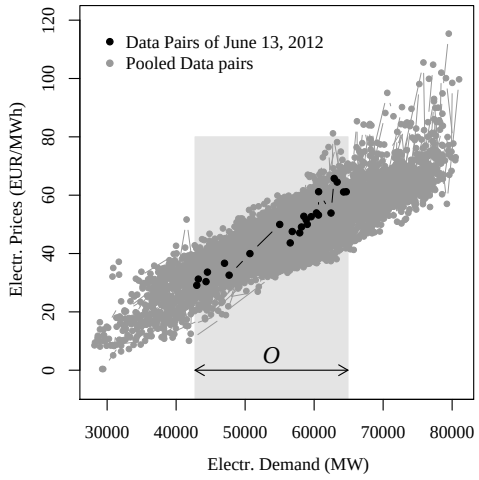

FIG 2. *Scatter plot of the observed data pairs* (Y_{ij}, U_{ij}).

3. Estimation. We typically do not observe a functional trajectory directly, but only its discretization with or without measurement errors. For instance, Figure 1 shows the *pre-smoothed* functions; however, the actual raw data is shown in Figure 2. Let $\mathbb{X}_i^O := ((Y_{i1}, U_{i1}), \ldots, (Y_{im}, U_{im}))$ denote the observable data pairs of a function X_i^O, where

$$Y_{ij} = X_i^O(U_{ij}) + \varepsilon_{ij}, \quad i \in \{1, \ldots, n\}, \quad j \in \{1, \ldots, m\}, \tag{16}$$

and $U_{ij} \in O_i = [A_i, B_i]$. We consider the case where $U_{i1}, \ldots, U_{im}$ are iid random variables with strictly positive density over the random subinterval $[A_i, B_i]$, which in practice can be approximated by $A_i \approx \min_{1 \le j \le m}(U_{ij})$ and $B_i \approx \max_{1 \le j \le m}(U_{ij})$. Let the error term ε_{ij} be a real iid random variable with mean zero and finite variance $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$, with $0 < \sigma^2 < \infty$. Motivated by our real data application we will concentrate on the case that n is considerably larger than m, which also holds in many other important applications.

So far, we have considered centered random functions X_i^O. Henceforth, we consider **non-centered** functions and will make the empirical centering explicit in all estimators. As already outlined in Section 1, we propose to estimate $[\mathcal{L}(X_i^O)](u)$ by the empirical counterpart of the truncated sum

$$[\mathcal{L}_K(X_i^O)](u) = \mu(u) + \sum_{k=1}^{K} \xi_{ik}^O \tilde{\phi}_k^O(u) = \mu(u) + \sum_{k=1}^{K} \xi_{ik}^O \frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O}, \tag{17}$$

where the unknown true values of ξ_{ik}^O and $\tilde{\phi}_k^O(u)$ are replaced by suitable estimates defined below.

Note, however, that our data structure (16) implies that we are faced with two simultaneous estimation problems. One is efficient estimation of $[\mathcal{L}(X_i^O)](u)$ for $u \in M$, the other one is a best possible estimation of the underlying function $X_i(u)$ for $u \in O$. There are two possible strategies which can be employed.

The first is motivated by the best basis property (8) and simply consists in using an FPCA-approximation of X_i on O. Recall that $[\mathcal{L}(X_i^O)](u)$ can be extended to an operator on $O \cup M$. For $u \in O$ we then obtain $\tilde{\phi}_k^O(u) = \langle \phi_k^O, \gamma_u \rangle_2 / \lambda_k^O = \phi_k^O(u)$, and thus $[\mathcal{L}(X_i^O)](u) = X_i(u)$. That is, estimates $[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u)$ of $[\mathcal{L}_K(X_i^O)](u)$ for $u \in O \cup M$ will simultaneously provide estimates of the true function $X_i(u)$ on the observed interval O and of the optimal reconstruction $[\mathcal{L}(X_i^O)](u)$ on the unobserved interval M.

The second approach is to rely on nonparametric curve estimation, e.g., local linear smoothers, to approximate X_i^O on O, while (17) is only used for reconstructing the unobserved part M. We then, however, run into the boundary problem already mentioned in the introduction. Let ϑ_u be the boundary point closest to the considered $u \in M$, i.e., $\vartheta_u = A_i$ if $|A_i - u| < |B_i - u|$ and $\vartheta_u = B_i$ else. Usually nonparametric estimates of X_i^O and reconstruction estimates based on (17) will not coincide for $u = A_i$ or $u = B_i$. A correction, leading to continuous function estimates on $O \cup M$ may then be based on the identity $[\mathcal{L}(X_i^O)](u) = X_i^O(\vartheta_u) + [\mathcal{L}(X_i^O)](u) - [\mathcal{L}(X_i^O)](\vartheta_u)$ and its truncated version

$$
\begin{aligned}
[\mathcal{L}_K^*(X_i^O)](u) &= X_i^O(\vartheta_u) + [\mathcal{L}_K(X_i^O)](u) - [\mathcal{L}_K(X_i^O)](u) \\
&= \mu(u) - \mu(\vartheta_u) \sum_{k=1}^{K} \xi_{ik}^O (\tilde{\phi}_k^O(u) - \tilde{\phi}_k^O(\vartheta_u), \quad \text{for } u \in M
\end{aligned} \tag{18}
$$

In this paper we propose to use the following empirical counterparts of $[\mathcal{L}_K(X_i^O)](u)$ or $[\mathcal{L}_K^*(X_i^O)](u)$:

$$
[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) := \hat{\mu}(u; h_\mu) + \sum_{k=1}^{K} \hat{\xi}_{ik}^O \hat{\tilde{\phi}}_k^O(u) \quad \text{for } u \in O \cup M, \tag{19}
$$

$$
\text{where } \hat{\tilde{\phi}}_k^O(u) := \frac{\langle \hat{\phi}_k^O, \hat{\gamma}_u \rangle_2}{\hat{\lambda}_k^O}, \quad k = 1, \ldots, K,
$$

$$[\widehat{\mathcal{L}}_K^*(\mathbb{X}_i^O)](u) := \widehat{X}_i^O(\vartheta; h_X) + [\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) - [\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](\vartheta)$$

$$= \widehat{X}_i^O(\vartheta_u; h_X) + \hat{\mu}(u; h_\mu) - \hat{\mu}(\vartheta_u; h_\mu) + \sum_{k=1}^{K} \hat{\xi}_{ik}^O \left(\hat{\tilde{\phi}}_k^O(u) - \hat{\phi}_k^O(\vartheta_u)\right). \tag{20}$$

Here, for $u \in O$ the LLK estimator $\widehat{X}_i^O(u; h_X)$ is defined by $\widehat{X}_i^O(u; h_X) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\arg\min} \sum_{j=1}^{m} [Y_{ij} - \beta_0 - \beta_1(U_{ij} - u)]^2 K_{h_X}(U_{ij} - u) \tag{21}$$

for $K_h(.) = \kappa(./h)/h$. The kernel function κ is assumed to be a univariate symmetric pdf with compact support $\text{supp}(\kappa) = [-1, 1]$ such as, e.g., the Epanechnikov kernel (see Assumption A5). The usual kernel constants are given by $\nu_2(\kappa) := \int v^2 \kappa(v) dv$, and $R(\kappa) := \int \kappa(v)^2 dv$.

The LLK mean estimator $\hat{\mu}(u; h_\mu)$ is defined by $\hat{\mu}(u; h_\mu) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\arg\min} \sum_{i=1}^{n} \sum_{j=1}^{m} [Y_{ij} - \beta_0 - \beta_1(U_{ij} - u)]^2 K_{h_\mu}(U_{ij} - u). \tag{22}$$

The estimator $\hat{\gamma}_u(v) = \hat{\gamma}(u, v)$ is defined as follows. Due to the symmetry of the covariance function, $\gamma(u, v) = \gamma(v, u)$, it suffices to estimate only the upper-diagonal part of $\gamma(u, v)$ with $u \leq v$. For all upper-diagonal points $(u, v) \in [a, b]^2$ with $u \leq v$, the LLK estimator $\hat{\gamma}(u, v; h_\gamma)$ is defined as $\hat{\gamma}(u, v; h_\gamma) = \hat{\beta}_0$, where

$$\begin{aligned}(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) =& \underset{\beta_0, \beta_1, \beta_2}{\arg\min} \sum_{i=1}^{n} \sum_{\substack{1 \leq j, l \leq m \\ U_{ij} < U_{jl}}} [\widehat{C}_{ijl} - \beta_0 - \beta_1(U_{ij} - u) - \beta_2(U_{il} - v)]^2 \\ & \times K_{h_\gamma}(U_{ij} - u) K_{h_\gamma}(U_{il} - v),\end{aligned} \tag{23}$$

with raw-covariance points $\widehat{C}_{ijl}$ defined as

$$\widehat{C}_{ijl} := (Y_{ij} - \hat{\mu}(U_{ij}))(Y_{il} - \hat{\mu}(U_{il})), \quad \text{for} \quad U_{ij} < U_{il}. \tag{24}$$

For estimating the lower-diagonal of $\gamma(u, v)$ with $u \geq v$ set $\hat{\gamma}(u, v; h_\gamma) = \hat{\gamma}(v, u; h_\gamma)$.

Like Yao, Müller and Wang (2005a), we do not include the diagonal raw-covariances $\widehat{C}_{ijj}$ for which $U_{ij} = U_{ij}$ as these would introduce an estimation

bias through taking squares of the error term ε_{ij} contained in Y_{ij}. Our LLK estimator $\hat{\gamma}(u,v)$ differs from that in Yao, Müller and Wang (2005a), since it avoids smoothing across the diagonal. This allows us to estimate also covariance functions that are not differentiable at the diagonal such as, for instance, the covariance function of a Brownian motion or many other stochastic processes.

Remark. Using only the upper-diagonal raw-covariance points does not mean that we are using less information, since the upper- and lower-diagonal raw-covariance points are only the mirrored points to each other. Yao, Müller and Wang (2005a) assume a more restrictive situation where all second order derivatives at the diagonal $\gamma(u,u)$ are continuous. Our approach is consistent there as well, without any loss in terms of convergence rates.

Estimates of the eigenvalues λ_k^O and the eigenfunctions ϕ_k^O are defined by the corresponding solutions of the empirical eigen-equations

$$\int_O \hat{\gamma}(u,v;h_\gamma)\hat{\phi}_k^O(v)\,dv = \hat{\lambda}_k^O\,\hat{\phi}_k^O(u), \quad u \in O. \tag{25}$$

Finally, the empirical pc-score $\hat{\xi}_{ik}^O$ is defined by the following integral approximation of ξ_{ik}^O:

$$\hat{\xi}_{ik}^O = \sum_{j=2}^{m} \hat{\phi}_k^O(U_{i(j)})(Y_{i(j)} - \hat{\mu}(U_{i(j)};h_\mu))(U_{i(j)} - U_{i,(j-1)}), \tag{26}$$

where $(Y_{i(j)}, U_{i(j)})$ are ordered data pairs for which the ordering is determined through the order sample $U_{i(1)} \leq \cdots \leq U_{i(m)}$

In our theoretical analysis we consider $K \equiv K_{nm} \to \infty$ as the sample size $nm \to \infty$. In practice, the truncation parameter K can be chosen by one of the usual procedures such as, for instance, Cross Validation or the Fraction of Variance Explained (FVE) criterion.

Alternatively, one can use an adapted version of the AIC-type criterion in Yao, Müller and Wang (2005a). For this let $(Y_{i(1)}, U_{i(1)}), \ldots, (Y_{i(m)}, U_{i(m)})$ denote the ordered data as in (26) and define the following quantities by partitioning the data into pseudo-missing (PM) and pseudo-observed (PO) halves:

$$\begin{aligned}
\mathbf{Y}_i^{PM} &= \big(Y_{i(1)}, \ldots, Y_{i(\lfloor m/2 \rfloor)}\big)^\top \\
\mathbb{X}_i^{PO} &= \big((Y_{i(\lfloor m/2 \rfloor+1)}, U_{i(\lfloor m/2 \rfloor+1)}), \ldots, (Y_{i(m)}, U_{i(m)})\big)^\top \\
\widehat{\mathbf{Y}}_{iK}^{PM} &= \big([\widehat{\mathcal{L}}_K(\mathbb{X}_i^{PO})](U_{i(\lfloor m/2 \rfloor+1)}), \ldots, [\widehat{\mathcal{L}}_K(\mathbb{X}_i^{PO})](U_{i(m)})\big)^\top
\end{aligned}$$

The information criterion is then defined as $\mathrm{AIC}_K = -\widehat{L}_K + K$ with

$$\widehat{L}_K = \sum_{i=1}^{n} \Big\{ -\frac{\lfloor m/2 \rfloor}{2} \log(2\pi) - \frac{\lfloor m/2 \rfloor}{2} \log \hat{\sigma}^2 - \frac{1}{2\hat{\sigma}^2} \big(\mathbf{Y}_i^{PM} - \widehat{\mathbf{Y}}_{iK}^{PM}\big)^\top \big(\mathbf{Y}_i^{PM} - \widehat{\mathbf{Y}}_{iK}^{PM}\big) \Big\}, \tag{27}$$

where $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \hat{\sigma}_i^2$ and $\hat{\sigma}_i^2$ is a nonparametric variance estimator such as, for instance, that of Rice (1984), i.e., $\hat{\sigma}_i^2 = (2(m-1))^{-1} \sum_{j=1}^{m-1} (Y_{i(j)} - Y_{i(j-1)})^2$.

4. Asymptotic results. Our theoretical analysis analyzes the reconstruction of an arbitrary sample function X_i satisfying $O \subset O_i$.

Our asymptotic results on the convergence of our nonparametric estimators are developed under the following assumptions which are generally close to those in Yao, Müller and Wang (2005b) and Hall, Müller and Wang (2006). We additionally allow for weakly dependent time series of random functions $(X_i)_i$, and we consider a different asymptotic setup excluding the case of sparse functional data. Only second-order kernels are employed.

A1 (Stochastic) For some $d_{\min} > 0$ the conditional random variables $U_{i1}|O_i, \ldots, U_{im}|O_i$ are iid with pdf $f_{U|O_i}(u) \geq d_{\min}$ for all $u \in O_i = [A_i, B_i]$ and zero else. For the marginal pdf f_U it is assumed that $f_U(u) > 0$ for all $u = [a, b]$ and zero else. The time series $(A_i)_{i=1,\ldots,n}$, $(B_i)_{i=1,\ldots,n}$, and $(X_i)_{i=1,\ldots,n}$ is a strictly stationary ergodic functional time series with finite fourth moments, i.e., $\mathbb{E}(||X_i||_2^4) < \infty$ and autocovariance functions with geometric decay. I.e., there are constants $C_A, C_B, C, \dot{C}, \iota_A, \iota_B, \iota, \dot{\iota}$ with $0 < C_A, C_B, C, \dot{C} < \infty$ and $0 < \iota_A, \iota_B, \iota, \dot{\iota} < 1$, such that $|\operatorname{Cov}(A_i, B_{i+h})| \leq C_A \iota_A^h$, $|\operatorname{Cov}(B_i, B_{i+h})| \leq C_B \iota_B^h$, $\sup_{(u,v)\in[a,b]^2} |\gamma_h(u,v)| \leq C\iota^h$, and $\sup_{(u_1,v_1,u_2,v_2)\in[a,b]^4} |\dot{\gamma}_h((u_1,v_1),(u_2,v_2))| \leq \dot{C}\dot{\iota}^h$ for all $h \geq 0$, where $\gamma_h(u,v) := \operatorname{Cov}(X_{i+h}(u), X_i(v))$ and $\dot{\gamma}_h((u_1,v_1),(u_2,v_2)) := \operatorname{Cov}(X_{i+h}(u_1)X_{i+h}(v_1), X_i(u_2)X_i(v_2))$. The error term ε_{ij} is assumed to be independent from all other random variables. The random variables U_{ij} and O_i are assumed to be independent from $(X_i)_{i=1,\ldots,n}$, which leads to the so-called "missing completely at random" assumption. The event $O_i \times O_i = [a,b]^2$ has a strictly positive probability and $B_i > A_i$ almost surely.

A2 (Asymptotic scenario) $nm \to \infty$, where $n \to \infty$ and $m = m_n \asymp n^\theta$ with $0 < \theta < \infty$. Here, $a_n \asymp b_n$ is used to denote that $\lim_{n\to\infty}(a_n/b_n) = c$ for some constant $0 < c < \infty$. It is assumed that $m_n \geq 2$ for all n.

A3 (Smoothness) For $\hat{\mu}$: All second order derivatives of $\mu(u)$ on $[a,b]$, $f_U(u)$ on $[a,b]$, $\gamma(u,v)$ on $[a,b]^2$, and of $f_{YU}(y,u)$ on $\mathbb{R}\times[a,b]$ are uniformly continuous and bounded, where f_{YU} is the joint pdf of (Y_{ij},U_{ij}), and where $\text{supp}(f_U)=[a,b]$. For $\hat{\gamma}$: All second order derivatives of $\gamma(u,v)$ on $\{(u,v)\in[a,b]^2 : u\neq v\}$, $f_{UU}(u,v)$ on $[a,b]^2$, $\dot{\gamma}((u_1,v_1),(u_2,v_2))$ on $[a,b]^4$, and of $f_{CUU}(c,u,v)$ on $\mathbb{R}\times[a,b]^2$ are uniformly continuous and bounded, where f_{CUU} is the joint pdf of (C_{ijl},U_{ij},U_{il}), and where $\text{supp}(f_{UU})=[a,b]^2$. Finally, all $f_{U|O_i}(u)$ are a.s. continuously differentiable, and $\mathbb{E}\left(\frac{|f'_{U|O_i}(u)|}{f_{U|O_i}(u)^2}\right)<\infty$.

A4 (Bandwidths) For estimating X_i^O: $h_X\equiv h_{m,X}\to 0$ and $(m\,h_X)^{-1/2}\to\infty$ as $m\to\infty$. For estimating μ: $h_\mu\equiv h_{nm,\mu}\to 0$ and $(nm\,h_\mu)^{-1/2}\to\infty$ as $nm\to\infty$. For estimating γ: $h_\gamma\equiv h_{nM,\gamma}\to 0$ and $(nM\,h_\gamma)^{-1/2}\to\infty$ as $nM\to\infty$, where $M=(m^2-m)/2$.

A5 (Kernel function) κ is a second-order kernel with compact support $\text{supp}(\kappa)=[-1,1]$.

We follow Hall, Müller and Wang (2006) and consider a deterministic sample size m; however, in case of a random m our results apply conditionally on m. In the case of different values m_i for each $i=1,\ldots,n$, our results can be adapted using the "OBS" and "SUBJ" weighting schemes proposed by Zhang and Wang (2016).

While (A1) - (A5) suffice to determine rates of convergence of mean and covariance estimators, it is well-known from the literature that rates of convergence of estimated eigenfunctions will depend on the rate of decay characterizing the convergence of λ_k^O to zero as $k\to\infty$.

We want to note that for a subinterval $O\subset[a,b]$ the decay of eigenvalues $\lambda_1^O,\lambda_2^O,\ldots$ will usually be faster than the rate of decay of the eigenvalues $\lambda_1^C,\lambda_2^C,\ldots$ of the complete covariance operator defined on $[a,b]^2\supset O^2$. This is easily seen. Let $\gamma_1^C,\gamma_2^C,\ldots$ denote the corresponding eigenfunctions on $[a,b]$, and define $\gamma_k^{C|O}\in\mathbb{L}^2(O)$ by $\gamma_k^{C|O}(u)=\gamma_k(u)$ for $u\in O$ and $k=1,2,\ldots$. For the special case $v_k=\gamma_k^{C|O}$, $k=1,\ldots,K$, inequality (8) then implies that for all $K\geq 1$ we have $\sum_{k=K+1}^\infty\lambda_k^O\leq\sum_{k=K+1}^\infty\lambda_k^C\int_O\gamma_k^{C|O}(u)^2du\leq\sum_{k=K+1}^\infty\lambda_k^C$, since $\int_O\gamma_k^{C|O}(u)^2du\leq\int_a^b\gamma_k^C(u)^2du=1$ for all $k=1,2,\ldots$.

To complete our asymptotic setup, we consider reconstruction of arbitrary sample functions X_i observed over an interval $O_i=[A_i,B_i]$ with length $B_i-A_i\geq\ell_{\min}$, where $0<\ell_{\min}<b-a$ is an (arbitrary) constant. We then impose the following additional assumptions.

A6 (Eigenvalues) For any subinterval $O=[A,B]\subset[a,b]$ with $B-A\geq\ell_{\min}$ the ordered eigenvalues $\lambda_1^O>\lambda_2^O>\cdots>0$ have all multiplicity one. Fur-

thermore, there exist some $a_O > 1$ and some $0 < c_O < \infty$, possibly depending on O, such that $\lambda_k^O - \lambda_{k+1}^O \geq c_O k^{-a_O - 1}$ with $0 < c_O < \infty$, and $\lambda_k^O = \mathcal{O}(k^{-a_O})$ as well as $1/\lambda_k^O = O(k^{a_O})$ as $k \to \infty$.

A7 (Eigenfunctions) For any subinterval $O = [A, B] \subset [a, b]$ with $B - A \geq \ell_{\min}$ there exists a constant $0 < D_O < \infty$ such that $\sup_{u \in [a,b]} \sup_{k \geq 1} |\tilde{\phi}_k^O(u)| \leq D_O$ (recall that $\tilde{\phi}_k^O(U) = \phi_k^O(u)$ for $u \in O$).

Assumption (A6) requires a polynomial decay of the sequence of eigenvalues. It cannot be tested, but it corresponds to the usual assumption characterizing a majority of work concerning eigenanalysis of functional data, although some authors also consider exponential decays. There exist various types of functional data, but this paper focuses on applications where the true sample functions are very smooth and all possess similar functional structure. This is quite frequent in practice, and in applied papers it is then often found that few functional principal components suffice to approximate sample functions with high accuracy. In view of the best basis property (8) one may then tend to assume that (A6) holds for some very large $a_O \gg 1$. Indeed, for increasing k eigenfunctions ϕ_k^O will become less and less "smooth" since the number of sign changes will necessarily tend to infinity. If observed trajectories are very smooth, then the influence of such high-frequency components must be very small, indicating a very small eigenvalue $\lambda_k^O = \mathbb{E}(\xi_k^O)$ for large k. This is of substantial interest, since the theorems below show that rates of convergence of our final estimators are better the larger a_O.

Assumption (A7) imposes a (typical) regularity condition on the structure of the eigenfunctions $\phi_k^O(u)$, since $\tilde{\phi}_k^O(u) = \phi_k^O(u)$ for $u \in O$. For $u \in M = [a, b] \setminus O$ condition $|\tilde{\phi}_k^O(u)| \leq D_O$ is much weaker than the standard assumption of a regression operator which would go along with the requirement $\sum_{k=1}^{\infty} \tilde{\phi}_k^O(u)^2 < \infty$. But, for $u \in M$, theory only ensures that $\sum_{k=1}^{\infty} \lambda_k (\tilde{\phi}_k^O(u))^2 < \infty$ (see Theorem 2.3 (a)) and (A7) is restrictive in so far as it excludes the possible case that for $u \in M$ we have $|\tilde{\phi}_k^O(u)| \to \infty$ as $k \to \infty$. We are not sure whether the latter excluded case constitutes a realistic scenario in practical applications, since by (15) it would correspond to the fairly odd situation that for large k the high-frequency components ξ_{ik}^O possess much larger influence on M than on O. Nevertheless, we want to emphasize that the arguments used in the proof of our theorems may easily be generalized to prove consistency of our estimators even in this excluded case; however, rates of convergence deteriorate and asymptotic expressions become much more involved.

THEOREM 4.1 (Preliminary consistency results).
Under Assumptions (A1)-(A5) we have that:

(a) $\sup_{u\in[a,b]} |\hat{\mu}(u;h_\mu) - \mu(u)| = \mathcal{O}_p\left(r_\mu\right)$
(ã) *Conditional on* X_i^O*:* $\sup_{u\in O} |\widehat{X}_i^O(u;h_\mu,h_X) - X_i^O(u)| = \mathcal{O}_p\left(r_X\right)$
(b) $\sup_{(u,v)\in[a,b]^2} |\hat{\gamma}(u,v;h_\gamma) - \gamma(u,v)| = \mathcal{O}_p\left(r_\mu + r_\gamma\right)$, *where*

$$\begin{aligned} r_\mu &\equiv r_\mu(h_\mu,n,m) := h_\mu^2 + 1/\sqrt{nm\,h_\mu} + 1/\sqrt{n} \\ r_X &\equiv r_X(h_X,m) := h_X^2 + 1/\sqrt{m\,h_X} \\ r_\gamma &\equiv r_\gamma(h_\gamma,n,M) := h_\gamma^2 + 1/\sqrt{nM\,h_\gamma^2} + 1/\sqrt{n}, \end{aligned}$$

and where $(\tilde{a})$ *is effectively only a corollary of result* (a).
If additionally Assumption (A6) and (A7) hold, we obtain for every subinterval $O = [A,B] \subset [a,b]$ *with* $B - A \geq \ell_{\min}$*:*

(c) $\sup_{k\geq 1} |\hat{\lambda}_k^O - \lambda_k^O| = \mathcal{O}_p\left(r_\mu + r_\gamma\right)$ *for all* $k \geq 1$
(d) $\sup_{1\leq k\leq K} \delta_k^O \|\hat{c}_k\hat{\phi}_k^O - \phi_k^O\|_2 = \mathcal{O}_p\left((r_\mu + r_\gamma)\right)$

where $\hat{c}_k := \text{sgn}(\langle\hat{\phi}_k^O, \phi_k^O\rangle_2)$ *and* $\delta_k^O := \min_{j\neq k}\{\lambda_j^O - \lambda_k^O\}$.

Related results can be found in Yao, Müller and Wang (2005a), Li and Hsing (2010), and Zhang and Wang (2016). Our proof of results (a)-(b) follows that of Yao, Müller and Wang (2005a), but is more restrictive as we allow only for compact second order kernels. Results (c) and (d) follow from standard arguments as used in Bosq (2000).

THEOREM 4.2 (Consistency results for $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)$). *Consider an arbitrary* $i \in 1,\ldots,n$ *and assume that* $O := O_i = [A_i,B_i]$ *satisfies* $B_i - A_i \geq \ell_{\min}$.
For some $0 < C < \infty$ *let* $\bar{K}_{mn} = C \cdot (\min\{n^{1/2},(nM)^{1/3}\})^{1/(a_O+3/2)}$.
The following results hold under Assumptions A1-A7, for $1 \leq K \leq \bar{K}_{mn}$, $h_X \asymp m^{-1/5}$, $h_\mu \asymp (nm)^{-1/5}$ *and* $h_\gamma \asymp (nM)^{-1/6}$, *as* $n \to \infty$ *and* $m \to \infty$ *with* $m \asymp n^\theta$, $0 < \theta < \infty$. *For any* $u \in [a,b]$*:*

$$[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) = [\mathcal{L}_K(X_i^O)](u) + \mathcal{O}_p\left(K\left(\frac{1}{m^{1/2}} + \frac{K^{a_O/2+3/2}}{\min\{n^{1/2},(nM)^{1/3}\}}\right)\right)$$

(28)

$$[\mathcal{L}(X_i^O)](u) - [\mathcal{L}_K(X_i^O)](u) = \mathcal{O}\left(\left(\sum_{k=K+1}^{\infty}\lambda_k^O\right)^{1/2}\right) = \mathcal{O}_p\left(K^{-(a_O-1)/2}\right)$$

Furthermore, for all $u \in M := [a,b] \setminus O$

$$[\widehat{\mathcal{L}}^*_K(\mathbb{X}^O_i)](u) = [\mathcal{L}^*_K(X^O_i)](u) + \mathcal{O}_p\left(m^{-2/5} + K\left(\frac{1}{m^{1/2}} + \frac{K^{a_O/2+3/2}}{\min\{n^{1/2},(nM)^{1/3}\}}\right)\right)$$

(29)

$$[\mathcal{L}(X^O_i)](u) - [\mathcal{L}^*_K(X^O_i)](u) = \mathcal{O}\left(\left(\sum_{k=K+1}^{\infty}\lambda^O_k\right)^{1/2}\right) = \mathcal{O}_p\left(K^{-(a_O-1)/2}\right)$$

The theorem tell us that for any $u \in [a,b]$ the estimator $[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)$ achieves the same rate of convergence. But recall that for $u \in O = O_i$ we have $[\mathcal{L}(X^O_i)](u) = X_i(u)$, and thus $[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)$ can be seen as a nonparametric estimator of X_i. In contrast, for $u \in M$ we have $[\mathcal{L}(X^O_i)](u) = X_i(u) + \mathcal{Z}_i(u)$, and therefore the distance between $X_i(u)$ and $[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)$ will additionally depend on the reconstruction error $\mathcal{Z}_i(u)$.

Note that by the above result the rates of convergence depend on m and n, and the optimal K depends on these quantities in a complex way. However, the situation simplifies if m is considerably smaller than n such that $m = m_n \asymp n^\theta$ for $\theta \le 1/2$. The following corollary then is a direct consequence of (28).

COROLLARY 4.1. *Under the conditions of Theorem 4.2 additionally assume that* $\theta \le 1/2$. *With* $K \equiv K_m \asymp m^{1/(a_O+2)}$ *we obtain for all* $u \in [a,b]$

$$|[\mathcal{L}(X^O_i)](u)) - [\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)| = \mathcal{O}_p\left(m^{-\frac{a_O-1}{2(a_O+2)}}\right) \tag{30}$$

Result (30) implies that if the true sample curves are very smooth in the sense that (A6) holds for a very large $a_O \gg 1$, then the rate of convergence of the individual curve estimators $[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)$ is very close to the parametric rate $m^{-1/2}$.

Recall that that the main difference between $\widehat{\mathcal{L}}_K$ and $\widehat{\mathcal{L}}^*_K$ consists in the way of estimating X_i on the observed interval O_i. $\widehat{\mathcal{L}}^*_K$ is based on local linear smoothing, and the associated estimation error of order $m^{-2/5}$ appears in result (29). When relying solely on $[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)|$, we approximate $X_i(u)$ for $u \in O_i$ by an FPCA-approximation, as used in hundreds of papers for different purposes. We can infer from (30) that at least under restrictive conditions (m considerably smaller than n; very large $a_O \gg 1$) FPCA-approximations of $X_i(u)$, $u \in O_i$, may provide better rates of convergence than any conventional smoothing method. We believe that this is an interesting result in

its own right, which to our knowledge has not yet been established in the literature.

5. Iterative reconstruction algorithm. So far we have focused on the regular situation where the covariance function $\gamma(u,v)$ is estimable for all points $(u,v) \in [a,b]^2$. Under this situation we can reconstruct the entire missing parts of the functions, such that the reconstructed functions $\tilde{X}_i$ with

$$\tilde{X}_i(u) = \begin{cases} [\mathcal{L}(X_i^O)](u) & \text{if} \quad u \in M \\ X_i^O(u) & \text{if} \quad u \in O \end{cases} \tag{31}$$

are identifiable for *all* $u \in [a,b]$.

In our application, however, we face the more restrictive situation where the mean function $\mu(u)$ can still be estimated for all $u \in [a,b]$, but where there is no information on $\gamma(u,v)$ for large values $|u-v|$; see Figure 6. This makes it impossible to reconstruct the entire missing part of a function, such that $\tilde{X}_i(u)$ cannot be identified for all $u \in [a,b]$.

In order to reconstruct functions $\tilde{X}_i$ that cover the total interval $[a,b]$, or a very large part of it, we propose successively plugging in the optimal reconstructions computed for subintervals. In the following we describe our iterative reconstruction algorithm:

ALGORITHM 5.1 (Iterative reconstruction algorithm).

1st Step *Denote the originally observed interval* O *as* O_1 *and compute*

$$\tilde{X}_{i,1}(u) = \begin{cases} [\mathcal{L}(X_i^{O_1})](u) & \textit{if} \quad u \in M_1 \\ X_i^{O_1}(u) & \textit{if} \quad u \in O_1 \end{cases}$$

rth Step $(r \geq 2)$ *Choose a new "observed" interval* $O_r \subset O_{r-1} \cup M_{r-1}$ *and use* $\tilde{X}_i^{O_r}(u) := \tilde{X}_{i,r-1}(u)$ *with* $u \in O_r$ *as the new "observed" fragment. Compute*

$$\tilde{X}_{i,r}(u) = \begin{cases} [\mathcal{L}(\tilde{X}_i^{O_r})](u) & \textit{if} \quad u \in M_r \\ \tilde{X}_i^{O_r}(u) & \textit{if} \quad u \in O_r. \end{cases}$$

Join the reconstructed fragments $\tilde{X}_{i,1}, \ldots, \tilde{X}_{i,r}$ *to form the new "observed" fragment* $\tilde{X}_{i,r-1}$ *on* $O_{r-1} \cup M_{r-1}$ *and repeat the* r*th step.*

Stopping *Stop if* $\bigcup_{l=1}^{r} O_l \cup M_l = [a,b]$ *or if* $r = r_{\max}$.

This algorithm has to be applied to every fragment X_i^O. An exemplary 1st Step of the reconstruction algorithm is shown in Figure 3. The subinterval

$O_1 \cup M_1$ is determined by the original interval O_1 and the extend to which γ can be estimated (see right panel). The function $\tilde{X}_{i,1}$ shown in the left panel still lacks the upper fragment for values $u \in [77362\ (\text{MW}), 82282(\text{MW})]$ such that a second step of the reconstruction algorithm is necessary.

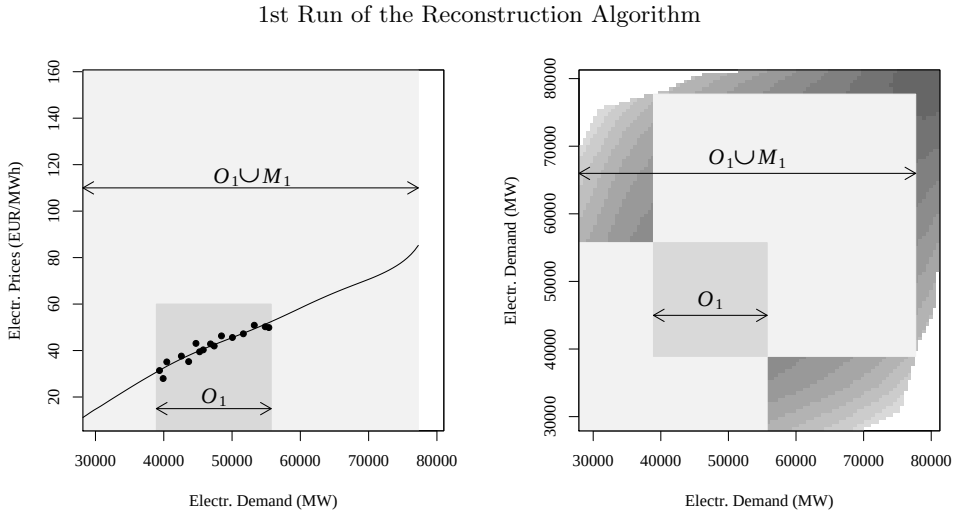

FIG 3. *Explanatory plots for the first run of the reconstruction algorithm.*

This 2nd Step is shown in Figure 4. There the new interval $O_2 \subseteq O_1 \cup M_1$ is chosen such that the still missing upper fragment becomes reconstructible. The new large interval $O_2 \cup M_2$ contains the missing upper fragments, such that we can stop the algorithm.

The choice of the subset O_r in the rth step is crucial. On the one hand, O_r should be chosen as large as possible to contain as much information as possible. On the other hand, O_r must be chosen such that M_r contains a still missing fragment which is—in tendency—met by smaller intervals O_r. That is, any efficient implementation of the algorithm and the choice of $r_{\max}$ depends on the extend to which γ can be estimated. A simple practical implementation is described in our application in Section 7.2.

In each iteration of the reconstruction algorithm we accumulate reconstruction errors. The following proposition provides a theoretical description of this accumulation of reconstruction errors:

PROPOSITION 5.1 (Accumulating reconstruction error). *For simplicity, let* $\mathbb{E}(X_i(u)) = 0$ *for all* $u \in [a, b]$ *and consider the second step of the reconstruction algorithm. Let* $X_i^{M_2}(u)$ *denote a missing value that we aim to reconstruct by* $[\mathcal{L}(\tilde{X}_i^{O_2})](u)$ *using* $\tilde{X}_i^{O_2}$ *which is taken from the reconstruc-*

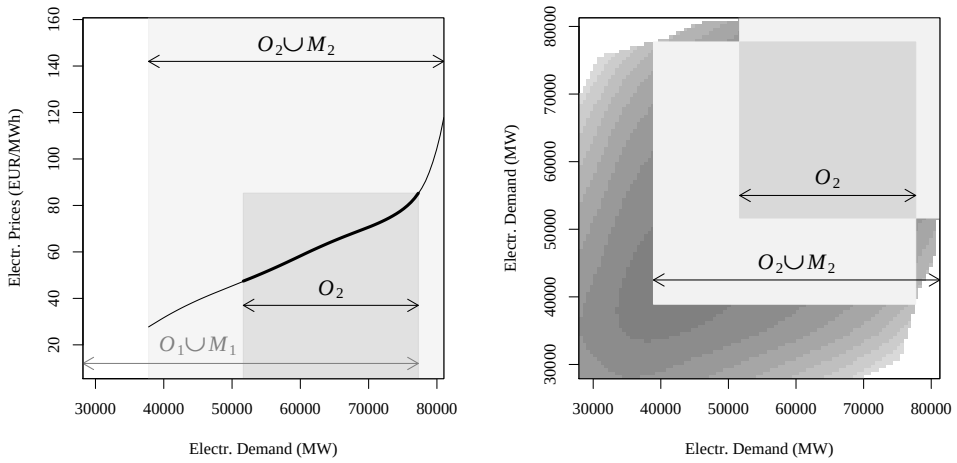

FIG 4. *Explanatory plots for the second run of the reconstruction algorithm.*

tion of the 1*st Step. The mean squared reconstruction error can then be approximated as following:*

$$\mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\tilde{X}_i^{O_2})](u)\right)^2\right) \leq \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_2})](u)\right)^2\right) + \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right),$$

where $\mathcal{L}(X_i^{O_1})$ *and* $\mathcal{L}(X_i^{O_2})$ *are the hypothetical reconstruction operators if* γ *were fully observed over* $[a,b]^2$*, and* $X_i^{O_2}$ *were observable.*

That is, the mean squared reconstruction error in the second run of the iterative algorithm is bounded from above by the two hypothetical mean squared reconstruction errors of $[\mathcal{L}(X_i^{O_1})](u)$ and $[\mathcal{L}(X_i^{O_2})](u)$.

6. Simulation study. For our simulation study we generate n iid normal and n iid exponential random functions $X_i(u) = \mu(u) + \xi_{i1}\phi_1(u) + \xi_{i2}\phi_2(u)$, where $\mu(u) = u + \sin(u)$, $\phi_1(u) = -\cos((\pi u)/(b-a))/\sqrt{5}$, $\phi_2(u) = -\cos(2(\pi u)/(b-a))/\sqrt{5}$, $a = 1$, and $b = 10$. For the normal case we let $\xi_{ik} \sim N(0, \lambda_k)$ with $\lambda_1 = 4$ and $\lambda_2 = 3$. For the exponential case we let ξ_{ik} be a centered exponential random variable with rate λ_k and centered by λ_k^{-1}. Furthermore, $U_{i1}, \ldots, U_{im} \overset{\text{iid}}{\sim} \text{Unif}[A_i, B_i]$, where $A_i \overset{\text{iid}}{\sim} \text{Unif}[a, a + (b-a)\cdot 0.25]$ and $B_i \overset{\text{iid}}{\sim} \text{Unif}[b-(b-a)\cdot 0.25, b]$. Finally, the observations Y_{ij} are generated

according to $Y_{ij} = X_t(U_{ij}) + \varepsilon_{ij}$ with $\varepsilon_{ij} \sim N(0, 0.2)$. The following data dimensions are investigated: $n \in \{100, 200\}$ and $m \in \{15, 25, 50, 75\}$. For reconstructing the missing parts $X_i^{M_i}(u)$ with $u \in M_i = [a, b] \setminus [A_i, B_i]$, we use the FPCA-based estimator $\widehat{\mathcal{L}}_K$ in (19). The estimate of the truncation parameter $\hat{K} := \arg\min \mathrm{AIC}_K$ is selected by minimizing the AIC criterion in (27). The bandwidths h_X, h_μ and h_γ are determined using generalized cross validation. In the jth ($j = 1, \ldots, B$) of in total $B = 100$ simulation runs we compute the Mean Absolute Reconstruction Error (MARE) as

$$\mathrm{MARE}_j = n^{-1} \sum_{i=1}^{n} \max_{u \in M_i} \left| X_i^{M_i}(u) - [\widehat{\mathcal{L}}_{\hat{K}}(\mathbb{X}_i^O)](u) \right|, \quad j = 1, \ldots, B$$

and finally report the averages $\mathrm{MARE} = B^{-1} \sum_{j=1}^{B} \mathrm{MARE}_j$. The implementation is done in R (R Core Team, 2017) and the R-codes are available from the authors.

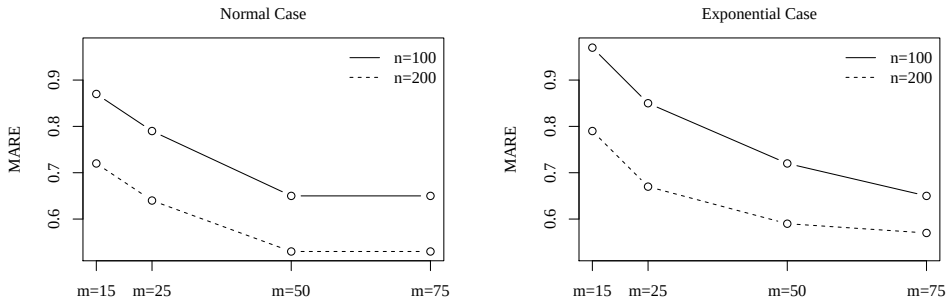

FIG 5. *Mean absolute reconstruction errors for the normal and exponential data generating processes.*

Figure 5 shows the MARE for the two different data generating processes and the different data dimensions. The simulation results demonstrate that the improvements in the MARE decline for larger values of m. This observation is in line with our asymptotic results. If m is small relatively to n, increasing m will significantly reduce the estimation error in $\widehat{\mathcal{D}}(X_i^O)$ (see Corollary 4.1). For relatively large values of m, however, the estimation error in $\widehat{\mathcal{D}}(X_i^O)$ does not depend anymore on m (see Theorem 4.2).

Table 1 shows the fractions of correctly estimated K. The AIC criterion correctly selects two eigen-components in at least 95% of the Monte Carlo repetitions within each simulation scenario, except for the cases with $m = 15$, where there are at leat 78% of correct selections.

TABLE 1
Fractions of correctly estimated K

		$m = 15$	$m = 25$	$m = 50$	$m = 75$
$n = 100$	Normal Case	0.90	0.99	1.00	1.00
	Exponential Case	0.78	0.95	0.98	0.99
$n = 200$	Normal Case	0.95	1.00	1.00	1.00
	Exponential Case	0.92	0.99	1.00	1.00

7. Application. The data for our analysis come from three different sources. Hourly spot prices of the German electricity market are provided by the European Energy Power Exchange (EPEX) (www.epexspot.com), hourly values of Germany's gross electricity demand and electricity exchanges with other countries are provided by the European Network of Transmission System Operators for Electricity (www.entsoe.eu), and German wind and solar power infeed data are provided by the transparency platform of the European energy exchange (www.eex-transparency.com). For academic, i.e., non-commercial, purposes the data can be accessed free of charge from these sources. The data dimensions are given by $m = 24$ hours and $n = 241$ working days between March 15, 2012 and March 14, 2013. Very few (0.4%) of the data pairs (Y_{ij}, U_{ij}) with prices $Y_{ij} > 120$ EUR/MWh and $U_{ij} > 82000$ MW are considered as outliers and reset to $Y_{ij} = 120$.

The German electricity market, like many other electricity markets, provides purchase guarantees for Renewable Energy Sources (RES). Therefore, the relevant variable for pricing at the energy exchange is electricity demand minus electricity infeeds from RES (Nicolosi, 2010). Correspondingly, the hourly values of electricity demand U_{ij} actually denote *residual* electricity demand, i.e., electricity demand minus infeeds from RES. That is, $U_{ij} := \texttt{Elect.Demand}_{ij} - \texttt{RES}_{ij}$, where $\texttt{RES}_{ij} = \texttt{Wind.Infeed}_{ij} + \texttt{Solar.Infeed}_{ij}$. The effect of further RES such as biomass is still negligible for the German electricity market.

7.1. *Checking for perfect reconstructions.* According to result (c) of Theorem 2.3 the reconstruction error can be zero if the structure of the random functions is simple enough. In the following, we asses empirically whether the functions can be reconstructed perfectly. The idea is to artificially partition the observed data $\mathbb{X}_i^O$ into an observed and a pseudo missing half. This way we can check whether the reconstructed part equals the artificially missing part, i.e., whether the functions are perfectly reconstructible. In practice, however, we have to take into account the estimation errors,

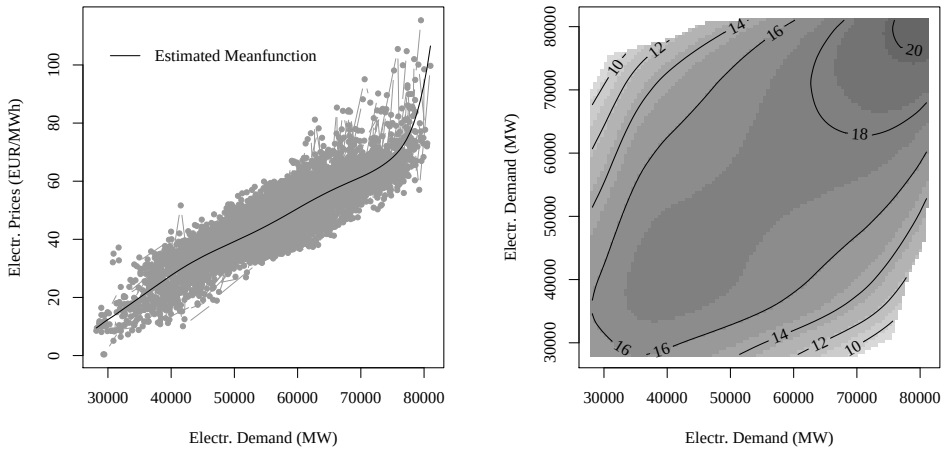

FIG 6. LEFT PANEL: *Estimated mean function plus a scatter plot of the data pairs* (Y_{ij}, U_{ij}). RIGHT PANEL: *Contour plot of the estimated covariance function. The white regions reflect the outer off-diagonal parts which are infeasible to estimate.*

which imply that the empirical reconstruction error $\widehat{\mathcal{Z}}_i(u) \neq 0$ even though the theoretical reconstruction error $\mathcal{Z}_i(u) = 0$ for all $u \in M$.

We take the lower halves of the data $\mathbb{X}_i^O$ as the observed parts and use them to reconstruct their pseudo missing upper halves and vice-versa which leads to, in total, $2n$ checks. To account for the pre-smoothing estimation error we use bootstrap $(1 - \alpha_{\text{Bonf}})$ confidence intervals with Bonferroni adjusted $\alpha_{\text{Bonf}} = 0.05/(2n)$ in order to account for the $2n = 482$ checks. Only if a reconstructed curve is within this confidence interval is the reconstruction considered "perfect". The function-wise pre-smoothing is done using the local linear kernel estimator, where the bandwidth is selected by cross-validation. For reconstructing the pseudo missing halves we use the FPCA-based estimator in (19). The left panel of Figure 7 demonstrates this approach for the case where we get a "perfect" reconstruction of the pseudo missing upper half given the information from the observed lower half of the data $\mathbb{X}_i^O$. Using this approach, 96.5% of the $2n$ reconstructions are considered perfect, which is roughly in line with the chosen significance level of $\alpha = 5\%$ from which we expect 5% of the reconstructions to be falsely classified as "imperfect". This suggests that we can expect good to very good reconstructions using our reconstruction operator, although we emphasize that our reconstructions are optimal among all linear reconstructions—no matter whether the functions are perfectly reconstructible or not.

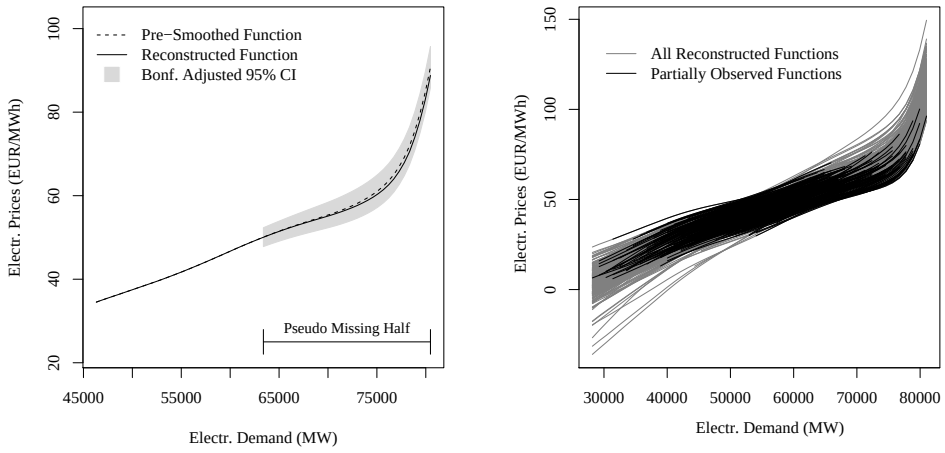

FIG 7. LEFT PANEL *Checking whether the reconstructed function is within the Bonferroni adjusted 95% confidence interval of the pre-smoothed function.* RIGHT PANEL *Recovered functions (gray) and the original, partially observed (pre-smoothed) functions (black).*

7.2. *Empirical results.* The estimated mean and covariance functions are shown in Figure 6. The outer off-diagonal parts of the covariance function γ cannot be estimated, since these parts of the domain are not covered by data pairs (U_{ij}, U_{il}), $j \neq l$. In order to reconstruct the entire missing parts X_i^M, we use the FPCA-based estimator in (19) and our iterative reconstruction algorithm 5.1, where our implementation is as follows. We use three iterations for each partially observed price function. In the first step, we use the information with respect to the original observations $\mathbb{X}_i^O$ in order to reconstruct the missing parts as far as possible. In the second step, we use the upper half of the empirically reconstructed curve $\hat{\hat{X}}_{i,1}$ and try to reconstruct possibly further missing upper fragments. In the final step we use the lower half of $\hat{\hat{X}}_{i,1}$ and try to reconstruct possibly further missing lower fragments. This simple approach allows us to recover 91% of the price functions over the total support. The reconstructed functions are shown in the right panel of Figure 7. Our preceding analysis suggests that the reconstructed functions are reliable estimates of the true price functions. Note that the price functions with negative electricity prices are perfectly plausible as the EPEX allows for negative prices (see, for instance, Nicolosi, 2010). Electricity producers are willing to sell electricity at negative prices (i.e., to pay for taking the electricity) if shutting off and restarting their power plants is more expensive than selling their electricity at negative prices.

References.

Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications* **149**. Springer Verlag.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34** 2159–2179.

Cardot, H., Mas, A. and Sarda, P. (2007). CLT in functional linear regression models. *Probability Theory and Related Fields* **138** 325–361.

Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *Journal of the American Statistical Association* **108** 1269–1283.

Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika* **00** 1–21.

Descary, M.-H. and Panaretos, V. M. (2017). Recovering covariance from functional fragments. *arXiv preprint arXiv:1708.02491*.

Goldberg, Y., Ritov, Y. and Mandelbaum, A. (2014). Predicting the continuation of a function with applications to call center data. *Journal of Statistical Planning and Inference* **147** 53–65.

Gromenko, O., Kokoszka, P., Sojka, J. et al. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics* **11** 898–918.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35** 70–91.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34** 1493–1517.

Hirth, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Economics* **38** 218–236.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications* **200**. Springer.

Kraus, D. (2015). Components and completion of partially observed functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77** 777–801.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38** 3321–3351.

Nicolosi, M. (2010). Wind power integration and power system flexibility – An empirical analysis of extreme events in Germany under the new negative price regime. *Energy Policy* **38** 7257–7268.

R Core Team, (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2. ed. *Springer Series in Statistics*. Springer.

Rice, J. (1984). Bandwidth Choice for Nonparametric Regression. *The Annals of Statistics* **12** 1215–1230.

Ruppert, D. and Wand, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of statistics* 1346–1370.

Weigt, H. (2009). Germany's wind energy: The potential for fossil capacity replacement and cost saving. *Applied Energy* **86** 1857–1863.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005b). Functional linear regression analysis

for longitudinal data. *The Annals of Statistics* **33** 2873–2903.
Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44** 2281–2321.

Alois Kneip and Dominik Liebl
Statistische Abteilung
University of Bonn
Adenauerallee 24-26
53113 Bonn, Germany
E-mail: akneip@uni-bonn.de
E-mail: dliebl@uni-bonn.de

Supplementary paper for:

On the Optimal Reconstruction of Partially Observed Functional Data

by Alois Kneip and Dominik Liebl

CONTENT

In the following we give the proofs of our theoretical results. The main steps in our proofs of Theorems 4.1 and 4.2 are as in Yao, Müller and Wang (2005a). Though, by contrast to Yao, Müller and Wang (2005a), we allow for a time series context (see Assumption A1), impose more restrictive assumptions on the kernel function (see Assumption A5), and consider a different asymptotic setup (see Assumption A2).

APPENDIX A: PROOFS

Proof of Theorem 2.1: For every linear operator $L : \mathbb{L}^2(O) \to \mathbb{L}^2(M)$ that is a reconstruction operator with respect to X_i^O according to Def. 2.1, we have that

$$\mathbb{V}([L(X_i^O)](u)) = \sum_{k=1}^{\infty} \lambda_k^O \left([L(\phi_k^O)](u)\right)^2 \quad \text{for every} \quad u \in M. \tag{32}$$

Existence: Writing $[L(X_i^O)](u)$ as $[L(X_i^O)](u) = \langle \alpha_u, X_i^O \rangle_H$ for some $\alpha_u \in H$ and computing again the variance of $[L(X_i^O)](u)$ yields that

$$\mathbb{V}([L(X_i^O)](u)) = \sum_{k=1}^{\infty} \lambda_k^O \left(\frac{\langle \alpha_u, \phi_k^O \rangle_2^2}{\lambda_k^O} \right)^2. \tag{33}$$

Since (32) and (33) must be equal, we have that $[L(\phi_k^O)](u) = \langle \alpha_u, \phi_k^O \rangle_2^2 / \lambda_k^O$ for all $k \geq 1$, which establishes that there exits a $\alpha_u \in H$ for every reconstruction $[L(X_i^O)](u)$.
Uniqueness: Assume that there is an alternative $\tilde{\alpha}_u \in H$ such that $[L(\phi_k^O)](u) = \langle \tilde{\alpha}_u, \phi_k^O \rangle_2^2 / \lambda_k^O$ for all $k \geq 1$. Then $\langle \alpha_u - \tilde{\alpha}_u, \phi_k^O \rangle_2^2 / \lambda_k^O = 0$ for all $k \geq 1$ or equivalently $\langle \alpha_u, \phi_k^O \rangle_2 - \langle \tilde{\alpha}_u, \phi_k^O \rangle_2 = 0$ for all $k \geq 1$ which shows that $\tilde{\alpha}_u - \alpha_u = 0$.

Proof of Theorem 2.2, part (a): First, note that continuity of $\gamma(u, v)$ implies continuity of $\mathbb{V}([\mathcal{L}(X_i^O)](u))$. Second, note that for any K and every

$u \in M$, we have

$$0 \leq \mathbb{E}\left(\left(X_i^M(u) - \sum_{k=1}^{K} \xi_{ik}^O \tilde{\phi}_k^O(u)\right)^2\right) = \gamma(u,u) - \sum_{k=1}^{K} \lambda_k^O \tilde{\phi}_k^O(u)^2. \tag{34}$$

But this implies that $\mathbb{V}(\mathcal{L}_{u,K}(X_i^O)) = \mathbb{V}(\sum_{k=1}^{K} \xi_{ik}^O \tilde{\phi}_k^O(u)) = \sum_{k=1}^{K} \lambda_k^O \tilde{\phi}_k^O(u)^2$ converges to a fixed limit $0 \leq \mathbb{V}([\mathcal{L}(X_i^O)](u)) < \infty$ as $K \to \infty$ for all $u \in M$.

Part (b): Follows directly from observing that $\mathbb{E}([\mathcal{L}(X_i^O)](u)) = 0$ for all $u \in M$.

Proof of Theorem 2.3, part (a): For all $v \in O$ and $u \in M$ we have that

$$\begin{aligned}
&\mathbb{E}\left(X_i^O(v)\mathcal{Z}_i(u)\right) = \mathbb{E}\left(X_i^O(v)\left(X_i^M(u) - [\mathcal{L}(X_i^O)](u)\right)\right) = \\
&= \mathbb{E}\left(\sum_{k=1}^{\infty} \xi_{ik}^O \phi_k^O(v)\Big(X_i^M(u) - \sum_{k=1}^{\infty} \xi_{ik}^O \tilde{\phi}_k^O(u)\Big)\right) = \\
&= \sum_{k=1}^{\infty} \phi_k^O(v)\Big(\mathbb{E}(\xi_{ik}^O X_i(u)) - \lambda_k \tilde{\phi}_k^O(u)\Big).
\end{aligned}$$

From the definition $\tilde{\phi}_k^O(u)$ in (6) we get that $\mathbb{E}(\xi_{ik}^O X_i^M(u)) = \lambda_k^O \tilde{\phi}_k^O(u)$, which leads to $\mathbb{E}(X_i^O(v)\mathcal{Z}_i(u)) = 0$ for all $u \in M$. This proves (12), while (13) directly follows from the definition of $\mathcal{Z}_i(u)$.

Part (b): By Theorem 2.1 there exists a unique $b_u \in H$ such that

$$[\ell(X_i^O)](u) = \langle b_u, X_i^O \rangle_H.$$

By (12) and the orthogonality property of the least squares projection we thus obtain

$$\begin{aligned}
&\mathbb{E}\left(\left(X_i(u) - [\ell(X_i^O)](u)\right)^2\right) = \\
&= \mathbb{E}\left(\left([\mathcal{L}(X_i^O)](u) + \mathcal{Z}_i(u) - \langle b_u, X_i^O \rangle_H\right)^2\right) = \\
&= \mathbb{E}\left(\left([\mathcal{L}(X_i^O)](u) - \langle b_u, X_i^O \rangle_H\right)^2\right) + \mathbb{E}(\mathcal{Z}_i(u)^2) + \\
&+ 2\left(\mathbb{E}([\mathcal{L}(X_i^O)](u)\,\mathcal{Z}_i(u)) - \mathbb{E}\left(\langle b_u, X_i^O \rangle_H \mathcal{Z}_i(u)\right)\right) = \\
&= \mathbb{E}\left(\left([\mathcal{L}(X_i^O)](u) - \langle b_u, X_i^O \rangle_H\right)^2\right) + \mathbb{E}(\mathcal{Z}_i^2(u)) \geq \mathbb{E}(\mathcal{Z}_i^2(u)).
\end{aligned}$$

Part (c): Observe that $\mathbb{V}(\mathcal{Z}_i(u) - \mathcal{Z}_j(u)) = \mathbb{V}(\mathcal{Z}_i(u)) + \mathbb{V}(\mathcal{Z}_j(u)) - 2\,\mathrm{Cov}(\mathcal{Z}_i(u), \mathcal{Z}_j(u)) = 2\,\mathbb{V}(\mathcal{Z}_i(u))$ for all $u \in M$ and $i \neq j$. Rearranging and using that $\mathbb{E}(\mathcal{Z}_i(u)) = \mathbb{E}(\mathcal{Z}_j(u)) = 0$ for all $u \in M$ and all $i, j \in \{1, \ldots, n\}$ yields $\mathbb{V}(\mathcal{Z}_i(u)) =$

$\frac{1}{2}\,\mathbb{E}((\mathcal{Z}_i(u)-\mathcal{Z}_j(u))^2)$. From result (a) we know that $\mathcal{Z}_i(u)$ and $X_i^O(v)$ are orthogonal and therefore uncorrelated for all $u \in M$ and all $v \in O$, that is, $\mathbb{E}\left(X_i^O(v)\mathcal{Z}_i(u)\right) = \mathrm{Cov}\left(X_i^O(v), \mathcal{Z}_i(u)\right) = 0$. Under the assumption of an independent Gaussian process, we have then independence between $\mathcal{Z}_i(u)$ and $X_i^O(v)$, such that

$$\mathbb{V}(\mathcal{Z}_i(u)) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left((\mathcal{Z}_i(u)-\mathcal{Z}_j(u))^2\right)|X_i^O = X_j^O\right),$$

where $X_i^O = X_j^O$ means that $X_i^O(u) = X_i^O(u)$ for all $u \in O$. For the two random functions $\mathcal{Z}_i(u)$ and $\mathcal{Z}_j(u)$ we can write $\mathcal{Z}_i(u) = X_i^M(u) - [\mathcal{L}(X_i^O)](u)$ and $\mathcal{Z}_j(u) = X_j^M(u) - [\mathcal{L}(X_j^O)](u)$. It follows from the definition of $\tilde{\phi}_k^O(u)$ in (6) that $[\mathcal{L}(X_i^O)](u) = [\mathcal{L}(X_j^O)](u)$ for all $u \in M$, if and only if $X_i^O = X_j^O$. Therefore,

$$\mathbb{V}(\mathcal{Z}_i(u)) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left((X_i^M(u)-X_j^M(u))^2\right)|X_i^O = X_j^O\right),$$

for all $u \in M$.

Proof of Theorem 4.1 For proofing the results in Theorem 4.1 we make use of the following two lemmas:

LEMMA A.1. *Define*

$$\Psi_{q,nm}(u;h_\mu) = \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)\psi_q\left(U_{ij}-u, Y_{ij}\right), \tag{35}$$

where

$$\psi_q\left(U_{ij}-u, Y_{ij}\right) = \begin{cases} (U_{ij}-u)^q & \text{for } q \in \{0,1,2\} \\ Y_{ij} & \text{for } q = 3 \\ (U_{ij}-u)\,Y_{ij} & \text{for } q = 4. \end{cases}$$

Then, under Assumptions A1-A5,

$$\tau_{q,nm} = \sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu) - m_q(u)| = \mathcal{O}_p\left(h_\mu^2 + \frac{1}{\sqrt{nm\,h_\mu}} + \frac{1}{\sqrt{n}}\right),$$

where $m_0(u) = f_U(u)$, $m_1(u) = 0$, $m_2(u) = f_U(u)\nu_2(\kappa)$, $m_3(u) = \mu(u)f_U(u) = \mathbb{E}(Y_{ij}|U_{ij} = u)f_U(u)$, *and* $m_4(u) = 0$.

LEMMA A.2. *Define*

(36)
$$\Theta_{q,nM}(u,v;h_\gamma) = \frac{1}{nMh_\gamma}\sum_{i,j\neq l}\kappa\left(\frac{U_{ij}-u}{h_\gamma}\right)\kappa\left(\frac{U_{il}-v}{h_\gamma}\right)\vartheta_q\left(U_{ij}-u, U_{il}-u, C_{ijl}\right),$$

where

$$\vartheta_q\left(U_{ij}-u, U_{il}-v, C_{ijl}\right) = \begin{cases} (U_{ij}-u)^q\,(U_{il}-v)^q & \text{for } q\in\{0,1,2\} \\ C_{ijl} & \text{for } q=3 \\ (U_{ij}-u)\,(U_{il}-v)\,C_{ijl} & \text{for } q=4. \end{cases}$$

Then, under Assumptions A1-A5,

$$\varrho_{q,nM} = \sup_{(u,v)\in[a,b]^2}|\Theta_{q,nM}(u,v;h_\gamma)-\eta_q(u,v)| = \mathcal{O}_p\left(h_\gamma^2+\frac{1}{\sqrt{nM\,h_\gamma^2}}+\frac{1}{\sqrt{n}}\right),$$

where $\eta_0(u,v) = f_{UU}(u,v)$, $\eta_1(u,v) = 0$, $\eta_2(u,v) = f_{UU}(u,v)(\nu_2(\kappa))^2$, $\eta_3(u,v) = \gamma(u,v)f_{UU}(u,v) = \mathbb{E}(C_{ijl}|(U_{ij},U_{il}) = (u,v))f_{UU}(u,v)$, *and* $m_4(u,v) = 0$.

Proof of Lemma A.1: Remember that $\mathbb{E}(|\tau_{q,nm}|) = \mathcal{O}(\texttt{rate}_{nm})$ implies that $\tau_{q,nm} = \mathcal{O}_p(\texttt{rate}_{nm})$, therefore, we focus in the following on $\mathbb{E}(|\tau_{q,nm}|)$, where $\mathbb{E}(|\tau_{q,nm}|) = \mathbb{E}(\tau_{q,nm})$. Adding a zero and applying the triangle inequality yields that $\mathbb{E}(\tau_{q,nm}) =$

$$\begin{aligned}\mathbb{E}(\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-m_q(u)|) \leq \sup_{u\in[a,b]}|\,\mathbb{E}(\Psi_{q,nm}(u;h_\mu))-m_p(u)|+ \\ +\,\mathbb{E}(\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-\mathbb{E}(\Psi_{q,nm}(u;h_\mu))|). \qquad (37)\end{aligned}$$

Let us first focus on the second summand in (37). The next steps will make use of the Fourier transformation of the kernel function κ (see, e.g., Tsybakov, 2008, Ch. 1.3):

$$\kappa^{\text{ft}}(x) := \mathcal{F}[\kappa](x) = \int_{\mathbb{R}}\kappa(z)\exp(-\mathrm{i}zx)dz = \int_{-1}^{1}\kappa(z)\exp(-\mathrm{i}zx)dz$$

with $\mathrm{i} = \sqrt{-1}$. By Assumption A5, $\kappa(.)$ has a compact support $[-1,1]$. The inverse transform gives then

$$\kappa\left(s\right) = \frac{1}{2\pi}\int_{\mathbb{R}}\kappa^{\text{ft}}(x)\exp\left(\mathrm{i}xs\right)dx = \frac{1}{2\pi}\int_{\mathbb{R}}\kappa^{\text{ft}}(x)\exp\left(\mathrm{i}xs\right)dx\;\mathbb{1}_{(|s|<1)}.$$

Furthermore, we can use that (see Tsybakov, 2008, Ch. 1.3, (1.34)) $\mathcal{F}[\kappa(./h_\mu)/h_\mu](x) = \mathcal{F}[\kappa](xh_\mu) = \kappa^{\text{ft}}(xh_\mu)$ which yields

$$\begin{aligned}\kappa\left(s/h_\mu\right)/h_\mu &= \frac{1}{2\pi}\int_{\mathbb{R}} \mathcal{F}[\kappa(./h_\mu)/h_\mu](x)\exp\left(\mathrm{i}xs\right)dx\ \mathbb{1}_{(|s|<h_\mu)} \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \kappa^{\text{ft}}(xh_\mu)\exp\left(\mathrm{i}xs\right)dx\ \mathbb{1}_{(|s|<h_\mu)}. \end{aligned} \tag{38}$$

Plugging (38) into (35) yields $\Psi_{q,nm}(u;h_\mu) =$

$$\begin{aligned}&= \frac{1}{nm}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)\frac{1}{h_\mu}\psi_q\left(U_{ij}-u,Y_{ij}\right) \\ &= \frac{1}{nm}\sum_{ij}\frac{1}{2\pi}\int_{\mathbb{R}}\kappa^{\text{ft}}(xh_\mu)\exp\big(\mathrm{i}x(U_{ij}-u)\big)dx\ \mathbb{1}_{(|U_{ij}-u|<h_\mu)}\ \psi_q\left(U_{ij}-u,Y_{ij}\right) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\left[\frac{1}{nm}\sum_{ij}\exp\left(\mathrm{i}xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right)\ \mathbb{1}_{(|U_{ij}-u|<h_\mu)}\right]\exp\big(\mathrm{i}xu\big)\kappa^{\text{ft}}(xh_\mu)dx.\end{aligned}$$

Using that $|\exp(ixu)| \leq 1$ leads to

$$\mathbb{E}(\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-\mathbb{E}(\Psi_{q,nm}(u;h_\mu))|) \leq \frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in[a,b]}\left|\int_{\mathbb{R}}\tilde{\omega}_{q,nm}(u,x)\cdot\kappa^{\text{ft}}(xh_\mu)dx\right|\right),$$

where

$$\begin{aligned}\tilde{\omega}_{q,nm}(u,x) =& \frac{1}{nm}\sum_{ij}\big[\exp\left(\mathrm{i}xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right)\ \mathbb{1}_{(|U_{ij}-u|<h_\mu)} - \\ &\mathbb{E}\left(\exp\left(\mathrm{i}xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right)\mathbb{1}_{(|U_{ij}-u|<h_\mu)}\right)\big].\end{aligned}$$

Using further that κ^{ft} is symmetric, since κ is symmetric by Assumption A5, and that $\exp\left(\mathrm{i}xU_{ij}\right) = \cos\left(xU_{ij}\right) + \mathrm{i}\sin\left(xU_{ij}\right)$ leads to

$$\frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in[a,b]}\left|\int_{\mathbb{R}}\tilde{\omega}_{q,nm}(u,x)\cdot\kappa^{\text{ft}}(xh_\mu)dx\right|\right) = \frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in[a,b]}\left|\int_{\mathbb{R}}\omega_{q,nm}(u,x)\cdot\kappa^{\text{ft}}(xh_\mu)dx\right|\right),$$

where

$$\begin{aligned}\omega_{q,nm}(u,x) =& \frac{1}{nm}\sum_{ij}\Big[\cos\left(xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right)\ \mathbb{1}_{(|U_{ij}-u|<h_\mu)} - \\ &\mathbb{E}\left(\cos\left(xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right)\mathbb{1}_{(|U_{ij}-u|<h_\mu)}\right)\Big],\end{aligned} \tag{39}$$

such that

$$\begin{aligned}
&\mathbb{E}(\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-\mathbb{E}(\Psi_{q,nm}(u;h_\mu))|)\\
&\leq \frac{1}{2\pi}\int_{\mathbb{R}}\mathbb{E}\left(\sup_{u\in[a,b]}\left|\omega_{q,nm}(u,x)\right|\right)\cdot\left|\kappa^{\mathrm{ft}}(xh_\mu)\right|dx\\
&\leq \frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\left(\sup_{u\in[a,b]}\left|\omega_{q,nm}(u,x)\right|\right)^2\right)}\cdot\left|\kappa^{\mathrm{ft}}(xh_\mu)\right|dx\\
&= \frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\sup_{u\in[a,b]}\left(\omega_{q,nm}(u,x)\right)^2\right)}\cdot\left|\kappa^{\mathrm{ft}}(xh_\mu)\right|dx. \qquad (40)
\end{aligned}$$

In order to simplify the notation we will denote

$$W_{ij}^q(x,u)=\cos\left(xU_{ij}\right)\psi_q\left(U_{ij}-u,Y_{ij}\right),$$

such that $\mathbb{E}\left(\sup_{u\in[a,b]}\left(\omega_{q,nm}(u,x)\right)^2\right)=$

$$\begin{aligned}
\mathbb{E}\Big(\sup_{u\in[a,b]}\Big(&\frac{1}{(nm)^2}\sum_{ij}\left[W_{ij}^q(x,u)\mathbb{1}_{(|U_{ij}-u|<h_\mu)}-\mathbb{E}(W_{ij}^q(x,u)\mathbb{1}_{(|U_{ij}-u|<h_\mu)})\right]^2+\\
&\frac{1}{(nm)^2}\sum_{(i,j)\neq(r,l)}\left[(W_{ij}^q(x,u)\mathbb{1}_{(|U_{ij}-u|<h_\mu)}-\mathbb{E}(W_{ij}^q(x,u)\mathbb{1}_{(|U_{ij}-u|<h_\mu)}))\cdot\right.\\
&\left.\cdot(W_{rl}^q(x,u)\mathbb{1}_{(|U_{rl}-u|<h_\mu)}-\mathbb{E}(W_{rl}^q(x,u)\mathbb{1}_{(|U_{rl}-u|<h_\mu)}))\right]\Big)\Big).
\end{aligned}$$

As u takes only values within the compact interval $[a,b]$, there exist constants C_1 and C_2 such that, uniformly for all $u\in[a,b]$, $\mathbb{P}(|U_{ij}-u|<h_\mu)\leq C_1h_\mu<\infty$, for all i,j, and $\mathbb{P}(|U_{ij}-u|<h_\mu \text{ AND } |U_{rl}-u|<h_\mu)\leq C_2h_\mu^2<\infty$, for all $(i,j)\neq(r,l)$. Together with the triangle inequality, this yields that $\mathbb{E}\left(\sup_{u\in[a,b]}\left(\omega_{q,nm}(u,x)\right)^2\right)\leq$

$$\begin{aligned}
&\frac{C_1h_\mu}{(nm)^2}\sum_{ij}\mathbb{E}\left(\sup_{u\in[a,b]}\left[W_{ij}^q(x,u)-\mathbb{E}(W_{ij}^q(x,u))\right]^2\right)+\\
&\frac{C_2h_\mu^2}{(nm)^2}\sum_{(i,j)\neq(r,l)}\mathbb{E}\left(\sup_{u\in[a,b]}\left[(W_{ij}^q(x,u)-\mathbb{E}(W_{ij}^q(x,u)))(W_{rl}^q(x,u)-\mathbb{E}(W_{rl}^q(x,u)))\right]\right)
\end{aligned}$$

From our moment assumptions (Assumption A1) and the fact that $[a,b]$ is compact, we can conclude that there must exist a constant C_3 such that, point-wise for every $x \in \mathbb{R}$,

$$\mathbb{E}\left(\left(\sup_{u\in[a,b]}\left|W_{ij}^q(x,u)-\mathbb{E}(W_{ij}^q(x,u))\right|\right)^2\right) \leq C_3 < \infty \tag{41}$$

for all i and j.

"Within function" dependencies: By the same reasoning there must exist a constant C_4 such that, point-wise for every $x \in \mathbb{R}$,

$$\begin{aligned}&\mathbb{E}\left(\sup_{u\in[a,b]}\left|W_{ij}^q(x,u)-\mathbb{E}(W_{ij}^q(x,u))\right|\cdot\sup_{u\in[a,b]}\left|W_{il}^q(x,u)-\mathbb{E}(W_{il}^q(x,u))\right|\right)\\ &\leq C_4 < \infty\end{aligned} \tag{42}$$

for all $j \neq l$ and all i.

"Between function" dependencies: Our weak dependency assumption (Assumption A1) and the fact that $[a,b]$ is compact yields that point-wise for every $x \in \mathbb{R}$

$$\begin{aligned}&\mathbb{E}\left(\sup_{u\in[a,b]}\left|W_{ij}^q(x,u)-\mathbb{E}(W_{ij}^q(x,u))\right|\cdot\sup_{u\in[a,b]}\left|W_{rl}^q(x,u)-\mathbb{E}(W_{rl}^q(x,u))\right|\right)\\ &\leq c_1\iota_1^{|i-r|}\end{aligned} \tag{43}$$

for all j,l and $|i-r| \geq 1$, where $0 < c_1 < \infty$ and $0 < \iota_1 < 1$.

Eq.s (41), (42), and (43) yield that $\mathbb{E}\left(\sup_{u\in[a,b]}\left(\omega_{q,nm}(u,x)\right)^2\right) \leq$

$$\begin{aligned}\leq&\frac{C_1 h_\mu}{(nm)^2}\sum_{ij}C_3+\frac{C_2 h_\mu^2}{(nm)^2}\sum_{i,j\neq l}C_4+\frac{C_2 h_\mu^2}{(nm)^2}\sum_{i\neq r,jl}c_1\iota_1^{|i-r|}\\ =&\mathcal{O}\left(\frac{h_\mu}{nm}+\frac{h_\mu^2(m-1)}{nm}+\frac{h_\mu^2}{n}\right)=\mathcal{O}\left(\frac{h_\mu}{nm}+\frac{h_\mu^2}{n}\right),\end{aligned}$$

such that

$$\sqrt{\mathbb{E}\left(\sup_{u\in[a,b]}\left(\omega_{q,nm}(u,x)\right)^2\right)}=\mathcal{O}\left(\sqrt{\frac{h_\mu}{nm}}+\frac{h_\mu}{\sqrt{n}}\right). \tag{44}$$

Plugging (44) into (40) and integration by substitution leads to

$$\mathbb{E}\Big(\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-\mathbb{E}(\Psi_{q,nm}(u;h_\mu))|\Big)\leq$$

(45)

$$\frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\sup_{u\in[a,b]}(\omega_{q,nm}(u,x))^2\right)}\cdot\left|\kappa^{\text{ft}}(xh_\mu)\right|dx=\mathcal{O}\left(\frac{1}{\sqrt{nm\,h_\mu}}+\frac{1}{\sqrt{n}}\right).$$

Let us now focus on the first summand in (37). From standard arguments in nonparametric statistics (see, e.g., Ruppert and Wand (1994)) we know that

$$\mathbb{E}(\Psi_{q,nm}(u;h_\mu))-m_q(u)=\mathcal{O}(h_\mu^2)$$

for each $u\in[a,b]$ and for all $q\in\{0,\ldots,4\}$. Under our smoothness Assumption A3, the "$\mathcal{O}(h_\mu^2)$" term becomes uniformly valid for all $u\in[a,b]$ and all $q\in\{0,1,2,4\}$, since all of the involved functions have uniformly bounded second order derivatives. We can conclude with respect to the first summand in (37) that

$$\sup_{u\in[a,b]}|\,\mathbb{E}(\Psi_{q,nm}(u;h_\mu))-m_p(u)|=\mathcal{O}(h_\mu^2)\quad\text{for all}\quad q\in\{0,\ldots,4\}. \tag{46}$$

Finally, plugging our results (45) and (46) into (37) leads to

$$\tau_{q,nm}=\sup_{u\in[a,b]}|\Psi_{q,nm}(u;h_\mu)-m_q(u)|=\mathcal{O}_p\left(h_\mu^2+\frac{1}{\sqrt{nm\,h_\mu}}+\frac{1}{\sqrt{n}}\right) \tag{47}$$

for all $q\in\{0,\ldots,4\}$.

Proof of Lemma A.2: Analogously to that of Lemma A.1.

Proof of Theorem 4.1, part (a): Let us rewrite the estimator $\hat{\mu}$ using matrix notation as in Ruppert and Wand (1994), i.e.,

$$\hat{\mu}(u;h_\mu)=e_1^\top\Big([\mathbf{1},\mathbf{U}_u]^\top\mathbf{W}_{\mu,u}[\mathbf{1},\mathbf{U}_u]\Big)^{-1}[\mathbf{1},\mathbf{U}_u]^\top\mathbf{W}_{\mu,u}\mathbf{Y}, \tag{48}$$

where $e_1=(1,0)^\top$, $[\mathbf{1},\mathbf{U}_u]$ is a $nm\times 2$ dimensional data matrix with typical rows $(1,U_{ij}-u)$, the $nm\times nm$ dimensional diagonal weighting matrix $\mathbf{W}_{\mu,u}$ holds the kernel weights $K_{\mu,h}(U_{ij}-u)=h_\mu^{-1}\,\kappa(h_\mu^{-1}(U_{ij}-u))$. The objects

$\mathbf{U}_u$ and $\mathbf{W}_{\mu,u}$ are filled in correspondence with the nm dimensional vector $\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{n,m-1}, Y_{n,m})^\top$.

This way we can decompose the estimator $\hat{\mu}(u; h_\mu)$ as

$$\hat{\mu}(u; h_\mu) = e_1^\top L_{1,nm,u}^{-1} L_{2,nm,u}, \tag{49}$$

with 2×2 matrix

$$\begin{aligned} L_{1,nm,u} &= (nm)^{-1}[\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u}[\mathbf{1}, \mathbf{U}_u] \\ &= \begin{pmatrix} \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right) & \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)(U_{ij}-u) \\ \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)(U_{ij}-u) & \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)(U_{ij}-u)^2 \end{pmatrix}, \end{aligned}$$

and 2×1 vector

$$L_{2,nm,u} = (nm)^{-1}[\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u}\mathbf{Y} = \begin{pmatrix} \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right) Y_{ij} \\ \frac{1}{nmh_\mu}\sum_{ij}\kappa\left(\frac{U_{ij}-u}{h_\mu}\right)(U_{ij}-u)\, Y_{ij} \end{pmatrix}.$$

Using the notation and the results from Lemma A.1 we have that

$$\begin{aligned} L_{1,nm,u} &= \begin{pmatrix} \Psi_{0,nm}(u; h_\mu) & \Psi_{1,nm}(u; h_\mu) \\ \Psi_{1,nm}(u; h_\mu) & \Psi_{2,nm}(u; h_\mu) \end{pmatrix} \\ &= \begin{pmatrix} f_U(u) & 0 \\ 0 & f_U(u)\nu_2(\kappa) \end{pmatrix} + \mathcal{O}_p^{\text{Unif}}\left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}}\right) \quad \text{and} \end{aligned} \tag{50}$$

$$L_{2,nm,u} = \begin{pmatrix} \Psi_{3,nm}(u; h_\mu) \\ \Psi_{4,nm}(u; h_\mu) \end{pmatrix} = \begin{pmatrix} \mu(u) f_U(u) \\ 0 \end{pmatrix} + \mathcal{O}_p^{\text{Unif}}\left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}}\right), \tag{51}$$

where we write $\Psi_{q,nm}(u; h_\mu) - m_q(u) = \mathcal{O}_p^{\text{Unif}}(\texttt{rate})$ in order to denote that $\sup_{u\in[a,b]} |\Psi_{q,nm}(u; h_\mu) - m_q(u)| = \mathcal{O}_p(\texttt{rate})$. Taking the inverse of (50) gives

$$L_{nm,u}^{-1} = \begin{pmatrix} 1/f_U(u) & 0 \\ 0 & 1/(f_U(u)\nu_2(\kappa)) \end{pmatrix} + \mathcal{O}_p^{\text{Unif}}\left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}}\right). \tag{52}$$

Plugging (52) and (51) into (49) leads to

$$\sup_{u\in[a,b]} |\hat{\mu}(u; h_\mu) - \mu(u)| = \mathcal{O}_p\left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{m}}\right).$$

Proof of Theorem 4.1, part (ã): Observe that

$$\sup_{u\in O}|\widehat{X}_i^O(u;h_\mu,h_X)-X_i^O(u)|\leq$$
$$\sup_{u\in O}|\hat{X}_i^{c,O}(u;h_X)-(X_i^O(u)-\mu(u))|+\sup_{u\in O}|\hat{\mu}(u;h_\mu)-\mu(u)|.$$

From Theorem 4.1, part (a), we have that $\sup_{u\in O}|\hat{\mu}(u;h_\mu)-\mu(u)|=\mathcal{O}_p(r_\mu)$ with $r_\mu=h_\mu^2+1/\sqrt{nm\,h_\mu}+1/\sqrt{m}$. From a simplified version of the proof of Theorem 4.1, part (a), with $n=1$, it follows that

$$\sup_{u\in O}|\hat{X}_i^{c,O}(u;h_X)-(X_i^O(u)-\mu(u))|=\mathcal{O}_p\left(h_X^2+\frac{1}{\sqrt{mh_X}}\right)$$

Proof of Theorem 4.1, part (b):

Let us rewrite the estimator $\hat{\gamma}$ using matrix notation as in Ruppert and Wand (1994), i.e., $\hat{\gamma}(u,v;h_\gamma)=$

$$=e_1^\top\left([\mathbf{1},\mathbf{U}_u,\mathbf{U}_v]^\top\mathbf{W}_{\gamma,u,v}[\mathbf{1},\mathbf{U}_u,\mathbf{U}_v]\right)^{-1}[\mathbf{1},\mathbf{U}_u,\mathbf{U}_v]^\top\mathbf{W}_{\gamma,u,v}\hat{\mathbf{C}}, \tag{53}$$

where $e_1=(1,0,0)^\top$, $[\mathbf{1},\mathbf{U}_u,\mathbf{U}_v]$ is a $nM\times 3$ dimensional data matrix with typical rows $(1,U_{ij}-u,U_{il}-v)$, the $nM\times nM$ dimensional diagonal weighting matrix $\mathbf{W}_{\gamma,u,v}$ holds the bivariate kernel weights $K_{\gamma,h}(U_{ij}-u,U_{il}-v)$. For the bivariate kernel weights $K_{\gamma,h}(z_1,z_2)=h_\gamma^{-2}\kappa_\gamma(z_1,z_2)$ we use a multiplicative kernel function $\kappa_\gamma(z_1,z_2)=\kappa(z_1)\kappa(z_2)$ with κ as defined above. The usual kernel constants are then $\nu_2(\kappa_\gamma):=(\nu_2(\kappa))^2$ and $R(\kappa_\gamma):=R(\kappa)^2$. The rows of the matrices $[\mathbf{1},\mathbf{U}_u,\mathbf{U}_v]$ and $\mathbf{W}_{\gamma,u,v}$ are filled in correspondence with the nM elements of the vector of raw-covariances $\hat{\mathbf{C}}=(\dots,\hat{C}_{ijl},\dots)^\top$.

Let us initially consider the infeasible estimator $\hat{\gamma}_C$ that is based on the infeasible "clean" raw-covariances $C_{ijl}=(Y_{ij}-\mu(U_{ij}))(Y_{il}-\mu(U_{il}))$ instead of the estimator $\hat{\gamma}$ in (53) that is based on the "dirty" raw-covariances $\hat{C}_{ijl}=(Y_{ij}-\hat{\mu}(U_{ij}))(Y_{il}-\hat{\mu}(U_{il}))$. Equivalently to the estimator $\hat{\mu}$ above, we can write the estimator $\hat{\gamma}_C$ as

$$\hat{\gamma}_C(u,v;h_\gamma)=e_1^\top\tilde{S}_{1,nM,(u,v)}^{-1}\tilde{S}_{2,nM,(u,v)}, \tag{54}$$

with

$$\tilde{S}_{1,nM,(u,v)}^{-1}=\begin{pmatrix}\Theta_{0,nM}(u,v;h_\gamma) & \Theta_{1,nM}(u,v;h_\gamma)\\ \Theta_{1,nM}(u,v;h_\gamma) & \Theta_{2,nM}(u,v;h_\gamma)\end{pmatrix}^{-1}$$
$$=\begin{pmatrix}1/f_{UU}(u,v) & 0\\ 0 & 1/f_{UU}(u,v)(\nu_2(\kappa))^2\end{pmatrix}+\mathcal{O}_p^{\text{Unif}}\left(h_\gamma^2+\frac{1}{\sqrt{nM\,h_\gamma^2}}+\frac{1}{\sqrt{n}}\right) \tag{55}$$

and $\tilde{S}_{2,nM,(u,v)} =$

(56)
$$= \begin{pmatrix} \Theta_{3,nM}(u,v;h_\gamma) \\ \Theta_{4,nM}(u,v;h_\gamma) \end{pmatrix} = \begin{pmatrix} \gamma(u,v) f_{UU}(u,v) \\ 0 \end{pmatrix} + \mathcal{O}_p^{\text{Unif}} \left(h_\gamma^2 + \frac{1}{\sqrt{nM\, h_\gamma^2}} + \frac{1}{\sqrt{n}} \right),$$

where we use the notation and the results from Lemma A.2, and where we write $\Theta_{q,nM}(u,v;h_\gamma) - \eta_q(u,v) = \mathcal{O}_p^{\text{Unif}}(\texttt{rate})$ in order to denote that $\sup_{(u,v)\in[a,b]^2} |\Theta_{q,nM}(u,v;h_\gamma) - \eta_q(u,v)| = \mathcal{O}_p(\texttt{rate})$.

Plugging (55) and (56) into (54) leads to

$$\sup_{(u,v)\in[a,b]^2} |\hat{\gamma}_C(u,v;h_\mu) - \gamma(u,v)| = \mathcal{O}_p \left(h_\gamma^2 + \frac{1}{\sqrt{nM\, h_\gamma^2}} + \frac{1}{\sqrt{n}} \right). \tag{57}$$

It remains to consider the additional estimation error, which comes from using the "dirty" response variables $\hat{C}_{ijl}$ instead of "clean" dependent variables C_{ijl}. Observe that we can expand $\hat{C}_{ijl}$ as following:

$$\begin{aligned} \hat{C}_{ijl} = C_{ijl} &+ (Y_{ij} - \mu(U_{ij}))(\mu(U_{il}) - \hat{\mu}(U_{il})) \\ &+ (Y_{il} - \mu(U_{il}))(\mu(U_{ij}) - \hat{\mu}(U_{ij})) \\ &+ (\mu(U_{ij}) - \hat{\mu}(U_{ij}))(\mu(U_{il}) - \hat{\mu}(U_{il})). \end{aligned}$$

Using our finite moment assumptions on Y_{ij} (Assumption A1) and our result in Theorem 4.1, part (a), we have that

$$\begin{aligned} \hat{C}_{ijl} = C_{ijl} &+ \mathcal{O}_p(1)\mathcal{O}_p \left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}} \right) \\ &+ \mathcal{O}_p(1)\mathcal{O}_p \left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}} \right) \\ &+ \left(\mathcal{O}_p \left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}} \right) \right)^2 = C_{ijl} + \mathcal{O}_p \left(h_\mu^2 + \frac{1}{\sqrt{nm\, h_\mu}} + \frac{1}{\sqrt{n}} \right), \end{aligned}$$

uniformly for all $j \neq l \in \{1,\ldots,m\}$ and $i \in \{1,\ldots,n\}$. Therefore

$$\sup_{(u,v)\in[a,b]^2} |\hat{\gamma}(u,v;h_\mu) - \gamma(u,v)| = \mathcal{O}_p \left(h_\gamma^2 + h_\mu^2 + \frac{1}{\sqrt{nM\, h_\gamma^2}} + \frac{1}{\sqrt{nm\, h_\mu^2}} + \frac{1}{\sqrt{n}} \right).$$

Proof of Theorem 4.1, parts (c) and (d): Part (c) follows directly from inequality $\sup_{k\geq 1} |\hat{\lambda}_k^O - \lambda_k^O| \leq ||\hat{\gamma} - \gamma||_2$; see inequality (4.43) in Bosq (2000). Part (d) follows directly from Lemma 4.3 in Bosq (2000).

In the following let $O := O_i = [A_i, B_i]$ for some $i \in 1, \ldots, n$. By assumption of Theorem 4.2 we have $B_i - A_i \geq \ell_{\min}$, and recall that by Assumption (A1) the structure of a function X_i, to be observed on O_i, does not depend on the specific interval O_i.

For the proof of Theorem 4.2 we need some additional lemmas. Generally note that under the assumed choice of bandwidths we have $r_\mu + r_\gamma \asymp r_{mn}$ for $r_{mn} = \frac{1}{\min\{n^{1/2},(nM)^{1/3}\}}$, since for all n and $m \asymp n^\theta$ sufficiently large we have $(mn)^{2/5} \geq \min\{n^{1/2}, (nM)^{1/3}\}$.

Recall that $\hat{\tilde{\phi}}_k^O(u) = \frac{\langle \hat{\phi}_k^O, \hat{\gamma}_u \rangle_2}{\hat{\lambda}_k^O}$ and $\tilde{\phi}_k^O(u) = \frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O}$ for $u \in O \cup M$, where in the particular case of $u \in O$ we additionally have $\hat{\tilde{\phi}}_k^O(u) = \hat{\phi}_k^O(u)$ and $\tilde{\phi}_k^O(u) = \phi_k^O(u)$. Also recall that $\bar{K}_{mn}^{a_O+3/2} r_{mn} = O(1)$ and that by (A6) we have $\delta_k^O = O(k^{-a_O-1})$ as well as $1/\delta_k^O = O(k^{a_O+1})$.

LEMMA A.3. *Under the assumptions of Theorem 4.2 we have for all* $1 \leq k \leq K \leq \bar{K}_{mn}$

$$\sup_{u \in O \cup M} \sup_{1 \leq k \leq K} \delta_k^O |\hat{\tilde{\phi}}_k^O(u) - \tilde{\phi}_k^O(u)| = \mathcal{O}_p\left(K^{1/2} r_{mn}\right), \tag{58}$$

Proof of Lemma A.3: Using results (b) and (c) of Theorem 4.1 we obtain

$$\hat{\tilde{\phi}}_k^O(u) = \frac{\langle \hat{\phi}_k^O, \gamma_u \rangle_2}{\hat{\lambda}_k^O} + \mathcal{R}_{1,k}(u), \quad \sup_{u \in O \cup M} \sup_{1 \leq k \leq K} \lambda_k^O \mathcal{R}_{1,k}(u) = \mathcal{O}_p\left(r_{mn}\right). \tag{59}$$

But by the established properties (in particular (12) in Theorem 2.3) of our operator we have $\gamma(u,v) = \sum_{j=1}^\infty \lambda_j^O \tilde{\phi}_j^O(u) \phi_j^O(v)$ for all $u \in O \cup M$ and $v \in O$. Hence

$$\frac{\langle \hat{\phi}_k^O, \gamma_u \rangle_2}{\hat{\lambda}_k^O} = \frac{1}{\hat{\lambda}_k^O} \sum_{j=1}^\infty \lambda_j^O \tilde{\phi}_j^O(u) \langle \hat{\phi}_k^O, \phi_j^O \rangle_2 \tag{60}$$

Now note that for all $j \geq 1$

$$\begin{aligned} \lambda_j^O \langle \hat{\phi}_k^O, \phi_j^O \rangle_2 &= \int_{O^2} \gamma(u,v) \hat{\phi}_k^O(u) \phi_j^O(v) du dv \\ &= \hat{\lambda}_k^O \langle \hat{\phi}_k^O, \phi_j^O \rangle_2 + \int_{O^2} (\gamma(u,v) - \hat{\gamma}(u,v)) \hat{\phi}_k^O(u) \phi_j^O(v) du dv \end{aligned} \tag{61}$$

Let $v_{\hat{\gamma},\gamma,k}(u) := \int_O (\gamma(u,v) - \hat{\gamma}(u,v)) \hat{\phi}_k^O(u) du$. By the orthonormality of the system $\phi_1^O, \phi_2^O, \ldots$ of eigenfunctions, the Cauchy-Schwarz inequality, and

Theorem 4.1 we have

$$\begin{aligned}
&\sup_{1\le k\le K}\sum_{j=1}^{\infty}\left(\int_{O^2}(\gamma(u,v)-\hat{\gamma}(u,v))\hat{\phi}_k^O(u)\phi_j^O(v)dudv\right)^2\\
&=\sup_{1\le k\le K}\sum_{j=1}^{\infty}\langle v_{\hat{\gamma},\gamma,k},\phi_j^O\rangle_2^2\le\sup_{1\le k\le K}\langle v_{\hat{\gamma},\gamma,k},v_{\hat{\gamma},\gamma,k}\rangle_2^2\\
&\le\sup_{1\le k\le K}\sup_{u,v\in O}|\gamma(u,v)-\hat{\gamma}(u,v)|^2=\mathcal{O}_p\left(r_{mn}^2\right). \qquad (62)
\end{aligned}$$

And since by Assumption (A7) $\sup_j \sup_{u\in O\cup M}|\tilde{\phi}_j^O(u)|\le D_O<\infty$ the Cauchy-Schwarz inequality yields

$$\sup_{u\in O\cup M}\sup_{1\le k\le K}\sum_{j=1}^{K}|\tilde{\phi}_j^O(u)\int_{O^2}(\hat{\gamma}(u,v)-\gamma(u,v))\hat{\phi}_k^O(u)\phi_j^O(v)dudv|=\mathcal{O}_p\left(K^{1/2}r_{mn}\right), \qquad (63)$$

By (c) of Theorem 4.1, (61), (63), $\langle\phi_k^O-\phi_k^O,\phi_j^O\rangle_2=0$ for $j\neq k$, and $\langle\phi_k^O-\phi_k^O,\phi_k^O\rangle_2=1$ relation (60) can thus be rewritten in the form

$$\begin{aligned}
\frac{\langle\hat{\phi}_k^O,\gamma_u\rangle_2}{\hat{\lambda}_k^O}&=\tilde{\phi}_k^O(u)+\sum_{j=1}^{K}\tilde{\phi}_j^O(u)\langle\hat{\phi}_k^O-\phi_k^O,\phi_j\rangle_2\\
&\quad+\frac{1}{\lambda_k^O}\sum_{j=K+1}^{\infty}\lambda_j^O\tilde{\phi}_j^O(u)\langle\hat{\phi}_k^O-\phi_k^O,\phi_j^O\rangle_2+\mathcal{R}_{2,k}(u), \qquad (64)\\
&\text{where }\sup_{u\in O\cup M}\sup_{1\le k\le K}\lambda_k^O\mathcal{R}_{2,k}(u)=\mathcal{O}_p\left(K^{1/2}r_{mn}\right),
\end{aligned}$$

Result (d) of Theorem 4.1 additionally implies

$$\sup_{1\le k\le K}(\delta_k^O)^2\sum_{j=1}^{\infty}\langle\hat{\phi}_k^O-\phi_k^O,\phi_j^O\rangle_2^2\le(\delta_k^O)^2\langle\hat{\phi}_k^O-\phi_k^O,\hat{\phi}_k^O-\phi_k^O\rangle_2^2=\mathcal{O}_p\left(r_{mn}^2\right), \qquad (65)$$

and, similar to (63), the Cauchy-Schwarz inequality leads to

$$\sup_{u\in O\cup M}\sup_{1\le k\le K}\delta_k^O|\sum_{j=1}^{K}\tilde{\phi}_j^O(u)\langle\hat{\phi}_k^O-\phi_k^O,\phi_j\rangle_2|=\mathcal{O}_p\left(K^{1/2}r_{mn}\right). \qquad (66)$$

By our assumptions on the sequence of eigenvalues we have for all $u \in O \cup M$

$$\sum_{j=K+1}^{\infty} (\lambda_j^O)^2 \tilde{\phi}_j^O(u)^2 \leq D_O^2 \sum_{j=K+1}^{\infty} (\lambda_j^O)^2$$
$$= O(\sum_{j=K+1}^{\infty} j^{-2a_O}) = O(K^{-2a_O+1}) = O(K(\lambda_K^O)^2)$$

When combining this result with (65), a further application of the Cauchy-Schwarz inequality yields

$$\sup_{u \in O \cup M} \sup_{1 \leq k \leq K} |\delta_k^O \frac{1}{\lambda_k^O} \sum_{j=K+1}^{\infty} \lambda_j^O \tilde{\phi}_j^O(u) \langle \hat{\phi}_k^O - \phi_k^O, \phi_j^O \rangle_2 |$$

(67)

$$\leq \delta_k^O \frac{1}{\lambda_k^O} (D_O^2 \sum_{K+1}^{\infty} (\lambda_j^O)^2)^{1/2} (\sum_{j=k+1}^{\infty} \langle \hat{\phi}_k^O - \phi_k^O, \phi_j^O \rangle_2^2)^{1/2} = \mathcal{O}_p\left(K^{1/2} r_{mn}\right).$$

Since $\frac{\delta_k^O}{\lambda_k^O} \to 0$ as $k \to \infty$, the desired result is an immediate consequence of (59) - (67).

A technical difficulty in the proof of Theorem 4.2 consists in the fact that $\hat{\phi}_k^O$ and the observations (Y_{ij}, U_{ij}) corresponding to the selected $i \in 1, \ldots, n$. But let $\hat{\gamma}_{-i}(t,s) \equiv \hat{\gamma}_{-i}(t,s,h_\gamma)$ denote the estimate of the covariance matrix when eliminating the m observations $\{(Y_{ij}, U_{ij})\}_{j=1,\ldots,m}$ from the sample, and let $\hat{\lambda}_{k,-i}^O$ and $\hat{\phi}_{k,-i}^O$, $k = 1, 2, \ldots$, denote eigenvalues and eigenfunctions of the corresponding covariance operator. Although in our time series context there may still exist dependencies between X_i and $\hat{\phi}_{k,-i}^O$, our assumptions imply that then all $\hat{\phi}_{k,-i}^O$ are independent of the particular samples $\{\epsilon_{ij}\}_{j=1,\ldots,m}$ and $\{U_{ij}\}_{j=1,\ldots,m}$. The following Lemma now provides bounds for the differences between $\hat{\tilde{\phi}}_k^O(u) = \frac{\langle \hat{\phi}_k^O, \hat{\gamma}_u \rangle_2}{\hat{\lambda}_k^O}$ and $\hat{\tilde{\phi}}_{k,-i}^O(u) = \frac{\langle \hat{\phi}_{k,-i}^O, \hat{\gamma}_{-i;u} \rangle_2}{\hat{\lambda}_{k,-i}^O}$, where $\hat{\gamma}_{-i;u}(v) := \hat{\gamma}_{-i}(u,v;h_\gamma)$.

LEMMA A.4. *Under the assumptions of Theorem 4.2 we have for all* $1 \leq k \leq K \leq \bar{K}_{mn}$

a) $\sup_{(u,v) \in [a,b]^2} \left| \hat{\gamma}(u,v;h_\gamma) - \hat{\gamma}_{-i}(u,v;h_\gamma) \right| = \mathcal{O}_p\left(\frac{1}{n^{1/2} r_{mn}}\right)$.

b) $\sup_{k \leq K} \delta_k^O \|\hat{\phi}_{k,-i}^O - \hat{\phi}_k^O\|_2 = \mathcal{O}_p\left(\frac{1}{n^{1/2} r_{mn}}\right)$, $\sup_{k \leq K} |\hat{\lambda}_{k,-i}^O - \hat{\lambda}_k^O| = \mathcal{O}_p\left(\frac{1}{n^{1/2} r_{mn}}\right)$

c) $\sup_{u \in O} \sup_{1 \leq k \leq K} \delta_k^O |\hat{\phi}_{k,-i}(u) - \hat{\phi}_k(u)| = \mathcal{O}_p\left(\frac{K^{a_O+3/2}}{n^{1/2} r_{mn}}\right) = \mathcal{O}_p(\frac{1}{n^{1/2}})$

Proof of Lemma A.4: Based on the definitions and techniques introduced in the proof of Theorem 4.1 and Lemma A.2 it is immediately seen that uniform rates of convergence of $\hat{\gamma}(u,v;h_\gamma) - \hat{\gamma}_{-i}(u,v;h_\gamma)$ can be derived by considering the following difference: For any $(u,v) \in O^2$ the estimator $\hat{\sigma}(u,v)$

$$\sup_{(u,v)\in[a,b]^2} \left| \Theta_{q,nM}(u,v;h_\gamma) - \Theta^{-i}_{q,nM}(u,v;h_\gamma) \right| \leq$$
$$\sup_{(u,v)\in[a,b]^2} \left| \frac{1}{nMh_\gamma^2} \sum_{j\neq l} \kappa\left(\frac{U_{ij}-u}{h_\gamma}\right) \kappa\left(\frac{U_{il}-v}{h_\gamma}\right) \vartheta_q\left(U_{ij}-u, U_{i^*l}-u, C_{i^*jl}\right) \right| +$$
$$\sup_{(u,v)\in[a,b]^2} \left| \frac{1}{n(n-1)} \frac{1}{Mh_\gamma^2} \sum_{k\neq i, j\neq l} \kappa\left(\frac{U_{kj}-u}{h_\gamma}\right) \kappa\left(\frac{U_{kl}-v}{h_\gamma}\right) \vartheta_q\left(U_{kj}-u, U_{kl}-u, C_{kjl}\right) \right|.$$

Using similar arguments as in the proof of Lemma A.1, leads to

$$\sup_{(u,v)\in[a,b]^2} \left| \Theta_{q,nM}(u,v;h_\gamma) - \Theta^{-i}_{q,nM}(u,v;h_\gamma) \right| = \mathcal{O}_p\left(\frac{1}{\sqrt{n^2 M h_\gamma^2}} + \frac{1}{\sqrt{n^2}} \right)$$

which implies

$$\sup_{(u,v)\in[a,b]^2} \left| \hat{\gamma}(u,v;h_\gamma) - \hat{\gamma}_{-i}(u,v;h_\gamma) \right| = \mathcal{O}_p\left(\frac{1}{\sqrt{n^2 M h_\gamma^2}} + \frac{1}{\sqrt{n^2}} \right),$$

and assertion a) of the Lemma is an immediate consequence.

The inequalities used to prove (c) and (d) of Theorem 4.1 now lead to $\sup_{k\geq 1} |\hat{\lambda}^O_{k,-i} - \hat{\lambda}^O_k| = \mathcal{O}_p\left(n^{-1/2} r_{mn}\right)$ and $\sup_{1\leq k\leq K} \hat{\delta}^O_{k,-i} \|\hat{\phi}^O_{k,-i} - \hat{\phi}^O_k\|_2 = \mathcal{O}_p\left(n^{-1/2} r_{mn}\right)$, where $\hat{\delta}^O_{k,-i} := \min_{j\neq k}\{\hat{\lambda}^O_j - \hat{\lambda}^O_k\}$. By (c) of Theorem 4.1, our assumptions on λ_j, and $k \leq K \leq \bar{K}_{mn}$ with $\bar{K}_{mn}^{a_O+3/2} r_{mn} = O(1)$ we additionally have $\hat{\delta}^O_{k,-i}(\delta^O_{k,-i})^{-1} = 1 + o_p(1)$. This proves assertion b) of the Lemma.

Furthermore, we have $\hat{\phi}^O_k(u) = \frac{\langle \hat{\phi}^O_k, \hat{\gamma}_u\rangle_2}{\hat{\lambda}^O_k}$ as well as $\hat{\phi}^O_{k,-i}(u) = \frac{\langle \hat{\phi}^O_{k,-i}, \hat{\gamma}_{-i;u}\rangle_2}{\hat{\lambda}^O_{k,-i}}$ for $u \in O$. The difference can be rewritten in the form

$$\begin{aligned} \hat{\phi}^O_k(u) - \hat{\phi}^O_{k,-i}(u) =& \frac{\langle \hat{\phi}^O_k - \hat{\phi}^O_{k,-i}, \gamma_u\rangle_2}{\hat{\lambda}^O_k} + \frac{\langle \hat{\phi}^O_k - \hat{\phi}^O_{k,-i}, \hat{\gamma}_u - \gamma_u\rangle_2}{\hat{\lambda}^O_k} \\ &+ \frac{\langle \hat{\phi}^O_{k,-i}, \hat{\gamma}_u - \hat{\gamma}_{-i;u}\rangle_2}{\hat{\lambda}^O_k} + \frac{(\hat{\lambda}^O_{k,-i} - \hat{\lambda}^O_k)\langle \hat{\phi}^O_{k,-i}, \hat{\gamma}_{-i;u}\rangle_2}{\hat{\lambda}^O_k \hat{\lambda}^O_{k,-i}}. \end{aligned} \tag{68}$$

When analyzing the terms in (68) first note that by assertions a) and b) of the lemma, and by (c) of Theorem 4.1

$$\sup_{u\in O}\sup_{1\le k\le K}\left|\frac{\langle\hat\phi^O_{k,-i},\hat\gamma_u-\hat\gamma_{-i;u}\rangle_2}{\hat\lambda^O_{k,-i}}\right|=\mathcal{O}_p\left(\frac{1}{\lambda^O_K}\frac{r_{mn}}{n^{1/2}}\right)=\mathcal{O}_p\left(\frac{K^{a_O}r_{mn}}{n^{1/2}}\right)$$

(69)

$$\sup_{u\in O}\sup_{1\le k\le K}\left|\frac{\langle\hat\phi^O_{k}-\hat\phi^O_{k,-i},\hat\gamma_u-\gamma_{u}\rangle_2}{\hat\lambda^O_{k}}\right|=\mathcal{O}_p\left(\frac{K^{2a_O+1}r^2_{mn}}{n^{1/2}}\right)=o_p\left(\frac{K^{a_O}r_{mn}}{n^{1/2}}\right)$$

The differences between the eigenfunctions $\hat\phi^O_{k,-i}(u)$ and $\hat\phi^O_k(u)$ reflect the elimination of one single curve, and it is immediately clear that the convergence results of Theorem 4.1 and all arguments of Lemma A.3 remain valid when considering estimated covariances $\hat\gamma_{-i}(u,v)$, eigenvalues $\hat\lambda^O_{k,-i}$, and eigenfunctions $\hat\phi^O_{k,-i}(u)$ of the reduced sample. It thus follows from Lemma A.3 and our assumption on $K\le\bar K_{mn}$ that $\frac{\langle\hat\phi^O_{k,-i},\hat\gamma_{-i;u}\rangle_2}{\hat\lambda^O_{k,-i}}$ is asymptotically uniformly bounded over all $u\in O\cup M$ and $k\le K$. Hence,

(70)

$$\sup_{u\in O\cup M}\sup_{1\le k\le K}\left|\frac{(\hat\lambda^O_{k,-i}-\hat\lambda^O_k)\langle\hat\phi^O_{k,-i},\hat\gamma_{-i;u}\rangle_2}{\hat\lambda^O_k\hat\lambda^O_{k,-i}}\right|=\mathcal{O}_p\left(\frac{1}{\lambda^O_K}\frac{r_{mn}}{n^{1/2}}\right)=\mathcal{O}_p\left(\frac{K^{a_O}r_{mn}}{n^{1/2}}\right)$$

The first term on the right side of (68) can now be analyzed by generalizing the arguments of Lemma A.3. Similar to (60) we obtain

$$\frac{\langle\hat\phi^O_k-\hat\phi^O_{k,-i},\gamma_u\rangle_2}{\hat\lambda^O_k}=\frac{1}{\hat\lambda^O_k}\sum_{j=1}^{\infty}\lambda^O_j\tilde\phi^O_j(u)\langle\hat\phi^O_k-\hat\phi^O_{k,-i},\phi^O_j\rangle_2,$$

while (61) becomes

$$\begin{aligned}\lambda^O_j\langle\hat\phi^O_k-\hat\phi^O_{k,-i},\phi^O_j\rangle_2&=(\hat\lambda^O_k-\hat\lambda^O_{k,-i})\langle\hat\phi^O_k-\hat\phi^O_{k,-i},\phi^O_j\rangle_2\\&+\int_{O^2}(\hat\gamma(u,v)-\gamma(u,v))(\hat\phi^O_k(u)-\hat\phi^O_{k,-i}(u))\phi^O_j(v)dudv\\&+\int_{O^2}(\hat\gamma(u,v)-\hat\gamma_{-i}(u,v))\hat\phi^O_{k,-i}(u)\phi^O_j(v)dudv\end{aligned}$$

Using result b), a straightforward generalization of the arguments given by

(62) and (63) then leads to

$$\frac{\langle\hat{\phi}_k^O - \hat{\phi}_{k,-i}^O, \gamma_u\rangle_2}{\hat{\lambda}_k^O} = \frac{(\hat{\lambda}_k^O - \hat{\lambda}_{k,-i}^O)}{\hat{\lambda}_k^O}\sum_{j=1}^{K}\tilde{\phi}_j^O(u)\langle\hat{\phi}_k^O - \hat{\phi}_{k,-i}^O, \phi_j\rangle_2$$
$$+ \frac{1}{\lambda_k^O}\sum_{j=K+1}^{\infty}\lambda_j^O\tilde{\phi}_j^O(u)\langle\hat{\phi}_k^O - \hat{\phi}_{k,-i}^O, \phi_j^O\rangle_2 + \mathcal{R}_{1,k}^{(-i)}(u),$$
$$\text{where } \sup_{u\in O\cup M}\sup_{1\le k\le K}\mathcal{R}_{1,k}^{(-i)}(u) = \mathcal{O}_p\left(\frac{K^{a_O}r_{mn}}{n^{1/2}}\right),$$

and proceeding similar to (65) - (67) we can conclude that

$$\sup_{u\in O\cup M}\sup_{1\le k\le K}\left|\frac{\langle\hat{\phi}_k^O - \hat{\phi}_{k,-i}^O, \gamma_u\rangle_2}{\hat{\lambda}_k^O}\right| = \mathcal{O}_p\left(\frac{K^{a_O+3/2}r_{mn}}{n^{1/2}}\right) \tag{71}$$

Assertion c) of the lemma is now an immediate consequence of (68) - (71).

Proof of Theorem 4.2.
We have to consider the asymptotic behavior of

$$[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) = \hat{\mu}(u;h_\mu) + \sum_{k=1}^{K}\hat{\xi}_{ik}^O\hat{\tilde{\phi}}_k^O(u), \quad u\in O\cup M.$$

Rates of convergence of $\hat{\mu}(u;h_\mu)$ are given by Theorem 4.1 (a), while Lemma A.3 provides rates of convergence for $\hat{\tilde{\phi}}_k^O(u)$. We therefore additionally have to consider convergence of the PC scores

$$\hat{\xi}_{ik}^O = \sum_{j=2}^{m}\hat{\phi}_k^O(U_{i(j)})(Y_{i(j)} - \hat{\mu}(U_{i(j)};h_\mu))(U_{i(j)} - U_{i,(j-1)}),$$

where $U_{i(j)}$, $j = 1,\dots,m$ is the order sample of observation points. By our assumption on h_μ result (a) of Theorem 4.1 directly implies that

$$\hat{\xi}_{ik}^O = \sum_{j=2}^{m}\hat{\phi}_k^O(U_{i(j)})(Y_{i(j)} - \mu(U_{i(j)}))(U_{i(j)} - U_{i,(j-1)}) + \mathcal{R}_{1,k,i},$$
$$\text{where } \sup_{1\le k\le K}|\mathcal{R}_{1,k,i}| = \mathcal{O}_p\left((mn)^{-2/5}\right) \tag{72}$$

A technical difficulty of subsequent analysis consists in the fact that $\hat{\phi}_k^O$ and $\{(Y_{ij},U_{ij})\}_{j=1,\dots,m}$ are correlated. As defined above, we thus eliminate the m

observations representing curve X_i and consider the eigenfunction $\hat{\phi}^O_{k,-i}(u)$ of the reduced sample. We can then infer from result c) of Lemma A.4 that

$$\mathcal{R}_{2,k,i} := |\sum_{j=2}^{m}(\hat{\phi}^O_{k,-i}(U_{i(j)}) - \hat{\phi}^O_k(U_{i(j)}))(Y_{i(j)} - \mu(U_{i(j)}))(U_{i(j)} - U_{i,(j-1)})|$$

$$\text{satisfies} \sup_{1\leq k\leq K} |\mathcal{R}_{2,k,i}| = \mathcal{O}_p(n^{-1/2}) \tag{73}$$

Recall that $Y_{ij} = X_i(U_{ij}) + \epsilon_{ij}$ and that $\hat{\phi}^O_{k,-i}$ is independent of ϵ_{ij}, U_{ij}. Since $k \leq K \leq \bar{K}_{mn}$, result c) of Lemma A.4 also implies that $\hat{\phi}^O_{k,-i}$, $k \leq K$, are asymptotically uniformly bounded over all $u \in O \cup M$ and $k \leq K$. By our assumptions on the error term we can now immediately infer that

$$\begin{aligned}
&\sum_{j=2}^{m} \hat{\phi}^O_{k,-i}(U_{i(j)})(Y_{i(j)} - \mu(U_{i(j)})(U_{i(j)} - U_{i,(j-1)}) \\
&= \sum_{j=2}^{m} \hat{\phi}^O_{k,-i}(U_{i(j)})(X_i(U_{i(j)}) - \mu(U_{i(j)}))(U_{i(j)} - U_{i,(j-1)}) + \mathcal{R}_{3,k,i},
\end{aligned}$$

$$\text{with } \mathbb{E}\left(\mathcal{R}^2_{3,k,i} |\ \hat{\gamma}_{-i}\right) \leq D_{k,i}\frac{1}{m} \text{ and } \sup_{1\leq k\leq K} |D_{k,i}| = \mathcal{O}_p(1) \tag{74}$$

Let $X^*_i := X_i - \mu$, and let $F_{U|O}$ denote the distribution function of U_{ij}. It is then well-known that the random variables $V_{ij} := F_{U|O}(U_{ij})$ are $U(0,1)$-distributed, and $\mathbb{E}(V_{i(j)}) - V_{i(j-1)})) = \frac{1}{m+1}$ while $\mathbb{V}(V_{i(j)}) - V_{i(j-1)})) = \frac{m}{(m+1)^2(m+2)}$. By our assumptions the density $f_{U|O}$ ($O \equiv O_i$) of U_{ij} a Taylor expansion now yields $F^{-1}_{U|O}(V_{i(j)}) - F^{-1}_{U|O}(V_{i(j-1)}) = \frac{1}{m f_{U|O}(F^{-1}_{U|O}(V_{i(j)}))} + \mathcal{R}^*_{4,k,i}$ with $\mathbb{E}(\mathcal{R}^*_{4,k,i}) = \mathcal{O}_p(1/m^2)$. This implies

$$\begin{aligned}
&\sum_{j=2}^{m} \hat{\phi}^O_{k,-i}(U_{i(j)})X^*_i(U_{i(j)})(U_{i(j)} - U_{i,(j-1)}) \\
&= \sum_{j=2}^{m} \hat{\phi}^O_{k,-i}(F^{-1}(V_{i(j)}))X^*_i(F^{-1}(V_{i(j)}))(F^{-1}(V_{i(j)}) - F^{-1}(V_{i(j-1)})) \\
&= \sum_{j=2}^{m} \hat{\phi}^O_{k,-i}(U_{i(j)}))X^*_i(U_{i(j)})\frac{1}{m f(U_{i(j)})} + \mathcal{R}^{**}_{4,k,i}
\end{aligned}$$

with $\mathbb{E}\left(\mathcal{R}^{**}_{4,k,i} |\ \hat{\gamma}_{-i}\right) \leq D^*_{ik}\frac{1}{m}$ for some $D^*_{ik} < \infty$ satisfying $\sup_{1\leq k\leq K} |D^*_{k,i}| =$

$\mathcal{O}_p(1)$. Obviously,

$$\mathbb{E}\left(\sum_{j=2}^{m}\hat{\phi}^O_{k,-i}(U_{i(j)}))X_i^*(U_{i(j)})\frac{1}{mf(U_{i(j)})}|\ \hat{\phi}^O_{k,-i},X_i\right)=\frac{m-1}{m}\int_O\hat{\phi}^O_{k,-i}(u)X_i^*(u)du,$$

and independent of k the conditional variance of this random variable can be bounded by $1/m$. We therefore arrive at

$$\sum_{j=2}^{m}\hat{\phi}^O_{k,-i}(U_{i(j)}))X_i^*(U_{i(j)})(U_{ij}-U_{i,j-1})=\int_O\hat{\phi}^O_{k,-i}(u)X_i^*(u)du+\mathcal{R}_{4,k,i},$$

(75)

$$\text{with }\ \mathbb{E}\left(\mathcal{R}^2_{4,k,i}|\ \hat{\gamma}_{-i}\right)\le D^{**}_{ik}\frac{1}{m},\ \text{ and }\ \sup_{1\le k\le K}|D^{**}_{k,i}|=\mathcal{O}_p(1).$$

Additionally note that Lemma A.3 together with $K^{a_O+3/2}r_{mn}=\mathcal{O}(1)$ implies that $\sup_{u\in O\cup M}\sup_{1\le k\le K}\hat{\tilde{\phi}}^O_k(u)=\mathcal{O}(1)$, while (a) of Theorem 4.1 yields $\sup_{u\in O\cup M}|\mu(u)-\hat{\mu}(u)|=\mathcal{O}_p((mn)^{-2/5})$ We can therefore infer from (72) - (75) that for $u\in O\cup M$

(76)

$$[\widehat{\mathcal{L}}_K(\mathbb{X}^O_i)](u)=\mu(u)+\sum_{k=1}^{K}(\int_O\hat{\phi}^O_{k,-i}(u)X_i^*(u)du)\hat{\tilde{\phi}}^O_k(u)+\mathcal{O}_p\left(\frac{K}{m^{1/2}}+\frac{K}{n^{1/2}}\right).$$

Since $[\mathcal{L}_K(X^O_i)](u)=\mu(u)+\sum_{k=1}^{K}\xi^O_k\tilde{\phi}^O_j(u)$ the next step is to consider the errors $\xi^O_{ik}-\int_O\hat{\phi}^O_{k,-i}(u)X_i^*(u)du$ for $k=1,\dots,K$.

The differences between the eigenfunctions $\hat{\phi}^O_{k,-i}(u)$ and $\hat{\phi}^O_k(u)$ reflect the elimination of one single curve, and it is immediately clear that the convergence results of Theorem 4.1 and all arguments of Lemma A.3 remain valid when considering estimated covariances $\hat{\gamma}_{-i}(u,v)$, eigenvalues $\hat{\lambda}^O_{k,-i}$, and eigenfunctions $\hat{\phi}^O_{k,-i}(u)$ of the reduced sample. Furthermore, note that since $\|\hat{\phi}^O_{k,-i}\|_2=\|\phi^O_j\|_2=1$ implies $\langle\hat{\phi}^O_{k,-i},\phi^O_k\rangle_2=1-\frac{1}{2}\|\hat{\phi}^O_{k,-i}-\phi^O_k\|^2_2$, and recall the Karhunen-Loève decomposition $X_i^*(u)=\sum_{j=1}^{\infty}\xi^O_{ik}\phi^O_k(u)$.

We can thus infer from (61) in Lemma A.3 that

$$\int_O \hat{\phi}^O_{k,-i} X^*_i(u)(u) = \sum_{j=1}^{\infty} \xi^O_{ij} \langle \hat{\phi}^O_{k,-i}, \phi^O_j \rangle_2 = \xi^O_{ik} \langle \hat{\phi}^O_{k,-i}, \phi^O_k \rangle_2 + \sum_{j \neq k} \xi^O_{ij} \langle \hat{\phi}^O_{k,-i}, \phi^O_j \rangle_2$$

$$= \xi^O_{ik} + \sum_{j=1}^{k-1} \xi^O_{ij} \frac{\hat{\lambda}^O_{k,-i}}{\lambda^O_j} \langle \hat{\phi}^O_{k,-i} - \phi^O_k, \phi^O_j \rangle_2 + \sum_{j=k+1}^{\infty} \xi^O_{ik} \langle \hat{\phi}^O_{k,-i} - \phi^O_k, \phi^O_j \rangle_2$$

(77)

$$+ \sum_{j=1}^{k-1} \xi^O_{ik} \frac{1}{\lambda^O_j} \int_{O^2} (\hat{\gamma}_{-i}(u,v) - \gamma(u,v)) \hat{\phi}^O_k(u) \phi^O_j(v) du dv - \frac{\xi^O_{ik}}{2} \| \hat{\phi}^O_{k,-i} - \phi^O_k \|_2^2.$$

By our assumptions on the sequence of eigenvalues we have

(78)

$$Q_{i,k} := \sum_{j=1}^{k-1} \frac{(\xi^O_{ik})^2}{(\lambda^O_j)^2} = \mathcal{O}_p \left(\sum_{j=1}^{k-1} \frac{\mathbb{E}((\xi^O_{ik})^2)}{(\lambda^O_j)^2} \right) = \mathcal{O}_p \left(\sum_{j=1}^{k-1} \frac{1}{\lambda^O_j} \right) = \mathcal{O}_p \left(\sum_{j=1}^{k-1} j^{a_O} \right) = \mathcal{O}_p(k^{a_O+1})$$

Using the Cauchy-Schwary inequality, (78), inequality (65) in Lemma A.3, and (d) of Theorem 4.1 lead to

$$\left| \sum_{j=1}^{k-1} \xi^O_{ij} \frac{\hat{\lambda}^O_{k,-i}}{\lambda^O_j} \langle \hat{\phi}^O_{k,-i} - \phi^O_k, \phi^O_j \rangle_2 \right| \leq \frac{\lambda^O_k Q^{1/2}_{i,k}}{\delta^O_k} R_{5,k,i},$$

(79) $$\text{where } R_{5,k,i} \geq 0 \text{ with } \sup_{1 \leq k \leq K} R_{5,k,i} = \mathcal{O}_p(r_{mn})$$

for all $1 \leq k \leq K$. Hence,

$$\sum_{k=1}^{K} |\hat{\tilde{\phi}}^O_k(u)| \left| \sum_{j=1}^{k-1} \xi^O_{ij} \frac{\hat{\lambda}^O_{k,-i}}{\lambda^O_j} \langle \hat{\phi}^O_{k,-i} - \phi^O_k, \phi^O_j \rangle_2 \right| \leq \left(\sum_{k=1}^{K} \frac{(\lambda^O_k)^2 Q_{i,k}}{(\delta^O_k)^2} \right)^{1/2} \left(\sum_{j=k}^{K} R^2_{5,k,i} \hat{\tilde{\phi}}^O_j(u)^2 \right)^{1/2}$$

(80) $$= \mathcal{O}_p \left(K^{a_O/2+5/2} r_{mn} \right)$$

Since $\mathbb{E}(Q^*_{i,k}) := \mathbb{E}(\sum_{j=k+1}^{\infty} (\xi^O_{ij})^2) = \sum_{j=k+1}^{\infty} \lambda^O_j = O(k^{-a_O+1})$, similar arguments based on inequalities (65) and (62) in Lemma A.3 yield $\sum_{k=1}^{K} |\hat{\tilde{\phi}}^O_k(u)| |\sum_{j=k+1}^{\infty} \xi^O_{ij} \langle \hat{\phi}^O_{k,-i} - \phi^O_k, \phi^O_j \rangle_2| = \mathcal{O}_p \left(K^{a_O/2+5/2} r_{mn} \right)$ as well as $\sum_{k=1}^{K} |\hat{\tilde{\phi}}^O_k(u)| |\sum_{j=1}^{k-1} \xi^O_{ij} \frac{1}{\lambda^O_j} \int_{O^2} (\hat{\gamma}_{-i}(u,v) - \gamma(u,v)) \hat{\phi}^O_k(u) \phi^O_j(v) du dv| = \mathcal{O}_p \left(K^{a_O/2+3/2} r_{mn} \right)$, while by (d) of Theorem 4.1 $\sum_{k=1}^{K} |\hat{\tilde{\phi}}^O_k(u)| \frac{|\xi^O_{ik}|}{2} \| \hat{\phi}^O_{k,-i} - \phi^O_k \|_2^2 = \mathcal{O}_p \left(K^{3a_O/2+3} r^2_{mn} \right)$.

Together with (76) and $K^{a_O+3/2}r_{mn} = \mathcal{O}(1)$ we therefore arrive at

$$[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) = \mu(u) + \sum_{k=1}^{K} \xi_{ik}^O \hat{\tilde{\phi}}_k^O(u) + \mathcal{O}_p\left(K\left(\frac{1}{m^{1/2}} + K^{a_O/2+3/2}r_{mn}\right)\right)$$

(81)

$$= \mathcal{L}_K(X_i^O) + \sum_{k=1}^{K} \xi_{ik}^O(\hat{\tilde{\phi}}_k^O(u) - \tilde{\phi}_k^O(u)) + \mathcal{O}_p\left(K\left(\frac{1}{m^{1/2}} + K^{a_O/2+3/2}r_{mn}\right)\right)$$

By Lemma A.3 we have $\sum_{k=1}^{K}(\delta_k^O)^2(\hat{\tilde{\phi}}_k^O(u) - \tilde{\phi}_k^O(u))^2 = \mathcal{O}(K^2 r_{mn}^2)$, and the Cauchy-Schwarz inequality implies

$$\sum_{k=1}^{K} \xi_{ik}^O(\hat{\tilde{\phi}}_k^O(u) - \tilde{\phi}_k^O(u)) = \mathcal{O}_p\left(\left(\sum_{k=1}^{K} \frac{\mathbb{E}((\xi_{ik}^O)^2)}{(\delta_k^O)^2}\right)^{1/2}\left(\sum_{k=1}^{K}(\delta_k^O)^2(\hat{\tilde{\phi}}_k^O(u) - \tilde{\phi}_k^O(u))^2\right)^{1/2}\right)$$

$$= \mathcal{O}_p\left(\left(\sum_{k=1}^{K} k^{a_O+2}\right)^{1/2}\left(K^2 r_{mn}^2\right)^{1/2}\right) = \mathcal{O}_p\left(K^{a_O/2+5/2}r_{mn}\right)$$

and therefore

$$[\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)](u) = \mathcal{L}_K(X_i^O)](u) + \mathcal{O}_p\left(K\left(\frac{1}{m^{1/2}} + K^{a_O/2+3/2}r_{mn}\right)\right) \tag{82}$$

We finally have to consider the truncation error. Recall that it is assumed that there is a constant $D_O < \infty$ such that $\sup_{u\in O\cup M}\sup_{k\geq 1}|\tilde{\phi}_k^O(u)| \leq D_O$ Since $[\mathcal{L}_K(X_i^O)](u) = \mu(u) + \sum_{k=1}^{\infty}\xi_{ik}^O\tilde{\phi}_k^O(u)$ we have

$$\mathbb{E}\left(\left([\mathcal{L}_K(X_i^O)](u) - \mu(u) - \sum_{k=1}^{K}\xi_{ik}^O\tilde{\phi}_k^O(u)\right)^2\right) = \sum_{k=K+1}^{\infty}\lambda_k\tilde{\phi}_k^O(u)^2 \leq D_O^2\sum_{k=K+1}^{\infty}\lambda_k^O$$

$$= \mathcal{O}\left(\sum_{k=K+1}^{\infty}k^{-a_O}\right) = \mathcal{O}\left(K^{-a_O+1}\right) \tag{83}$$

Result (28) now follows from (82) and (83). When additionally noting that standard arguments imply that the local linear estimator of X_i with bandwidth $h_X \asymp m^{-1/5}$ satisfies $|\widehat{X}_i^O(\vartheta_u; h_X) - X_i(\vartheta_u)| = \mathcal{O}\left(m^{-2/5}\right)$, result (29) follows from (28) and definition of $\widehat{\mathcal{L}}_K^*(\mathbb{X}_i^O)$.

Proof of Proposition 5.1:
For $u \in O_2$ the optimal linear reconstruction of $X_i^{O_2}(u)$, given $X_i^{O_1}$, is $[\mathcal{L}(X_i^{O_1})](u)$, such that $X_i^{O_2}(u) = [\mathcal{L}(X_i^{O_1})](u) + \mathcal{Z}_i^{O_2}(u)$, with $\tilde{X}_i^{O_2}(u) = [\mathcal{L}(X_i^{O_1})](u)$. From result (a) of Theorem 2.3, we know that $\mathcal{Z}_i^{O_2}(u)$ and $X_i^{O_1}(v)$ are uncorrelated for all $u \in O_2$ and all $v \in O_1$. Consequently, by linearity of $\mathcal{L}$, also

$$\mathbb{E}([\mathcal{L}(X_i^{O_1})](u)\,[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)) = 0 \quad \text{for all} \quad u \in O_2,$$

where in $[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)$ we are using that $\mathcal{L}$ is also well defined as a linear mapping from $\mathbb{L}^2(O_2)$ to O_2; see remark to (7).

Therefore,

$$\begin{aligned}
&0 \leq \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u) - [\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) = \\
&= \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right) - 2\,\mathbb{E}\left(X_i^{M_2}(u)[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right) + \mathbb{E}\left(\left([\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) \\
&\Rightarrow -\mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right) \leq \\
&\leq -2\,\mathbb{E}\left(X_i^{M_2}(u)[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right) + \mathbb{E}\left(\left([\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right).
\end{aligned}$$

But then,

$$\begin{aligned}
&\mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) = \\
&= \mathbb{E}\left(\left(X_i^{M_2}(u)\right)^2\right) - 2\,\mathbb{E}\left(X_i^{M_2}(u)[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right) + \mathbb{E}\left(\left([\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) \\
&\geq \mathbb{E}\left(\left(X_i^{M_2}(u)\right)^2\right) - \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right)
\end{aligned}$$

On the other hand, we have that $\mathbb{E}([\mathcal{L}(\tilde{X}_i^{O_2})](u)[\mathcal{L}(\mathcal{Z}_i^{O_2})](u)) = 0$ for all $u \in M_2$, which follows by the same reasoning as used above, since $\tilde{X}_i^{O_2}(u) = [\mathcal{L}(X_i^{O_1})](u)$, with $u \in O_2$, is just another linear transformation of $X_i^{O_1}$ and $X_i^{O_1}(v)$ is known to be uncorrelated with $\mathcal{Z}_i^{O_2}(u)$ for all $u \in O_2$ and $v \in O_1$

by result (a) of Theorem 2.3. Therefore, using also the latter inequality,

$$\mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_2})](u)\right)^2\right) = \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\tilde{X}_i^{O_2})](u) - [\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) =$$
$$\mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\mathcal{Z}_i^{O_2})](u)\right)^2\right) - 2\,\mathbb{E}\left(X_i^{M_2}(u)\,[\mathcal{L}(\tilde{X}_i^{O_2})](u)\right) + \mathbb{E}\left(\left([\mathcal{L}(\tilde{X}_i^{O_2})](u)\right)^2\right)$$
$$\geq \mathbb{E}\left(\left(X_i^{M_2}(u)\right)^2\right) - \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right)$$
$$- 2\,\mathbb{E}\left(X_i^{M_2}(u)\,[\mathcal{L}(\tilde{X}_i^{O_2})](u)\right) + \mathbb{E}\left(\left([\mathcal{L}(\tilde{X}_i^{O_2})](u)\right)^2\right) =$$
$$= \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\tilde{X}_i^{O_2})](u)\right)^2\right) - \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right)$$
$$\Rightarrow \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(\tilde{X}_i^{O_2})](u)\right)^2\right) \leq$$
$$\leq \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_1})](u)\right)^2\right) + \mathbb{E}\left(\left(X_i^{M_2}(u) - [\mathcal{L}(X_i^{O_2})](u)\right)^2\right)$$

REFERENCES

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*, 1. ed. *Springer Series in Statistics*. Springer.

Alois Kneip and Dominik Liebl
Statistische Abteilung
University of Bonn
Adenauerallee 24-26
53113 Bonn, Germany
E-mail: akneip@uni-bonn.de
E-mail: dliebl@uni-bonn.de