

ON THE OPTIMAL RECONSTRUCTION OF PARTIALLY OBSERVED FUNCTIONAL DATA

BY ALOIS KNEIP* AND DOMINIK LIEBL**

Statistische Abteilung, University of Bonn, *akneip@uni-bonn.de; **dliabl@uni-bonn.de

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_i , $i = 1, \dots, n$, discretization points. We derive rates of consistency for our nonparametric estimation procedures using a double asymptotic. For data situations, as in our real data application where m_i is considerably smaller than n , we show that our functional principal components based estimator can provide better rates of convergence than conventional nonparametric smoothing methods.

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, that is, a ceteris-paribus analysis of price effects with respect to changes in electricity demand (cf. Hirth (2013), Weigt (2009)). 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, \dots, 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 $\|X_i\|_2^2 = \int_a^b (X_i(x))^2 dx$. We denote the observed and missing parts of X_i by $X_i^{O_i}$ and $X_i^{M_i}$, where

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

and where $O_i = [A_i, B_i] \subseteq [a, b]$ is a random subinterval, independent from X_i with $B_i - A_i > 0$ almost surely. In our theoretical part (Section 2) we also allow for the general case, where O_i consists of multiple subintervals of $[a, b]$. 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 :

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

Received September 2018; revised May 2019.

MSC2010 subject classifications. 62M20, 62H25, 62G05, 62G08.

Key words and phrases. Functional data analysis, functional principal components, incomplete functions.

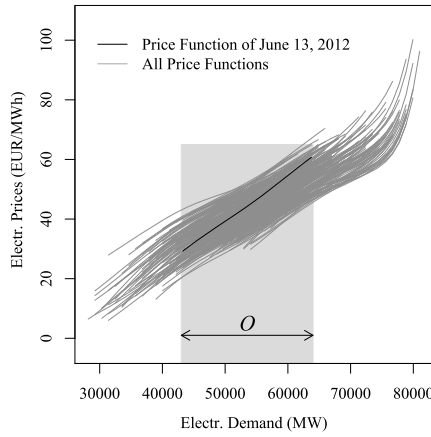


FIG. 1. Partially observed electricity price functions $X_i^O(u)$ with $u \in O_i \subseteq [a, b]$.

typical realizations of X_i^O and O are shown in Figure 1. We denote the inner product and norm of $\mathbb{L}^2(O)$ as $\langle \cdot, \cdot \rangle_2$ and $\| \cdot \|_2$; 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 $\langle x^O, y^O \rangle_2 = \int_O x(u)y(u) du$. Throughout the introduction and Section 2, we consider centered random functions, that is, $\mathbb{E}(X_i(u)) = \mu(u)$ with $\mu(u) = 0$ for all $u \in [a, b]$.

Our object of interest is the following linear reconstruction problem:

$$(1) \quad X_i^M = L(X_i^O) + Z_i,$$

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 $L : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(M)$ which minimizes the mean squared error loss $\mathbb{E}[(X_i^M(u) - L(X_i^O)(u))^2]$ at any $u \in M$.

The case of partially observed functional data was initially considered in the applied work of Liebl (2013) and the theoretical works of Goldberg, Ritov and Mandelbaum (2014) and Kraus (2015). The work of Gromenko, Kokoszka and Sojka (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 Fréchet-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, that is, 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 conceptual work. Hilbert–Schmidt operators on \mathbb{L}^2 spaces correspond to linear regression operators,

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

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 points¹ $\vartheta \in \partial M$ of M . For any optimal reconstruction

¹The 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 .

operator L , it must hold that the “first reconstructed” value, $L(X_i^O)(\vartheta)$, connects with the “last observed” value, $X_i^O(\vartheta)$, that is, that

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

There is no hope, though, of finding a slope function $\beta(\vartheta, \cdot) \in \mathbb{L}^2(O)$ that fulfills the equation $X_i^O(\vartheta) = \int_O \beta(\vartheta, v) X_i^O(v) dv$ (the Dirac- δ function is not an element of $\mathbb{L}^2(O)$). It is therefore 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 finite variance $\mathbb{V}(L(X_i^O)(u)) < \infty$ and thus $\mathbb{P}(|L(X_i^O)(u)| < \infty) = 1$ 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

$$(3) \quad \mathcal{L}(X_i^O)(u) = \sum_{k=1}^{\infty} \frac{\xi_{ik}^O \mathbb{E}[X_i^M(u) \xi_{ik}^O]}{\lambda_k^O} = \sum_{k=1}^{\infty} \xi_{ik}^O \frac{\langle \phi_k^O, \gamma_u \rangle_2}{\lambda_k^O},$$

where $(\phi_k^O, \lambda_k^O)_{k \geq 1}$ denote the pairs of orthonormal eigenfunctions and nonzero eigenvalues of the covariance operator $\Gamma^O(x)(u) = \int \gamma^O(u, v)x(v) dv$ with $x \in \mathbb{L}^2(O)$, while $\xi_{ik}^O := \langle \phi_k^O, X_i^O \rangle_2$. Here $\gamma^O(u, v) = \text{Cov}(X_i^O(u), X_i^O(v))$ denotes the covariance function of X_i^O , and $\gamma_u(v) = \gamma(u, v)$ the covariance function $\gamma(u, v) = \text{Cov}(X_i^M(u), X_i^O(v))$.

The general structure of \mathcal{L} in (3) is similar to 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, however, no reason to expect that the optimal reconstruction operator \mathcal{L} satisfies (2). To see the point note that $\mathcal{L}(X_i^O)(u)$ can be represented in the form (2) if and only if the additional square summability condition $\sum_{k=1}^{\infty} \langle \phi_k^O, \gamma_u \rangle_2^2 / (\lambda_k^O)^2 < \infty$ is satisfied for $u \in M$. Only then the series $\sum_{k=1}^L (\langle \phi_k^O, \gamma_u \rangle_2 / \lambda_k^O) \phi_k^O(v)$, $v \in O$, converge as $L \rightarrow \infty$ and define a function $\beta_u := \sum_{k=1}^{\infty} (\langle \phi_k^O, \gamma_u \rangle_2 / \lambda_k^O) \phi_k^O(\cdot) \in \mathbb{L}^2(O)$ such that $\int_O \beta_u(v) X_i^O(v) dv = \sum_{k=1}^{\infty} \xi_{ik}^O \langle \phi_k^O, \beta_u \rangle_2 = \sum_{k=1}^{\infty} \xi_{ik}^O \langle \phi_k^O, \gamma_u \rangle_2 / \lambda_k^O$.

But 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 $\gamma_{\vartheta}^O(\cdot) = \gamma^O(\vartheta, \cdot) = \sum_{k=1}^{\infty} \lambda_k^O \phi_k^O(\vartheta) \phi_k^O(\cdot)$. Plugging this simplification into (3) and using the Karhunen–Loève decomposition of X_i^O implies that $\mathcal{L}(X_i^O)(\vartheta) = \sum_{k=1}^{\infty} \xi_{ik}^O \phi_k^O(\vartheta) = X_i^O(\vartheta)$. This means that our reconstruction operator \mathcal{L} indeed connects the “last observed” value $X_i^O(\vartheta)$ with the “first reconstructed” value $\mathcal{L}(X_i^O)(\vartheta)$. On the other hand, the sum $\sum_{k=1}^L \langle \phi_k^O, \gamma_{\vartheta} \rangle_2^2 / (\lambda_k^O)^2 = \sum_{k=1}^L (\phi_k^O(\vartheta))^2$ will generally tend to infinity as $L \rightarrow \infty$, which violates the additional condition necessary for establishing (2). Therefore, in general, \mathcal{L} does not constitute a regression operator.²

²A frequently used justification of the use of regression operators relies on the Riesz representation theorem which states that any continuous linear functional $L(X_i^O)(u)$ can be represented in the form (2). This argument, however, does not necessarily apply to the optimal linear functional $\mathcal{L}(X_i^O)(u)$ which may not be a continu-

The above arguments indicate that methods for estimating \mathcal{L} should not be based on (2). Any theoretical justification of such procedures has to rely on nonstandard asymptotics avoiding the restrictive assumption that $\sum_{k=1}^\infty \langle \phi_k^O, \gamma_u \rangle_2^2 / (\lambda_k^O)^2 < \infty$. This constitutes a major aspect of our asymptotic theory given in Section 4.

The problem of estimating $\mathcal{L}(X_i^O)$ from real data is considered in Section 3. Motivated by our application, the estimation theory allows for an autocorrelated time series of functional data and considers the practically relevant case where the function parts X_i^O are only observed at m_i many discretization points $(Y_{i1}, U_{i1}), \dots, (Y_{im_i}, U_{im_i})$ with $Y_{ij} = X_i^O(U_{ij}) + \varepsilon_{ij}$, $i = 1, \dots, n$, and $j = 1, \dots, m_i$.

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 $(Y_{i1}, U_{i1}), \dots, (Y_{im_i}, U_{im_i})$. 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 $\gamma_u(v) = \text{Cov}(X_i^M(u), X_i^O(u))$ to $\gamma_u(v) = \text{Cov}(X_i(u), X_i(u))$. This way the empirical counterpart of the truncated sum

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

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 $(Y_{i1}, U_{i1}), \dots, (Y_{im_i}, U_{im_i})$ 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 $u \in M$, that is, $\vartheta_u = A_i$ 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

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

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

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

ous functional $\mathbb{L}^2(O) \rightarrow \mathbb{R}$. In particular, although being a well-defined linear functional, the point evaluation $\mathcal{L}(X_i^O)(\vartheta) = X_i^O(\vartheta)$ is *not continuous*, since for two functions $f, g \in \mathbb{L}^2(O)$ an arbitrarily small \mathbb{L}^2 -distance $\|f - g\|_2$ may go along with a very large pointwise distance $|f(\vartheta) - g(\vartheta)|$ (see the example in Appendix B.1 of the supplementary paper Kneip and Liebl (2019)).

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 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 all sample curves are of similar structure, we show that our functional principal components based estimator achieves almost parametric convergence rates and can provide better rates of convergence than conventional nonparametric smoothing methods, such as, for example, local linear regression.

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 $(u, v) \in [a, b]^2$. 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, 2016), Descary and Panaretos (2019), and Delaigle et al. (2018). 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 (Kneip and Liebl (2019)).

2. Optimal reconstruction of partially observed functions. Let our basic setup be as described in Section 1. Any (centered) random function $X_i^O \in \mathbb{L}^2(O)$ then adopts the well-known Karhunen–Loève (KL) representation

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

with the principal component (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 $\lambda_1^O > \lambda_2^O > \dots > 0$. We want to note that all arguments in this section also apply to the more general case where the observed subdomain $O = \bigcup_{j=1}^J [A_j, B_j]$ consists of a finite number $1 \leq J < \infty$ of mutually disjoint subintervals $[A_j, B_j] \subseteq [a, b]$.

By the classical eigen-equations we have that

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

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” k th basis function:

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

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):

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

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

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

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 $L : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(M)$ 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 reconstruction 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 $L : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(M)$ 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 $H := \{f \in \mathbb{L}^2(O) : \|f\|_H^2 < \infty\}$ 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 , that is, $\mathbb{V}(L(X_i^O)(u)) = \|\alpha_u\|_H^2$.

Let us consider two examples of possible reconstruction operators. While the first example does not belong the class of regression operators, the second example is a regression operator demonstrating the more restrictive model assumptions.

EXAMPLE 1 (Point of impact). Consider $L(X_i^O)(u) = X_i^O(\tau)$, that is, 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

$$\begin{aligned}
 (9) \quad L(X_i^O)(u) &= X_i^O(\tau) = \sum_{k=1}^\infty \xi_{ik}^O \phi_k^O(\tau) = \sum_{k=1}^\infty \frac{\langle X_i^O, \phi_k^O \rangle_2 \lambda_k^O \phi_k^O(\tau)}{\lambda_k^O} \\
 &= \sum_{k=1}^\infty \frac{\langle X_i^O, \phi_k^O \rangle_2 \langle \gamma_\tau, \phi_k^O \rangle_2}{\lambda_k^O} = \langle \gamma_\tau, X_i^O \rangle_H,
 \end{aligned}$$

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

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

$$\begin{aligned}
 (10) \quad L(X_i^O)(u) &= \langle \beta_u, X_i^O \rangle_2 = \sum_{k=1}^\infty \xi_{ik}^O \beta_{u,k} = \sum_{k=1}^\infty \frac{\langle X_i^O, \phi_k^O \rangle_2 \lambda_k^O \beta_{u,k}}{\lambda_k^O} \\
 &= \sum_{k=1}^\infty \frac{\langle X_i^O, \phi_k^O \rangle_2 \langle \alpha_u, \phi_k^O \rangle_2}{\lambda_k^O} = \langle \alpha_u, X_i^O \rangle_H,
 \end{aligned}$$

where $\alpha_u(\cdot) := \sum_{j=1}^\infty \lambda_k^O \beta_{u,k} \phi_k^O(\cdot) \in H$ with $\|\alpha\|_H^2 = \sum_{k=1}^\infty \frac{(\lambda_k^O)^2 \beta_{u,k}^2}{\lambda_k^O} = \sum_{k=1}^\infty \lambda_k^O \beta_{u,k}^2 < \infty$.

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 $\|\alpha_u\|_H^2 = \sum_{k=1}^\infty \langle \alpha_u, \phi_k^O \rangle_2^2 / \lambda_k^O < \infty$ 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 $L : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(M)$ which for all $u \in M$ minimizes

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

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

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

where $\gamma_u(v) = \gamma(u, v)$ for $v \in O$ and $u \in M$. In particular, Theorem 2.2 below shows that $\mathbb{V}(\mathcal{L}(X_i^O)(u)) = \|\gamma_u\|_H^2 < \infty$ for any $u \in M$, that is, 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 $\tilde{L} : \mathbb{L}^2(O) \rightarrow \mathbb{R}$ according to the model $Y_i = \tilde{L}(X_i^O) + \varepsilon_i$. Straightforward generalizations of Theorems 2.2 and 2.3 below then show 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, that is, $\mathbb{V}(\mathcal{L}(X_i^O)(u)) < \infty$ 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 $u \in M$,
- (12) $\mathbb{E}(X_i^O(v)\mathcal{Z}_i(u)) = 0$ and
- (13) $\mathbb{V}(\mathcal{Z}_i(u)) = \mathbb{E}((\mathcal{Z}_i(u))^2) = \gamma(u, u) - \sum_{k=1}^{\infty} \lambda_k^O (\tilde{\phi}_k^O(u))^2$.
- (b) For any linear operator $L : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(M)$ that is a reconstruction operator with respect to X_i^O , according to Definition 2.1,

$$\mathbb{E}((X_i^M(u) - L(X_i^O)(u))^2) \geq \mathbb{V}(\mathcal{Z}_i(u)) \quad \text{for all } u \in M.$$

(c) Assume that the underlying process X_i is Gaussian, and let $X_{i,1}$ and $X_{i,2}$ 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

$$(14) \quad \mathbb{V}(\mathcal{Z}_i(u)) = \frac{1}{2} \mathbb{E}(\mathbb{E}((X_{i,1}(u) - X_{i,2}(u))^2 | X_{i,1}^O = X_{i,2}^O))$$

where $X_{i,1}^O = X_{i,2}^O$ means that $X_{i,1}(v) = X_{i,2}(v)$ 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” (and 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 $u \in [\vartheta, 1]$. Although in this paper we focus on processes that lead to smooth, regularly shaped sample curves, the Brownian motion is of some theoretical interest since it defines a reconstruction operator which obviously does not constitute a regression operator. Also note that $\mathcal{L}(X_i^O)(u) = X_i^O(\vartheta)$ will provide perfect reconstructions if a.s. sample functions $X_i(u)$ are constant for all $u \in M$.

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, that is, $\mathbb{V}(\mathcal{Z}_i(u)) = 0$ 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 $X_i^K(u) = \sum_{k=1}^K \xi_{ik} \phi_k(u)$, $\lambda_{K+1} = \lambda_{K+2} = \dots = 0$, as long as the basis functions ϕ_1, \dots, ϕ_K are linearly 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 $\mathcal{L} : \mathbb{L}^2(O) \rightarrow \mathbb{L}^2(O \cup M)$. 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

$$\begin{aligned} \lim_{u \in M, u \rightarrow \vartheta_u} \mathcal{L}(X_i^O)(u) &= X_i(\vartheta_u) \quad \text{as well as} \\ \lim_{u \in M, u \rightarrow \vartheta_u} \tilde{\phi}_k^O(u) &= \phi_k^O(\vartheta_u), \quad k = 1, 2, \dots \end{aligned}$$

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

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

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

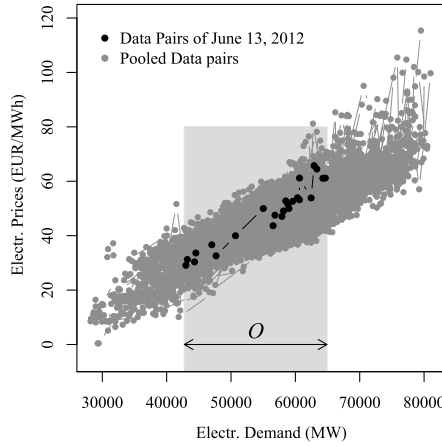


FIG. 2. Scatter plot of the observed data pairs (Y_{ij}, U_{ij}) .

with smooth, similarly shaped trajectories. Note that even if $X_i(u) = \sum_{k=1}^{\infty} \xi_{ik}^O \tilde{\phi}_k^O(u)$ for $u \in M$, the eigenfunctions of X_i^M will usually not coincide with $\tilde{\phi}_k^O$ for $u \in M$, since there is no reason to expect that these functions are mutually orthogonal.

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 $\mathbb{X}_i^O := ((Y_{i1}, U_{i1}), \dots, (Y_{im_i}, U_{im_i}))$ denote the observable data pairs of a function X_i^O , where

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

and $U_{ij} \in O_i$.

For the rest of the paper, we focus on the case where $O_i = [A_i, B_i]$ as in our real data application. However, we give detailed descriptions on how to use our methods in the more general cases where O_i consists of several mutually disjoint subintervals. We consider the case where U_{i1}, \dots, U_{im_i} are iid random variables with strictly positive density over the random subinterval $[A_i, B_i]$, which in practice can be approximated by $A_i \approx \min_{1 \leq j \leq m_i}(U_{ij})$ and $B_i \approx \max_{1 \leq j \leq m_i}(U_{ij})$. Let the error term ε_{ij} be a real iid random variable that is independent from all other stochastic model components and has mean zero and finite (possibly zero) variance $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$ with $0 \leq \sigma^2 < \infty$. Motivated by our real data application we will concentrate on the case that n is considerably larger than m_i , 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

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

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

Remember, however, that our data structure in (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 the 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, for example, 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 $u \in M$, that is, $\vartheta_u = A_i$ 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

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

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

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

(19)

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

$$\begin{aligned} \widehat{\mathcal{L}}_K^*(\mathbb{X}_i^O)(u) &:= \widehat{X}_i^O(\vartheta_u; h_X) + \widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u) - \widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(\vartheta_u) \\ (20) \quad &= \widehat{X}_i^O(\vartheta_u; h_X) + \hat{\mu}(u; h_\mu) - \hat{\mu}(\vartheta_u; h_\mu) \\ &\quad + \sum_{k=1}^K \hat{\xi}_{ik}^O(\hat{\phi}_k^O(u) - \hat{\phi}_k^O(\vartheta_u)), \end{aligned}$$

where \widehat{X}_i^O denotes the LLK estimator of X_i^O (see (21)), $\hat{\mu}$ denotes the LLK estimator of μ (see (22)), $\hat{\gamma}_u$ denotes the LLK estimator of the covariance function (see (23)), $\hat{\phi}_k^O$ and $\hat{\lambda}_k^O$ denote the estimators of the eigenfunctions and eigenvalues (see (24)), and $\hat{\xi}_{ik}$ denote the estimators of the pc-scores (see (25)).

REMARK. Estimator (19) can be directly applied in the general case, where $O = \bigcup_{j=1}^J [A_j, B_j]$ consists of a union of finitely many mutually disjoint subintervals $[A_j, B_j] \subseteq [a, b]$. Estimator (20), however, must be adjusted for this general case as follows. First, consider a point $u \in M$ located between the observed intervals $[A_j, B_j]$ and $[A_{j+1}, B_{j+1}]$ for any $j = 1, \dots, J - 1$. In this case the quantities $\widehat{X}_i^O(\vartheta_u; h_X)$ and $\hat{\phi}_k^O(\vartheta_u)$ in (20) have to be replaced by the linear interpolations $(1 - w_u)\widehat{X}_i^O(B_j; h_X) + w_u\widehat{X}_i^O(A_{j+1}; h_X)$ and $(1 - w_u)\hat{\phi}_k^O(B_j) + w_u\hat{\phi}_k^O(A_{j+1})$ with $w_u = (u - B_j)/(A_{j+1} - B_j)$. Second, for $0 \leq u < A_1$ replace $\widehat{X}_i^O(\vartheta_u; h_X)$ and $\hat{\phi}_k^O(\vartheta_u)$ by $\widehat{X}_i^O(A_1; h_X)$ and $\hat{\phi}_k^O(A_1)$. Third, for $B_J < u \leq 1$ replace $\widehat{X}_i^O(\vartheta_u; h_X)$ and $\hat{\phi}_k^O(\vartheta_u)$ by $\widehat{X}_i^O(B_J; h_X)$ and $\hat{\phi}_k^O(B_J)$.

In our asymptotic analysis (Section 4) we focus on the case of single subintervals $O_i = [A_i, B_i]$ which leads to comprehensible theorems and proofs.

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

$$(21) \quad (\widehat{\beta}_0, \widehat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{j=1}^{m_i} [Y_{ij} - \beta_0 - \beta_1(U_{ij} - u)]^2 K_{h_X}(U_{ij} - u)$$

for $K_h(\cdot) = \kappa(\cdot/h)/h$. The kernel function κ is assumed to be a univariate symmetric pdf with compact support $\text{supp}(\kappa) = [-1, 1]$ such as, for example, the Epanechnikov kernel (see Assumption A5). The usual kernel constants are given by $v_2(\kappa) := \int v^2 \kappa(v) dv$, and $R(\kappa) := \int \kappa(v)^2 dv$.

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

$$(22) \quad (\widehat{\beta}_0, \widehat{\beta}_1) = \arg \min_{\beta_0, \beta_1} \sum_{i=1}^n \sum_{j=1}^{m_i} [Y_{ij} - \beta_0 - \beta_1(U_{ij} - u)]^2 K_{h_\mu}(U_{ij} - u).$$

The LLK estimator $\widehat{\gamma}_u(v) = \widehat{\gamma}(u, v; h_\gamma)$ is defined as $\widehat{\gamma}(u, v; h_\gamma) = \widehat{\beta}_0$, where

$$(23) \quad (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2) = \arg \min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^n \sum_{1 \leq j, l \leq m_i} [\widehat{C}_{ijl} - \beta_0 - \beta_1(U_{ij} - u) - \beta_2(U_{il} - v)]^2 \times K_{h_\gamma}(U_{ij} - u) K_{h_\gamma}(U_{il} - v),$$

with raw-covariance points $\widehat{C}_{ijl} = (Y_{ij} - \widehat{\mu}(U_{ij}))(Y_{il} - \widehat{\mu}(U_{il}))$. 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} .

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

$$(24) \quad \int_O \widehat{\gamma}(u, v; h_\gamma) \widehat{\phi}_k^O(v) dv = \widehat{\lambda}_k^O \widehat{\phi}_k^O(u), \quad u \in O.$$

REMARK. The implementation of (24) can be done as usually by discretizing the smoothed covariance $\widehat{\gamma}(u_r, v_s)$ using regular grid points $(u_r, v_s) \in [a, b]^2$, $r, s \in \{1, \dots, L\}$ (see, for instance, Rice and Silverman (1991)). For approximating the eigenvalues and eigenfunctions of $\Gamma^O(x)(u) = \int \gamma^O(u, v)x(v) dv$ one needs to construct the matrix $(\widehat{\gamma}^O(u_r, v_s))_{r,s}$ from the grid points falling into $[A, B] \times [A, B]$. In the case of several disjoint intervals the matrix must be assembled from the grid points falling into the intervals $[A_j, B_j] \times [A_{j'}, B_{j'}]$, $j, j' \in \{1, \dots, J\}$.

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

$$(25) \quad \widehat{\xi}_{ik}^O = \sum_{j=2}^{m_i} \widehat{\phi}_k^O(U_{i(j)}) (Y_{i(j)} - \widehat{\mu}(U_{i(j)}; h_\mu))(U_{i(j)} - U_{i,(j-1)}),$$

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

In our theoretical analysis we consider $K \equiv K_{nm} \rightarrow \infty$ as the sample size $nm \rightarrow \infty$, where $m \leq m_i$ for all $i = 1, \dots, n$. 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 GCV criterion in Kraus (2015) in order to define an M -specific GCV criterion. For this let \mathcal{C} denote the index set of the completely observed functions $\mathbb{X}_l, l \in \mathcal{C}$, with $[a, b] \approx [\min_{1 \leq j \leq m_i}(U_{lj}), \max_{1 \leq j \leq m}(U_{lj})]$, for instance, with $\min_{1 \leq j \leq m_i}(U_{lj}) \in [a, a + (b - a)/10]$ and $\max_{1 \leq j \leq m_i}(U_{lj}) \in [b - (b - a)/10, b]$ and define the following vectors by partitioning the complete data-vectors into pseudo-missing and pseudo-observed parts:

$$\begin{aligned} \mathbf{Y}_l^M &= (Y_{lj} : j = 1, \dots, m_l; U_{lj} \in M)^\top, \\ \mathbb{X}_l^O &= ((Y_{lj}, U_{lj}) : j = 1, \dots, m_l; U_{lj} \in O)^\top \quad \text{and} \\ \widehat{\mathbf{Y}}_{lK}^M &= (\widehat{\mathcal{L}}_K(\mathbb{X}_l^O)(U_{lj}) : j = 1, \dots, m_l; U_{lj} \in M)^\top. \end{aligned}$$

This allows us to compute the weighted sum of the residual sum of squares $\|\mathbf{Y}_l^M - \widehat{\mathbf{Y}}_{lK}^M\|^2$ for reconstructions over M

$$\text{RSS}_M(K) = \sum_{l \in \mathcal{C}} \|\mathbf{Y}_l^M - \widehat{\mathbf{Y}}_{lK}^M\|^2 / |\mathbf{Y}_l^M|,$$

where $|\mathbf{Y}_l^M|$ is the number of elements in \mathbf{Y}_l^M . The GCV criterion for reconstructing functions over M is

$$(26) \quad \text{GCV}_M(K) = \frac{\text{RSS}_M(K)}{(1 - K/|\mathcal{C}|)^2},$$

where $|\mathcal{C}|$ is the number of elements in \mathcal{C} , that is, the number of complete functions.

4. Asymptotic results. Our theoretical analysis analyzes the reconstruction of an arbitrary sample function X_i satisfying $O \subseteq O_i = [A_i, B_i]$.

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 $U_{i1}|O_i, \dots, U_{im}|O_i$ 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 $u \in [a, b]$ and zero else. The time series $(A_i)_{i=1, \dots, n}$, $(B_i)_{i=1, \dots, n}$, and $(X_i)_{i=1, \dots, n}$ are strictly stationary ergodic (functional) time series with finite fourth moments (i.e., $\mathbb{E}(\|X_i\|_2^4) < \infty$ in the functional case) and autocovariance functions with geometric decay. That is, there are constants $C_A, C_B, C, \dot{C}, \iota_A, \iota_B, \iota, i$ with $0 < C_A, C_B, C, \dot{C} < \infty$ and $0 < \iota_A, \iota_B, \iota, i < 1$, such that $|\text{Cov}(A_i, B_{i+h})| \leq C_A \iota_A^h$, $|\text{Cov}(B_i, B_{i+h})| \leq C_B \iota_B^h$,

$$\sup_{(u,v) \in [a,b]^2} |\gamma_h(u, v)| \leq C \iota^h,$$

and

$$\sup_{(u_1, v_1, u_2, v_2) \in [a,b]^4} |\dot{\gamma}_h((u_1, v_1), (u_2, v_2))| \leq \dot{C} i^h$$

for all $h \geq 0$, where $\gamma_h(u, v) := \text{Cov}(X_{i+h}(u), X_i(v))$ and $\dot{\gamma}_h((u_1, v_1), (u_2, v_2)) := \text{Cov}(X_{i+h}(u_1)X_{i+h}(v_1), X_i(u_2)X_i(v_2))$.

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 $(X_i)_{i=1,\dots,n}$, 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) $nm \rightarrow \infty$ with $m \leq m_i$ for all $i = 1, \dots, n$, where $n \rightarrow \infty$ and $m = m(n) \asymp n^\theta$ with $0 < \theta < \infty$. Here, $a(n) \asymp b(n)$ is used to denote that $(a(n)/b(n)) \rightarrow c$ as $n \rightarrow \infty$, where c is some constant $0 < c < \infty$.

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}) . For $\hat{\gamma}$: All second order derivatives of $\gamma(u, v)$ on $[a, b]^2$, $f_{UU}(u, v)$ 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 $(C_{ijl}, U_{ij}, U_{il})$. Finally, $f_{U|O_i}(u)$ is a.s. continuously differentiable, and $\mathbb{E}(|f'_{U|O_i}(u)|/f_{U|O_i}(u)^2) < \infty$, and X_i is a.s. twice continuously differentiable.

A4 (Bandwidths) For estimating X_i^O : $h_X \rightarrow 0$ and $(mh_X) \rightarrow \infty$ as $m \rightarrow \infty$. For estimating μ : $h_\mu \rightarrow 0$ and $(nmh_\mu) \rightarrow \infty$ as $nm \rightarrow \infty$. For estimating γ : $h_\gamma \rightarrow 0$ and $(n\mathcal{M}h_\gamma) \rightarrow \infty$ as $n\mathcal{M} \rightarrow \infty$, where $\mathcal{M} = m^2 - m$.

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

In Assumption A2, we follow Zhang and Chen (2007) and consider a deterministic sample size $m \rightarrow \infty$, where $m \leq m_i$ for all $i = 1, \dots, n$. As Hall, Müller and Wang (2006), Zhang and Chen (2007) and Zhang and Wang (2016) we do not consider random numbers m_i , but if m_i are random, our theory can be considered as conditional on m_i .

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 $k \rightarrow \infty$.

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

To complete our asymptotic setup, we consider the 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 $\lambda_1^O > \lambda_2^O > \dots > 0$ have all multiplicity one. Furthermore, 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 $1/\lambda_k^O = \mathcal{O}(k^{a_O})$ as $k \rightarrow \infty$.

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 smooth and all possess a 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 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 $|\tilde{\phi}_k^O(u)| \rightarrow \infty$ as $k \rightarrow \infty$. 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) $\sup_{u \in [a, b]} |\hat{\mu}(u; h_\mu) - \mu(u)| = \mathcal{O}_p(r_\mu)$
- (ã) *Conditional on X_i^O :* $\sup_{u \in O} |\hat{X}_i^O(u; h_\mu, h_X) - X_i^O(u)| = \mathcal{O}_p(r_X)$
- (b) $\sup_{(u, v) \in [a, b]^2} |\hat{\gamma}(u, v; h_\gamma) - \gamma(u, v)| = \mathcal{O}_p(r_\mu + r_\gamma)$, where

$$r_\mu \equiv r_\mu(h_\mu, n, m) := h_\mu^2 + 1/\sqrt{nmh_\mu} + 1/\sqrt{n},$$

$$r_X \equiv r_X(h_X, m) := h_X^2 + 1/\sqrt{mh_X},$$

$$r_\gamma \equiv r_\gamma(h_\gamma, n, \mathcal{M}) := h_\gamma^2 + 1/\sqrt{n\mathcal{M}h_\gamma^2} + 1/\sqrt{n},$$

and where $\mathcal{M} = m^2 - m$ and $m \leq m_i$ for all $i = 1, \dots, n$ (see A2 and A4).

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

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 $\hat{\mathcal{L}}_K(\mathbb{X}_i^O)$). *Consider an arbitrary $i \in 1, \dots, n$ and assume that $O = [A, B] \subset [a, b]$ satisfies $B - A \geq \ell_{\min} > 0$. For some $0 < C < \infty$ let $\bar{K}_{mn} = C \cdot (\min\{n^{1/2}, (n\mathcal{M})^{1/3}\})^{1/(a_O+3/2)}$. The following results hold then 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 $h_\gamma \asymp (n\mathcal{M})^{-1/6}$, as*

$n \rightarrow \infty$ and $m \rightarrow \infty$ with $m \asymp n^\theta, 0 < \theta < \infty$. For any $u \in [a, b]$:

$$\begin{aligned}
 & \widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u) \\
 &= \mathcal{L}_K(X_i^O)(u) + \mathcal{O}_p\left(K\left(\frac{1}{m^{1/2}} + \frac{K^{a_O/2+3/2}}{\min\{n^{1/2}, (n\mathcal{M})^{1/3}\}}\right)\right), \\
 (27) \quad & \mathcal{L}(X_i^O)(u) - \mathcal{L}_K(X_i^O)(u) \\
 &= \mathcal{O}\left(\left(\sum_{k=K+1}^\infty \lambda_k^O\right)^{1/2}\right) = \mathcal{O}(K^{-(a_O-1)/2}).
 \end{aligned}$$

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

$$\begin{aligned}
 & \widehat{\mathcal{L}}_K^*(\mathbb{X}_i^O)(u) \\
 &= \mathcal{L}_K^*(X_i^O)(u) + \mathcal{O}_p\left(m^{-2/5} + K\left(\frac{1}{m^{1/2}} + \frac{K^{a_O/2+3/2}}{\min\{n^{1/2}, (n\mathcal{M})^{1/3}\}}\right)\right), \\
 (28) \quad & \mathcal{L}(X_i^O)(u) - \mathcal{L}_K^*(X_i^O)(u) \\
 &= \mathcal{O}\left(\left(\sum_{k=K+1}^\infty \lambda_k^O\right)^{1/2}\right) = \mathcal{O}(K^{-(a_O-1)/2}).
 \end{aligned}$$

The theorem tells us that for any $u \in [a, b]$ the estimator $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u)$ achieves the same rate of convergence. But recall that for $u \in O = O_i$ we have $\mathcal{L}(X_i^O)(u) = X_i(u)$, and thus $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u)$ can be seen as a nonparametric estimator of X_i . In contrast, for $u \in M$ we have $\mathcal{L}(X_i^O)(u) = X_i(u) + \mathcal{Z}_i(u)$, and therefore the distance between $X_i(u)$ and $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u)$ 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 $\theta \leq 1/2$. The following corollary then is a direct consequence of (27).

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

$$(29) \quad |\mathcal{L}(X_i^O)(u) - \widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u)| = \mathcal{O}_p(m^{-\frac{a_O-1}{2(a_O+2)}}).$$

Let us consider the simple case where $m_i = m$ for all $i = 1, \dots, n$, and recall 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 $O := O_i$. $\widehat{\mathcal{L}}_K^*$ is based on local linear smoothing of the individual data (Y_{ij}, U_{ij}) , $j = 1, \dots, m$, and the associated estimation error of order $m^{-2/5}$ appears in result (28). Twice continuously differentiable functions are assumed, and using only individual data it is well-known that $m^{-2/5}$ constitutes the optimal rate of convergence of nonparametric function estimators with respect to this smoothness class.

In contrast, $\widehat{\mathcal{L}}_K(X_i^O)(u)$ combines information from all n sample curves in order to estimate $X_i(u)$ for $u \in O$. If all samples curves are structurally similar in the sense that A6 holds for a very large $a_O \gg 1$, then (29) implies that the rate of convergence of $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)(u)$ is very close to the parametric rate $m^{-1/2}$. That is, under the conditions of Corollary 4.1 (m smaller than \sqrt{n} and $a_O \gg 1$) it becomes advantageous to use $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)$ instead of $\widehat{\mathcal{L}}_K^*(\mathbb{X}_i^O)$

for estimating X_i on the observed interval, since $\widehat{\mathcal{L}}_K(\mathbb{X}_i^O)$ may provide faster rates of convergence than the rate $m^{-2/5}$ achieved by nonparametric smoothing of individual data.. 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

$$(30) \quad \tilde{X}_i(u) = \begin{cases} \mathcal{L}(X_i^O)(u) & \text{if } u \in M, \\ X_i^O(u) & \text{if } u \in O \end{cases}$$

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 5. 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 at least 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

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

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

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

Join the reconstructed fragments $\tilde{X}_{i,1}, \dots, \tilde{X}_{i,r}$ 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 first 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 extent 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.

This second 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 r th 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

1st Step of the Reconstruction Algorithm

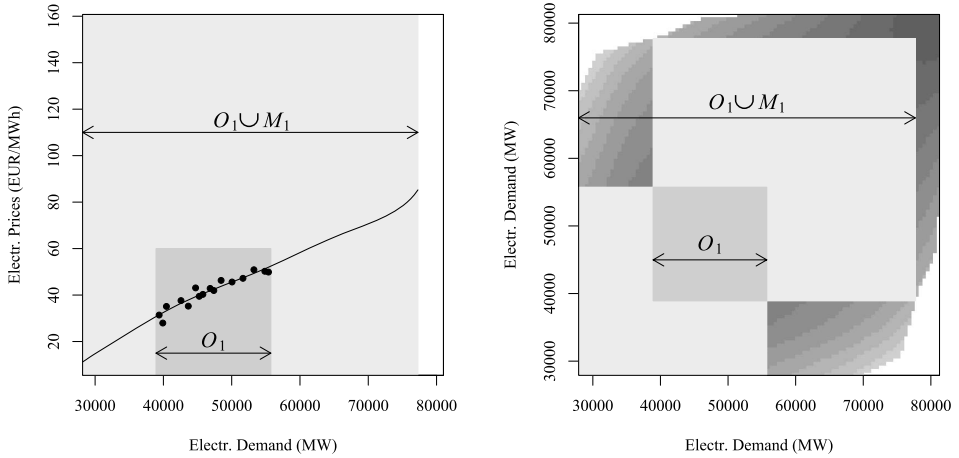


FIG. 3. Explanatory plots for the first run of the reconstruction algorithm.

r_{\max} depends on the extend to which γ can be estimated. A simple practical implementation is described in our application in Section 7.

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

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

2nd Step of the Reconstruction Algorithm

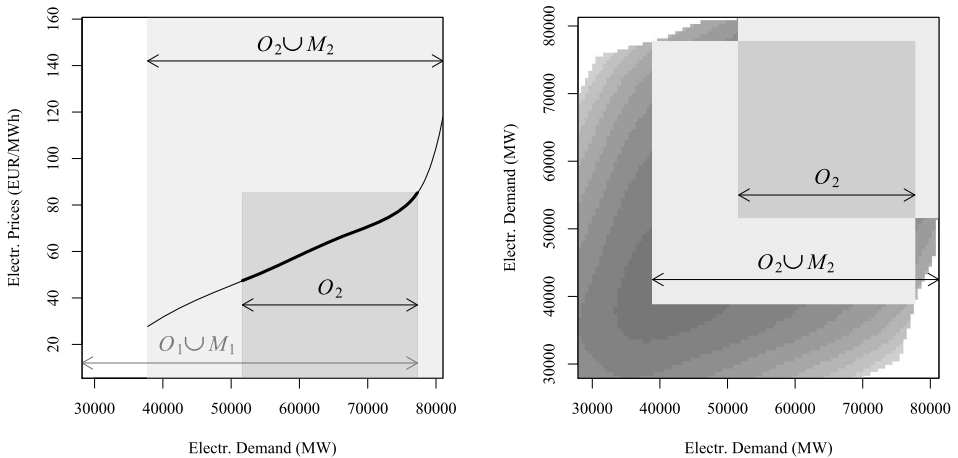


FIG. 4. Explanatory plots for the second run of the reconstruction algorithm.

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. We compare the finite sample performance of our reconstruction operators (19) and (20) with that of the PACE method proposed by Yao, Müller and Wang (2005a) and the functional linear ridge regression model proposed by Kraus (2015). A further interesting comparison method might be the functional linear regression model for sparse functional data as considered by Yao, Müller and Wang (2005b). Note, however, that this regression model becomes equivalent to the PACE method of Yao, Müller and Wang (2005a), when used to predict the trajectory of X_i given its own sparse, that is, irregular and noise contaminated measurements (see Appendix B.2 in the supplementary paper Kneip and Liebl (2019) for more detailed explanations regarding this equivalence).

The following acronyms are used to refer to the different reconstruction methods considered in this simulation study:

ANo: $\widehat{\mathcal{L}}_{\widehat{K}_{iM}}(\mathbb{X}_i^O)$ in (19) is denoted as ANo to indicate that this method involves No Alignment of the estimate of X_i^O and the reconstruction of X_i^M .

ANoCE: Equivalent to ANo, but with replacing the integral scores (25) using the following Conditional Exactions (CE) scores adapted from Yao, Müller and Wang (2005a)

$$(31) \quad \widehat{\xi}_{ik,CE}^O = \widehat{\lambda}_k^O \widehat{\phi}_{ik}^{O\top} \widehat{\Sigma}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i),$$

where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{im_i})^\top$, $\widehat{\phi}_{ik}^O = (\widehat{\phi}_k^O(U_{i1}), \dots, \widehat{\phi}_k^O(U_{im_i}))^\top$, $[\widehat{\Sigma}_{\mathbf{Y}_i}]_{1 \leq j, k \leq m_i} = \widehat{\gamma}(U_{ij}, U_{ik}) + \widehat{\sigma}^2 \delta_{jk}$, with $\delta_{jk} = 1$ if $j = k$ and zero else, and with $\widehat{\lambda}_k^O$ and $\widehat{\phi}_k^O$ as defined in (24). The estimate of the error variance, $\widehat{\sigma}^2$, is computed using LLK estimators as described in equation (2) of Yao, Müller and Wang (2005a).

AYes: $\widehat{\mathcal{L}}_{\widehat{K}_{iM}}^*(\mathbb{X}_i^O)$ in (20) is denoted as AYes to indicate that this method involves an alignment of the estimate of X_i^O and the reconstruction of X_i^M .

AYesCE: Equivalent to AYes, but with replacing the integral scores (25) by the conditional exaction scores of (31).

PACE: The method of Yao, Müller and Wang (2005a), who approximate the missing X_i^M and observed X_i^O parts jointly using the truncated Karhunen–Loève decomposition $\widehat{X}_i(t) = \widehat{\mu}(t) + \sum_{k=1}^{\widehat{K}_{iM}} \widehat{\xi}_{ik}^{\text{PACE}} \widehat{\phi}_k(t)$ with conditional expectation scores

$$(32) \quad \widehat{\xi}_{ik}^{\text{PACE}} = \widehat{\lambda}_k \widehat{\phi}_{ik}^\top \widehat{\Sigma}_{\mathbf{Y}_i}^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i),$$

where $\widehat{\lambda}_k$ and $\widehat{\phi}_k$ are as defined in (24), but with $O = [a, b]$.

KRAUS: The functional linear ridge regression model of Kraus (2015).

The idea of using the conditional expectation scores (31) in ANoCE and AYesCE as an alternative to the integral scores (25) in ANo and AYes is inspired by a comment of one of the anonymous referees, who correctly pointed out that the integral scores (25) might be instable for irregular and noisy data. PACE also uses condition expectation scores, but is fundamentally different from ANoCE and AYesCE. While PACE uses approximations of the classical eigenfunctions ϕ_k , the classical eigenvalues λ_k , and the classical scores ξ_{ik} , ANoCE and AYesCE use approximations of the reconstructive eigenfunctions $\tilde{\phi}_k^O$, the eigenvalues λ_k^O , and the scores ξ_{ik}^O with respect to the partially observed domain O .

The truncation parameters \hat{K}_{iM} for ANo, ANoCE, AYes, AYesCE, and PACE are selected by minimizing the GCV criterion in (26). For PACE, we do not use the AIC-type criterion as proposed by Yao, Müller and Wang (2005a), since this criterion determines a “global” truncation parameter \hat{K} , which performs worse than our local, that is, M -specific truncation parameter \hat{K}_{iM} . The ridge regularization parameter for KRAUS is determined using the GCV criterion as described in Kraus (2015).

We consider four different Data Generating Processes (DGPs). DGP1 and DGP2 comprise irregular evaluation points and measurement errors which facilitates the comparison of ANo, ANoCE, AYes, AYesCE and the PACE method. DGP3 and DGP4 comprise regular evaluation points and no measurements errors which facilitates the comparison of ANo, AYes, PACE and the KRAUS method. For all simulations we set $[a, b] = [0, 1]$.

DGP1: The data points (Y_{ij}, U_{ij}) are generated according to $Y_{ij} = X_i(U_{ij}) + \varepsilon_{ij}$ with error term $\varepsilon_{ij} \sim N(0, 0.0125)$ and random function $X_i(u) = \mu(u) + \sum_{k=1}^{50} \xi_{ik,1} \cos(k\pi u) / \sqrt{5} + \xi_{ik,2} \sin(k\pi u) / \sqrt{5}$, where $\mu(u) = u + \sin(2\pi u)$, $\xi_{ik,1} = 50\sqrt{\exp(-(k-1)^2/5)}Z_{i,1}$, and $\xi_{ik,2} = 50\sqrt{\exp(-k^2/5)}Z_{i,2}$ with $Z_{i,1}, Z_{i,2} \sim N(0, 1)$. The evaluation points are generated as $U_{ij} \sim \text{Unif}[A_i, B_i]$, where with probability 1/2, $A_i \sim \text{Unif}[0, 0.45]$ and $B_i \sim \text{Unif}[0.55, 1]$ and with probability 1/2, $[A_i, B_i] = [0, 1]$. That is, about one half of the sample consists of partially observed functions with mean interval-width 0.55.

DGP2: Equivalent to DGP1, except for a larger noise component with $\varepsilon_{ij} \sim N(0, 0.125)$.

DGP3: The data points (Y_{ij}, U_{ij}) are generated according to $Y_{ij} = X_i(U_{ij})$ with random function $X_i(u) = \mu(u) + \sum_{k=1}^{50} \xi_{ik,1} \cos(k\pi u) + \xi_{ik,2} \sin(k\pi u)$, where $\mu(u) = u^2 + \sin(2\pi u)$, $\xi_{ik,1} = 50\sqrt{\exp(-(k-1)^2)}Z_{i,1}$, and $\xi_{ik,2} = 50\sqrt{\exp(-k^2)}Z_{i,2}$ with $Z_{i,1}, Z_{i,2} \sim N(0, 1)$. The evaluation points are equidistant grid points $U_{ij} = j/51$, with $j = 1, \dots, 51$, where all points $U_{ij} \notin [A_i, B_i]$ are set to NA. With probability 3/4, $A_i \sim \text{Unif}[0, 1/3]$ and $B_i = A_i + 1/2$ and with probability 1/4, $[A_i, B_i] = [0, 1]$.

DGP4: Equivalent to DGP3, but with $A_i \sim \text{Unif}[0, 2/3]$ and $B_i = A_i + 1/3$. That is, DGP4 has smaller and therefore more challenging fragments than DGP3.

For each DGP, we generate 50 different targets X_ℓ , $\ell = 1, \dots, 50$, where each target is partitioned into a (nondegenerated) missing part X_ℓ^M and an observed part X_ℓ^O . Each of these targets X_ℓ are reconstructed in each of the $b = 1, \dots, 100$ simulation runs with sample sizes $n \in \{50, 100\}$ for DGP1–DGP4 and $m \in \{15, 30\}$ for DGP1 and DGP2.

Let $\hat{X}_{\ell,b}$ denote the reconstructed function in simulation run b using one of the reconstruction methods ANo, ANoCE, AYes, AYesCE, PCAE, or KRAUS. For each target X_ℓ , we compute the integrated mean squared error, the integrated squared bias, and the integrated variance,

$$\text{MSE}_\ell = \text{Var}_\ell + \text{Bias}_\ell^2, \quad \text{Bias}_\ell^2 = \int_0^1 (\bar{X}_\ell(u) - X_\ell(u))^2 dt,$$

$$\text{and } \text{Var}_\ell = \int_0^1 100^{-1} \sum_{b=1}^{100} (\hat{X}_{\ell,b}(u) - \bar{X}_\ell(u))^2 dt,$$

where $\bar{X}_\ell(u) = 100^{-1} \sum_{r=1}^{100} \hat{X}_{\ell,b}(u)$. The finite sample performance is evaluated using the averages over all 50 targets,

$$\text{Var} = \frac{1}{50} \sum_{\ell=1}^{50} \text{Var}_\ell, \quad \text{Bias}^2 = \frac{1}{50} \sum_{\ell=1}^{50} \text{Bias}_\ell^2 \quad \text{and} \quad \text{MSE} = \frac{1}{50} \sum_{\ell=1}^{50} \text{MSE}_\ell.$$

The simulation study is implemented using the R-package `ReconstPoFD` which can be downloaded and installed from the second author’s GitHub account.

TABLE 1
Simulation results for DGP1

DGP	n	m	Method	MSE_{ratio}	MSE	Bias ²	Var
DGP1	50	15	AYesCE	1.00	0.161	0.135	0.025
DGP1	50	15	AYes	1.02	0.164	0.139	0.025
DGP1	50	15	ANoCE	1.38	0.222	0.199	0.023
DGP1	50	15	ANo	1.39	0.224	0.200	0.024
DGP1	50	15	PACE	10.49	1.685	0.259	1.426
DGP1	50	30	AYesCE	1.00	0.136	0.112	0.024
DGP1	50	30	AYes	1.00	0.137	0.113	0.024
DGP1	50	30	ANoCE	1.48	0.202	0.173	0.029
DGP1	50	30	ANo	1.53	0.209	0.180	0.029
DGP1	50	30	PACE	5.19	0.707	0.131	0.576
DGP1	100	15	AYesCE	1.00	0.131	0.112	0.018
DGP1	100	15	AYes	1.00	0.131	0.114	0.017
DGP1	100	15	ANoCE	1.58	0.207	0.191	0.017
DGP1	100	15	ANo	1.61	0.211	0.194	0.017
DGP1	100	15	PACE	8.74	1.145	0.154	0.991
DGP1	100	30	AYes	1.00	0.125	0.108	0.017
DGP1	100	30	AYesCE	1.01	0.126	0.109	0.017
DGP1	100	30	ANoCE	1.36	0.170	0.146	0.023
DGP1	100	30	ANo	1.45	0.181	0.158	0.023
DGP1	100	30	PACE	3.59	0.448	0.123	0.325

$$MSE_{\text{ratio}} = \text{MSE} / \min(\text{MSE}).$$

Table 1 shows the simulation results for DGP1. The methods (ANo, ANoCE, AYes, AYesCE and PACE) are ranked according to their MSE_{ratio} which is defined by the method's MSE-value relative to the lowest MSE-value within the comparison group. The rankings are stable for all sample sizes m and n . The AYesCE reconstruction method shows the best performance. The AYes method, which uses integral scores instead of conditional expectation scores, is only marginally less efficient than AYesCE. Our nonalignment methods ANoCE and ANo are ranked third and fourth. The PACE method of Yao, Müller and Wang (2005a), originally proposed for sparse functional data analysis, shows a rather poor performance. The reason for this is that PACE adds the variance of the measurement error to the diagonal of the discretized covariance matrix, which has a regularization effect on the generally ill-posed inversion problem. For DGP1, however, the variance of the error term is rather small which results in a too small regularization of the inverse.

Table 2 shows the simulation results for DGP2. DGP2 is equivalent to DGP1 except for a larger variance of the error term. Our alignment methods AYesCE and AYes still show the best performance. However, having a larger variance leads to a better regularization of the inverse problem involved in the PACE method, such that PACE is ranked third. Our nonalignment methods ANoCE and ANo are ranked fourth and fifth. Figures 7 and 8 in Appendix C of the supplementary paper Kneip and Liebl (2019) provide graphical illustrations of the different reconstruction results as well as a visual impression of the different signal-to-noise ratios in DGP1 and DGP2.

Table 3 shows the simulation results for DGP3 and DGP4 comparing the methods ANo, AYes, PACE and KRAUS. Here, the alignment method AYes shows by far the best performance for all sample sizes and for both DGPs. The partially very bad performance of PACE is due to the missing measurement error in DGP3 and DGP4, which results in a missing

TABLE 2
Simulation results for DGP2

DGP	n	m	Method	MSE _{ratio}	MSE	Bias ²	Var
DGP2	50	15	AYesCE	1.00	0.198	0.173	0.025
DGP2	50	15	AYes	1.04	0.207	0.179	0.027
DGP2	50	15	PACE	1.07	0.212	0.174	0.039
DGP2	50	15	ANoCE	1.14	0.227	0.203	0.023
DGP2	50	15	ANo	1.16	0.230	0.204	0.026
DGP2	50	30	AYesCE	1.00	0.189	0.167	0.022
DGP2	50	30	AYes	1.01	0.192	0.169	0.023
DGP2	50	30	PACE	1.09	0.206	0.167	0.039
DGP2	50	30	ANoCE	1.14	0.215	0.188	0.027
DGP2	50	30	ANo	1.16	0.219	0.190	0.028
DGP2	100	15	AYesCE	1.00	0.178	0.161	0.017
DGP2	100	15	AYes	1.01	0.180	0.162	0.018
DGP2	100	15	PACE	1.08	0.193	0.165	0.028
DGP2	100	15	ANoCE	1.20	0.213	0.198	0.015
DGP2	100	15	ANo	1.21	0.216	0.199	0.018
DGP2	100	30	AYesCE	1.00	0.177	0.159	0.018
DGP2	100	30	AYes	1.03	0.181	0.162	0.020
DGP2	100	30	PACE	1.03	0.183	0.153	0.029
DGP2	100	30	ANoCE	1.07	0.189	0.167	0.023
DGP2	100	30	ANo	1.12	0.197	0.174	0.024

MSE_{ratio} = MSE/ min(MSE).

TABLE 3
Simulation results for DGP3 and DGP4

DGP	n	Method	MSE _{ratio}	MSE	Bias ²	Var
DGP3	50	AYes	1.00	0.168	0.131	0.037
DGP3	50	PACE	1.33	0.223	0.099	0.124
DGP3	50	ANo	1.40	0.234	0.178	0.056
DGP3	50	KRAUS	1.52	0.254	0.205	0.049
DGP3	100	AYes	1.00	0.142	0.120	0.022
DGP3	100	PACE	1.26	0.179	0.081	0.098
DGP3	100	KRAUS	1.29	0.184	0.151	0.033
DGP3	100	ANo	1.36	0.194	0.158	0.035
DGP4	50	AYes	1.00	0.276	0.220	0.056
DGP4	50	ANo	1.11	0.307	0.247	0.060
DGP4	50	KRAUS	1.20	0.330	0.269	0.061
DGP4	50	PACE	41.93	11.564	0.313	11.252
DGP4	100	AYes	1.00	0.232	0.202	0.030
DGP4	100	KRAUS	1.11	0.258	0.222	0.035
DGP4	100	ANo	1.12	0.261	0.227	0.034
DGP4	100	PACE	3.59	0.834	0.151	0.682

MSE_{ratio} = MSE/ min(MSE).

regularization of the inverse problem involved in the PACE method. Furthermore, PACE is designed for the case where one observes only a few noisy discretization points per function, but these points should be distributed over the total domain $[a, b]$. For the considered DGPs, however, the discretization points are only observed within challenging small subdomains $[A_i, B_i] \subset [0, 1]$. Graphical illustrations of the different reconstruction results for DGP3 and DGP4 are provided in Figures 9 and 10 in Appendix C of the supplementary paper [Kneip and Liebl \(2019\)](#).

Summing up, in all DGPs the best performing reconstruction method are our alignment methods AYesCE and AYes. For discretized functional data plus measurement errors it is advantageous to use the alignment method AYesCE which involves conditional expectation scores.

7. Application. Our functional data point of view on electricity spot prices provides a practical framework that is useful for forecasting electricity spot prices ([Liebl \(2013\)](#), [Weron \(2014\)](#)) and for testing price differences ([Liebl \(2019\)](#)). In the following, we focus on the problem of reconstructing the partially observed price-functions, which is highly relevant for practitioners who need complete price functions for doing comparative statics, that is, a ceteris-paribus analysis of price effects with respect to changes in electricity demand (cf. [Hirth \(2013\)](#), [Weigt \(2009\)](#)).

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, D_{ij} , and net-imports of electricity from other countries, N_{ij} , 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). 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%) data pairs (Y_{ij}, U_{ij}) with prices $Y_{ij} > 120$ EUR/MWh and $U_{ij} > 82,000$ 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 relevant values of electricity demand U_{ij} are defined as electricity demand minus infeeds from RES and plus net-imports from other countries, that is, $U_{ij} := D_{ij} - \text{RES}_{ij} + N_{ij}$, where $\text{RES}_{ij} = \text{Wind.Infeed}_{ij} + \text{Solar.Infeed}_{ij}$. The effect of further RES such as biomass is still negligible for the German electricity market.

The estimated mean and covariance functions are shown in Figure 5. 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 AYesCE estimator, which showed a very good performance in our simulation studies, and our iterative reconstruction Algorithm 5.1 implemented 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 reconstructed curve $\hat{X}_{i,1}$ and try to reconstruct possibly further missing upper fragments. In the final step we use the lower half of $\hat{X}_{i,1}$ and try to reconstruct possibly further missing lower fragments.

This approach allows us to recover 91% of the price functions over the total support (Figure 6). Note that the price functions with negative electricity prices are perfectly plausible. Negative prices are an important market-feature of the EPEX (see, for instance, [Cludius et al.](#)

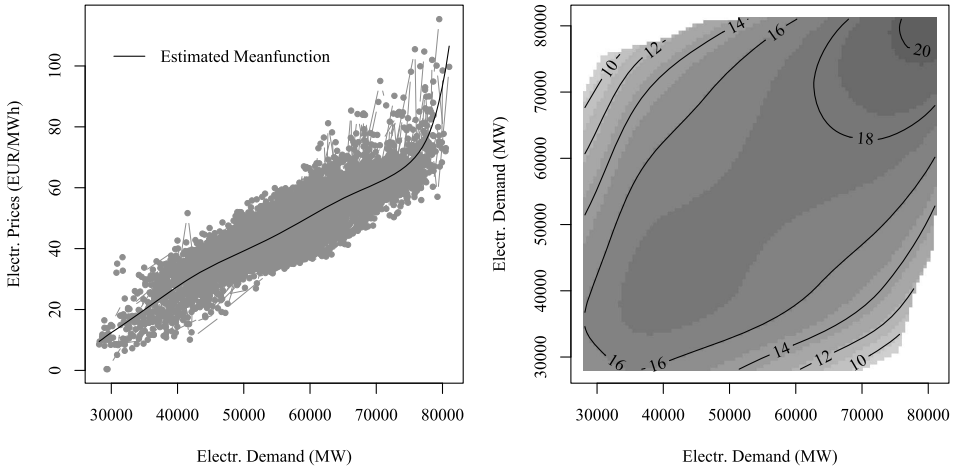


FIG. 5. LEFT PANEL: *Estimated mean function plus a scatter plot of the data pairs (Y_{ij}, U_{ij}) .* RIGHT PANEL: *Contour plot of the estimated covariance function. The white regions reflect the outer off-diagonal parts which are infeasible to estimate.*

(2014), Fanone, Gamba and Prokopczuk (2013), Nicolosi (2010)). Electricity producers are willing to sell electricity at negative prices (i.e., to pay for delivering electricity) if shutting off and restarting their power plants is more expensive than selling their electricity at negative prices. That is, the reconstructed price functions are conform with the specific market design of the EPEX and may be useful for a variety of further subsequent analysis using classical methods of functional data analysis.

Acknowledgements. We would like to thank the referees and the editors for their constructive feedback which helped to improve this research work.

SUPPLEMENTARY MATERIAL

Supplemental paper (DOI: [10.1214/19-AOS1864SUPP](https://doi.org/10.1214/19-AOS1864SUPP); .pdf). The supplemental paper contains the proofs of our theoretical results.

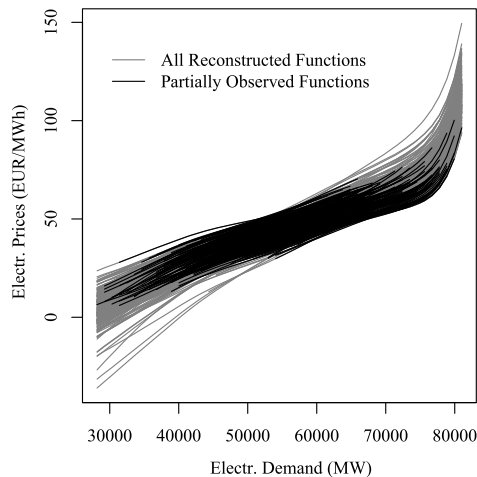


FIG. 6. *Recovered functions (gray) and the original partially observed functions (black).*

REFERENCES

- BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications. Lecture Notes in Statistics* **149**. Springer, New York. MR1783138 <https://doi.org/10.1007/978-1-4612-1154-9>
- CAI, T. T. and HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34** 2159–2179. MR2291496 <https://doi.org/10.1214/009053606000000830>
- CARDOT, H., MAS, A. and SARDA, P. (2007). CLT in functional linear regression models. *Probab. Theory Related Fields* **138** 325–361. MR2299711 <https://doi.org/10.1007/s00440-006-0025-2>
- CLUDIUS, J., HERMANN, H., MATTHES, F. C. and GRAICHEN, V. (2014). The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications. *Energy Econ.* **44** 302–313.
- DELAIGLE, A. and HALL, P. (2013). Classification using censored functional data. *J. Amer. Statist. Assoc.* **108** 1269–1283. MR3174707 <https://doi.org/10.1080/01621459.2013.824893>
- DELAIGLE, A. and HALL, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika* **103** 779–799. MR3620439 <https://doi.org/10.1093/biomet/asw040>
- DELAIGLE, A., HALL, P., HUANG, W. and KNEIP, A. (2018). Estimating the covariance of fragmented and other incompletely observed functional data. Working Paper.
- DESCARY, M.-H. and PANARETOS, V. M. (2019). Recovering covariance from functional fragments. *Biometrika* **106** 145–160. MR3912388 <https://doi.org/10.1093/biomet/asy055>
- FANONE, E., GAMBA, A. and PROKOPCZUK, M. (2013). The case of negative day-ahead electricity prices. *Energy Econ.* **35** 22–34.
- GOLDBERG, Y., RITOV, Y. and MANDELBAUM, A. (2014). Predicting the continuation of a function with applications to call center data. *J. Statist. Plann. Inference* **147** 53–65. MR3151845 <https://doi.org/10.1016/j.jspi.2013.11.006>
- GROMENKO, O., KOKOSZKA, P. and SOJKA, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *Ann. Appl. Stat.* **11** 898–918. MR3693551 <https://doi.org/10.1214/17-AOAS1022>
- HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35** 70–91. MR2332269 <https://doi.org/10.1214/0090536060000000957>
- HALL, P., MÜLLER, H.-G. and WANG, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34** 1493–1517. MR2278365 <https://doi.org/10.1214/009053606000000272>
- HIRTH, L. (2013). The market value of variable renewables: The effect of solar wind power variability on their relative price. *Energy Econ.* **38** 218–236.
- HORVÁTH, L. and KOKOSZKA, P. (2012). *Inference for Functional Data with Applications. Springer Series in Statistics*. Springer, New York. MR2920735 <https://doi.org/10.1007/978-1-4614-3655-3>
- KNEIP, A. and LIEBL, D. (2020). Supplement to “On the optimal reconstruction of partially observed functional data.” <https://doi.org/10.1214/19-AOS1864SUPP>.
- KRAUS, D. (2015). Components and completion of partially observed functional data. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 777–801. MR3382597 <https://doi.org/10.1111/rssb.12087>
- LI, Y. and HSING, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38** 3321–3351. MR2766854 <https://doi.org/10.1214/10-AOS813>
- LIEBL, D. (2013). Modeling and forecasting electricity spot prices: A functional data perspective. *Ann. Appl. Stat.* **7** 1562–1592. MR3127959 <https://doi.org/10.1214/13-AOAS652>
- LIEBL, D. (2019). Nonparametric testing for differences in electricity prices: The case of the Fukushima nuclear accident. *Ann. Appl. Stat.* **13** 1128–1146. MR3963565 <https://doi.org/10.1214/18-AOAS1230>
- 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.
- RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993
- RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283
- WEIGT, H. (2009). Germany’s wind energy: The potential for fossil capacity replacement and cost saving. *Appl. Energy* **86** 1857–1863.
- WERON, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **30** 1030–1081.
- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 <https://doi.org/10.1198/016214504000001745>

- YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *Ann. Statist.* **33** 2873–2903. MR2253106 <https://doi.org/10.1214/009053605000000660>
- ZHANG, J.-T. and CHEN, J. (2007). Statistical inferences for functional data. *Ann. Statist.* **35** 1052–1079. MR2341698 <https://doi.org/10.1214/009053606000001505>
- ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 <https://doi.org/10.1214/16-AOS1446>