ON THE OPTIMAL RECONSTRUCTION OF PARTIALLY OBSERVED FUNCTIONAL DATA

By Dominik Liebl and Alois Kneip

University of Bonn

If only fragments of functions are observable and none of these fragments cover the total domain, then it is impossible to estimate the covariance function over its total support. We propose a new functional PCA based prediction procedure that allows to reconstruct functional data from their fragmental observations under this challenging setup. By contrast to functional prediction models in the literature, we can proof the optimality of our prediction model without making use of the crucial, but unverifiable, orthogonality assumption on the prediction error. This also allows us to identify situations under which our prediction procedure leads to a perfect reconstruction of the fragmental observations, i.e., without any prediction error. Finite sample properties are investigated through simulations and a real data application.

1. Introduction. Let $(X_t)_t$ be a stationary weakly dependent time series of random functions $t \in \{1, \dots, T\}$, where each random function X_t is square integrable, i.e., $X_t \in L^2(I_0)$ with $I_0 = [a, b]$. Though, instead of observing realizations of X_t, we only observe "small" fragments X_t^{S}, where $X_t^{\mathrm{S}}(u) = X_t(u)$ with $u \in I_{\mathrm{S},t}$ and $I_{\mathrm{S},t} \subset I_0$. This situation is illustrated in the left panel of Figure 1. There the (electricity) price function of March 22, 2011, covers only a subset of the total support $I_0 = [a, b]$. See our real data application in Section 7 for more information on these price functions.

The situation becomes unpleasant, if $I_{\mathrm{S},t} \subset I_0$ for all $t \in \{1, \dots, T\}$, which is the case in our real data application. If the functions X_t are observable only over *proper* subsets $I_{\mathrm{S},t} \subset I_0$, then it is impossible to estimate the covariance function $\gamma(u, v) = \mathrm{Cov}(X_t(u), X_t(v))$ over its total support $I_0^2 = [a, b]^2$. For instance, we cannot estimate the value $\gamma(a, b)$, since we cannot observe $X_t(a)$ and $X_t(b)$ for the same time point t. The covariance function γ is then only estimable for points $(u, v) \in \bigcup_t I_{\mathrm{S},t}^2$. The empirical counterpart of the feasible region $\bigcup_t I_{\mathrm{S},t}^2$ is illustrated by the gray area in the right panel of Figure 1. The white areas at the outer off-diagonal parts of the square $[a, b]^2$ describe the infeasible regions.

AMS 2000 subject classifications: Primary 62M20, 62H25, 62G05, 62G08, 62M10

Keywords and phrases: functional data analysis, functional linear prediction, functional principal components, time series analysis

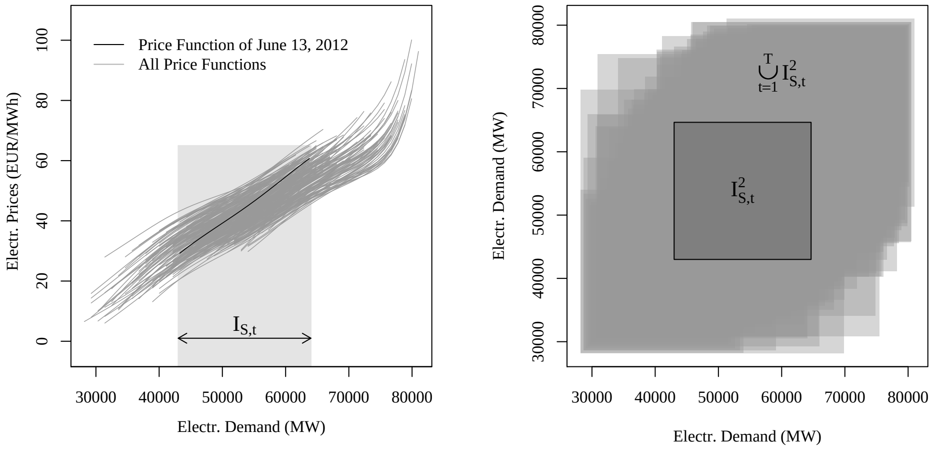

Fig 1. Left Panel: *Fragmental random functions observed only over proper subsets* $I_{\mathrm{S},t}$ *of the total support* $[a, b]$. Right Panel: *Visualization of the set* $\bigcup_t I_{\mathrm{S},t}^2$ *over which the covariance function is estimable.*

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, "electricity demand" in Figure 1 actually denotes *residual* electricity demand, i.e., electricity demand minus infeeds from RES. Practitioners in energy economics are interested in comparative statics with respect to changes in residual electricity demand (cf. Weigt 2009 and Hirth 2013). As these changes can be drastic, it is necessary to predict the partially observed price functions over a large as possible part of the total support. Typically, this is done through extrapolation based on rigid parametric model assumptions.

Our main contribution is a nonparametric functional PCA (FPCA) based prediction model

$$X_t^{\mathrm{L}}(u) = \left[S X_t^{\mathrm{S}}\right](u) + Z_t(u), \quad u \in I_{\mathrm{L},t}$$

that allows to reconstruct the partially observed, i.e., "small" random functions X_t^{S} over a "large" as possible part of the total domain $I_{\mathrm{L},t} \subseteq [a, b]$ with $I_{\mathrm{S},t} \subseteq I_{\mathrm{L},t}$, where $S : L^2(I_{\mathrm{S},t}) \to L^2(I_{\mathrm{L},t})$ denotes the prediction operator and $Z_t \in L^2(I_{\mathrm{L},t})$ the prediction error. Whether the functions can be reconstructed over the total domain, i.e., $I_{\mathrm{L},t} = [a, b]$, eventually depends on how much of the covariance function is estimable. In Section 4 we propose an prediction algorithm which typically allows to reconstruct the total trajectories.

In our theoretical part, we show that our predictor $\tilde{X}_t^{\mathrm{L}} = SX_t^{\mathrm{S}}$ is the best linear predictor under the L2 loss. Under similar setups, this property is already claimed by two other functional prediction models, namely that of Goldberg, Ritov and Mandelbaum (2014) and Kraus (2015). Though, these approaches do not consider the case of truly fragmental observations. I.e., they assume that the covariance function can still be estimated over its total support $[a, b]^2$, which simplifies the prediction problem and constitutes quite a restrictive assumption within this problem set. Beyond this, the work of Goldberg, Ritov and Mandelbaum (2014) considers only the case of finite dimensional functional data. Their results have well known counterparts in multivariate statistics. Kraus (2015) only informally claims the optimality property, but lacks of a formal statement which is common in the literature on the classical functional prediction model (cf. Yao, Müller and Wang (2005b), Crambes and Mas (2013), and Hörmann and Kidziński (2015)). In fact, under the common (unverifiable) assumption that the prediction error Z_t is orthogonal to the predictor X_t^{S}, this optimality result is trivial and does not demand for a formal statement. It follows directly from arguments as in He, Müller and Wang (2003). Delaigle and Hall (2013) consider the case of truly fragmental functional observations, though, within a classification framework. The FPCA based model of James, Hastie and Sugar (2000) and James and Hastie (2001) allows for sparse functional data that are observed over proper subintervals, too. This approach makes use of rigid parametric (mixed effects) model assumptions in order to cope with the sparseness in the data. By contrast, we consider a less rigid nonparametric model setup. A further reconstruction method is the PACE estimation procedure (cf. Yao, Müller and Wang (2005a) and Yao, Müller and Wang (2005b)). Though, this method assumes too that covariance function is estimable over its total support $[a, b]^2$.

Our prediction model is fundamentally different to the above cited classical functional prediction models, since the predictor function X_t^{S} and the function X_t^{L}, that we aim to predict, are fragments of the *same* underlying function X_t. This specific situation allows us to propose an prediction operator S for which it can be shown that the prediction error Z_t is orthogonal to the predictor X_t^{S}. That is, in our case orthogonality is not an unverifiable model assumption, but a proven fact that applies to any square integrable process X_t. This orthogonality result makes our optimality result much more substantial than the optimality results of the classical functional prediction models. Furthermore, we are able to derive an expression for the variance of the prediction error which allows to identify situations under which it is possible to recover the partially observed functions perfectly, i.e., without any prediction error. These results go beyond that of Kraus (2015), who consider a similar prediction model as we do. Though, the author treats his model as a classical functional prediction model and therefore misses to exploit the specific prediction context.

The rest of this paper is structured as follows: The next section introduces our theoretical setup and our statistical prediction model. Section 2 presents our main theoretical results with respect our prediction model. Section 3 contains our estimation procedure and Section 4 introduces our prediction algorithm. Our asymptotic results are presented in Section 5 and section 6 contains the simulation study. The real data application can be found Section 7 and Section 8 concludes. All proofs can be found in the supplemental paper.

2. Optimal prediction of partially observed functions. Let us first fix our basic setup. Each random function $X_t \in H$ is an element of the Hilbert space $H = L^2(I_0)$, with $I_0 = [a, b] \subset \mathbb{R}$, $0 < a < b < \infty$, inner product $\langle x, y\rangle_H = \int_{I_0} x(s)y(s)ds$, and norm $||y||_H = (\int_{I_0}(y(s))^2 ds)^{1/2}$ for all $x, y \in H$. Furthermore, we need to assume that $\mathbb{E}(||X_t||_H^4) < \infty$, which assures existence (and estimability) of the mean function $\mu(u) = \mathbb{E}(X_t(u))$, with $u \in I_0$, and of the covariance function $\gamma(u, v) = \mathrm{Cov}(X_t(u), X_t(v))$, with $(u, v) \in I_0^2$. In order to simplify the notation within this theoretical section, we assume centered random functions, i.e., $\mu(u) = 0$ for all $u \in I_0$. This is, of course, without loss of generality.

As outlined in the introduction, we do not observe the random functions X_t, but only "small" fragments X_t^{S}, where $X_t^{\mathrm{S}}(u) = X_t(u)$ with $u \in I_{\mathrm{S},t}$ and $I_{\mathrm{S},t} \subset I_0$. The set $I_{\mathrm{S},t}$ denotes a random set and can be thought of as some censoring process. In principle our theory applies to very general cases of random subsets $I_{\mathrm{S},t}$, i.e., containing, e.g., "wholes". Though, for simplicity we treat $I_{\mathrm{S},t}$ as a random subinterval, i.e., $I_{\mathrm{S},t} = [A_t, B_t] \subset I_0 = [a, b]$ with $a \leq A_t < B_t \leq b$.

In the following we list our regularity assumptions with respect to the random subsets $I_{\mathrm{S},t}$:

RS Let $(I_{\mathrm{S},t})_t$ denote an strictly stationary, weakly dependent series of random sets with $I_{\mathrm{S},t} \subset I_0 = [a, b]$, where $I_{\mathrm{S},t}$ is independent from X_s and ε_{jk} for all $t, s, k \in \{1, \ldots, T\}$ and all $j \in \{1, \ldots, n\}$. Further, let for all t, (i) $\mathbb{P}(I_{\mathrm{S},t} = \emptyset) = 0$, (ii) $\mathbb{P}(I_0 \setminus I_{\mathrm{S},t} = \emptyset) = 0$, and (iii) $\mathbb{P}(u \in I_{\mathrm{S},t}) > 0$ for all $u \in I_0$.

Remarks on Assumption RS. Property (i) excludes the degenerated cases and property (ii) assures that $I_{\mathrm{S},t}$ is a *proper* subset, i.e., that $I_{\mathrm{S},t} \subset I_0$. The latter implies that the composed set $\bigcup_t I_{\mathrm{S},t}^2$ does not cover the total square

$I_0^2 = [a,b]^2$. This has far-reaching consequences, since then it is impossible to estimate the covariance $\gamma(u,v)$ for all $(u,v) \in I_0^2$; see also the right panel of Figure 1. Property (iii) assures that every point $u \in I_0$ is covered by the random subset $I_{\mathrm{S},t}$ with a strictly positive probability. Under the latter property we can estimate the mean function $\mu(u)$ for all $u \in I_0$.

Now we can consider the prediction problem from a general theoretical point of view, where we seek to reconstruct a partially observed random function X_t^{S} over a large as possible interval $I_{\mathrm{L},t}$, given the information with respect to the observed "small" fraction X_t^{S}. The large interval $I_{\mathrm{L},t}$ must be chosen sufficiently small, such that $\mathbb{P}(I_{\mathrm{L},t} \subseteq I_{\mathrm{S},s}) > 0$ for $s \neq t$. That is, with positive probability there are random subintervals $I_{\mathrm{S},s}$ that contain the whole large interval $I_{\mathrm{L},t}$. Otherwise, we cannot consistently estimate the underlying covariance function γ_t^{L} over its total support, where $\gamma_t^{\mathrm{L}}(u,v) := \gamma(u,v)$ for all $(u,v) \in I_{\mathrm{L},t}^2$. In order to simplify the notation, we suppress in the following the index t for covariances and intervals.

Let us first assume the ideal situation that we know γ^{L}. But if $\gamma^{\mathrm{L}}(u,v)$ is known over I_{L}^2, we also know $\gamma^{\mathrm{S}}(u,v)$ over $I_{\mathrm{S}}^2 \subset I_{\mathrm{L}}^2$ and therefore γ^{L} defines a unique covariance operator Γ^{S} of $X_t^{\mathrm{S}} \in L^2(I_{\mathrm{S}})$ as

$$\Gamma^{\mathrm{S}}(x)(u) = \int_{I_{\mathrm{S}}} \gamma^{\mathrm{S}}(u,v)x(v)dv, \quad x \in L^2(I_{\mathrm{S}}),$$

where $\lambda_1^{\mathrm{S}} \geq \lambda_2^{\mathrm{S}} \geq \dots$ and $\phi_1^{\mathrm{S}}, \phi_2^{\mathrm{S}}, \dots \in L^2(I_{\mathrm{S}})$ denote eigenvalues and eigenfunctions of Γ^{S}, and where $(\phi_k^{\mathrm{S}})_{k\geq 1}$ forms an orthonormal basis system that spans the space $L^2(I_{\mathrm{S}})$. The covariance operators Γ and Γ^{L} of the total random function X_t and the large random fraction X_t^{L}, as well as their eigenvalues and eigenfunctions, are defined correspondingly.

Any centered fragment X_t^{S} adopts then the well known Karhunen-Loéve (KL) representation

$$(1) \qquad X_t^{\mathrm{S}}(u) \;=\; \sum_{k=1}^{\infty} \xi_{tk}^{\mathrm{S}} \phi_k^{\mathrm{S}}(u), \quad u \in I_{\mathrm{S}},$$

where the Functional Principal Component (FPC) scores are defined by $\xi_{tk}^{\mathrm{S}} = \langle X_t^{\mathrm{S}}, \phi_k^{\mathrm{S}} \rangle$, with $\langle X_t^{\mathrm{S}}, \phi_k^{\mathrm{S}} \rangle = \int_{I_{\mathrm{S}}} X_t^{\mathrm{S}}(x)\phi_k^{\mathrm{S}}(x)dx$, and where $\mathbb{E}(\xi_{tk}^{\mathrm{S}}) = 0$ and $\mathbb{E}(\xi_{tk}^{\mathrm{S}} \cdot \xi_{tl}^{\mathrm{S}}) = \lambda_k^{\mathrm{S}}$ for all $k = l$ and zero else. The KL representations of the total random function X_t and the large random fragment X_t^{L} are defined correspondingly.

The properties of the KL representation imply that if $\lambda_k^{\mathrm{S}} > 0$, the value $\phi_k^{\mathrm{S}}(u)$ can be obtained as the slope-coefficient of a simple linear regression of the small fragment $X_t^{\mathrm{S}}(u)$ on the small FPC score ξ_{tk}^{S}:

$$\phi_k^{\mathrm{S}}(u) = \frac{\mathbb{E}(\xi_{tk}^{\mathrm{S}}\, X_t^{\mathrm{S}}(u))}{\lambda_k^{\mathrm{S}}} = \frac{1}{\lambda_k^{\mathrm{S}}} \int_{I_{\mathrm{S}}} \phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{S}}(u,v)dv,$$

for $u \in I_{\mathrm{S}}$, where the latter equation is a direct result form the definition of the kth FPC score ξ_{tk}^{S}.

This can obviously be generalized to an enlarged set of points $u \in I_{\mathrm{L}}$ by regressing the large fragment $X_t^{\mathrm{L}}(u)$ on the small FPC score ξ_{tk}^{S}, which leads to our definition of the predictive kth eigenfunction:

$$\tilde{\phi}_k^{\mathrm{L}}(u) \;:=\; \frac{\mathbb{E}(\xi_{tk}^{\mathrm{S}}\, X_t^{\mathrm{L}}(u))}{\lambda_k^{\mathrm{S}}} = \frac{1}{\lambda_k^{\mathrm{S}}} \int_{I_{\mathrm{S}}} \phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{L}}(u,v)dv \tag{2}$$

for $u \in I_{\mathrm{L}}$ if $\lambda_k^{\mathrm{S}} > 0$, while $\tilde{\phi}_k^{\mathrm{L}}(x) := 0$ if $\lambda_k^{\mathrm{S}} = 0$. The latter equality in Eq. (2) follows from the definition of the kth FPC score ξ_{tk}^{S} and the KL representation of X_t^{L}. In fact, $\tilde{\phi}_k^{\mathrm{L}}(u)$ is predictive only with respect to the remainder fragments $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$, but for $v \in I_{\mathrm{S}}$ we have that $\tilde{\phi}_k^{\mathrm{L}}(v) = \phi_k^{\mathrm{S}}(v)$. Furthermore, continuity of γ^{L} implies continuity of $\tilde{\phi}^{\mathrm{L}}$.

Given the observable centered small fragment X_t^{S} we now propose to predict the remainder fragments by applying Eq. (1) with respect to the predictive eigenfunctions $\tilde{\phi}_k^{\mathrm{L}}$:

$$\tilde{X}_t^{\mathrm{L}}(u) \;:=\; \sum_{k=1}^{\infty} \xi_{tk}^{\mathrm{S}}\, \tilde{\phi}_k^{\mathrm{L}}(u), \quad u \in I_{\mathrm{L}}. \tag{3}$$

Note that $\tilde{X}_t^{\mathrm{L}}$ is continuous and $\tilde{X}_t^{\mathrm{L}}(v) = X_t^{\mathrm{S}}(v)$ for all $v \in I_{\mathrm{S}}$, due to the above described properties of $\tilde{\phi}_k^{\mathrm{L}}(u)$.

2.1. *Theoretical properties.* First of all, we need to address the question whether Eq. (3) actually determines a well-defined random process $\tilde{X}_t^{\mathrm{L}}$ on $L^2(I_{\mathrm{L}})$, i.e., whether the infinite sum Eq. (3) converges in the L2 sense and whether the random functions $\tilde{X}_t^{\mathrm{L}}$ are continuous. Both is in fact true and for ease of reference we state this in the following theorem:

THEOREM 2.1. *The random function $\tilde{X}_t^{\mathrm{L}}$ defined in Eq. (3) has a continuous and finite variance function $\mathbb{V}(\tilde{X}_t^{\mathrm{L}}(u)) < \infty$ for all $u \in I_{\mathrm{L}}$.*
The proof can be found in Appendix B of the supplemental paper.

In the following we analyze the theoretical properties of the prediction error

$$Z_t(u) = X_t^{\mathrm{L}}(u) - \tilde{X}_t^{\mathrm{L}}(u), \quad \text{for all} \quad u \in I_{\mathrm{L}}, \tag{4}$$

but with a particular interest for all $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$, since anyways $Z_t(v) = 0$ for all $v \in I_{\mathrm{S}}$; see our discussion of Eq. (3). An obvious consequence of Eq. (3) is that $Z_t(u)$ has mean zero for all $u \in I_{\mathrm{L}}$, i.e., the prediction procedure is unbiased. Result (a) in the following theorem shows that the prediction error $Z_t(u)$ is orthogonal to the predictor $X_t^{\mathrm{S}}(u)$ which serves as an auxiliary result for result (b). Result (b) shows that $\tilde{X}_t^{\mathrm{L}}(u)$ is the optimal linear predictor of the true value $X_t^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{L}}$, i.e., having the lowest prediction error variance among all linear predictors. Finally, result (c) allows us to identify specifications of random functions X_t that can be predicted without any prediction error.

Theorem 2.2 (Optimal linear prediction).

(a) Under our setup we obtain that for every $v \in I_{\mathrm{S}}$ and $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$

$$\mathbb{E}\left(X_t^{\mathrm{S}}(v) Z_t(u)\right) = 0 \quad \textit{as well as} \tag{5}$$

$$s_Z^2(u) := \mathbb{E}\left(Z_t(u)^2\right) = \gamma^{\mathrm{L}}(u,u) - \sum_{k=1}^{\infty} \lambda_k^{\mathrm{S}} \tilde{\phi}_k^{\mathrm{L}}(u)^2. \tag{6}$$

(b) For any continuous linear functional $\ell_u : L^2(I_{\mathrm{S}}) \to \mathbb{R}$ we have that

$$\mathbb{E}\left((X_t^{\mathrm{L}}(u) - \ell_u(X_t^{\mathrm{S}}))^2\right) \geq \mathbb{E}\left((X_t^{\mathrm{L}}(u) - \tilde{X}_t^{\mathrm{L}}(u))^2\right) = s_Z^2(u)$$

for all $u \in I_{\mathrm{L}}$.

(c) Let X_t be a Gaussian process, i.e., all FPC scores are Gaussian random variables. Then the variance of the prediction error can be written as

$$s_Z^2(u) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left(\left(X_t^{\mathrm{L}}(u) - X_s^{\mathrm{L}}(u)\right)^2 \middle| X_t^{\mathrm{S}} = X_s^{\mathrm{S}}\right)\right) + \mathcal{O}(r^{|t-s|}), \tag{7}$$

where $X_t^{\mathrm{S}} = X_s^{\mathrm{S}}$ denotes point-wise equality, i.e., $X_t^{\mathrm{S}}(v) = X_s^{\mathrm{S}}(v)$ for all $v \in I_{\mathrm{S}}$.

The proofs can be found in Appendix B of the supplemental paper.

In order to explain result (c) of Theorem 2.2 consider first the case of i.i.d. random functions X_t, i.e., with $r = 0$ in Assumption A2. Then the variance of the prediction error can be written as

$$s_Z^2(u) = \frac{1}{2}\ \mathbb{E}\left(\mathbb{E}\left(\left(X_t^{\mathrm{L}}(u) - X_s^{\mathrm{L}}(u)\right)^2 \middle| X_t^{\mathrm{S}} = X_s^{\mathrm{S}}\right)\right), \quad t \neq s.$$

This result tells us that there is no prediction error, i.e., $s_Z^2(u) = 0$ for all $u \in I_{gr}$, if the structure of X_t is such that the event $X_t^{\mathrm{S}}(v) = X_s^{\mathrm{S}}(v)$ for all $v \in I_{\mathrm{S}}$ and $t \neq s$ implies that also $X_t^{\mathrm{L}}(u) = X_s^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$.

The same interpretation applies to weakly dependent functional time series $(X_t)_t$, but with the additional requirement that the distance between the time points t and s needs to be large enough such that $\mathcal{O}(r^{|t-s|})$ is negligible. This additional requirement assures that the implication, "$X_t^{\mathrm{S}} = X_s^{\mathrm{S}} \Rightarrow X_t^{\mathrm{L}} = X_s^{\mathrm{L}}$", is not just a mere consequence of a (positive) auto correlation when $|t-s|$ is small.

To give an example, assume that for some integer $K < \infty$ we have $X_t^{\mathrm{L}}(u) = \sum_{r=1}^K \theta_{tr}^{\mathrm{L}} f_r^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{L}}$, where $f_1^{\mathrm{L}}, \dots, f_K^{\mathrm{L}}$ are continuous orthonormal functions on I_{L}, and where $\theta_{tr}^{\mathrm{L}} \in \mathbb{R}$ are Gaussian random variables with mean zero and $\mathbb{E}(\theta_{tk}^{\mathrm{L}} \theta_{tl}^{\mathrm{L}}) = v_k < \infty$ for all $k = l \in \{1, \dots, K\}$ and zero else. Now consider the smaller interval I_{S}, and suppose that also the K truncated functions $f_1^{\mathrm{S}}, \dots, f_K^{\mathrm{S}}$, with $f_k^{\mathrm{S}}(u) := f_k^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{S}}$ and all $k \in \{1, \dots, K\}$, are linearly independent. This is equivalent to require that

$$\int_{I_{\mathrm{S}}} \left(\sum_{k=1}^K (\theta_{tk}^{\mathrm{L}} - \alpha_k) f_k^{\mathrm{L}}(u) \right)^2 du = 0, \text{ if and only if, } \theta_{t1}^{\mathrm{L}} = \alpha_1, \dots, \theta_{tK}^{\mathrm{L}} = \alpha_K.$$

In other words, in this situation the knowledge of the structure of X_t^{L} on I_{S} is sufficient to determine the corresponding coefficients $\theta_{t1}^{\mathrm{L}}, \dots, \theta_{tK}^{\mathrm{L}}$, and thus uniquely determines the function X_t^{L} on the larger set I_{L}. By (7) we can then conclude that there is no prediction error, i.e., $s_Z^2(u) = 0$ for all $u \in I_{\mathrm{L}}$. This can be made more precisely using change of basis arguments. If the orthonormal functions $f_1^{\mathrm{L}}, \dots, f_K^{\mathrm{L}}$ are linear independent over I_{S}, we have that $\sum_{k=1}^K \theta_{tk}^{\mathrm{L}} f_k^{\mathrm{L}}(u) = \sum_{k=1}^K \xi_{tk}^{\mathrm{S}} \phi_k^{\mathrm{S}}(u)$ for all $u \in I_{\mathrm{S}}$, where $\lambda_K^{\mathrm{S}} > 0$ and $\lambda_{K+1}^{\mathrm{S}} = 0$. Furthermore, there exist unique coefficients $\beta_1, \dots, \beta_K$ such that $\phi_k^{\mathrm{S}}(u) = \sum_{r=1}^K \beta_r f_r^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{S}}$. This then implies that $\tilde{\phi}_k^{\mathrm{L}}(u) = \sum_{r=1}^K \beta_r f_r^{\mathrm{L}}(u)$ as well as $X_t^{\mathrm{L}}(u) = \sum_{r=1}^K \theta_{tk}^{\mathrm{L}} f_k^{\mathrm{L}}(u) = \sum_{k=1}^K \xi_{tk}^{\mathrm{S}} \tilde{\phi}_k^{\mathrm{L}}(u) = \tilde{X}_t^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{L}}$, i.e., $s_Z^2(u) = 0$ for all $u \in I_{\mathrm{L}}$.

Note that the definition of the predictor in Eq. (3) as well as $s_Z^2(u) = \mathbb{V}(Z_t(u))$ are uniquely determined by the structure of the covariance function γ, and do not depend on whether the process is Gaussian or not. Hence, result (c) of Theorem 2.2 tells us that $s_Z^2(u) = 0$ if the structure of X_t is such that for two independent realizations X_t and X_s the event $X_t^{\mathrm{S}} = X_s^{\mathrm{S}}$ implies that also $X_t^{\mathrm{L}}(u) = X_s^{\mathrm{L}}(u)$ for all $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$. This may be fulfilled for simple structured functional data.

2.2. *The common features and differences regarding the classical functional prediction model.* Eq. (3), together with Eq. (2), defines a linear operator mapping $S : X_t^{\mathrm{S}} \mapsto \tilde{X}_t^{\mathrm{L}}$, such that

$$\tilde{X}_t^{\mathrm{L}}(u) = \left[S X_t^{\mathrm{S}}\right](u) \;\Leftrightarrow\; X_t^{\mathrm{L}}(u) = \left[S X_t^{\mathrm{S}}\right](u) + Z_t(u),\; u \in I_{\mathrm{L}}, \tag{8}$$

where the prediction error Z_t is as defined in Eq. (4), and where the linear operator $S : H_{\mathrm{S}} \to H_{\mathrm{L}}$ is defined as $S = \sum_{k=1}^{\infty} \tilde{\phi}_k^{\mathrm{L}} \otimes \phi_k^{\mathrm{S}}$ by using the tensor product notation $x^{\mathrm{L}} \otimes y^{\mathrm{S}}$ to denote $(x^{\mathrm{L}} \otimes y^{\mathrm{S}})(z^{\mathrm{S}}) = \langle y^{\mathrm{S}}, z^{\mathrm{S}} \rangle_{\mathrm{S}} x^{\mathrm{L}}$, for all $y^{\mathrm{S}}, z^{\mathrm{S}} \in H_{\mathrm{S}}$ and all $x^{\mathrm{L}} \in H_{\mathrm{L}}$.

Identification Restriction: As in the case of classical functional prediction models, we need to address the identifiability issue as dealing with infinite dimensional functional data generally involves always the issue of (in-principle infinitely many) zero-valued eigenvalues. I.e., we need to address the problem that $SX_t^{\mathrm{S}} = S(X_t^{\mathrm{S}} + W)$, where $W \in \ker(\Gamma^{\mathrm{S}})$. We follow the typical approach and focus only on that part of SX_t^{S} which is identifiable, i.e., for which the full rank condition $\ker(\Gamma^{\mathrm{S}}) = \{0\}$ holds, which is equivalent to the condition that $\lambda_k^{\mathrm{S}} > 0$ for all $k \in \{1, 2, \ldots\}$. The latter explains our definition in (2), where we set $\tilde{\phi}_k^{\mathrm{L}}(x) := 0$ if $\lambda_k^{\mathrm{S}} = 0$.

By plugging the definition of $\tilde{\phi}_k^{\mathrm{L}}$ into the definition of S and replacing X_t^{L} by its KL representation we can write Eq. (8) as

$$\begin{aligned} X_t^{\mathrm{L}}(u) &= \int_{I_{\mathrm{S}}} \left[\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{S}}} \phi_m^{\mathrm{L}}(u) \phi_k^{\mathrm{S}}(v)\right] X_t^{\mathrm{S}}(v)\, dv + Z_t(u) \\ \Leftrightarrow \tilde{X}_t^{\mathrm{L}}(u) &= \int_{I_{\mathrm{S}}} \left[\sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{S}}} \phi_m^{\mathrm{L}}(u) \phi_k^{\mathrm{S}}(v)\right] X_t^{\mathrm{S}}(v)\, dv, \end{aligned} \tag{9}$$

where $\sigma_{km} = \mathbb{E}(\xi_{tk}^{\mathrm{S}} \xi_{tm}^{\mathrm{L}})$. For general dependent variables Y_t, instead of X_t^{L}, and under the additional (restrictive) assumption that the limit of the infinite double sum within the square brackets in Eq. (9) exits, our functional prediction model becomes equivalent to the classical functional prediction model as considered by Bosq (2000), Yao, Müller and Wang (2005b), Crambes and Mas (2013), and Kraus (2015). These classical functional prediction models use a methodology that is motivated from generalizing the multivariate linear prediction model with its typical focus on parameter estimation to an infinite dimensional functional version (see, e.g., He, Müller and Wang, 2003). In fact, the summability assumption essentially means a particular, however restrictive, focus on the so-called integral kernel parameter $\beta(u, v) = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{\sigma_{km}}{\lambda_k^{\mathrm{S}}} \phi_m^{\mathrm{L}}(u) \phi_k^{\mathrm{S}}(v)$ and assuming the existence of $\beta(u, v)$ is, under the classical setup with $X_t \in L^2([a, b])$, equivalent to the assumption that S is a Hilbert-Schmidt operator.

By contrast to the above cited classical prediction approaches, we do not require the restrictive existence of $\beta(u, v)$, since we are anyways only interested in estimating the integral value $\tilde{X}_t^{\mathrm{L}}$. Indeed, a Hilbert-Schmidt assumption on S is particularly inappropriate in our rather specific prediction context, since it rules out the identity operator. I.e., it would rule out the in (9) necessary functional identity mapping of an observed fragment X_t^{S} to itself (remember that $I_{\mathrm{S}} \subset I_{\mathrm{L}}$). Our Theorem 2.1 assures that the integral value $\tilde{X}_t^{\mathrm{L}}$ is well defined even though the infinite double sum within the square brackets in Eq. (9) does not necessarily converge.

Essentially, Theorems 2.1 and 2.2 follow from the very specific nature of our prediction problem where we use the observed parts of X_t to predict the missing ones. This is fundamentally, different to the classical prediction models which aim to predict some general function Y. Within the classical prediction setup, one has to postulate a model like $Y = SX + \epsilon$ and then aims to identify and estimate the operator S, or the integral value SX, which requires the unverifiable model assumption of orthogonality between the predictor X and the prediction error ϵ (see, e.g., Crambes and Mas 2013, Section 1.2, or Hörmann and Kidziński 2015, Section 2.3).

By contrast, we derive our prediction model in Eq. (8) on basis of the heuristic considerations that lead to Eq. (2). Through this approach we get the explicit expression for the prediction error Z_t in Eq. (4). The orthogonality between the predictor X_t^{S} and the prediction error Z_t is assured by result (a) of Theorem 2.2, i.e., it is not just an unverifiable model assumption, but a direct consequence of our specific prediction context. The orthogonality result (a) of Theorem 2.2 implies the optimal linear prediction result (b) of Theorem 2.2 which is much stronger than the non-formal optimality statement in Kraus (2015) that is based on the usual (unverifiable) orthogonality assumption on the prediction error and the restrictive Hilbert-Schmidt assumption.

3. Estimation.

We rarely observe the functional trajectories directly, but typically only their noisy discretization points. In fact, the left panel of Figure 1 shows the *pre-smoothed* functions, however, the actual raw data is shown in Figure 2. In our estimation theory we take into account these additional discretization and measurement errors.

Let (Y_{it}, U_{it}) denote the observed noisy discretization pairs of a random functions X_t with

$$Y_{it} = X_t(U_{it}) + \varepsilon_{it}, \quad t \in \{1, \dots, T\}, \quad i \in \{1, \dots, n\}, \tag{10}$$

where $U_{it} \in \mathbb{R}$ is a stationary weakly dependent time series with finite second moments and ε_{it} a real i.i.d. random error term with mean zero, finite

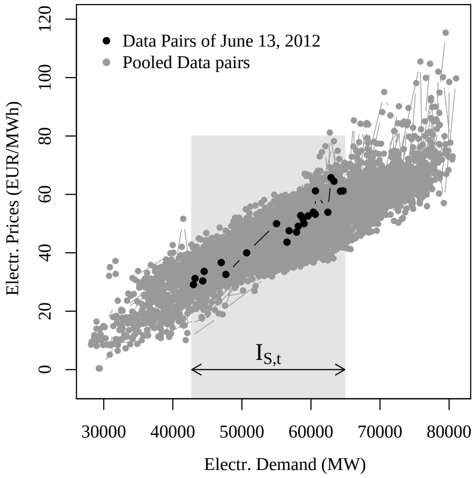

Fig 2. *Scatter plot of the observed data pairs* (Y_{it}, U_{it}).

variance $\mathbb{V}(\varepsilon_{it}) = \sigma_\varepsilon^2$, and finite fourth moment. The random variables U_{it} and ε_{js} are assumed to be independent from each other and from the random functions X_k for all $i, j \in \{1, \ldots, n\}$ and all $s, t, k \in \{1, \ldots, T\}$.

In classical functional data scenarios it is assumed that the random function argument U_{it} is distributed as $U_0 \sim f_{U_0}$, where $f_{U_0}(u) > 0$ for all $u \in I_0$ and zero else. This applies to the case of sparse functional data (see, e.g., Yao, Müller and Wang, 2005a), but essentially also includes the classical case where each function can be pre-smoothed in a pre-processing step (see Ramsay and Silverman, 2005, Ch. 3). In order to model the case of partially observed functional data, we assume that U_{it} is distributed as $U_t \sim f_{U_t}$, where $f_{U_t}(u) > 0$ for all $u \in I_{\mathrm{S},t}$ with $I_{\mathrm{S},t} \subset I_0$ as specified in Assumption RS; see also Assumption A3 below.

Our empirical predictor is the truncated, empirical version of the definition of $\tilde{X}_t^{\mathrm{L}}$ in Eq. (3):

$$\hat{\tilde{X}}^{\mathrm{L}}_{t,K}(u) = \hat{\mu}(u; h_\mu) + \sum_{k=1}^{K} \frac{\hat{\xi}^{\mathrm{S}}_{tk}}{\hat{\lambda}^{\mathrm{S}}_k} \int_{I_{\mathrm{S}}} \hat{\phi}^{\mathrm{S}}_k(v) \hat{\gamma}^{\mathrm{L}}(u, v) dv, \tag{11}$$

where $\hat{\xi}^{\mathrm{S}}_{kt}$ can be estimated, e.g., by the following integral approximation:

$$\hat{\xi}^{\mathrm{S}}_{kt} = \sum_{i=1}^{n} \hat{\phi}^{\mathrm{S}}_k(U^{\mathrm{S}}_{it})(Y_{it} - \hat{\mu}(U^{\mathrm{S}}_{it}; h_\mu))(U^{\mathrm{S}}_{it} - U^{\mathrm{S}}_{i-1,t}) \tag{12}$$

with $U^{\mathrm{S}}_{0,t} = A_t$. The parameter K can be chosen, e.g., by the usual fraction-of-explained-variance approach.

This integral approximation, however, may have bad finite sample properties. The discretization of the empirical eigenfunctions at the design points U_{it} may lead to a lack of orthogonality between the empirical eigenfunctions. More stable and asymptotically equivalent to (12) is to estimate the FPC scores $(\xi^{\mathrm{S}}_{1t}, \dots, \xi^{\mathrm{S}}_{Kt})^\top$ through a multiple linear regression of $Y^{\mathrm{S}}_{it} - \hat{\mu}(U^{\mathrm{S}}_{it}; h_\mu)$ on the K regressors $\hat{\phi}^{\mathrm{S}}_1(U^{\mathrm{S}}_{it}), \dots, \hat{\phi}^{\mathrm{S}}_K(U^{\mathrm{S}}_{it})$, with $(Y^{\mathrm{S}}_{it}, U^{\mathrm{S}}_{it})$ denoting the noisy observations from the partially observed function X^{S}_t; note that this regression must not contain a constant. We use this regression approach in our implementations.

For estimating μ and γ we use classical local linear estimators, where the below definitions are based on matrix notation as proposed, e.g., in Ruppert and Wand (1994). For given data (Y_{it}, U_{it}), with $t \in \{1, \dots, T\}$ and $i \in \{1, \dots, n\}$, we estimate the mean function $\mu(u)$ by

$$\hat{\mu}(u; h_\mu) = e_1^\top \left([\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u} [\mathbf{1}, \mathbf{U}_u] \right)^{-1} [\mathbf{1}, \mathbf{U}_u]^\top \mathbf{W}_{\mu,u} \mathbf{Y}, \tag{13}$$

where $e_1 = (1, 0)^\top$, $[\mathbf{1}, \mathbf{U}_u]$ is a $nT \times 2$ dimensional data matrix with typical rows $(1, U_{it} - u)$, the $nT \times nT$ dimensional diagonal weighting matrix $\mathbf{W}_{\mu,u}$ holds the kernel weights $K_{\mu,h}(U_{it} - u) = h_\mu^{-1}\, \kappa(h_\mu^{-1}(U_{it} - u))$. The kernel function κ is assumed to be a univariate, symmetric, pdf with compact support $\mathrm{supp}(\kappa) = [-1, 1]$, such as, e.g., the univariate Epanechnikov kernel (Assumption A7). 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 objects $\mathbf{U}_u$ and $\mathbf{W}_{\mu,u}$ are filled in correspondence with the nT dimensional vector $\mathbf{Y} = (Y_{11}, Y_{21}, \dots, Y_{n-1,T}, Y_{n,T})^\top$.

For given finite data $(\hat{C}_{ijt}, U_{it}, U_{jt})$, with $t \in \{1, \dots, T\}$ and $i \neq j \in \{1, \dots, n\}$, we estimate the covariance function $\gamma(u, v)$ by $\hat{\gamma}(u, v; h_\gamma) =$

$$= e_1^\top \left([\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]^\top \mathbf{W}_{\gamma,u,v} [\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v] \right)^{-1} [\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]^\top \mathbf{W}_{\gamma,u,v} \hat{\mathbf{C}}, \tag{14}$$

where $e_1 = (1, 0, 0)^\top$, $[\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]$ is a $NT \times 3$ dimensional data matrix with typical rows $(1, U_{it} - u, U_{jt} - v)$, $N = (n^2 - n)$, the $NT \times NT$ dimensional diagonal weighting matrix $\mathbf{W}_{\gamma,u,v}$ holds the bivariate kernel weights $K_{\gamma,h}(U_{it} - u, U_{jt} - v)$. 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 objects $[\mathbf{1}, \mathbf{U}_u, \mathbf{U}_v]$ and $\mathbf{W}_{\gamma,u,v}$ are filled in correspondence with the NT dimensional vector of empirical raw covariances with

$$\hat{\mathbf{C}} = (\hat{C}_{211}, \hat{C}_{311}, \dots, \hat{C}_{n-2,n,T}, \hat{C}_{n-1,n,T})^\top$$
$$\hat{C}_{ijt} = (Y_{it} - \hat{\mu}(U_{it}))(Y_{jt} - \hat{\mu}(U_{jt})) \tag{15}$$

for all t and all $i \neq j$. Raw covariances $\hat{C}_{ijt}$ with $i = j$ need to be removed as these would introduce an estimation bias at the diagonal of γ through taking squares of the error term ε_{it} that is contained in Y_{it}. This removal explains the number of raw covariance points $(n^2 - n)T = NT$.

Estimates of the eigenvalues λ_k^{S} and the eigenfunctions ϕ_k^{S} are defined by the corresponding solutions of the empirical eigenequations:

$$\int_{I_{\mathrm{S}}} \hat{\gamma}(u, v; h_\gamma) \hat{\phi}_k^{\mathrm{S}}(v)\, dv = \hat{\lambda}_k^{\mathrm{S}}\, \hat{\phi}_k^{\mathrm{S}}(u), \text{ for } u \in I_{\mathrm{S}}. \tag{16}$$

4. Prediction algorithm.

In practice our prediction algorithm usually needs to be repeated. In the following we describe the first run:

1st Run: For each $t \in \{1, \dots, T\}$, compute $\hat{\hat{X}}_{t,K}^{\mathrm{L}}$ according to Eq. (11) with $I_{\mathrm{S},t} = [A_t, B_t]$, where $A_t = \min_{1 \le i \le n}(U_{it})$ and $B_t = \max_{1 \le i \le n}(U_{it})$.

An exemplary first run is visualized in Figure 3. The right panel of Figure 3 demonstrates how the resulting large intervals $I_{\mathrm{L},t}$ depend on $I_{\mathrm{S},t}$ and the non-missing parts of the covariance function $\hat{\gamma}$; an uncovered plot of $\hat{\gamma}$ can be seen in the right panel of Figure 7.

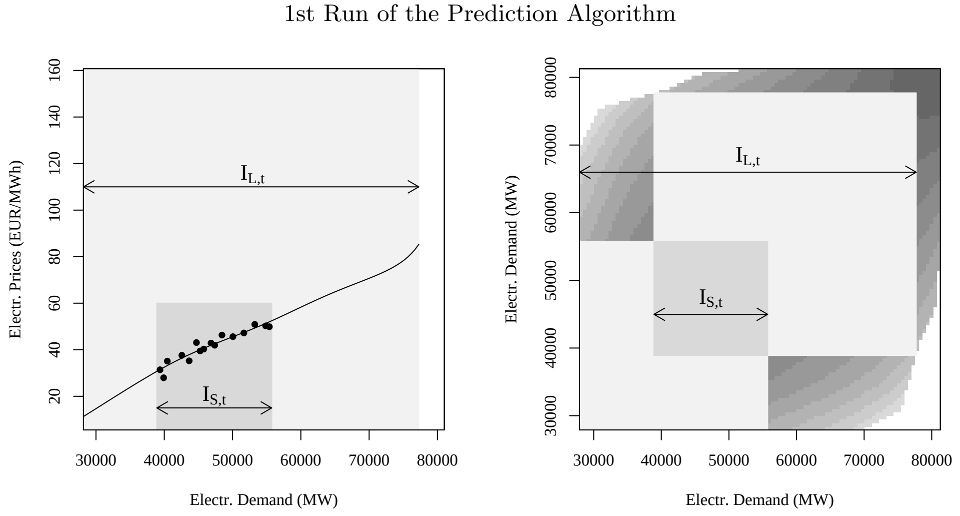
1st Run of the Prediction Algorithm

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

The first run of the prediction algorithm my not allow to predict all the missing fragments. For instance, the predicted price function in the left panel of Figure 3 still lacks the upper fragment for values of electricity demand $u \in [77362\text{ (MW)}, 82282\text{ (MW)}]$. In order to predict also the further missing fragments, we suggest to repeat the prediction algorithm. In the following we describe the ℓth, $\ell \geq 2$, run of the algorithm:

$\boldsymbol{\ell}$th Run: For each $t \in \{1, \dots, T\}$, compute $\hat{\hat{X}}^{\mathrm{L}}_{t,K}$ according to Eq. (11), but with $I_{\mathrm{S},t} \subseteq \bigcup_{m<\ell} I^{m\text{th Run}}_{\mathrm{L},t}$, such that:

1. A missing fragment can be predicted.
2. The width of $I_{\mathrm{S},t}$ is maximized.

That is, the positioning of $I_{\mathrm{S},t}$ needs to enable the prediction of further missing fragments, but otherwise we need to maximize the width of $I_{\mathrm{S},t}$ in order to maximize the information contained in the "small" fragment.

An exemplary second run is visualized in Figure 4. There the new small interval $I_{\mathrm{S},t} \subseteq I^{\text{1st Run}}_{\mathrm{L},t}$ is chosen such that the missing upper fragment can be predicted, but otherwise such that its width is maximized. The new large interval $I_{\mathrm{L},t}$ contains the missing upper fragment, but also the redundant region $I_{\mathrm{L},t} \cap I^{\text{1st Run}}_{\mathrm{L},t}$. We keep only the newly predicted fragment with respect to the upper interval $I_{\mathrm{L},t} \setminus \{I_{\mathrm{L},t} \cap I^{\text{1st Run}}_{\mathrm{L},t}\} = [77362\text{ (MW)}, 82282\text{ (MW)}]$, and join it with the prediction from the first run.

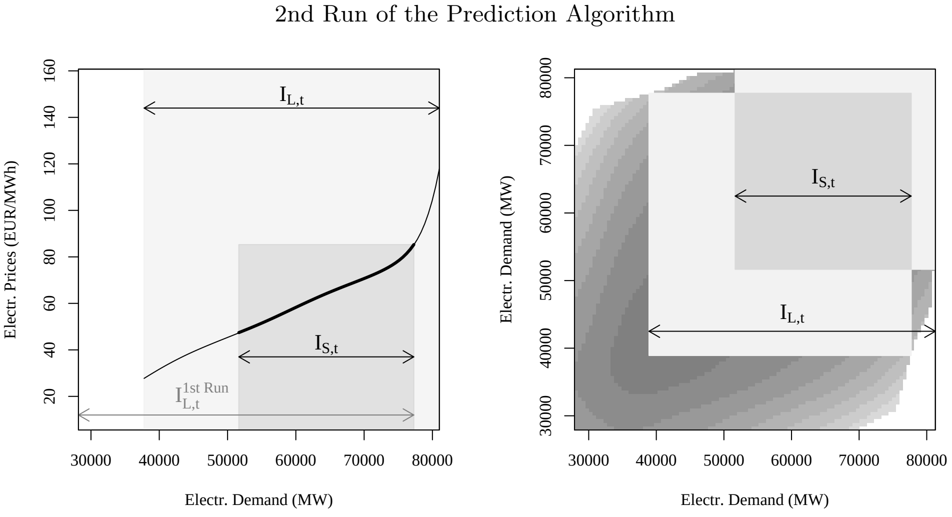
2nd Run of the Prediction Algorithm

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

Practical remarks:

- After each repetition, the newly predicted fragments need to be joined with the predicted fragments from the preceding runs. Theoretically, as long as the parameter K remains fix, the single fragments will be perfectly aligned. Empirical discontinuities at the junction points are typically very small and can be balanced by adding a constant to the newly predicted fragment.
- In the ℓth, $\ell \geq 2$, run of the prediction algorithm, we can make use of the predicted smooth function that is constructed from the preceding $(\ell-1)$ runs, say, $\hat{\hat{X}}_t^{(\ell-1)}$. In particular, we can estimate the new "small" FPC scores by the classical integral method, i.e., by
$$\hat{\xi}_{kt}^{\mathrm{S}} = \int_{I_{\mathrm{S},t}} (\hat{\hat{X}}_t^{(\ell-1)}(u) - \hat{\mu}(u))\hat{\phi}_k^{\mathrm{S}}(u)du.$$
- We can predict a lower (or upper) fragment completely if the small interval $I_{\mathrm{S},t}$ is chosen such that the covariances $\gamma(v,a)$ (or $\gamma(v,b)$) are estimable for all $v \in I_{\mathrm{S},t}$. Compare to this the position of the small interval $I_{\mathrm{S},t}$ in the right panel of Figure 3. There we have estimates $\hat{\gamma}(v,a)$ for all $v \in I_{\mathrm{S},t}$ which allow us to predict the complete lower fragment, but we have not the estimates $\hat{\gamma}(v,b)$ for all $v \in I_{\mathrm{S},t}$, such that we cannot predict the complete upper fragment. The right panel of Figure 4 shows just the opposite case.

5. Asymptotic results.

We refer to our model assumptions, as specified in the first paragraph of Section 2 and in the second paragraph of Section 3, as Assumption **A1**. Due to space limitations we list our further Assumptions A2-A9 in Appendix A of the supplemental paper. Our assumptions are generally close to those in Yao, Müller and Wang (2005b) and Hall, Müller and Wang (2006). Though, we additionally allow for weakly dependent time series of random function $(X_t)_t$ (Assumption A2), consider the case of proper (random) subsets (Assumption A3), and a more general asymptotic setup as specified by the following assumption:

A4 (Asymptotic Scenario) $Tn \to \infty$, where $n = n(T) \geq 2$ such that $n(T) \sim T^\theta$ with $0 \leq \theta < \infty$. Here "$a_T \sim b_T$" denotes that two sequences a_T and b_T are asymptotically equivalent up to some positive constant $0 < c < \infty$, i.e., that $\lim_{T\to\infty}(a_T/b_T) = c$.

Assumptions A5-A7 contain typical smoothness, bandwidth and kernel assumptions. In order to make use of the eigenvalue and eigenfunction expansion of Hall and Hosseini-Nasab (2006), we impose the regularity Assumptions A8 and A9.

Remarks on Assumption A4. For the estimation of the mean and covariance functions, the eigenvalues, and eigenfunctions we allow also for the case of very sparsely sampled prediction points U_{it}, i.e., with $\theta = 0$ corresponding to, e.g., $n \leq 5$. However, for estimating the pc-scores, we only focus on the case of $\theta > 0$, since otherwise the pc-scores cannot be consistently estimated. Under truly sparse sampling schemes one should use the PACE procedure of Yao, Müller and Wang (2005a) and Yao, Müller and Wang (2005b) to get approximations for the pc-scores. We follow Hall, Müller and Wang (2006) and consider n as deterministic, though, under some minor modifications n can be assumed to be random. For instance, our results directly apply to the case with n being replaced by N_T, where N_T is an independent random variable defined as $N_T = n(T) + W_T$, where $n(T)$ is deterministic function of T with $n(T) \sim T^\theta$ and the random variable W_T has realizations in some appropriate subset of $\mathbb{N}$, such that $W_T \geq 2$ wp1 and $\mathbb{E}(N_T) = n(T)$ for all T as $T \to \infty$.

THEOREM 5.1 (Preliminary uniform consistency results). *Under Assumptions A1-A7 we have the following weak consistency results:*

(a) *Estimator of the mean function:*

$$\sup_{u \in I_0} |\hat{\mu}(u; h_\mu) - \mu(u)| = \mathcal{O}_p\left(h_\mu^2 + \frac{1}{\sqrt{Tn\, h_\mu}} + \frac{1}{\sqrt{T}}\right). \tag{17}$$

(b) *Estimator of the covariance function:*

$$\sup_{(u,v) \in I_0^2} |\hat{\gamma}(u, v; h_\gamma) - \gamma(u, v)| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\, h_\gamma^2}} + \frac{1}{\sqrt{T}}\right). \tag{18}$$

The following results (c) and (d) are valid under the additional Assumption A9, our identification restriction in Section 2.2, i.e., that $\lambda_k > 0$ for all $k \geq 1$, and under the assumption that all eigenvalues $\{\lambda_k\}_{k \geq 1}$ are of multiplicity one. Furthermore, we assume that $\hat{\phi}_k$ is chosen such that $\langle \hat{\phi}_k, \phi_k \rangle_H > 0$.

(c) *Estimators of the eigenvalues:*

$$\sup_{k \geq 1} |\hat{\lambda}_k - \lambda_k| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\, h_\gamma^2}} + \frac{1}{\sqrt{T}}\right). \tag{19}$$

(d) *Estimators of the eigenfunctions:*

$$\sup_{u\in I_0}|\hat{\phi}_k(u)-\phi_k(u)| = \mathcal{O}_p\left(\frac{1}{\delta_k}\left(h_\gamma^2+\frac{1}{\sqrt{TN\,h_\gamma^2}}+\frac{1}{\sqrt{T}}\right)\right), \tag{20}$$

for all $1\leq k\leq \bar{K}_{TN}-1$ *with* $\bar{K}_{TN}=\inf\{k\geq 1: \lambda_k-\lambda_{k+1}\leq 2\hat{\Delta}_{TN}\}$, *where* $\hat{\Delta}_{TN} = (\int_{(u,v)\in I_0^2}(\hat{\gamma}(u,v)-\gamma(u,v))^2 d(u,v))^{1/2}$, *with* $\hat{\Delta}_{TN} = \mathcal{O}_p(h_\gamma^2+(TN\,h_\gamma^2)^{-1/2}+T^{-1/2})$, *and where* $\delta_k=\min_{1\leq i\leq k}\{\lambda_i-\lambda_{i+1}\}$.

The proofs can be found in Appendix B of the supplemental paper.

Remarks on Theorem 5.1. By contrast to Yao, Müller and Wang (2005b) our rates of the estimators $\hat{\mu}$ and $\hat{\gamma}$ are optimal. Note that it is generally advisable use an under-smoothed mean function to construct the raw covariance points $\hat{C}_{ijt}$ for estimating the covariance function.

Theorem 5.2 (Main uniform consistency results). *Under Assumptions A1-A9 and our identification restriction in Section 2.2 the following results (a)-(c) hold for* $1\leq K\leq \bar{K}_{TN}-1$, $a>1$, *and for optimal bandwidth choices* $h_\mu\sim(Tn)^{-1/5}$ *and* $h_\gamma\sim(TN)^{-1/6}$*:*

(a) *For* $0<\theta<1/4$ *(*$\approx T/n$ *large)*

$$\sup_{u\in I_0}\left|\tilde{X}_t^{\mathrm{L}}(u)-\hat{\hat{X}}_{t,K}^{\mathrm{L}}(u)\right| = \mathcal{O}_p\left(K\,n^{-1/2}+(TN)^{-1/3}K^{a+2}+\sum_{k=K+1}^{\infty}k^{-a/2}\right).$$

(b) *For* $1/4\leq\theta<1$ *(*$\approx T/n$ *moderate)*

$$\sup_{u\in I_0}\left|\tilde{X}_t^{\mathrm{L}}(u)-\hat{\hat{X}}_{t,K}^{\mathrm{L}}(u)\right| = \mathcal{O}_p\left(K\,n^{-1/2}+T^{-1/2}K^{a+2}+\sum_{k=K+1}^{\infty}k^{-a/2}\right).$$

(c) *For* $1\leq\theta<\infty$ *(*$\approx T/n$ *small)*

$$\sup_{u\in I_0}\left|\tilde{X}_t^{\mathrm{L}}(u)-\hat{\hat{X}}_{t,K}^{\mathrm{L}}(u)\right| = \mathcal{O}_p\left(T^{-1/2}K^{a+2}+\sum_{k=K+1}^{\infty}k^{-a/2}\right).$$

The proofs can be found in Appendix B of the supplemental paper.

Remarks on Theorem 5.2. The $\mathcal{O}_p(\sum_{k=K+1}^{\infty}k^{-a/2})$ terms quantify the cut-off regularization error, while the other terms quantify the different estimation errors. Eq. (50) in Appendix B of the supplemental paper contains a more general consistency result without imposing the eigenvalue Assumption A8 and without assuming optimal bandwidth choices.

6. Simulation study. For our simulation study we generate $T = 200$ many iid normal and $T = 200$ many iid exponential random functions $X_t(u) = \mu(u) + \xi_{1,t}\phi_1(u) + \xi_{2,t}\phi_2(u)$, 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 $\xi_{k,t} \sim N(0, \lambda_k)$ with $\lambda_1 = 4$ and $\lambda_2 = 3$. For the exponential case we let $\xi_{k,t}$ be a centered exponential random variable with rate λ_k and centered by λ_k^{-1}. Furthermore, $U_{1t}, \dots U_{nt} \overset{\text{iid}}{\sim} \text{Unif}[A_t, B_t]$, where $A_1, \dots, A_t \overset{\text{iid}}{\sim} \text{Unif}[a, a + (b-a) \cdot 0.25]$ and $B_t = A_t + (b-a) \cdot 0.75$. That is, the discretization points of a function X_t cover at most three-quarter of the total domain $[a, b]$. Finally, the observations Y_{it} are generated according to $Y_{it} = X_t(U_{it}) + \varepsilon_{it}$ with $\varepsilon_{it} \sim N(0, 0.2)$.

The prediction procedure is implemented as described in Section 4. We allowed for a maximum of five repetitions of the prediction algorithm. The implementation is done in R and the codes are available from the authors. The parameter K is selected by the Fraction of Variance Explained (FVE) criterion with FVE $= 0.95$, i.e., the value of K is the highest K such that $(\sum_{l=1}^{K} \hat{\lambda}_l^{\text{L}})/(\sum_{j \geq 1} \hat{\lambda}_j^{\text{L}}) \leq \text{FVE}$.

In the jth simulation run we compute the Mean Absolute Prediction Error (MAPE) as

$$\text{MAPE}_j = T^{-1} \sum_{t=1}^{T} \max_u |\hat{\hat{X}}_{j,t,K}^{(5)}(u) - X_t(u)|, \quad j = \{1, \dots, B\}$$

with in total $B = 100$ simulation runs, where $\hat{\hat{X}}_{j,t,K}^{(5)}$ denotes the sequentially recovered function from five runs of our prediction algorithm described in Section 4. Only five runs are used in order to reduce the computational costs. Bandwidth selection is done using cross validation.

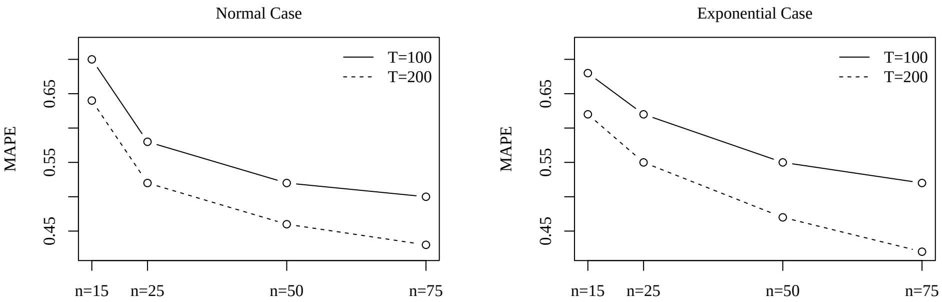

Fig 5. *Mean absolute prediction errors for the normal and exponential data generating processes.*

In each simulation run the percentage of recovered functions that cover the total domain is at least 80%. The total averages $\text{MAPE} = B^{-1}\sum_{j=1}^{B}\text{MAPE}_j$ for the two different data generating processes are shown in Figure 5. The simulation results demonstrate that the improvements in the MAPE are larger for increasing n from 15 to 25 than for increasing n from 50 to 75. This is in line with our asymptotic results in Theorem 5.2, where result (a) shows that if n is relatively small, then increasing n will reduce the estimation error. However, result (c) of Theorem 5.2 shows that the positive effect of a larger n vanishes if n is already large relatively to T. In this case only increasing T can effectively reduce the estimation error.

7. Application.

7.1. *Data sources and preprocessing.* 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`).

The data dimensions are given by $n = 24$ hours and $T = 241$ working days between March 15, 2012 and March 14, 2013. Very few (0.4%) of the data pairs (Y_{it}, U_{it}) with prices $Y_{it} > 120$ EUR/MWh and $U_{it} > 82000$ MW are considered as outliers and removed. This leads to cases of $n < 24$, but our estimation procedure can directly deal with such cases. As mentioned in the introduction, "electricity demand" actually denotes *residual* electricity demand, i.e., electricity demand minus infeeds from RES. defined as $X_{it} = \texttt{Elect.Demand}_{it} - \texttt{RES}_{it}$, where $\texttt{RES}_{it} = \texttt{Wind.Infeed}_{it} + \texttt{Solar.Infeed}_{it}$. The effect of further RES such as biomass is still negligible for the German electricity market.

7.2. *Checking for perfect predictability.* According to result (c) of Theorem 2.2, the prediction error can be zero if the structure of the random function X_t is simple enough. In the following, we check empirically whether the functions are perfectly predictable. The idea is to use the observed functions and to partition them into pseudo-observed and pseudo-missing fragments. If the functions are perfectly predictable, and if there are no estimation errors, then the predictions of the pseudo-missing fragments equal the observed pseudo-missing fragments. The practical problem is, of course, to deal

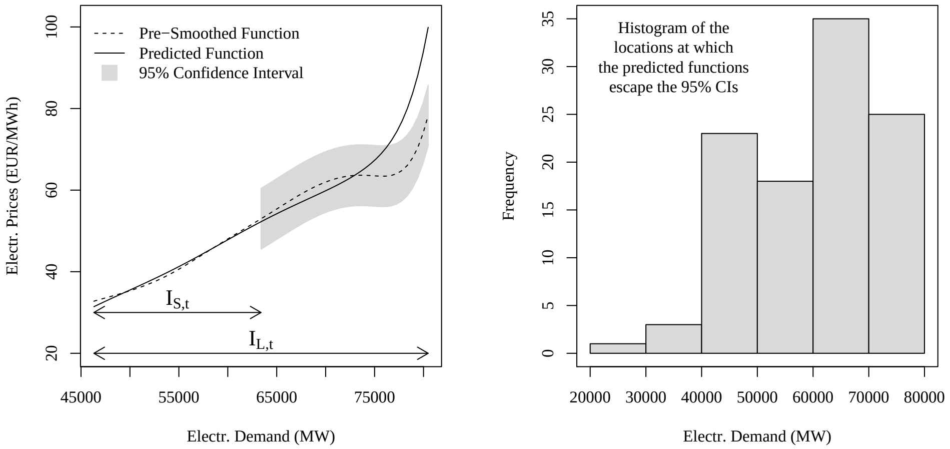

FIG 6. LEFT PANEL: *Checking whether the predicted function is within the 95% confidence interval of the pre-smoothed function.* RIGHT PANEL: *Histogram of the locations at which the predicted functions escape the 95% confidence intervals of the pre-smoothed functions.*

with the estimation errors; particularly, as we do not observe the functions directly.

We choose a rather pragmatic and conservative approach to deal with this problem. We take the lower halfs of individually pre-smoothed functions as the pseudo-observed fragments and use them to predict their pseudo-missing upper halfs. In order to account for the pre-smoothing estimation error and the multiple comparisons we consider a Bonferroni adjusted Gaussian 95% confidence intervals for the pre-smoothed functions. Only if the prediction result is within the Bonferroni adjusted 95% confidence interval, the prediction is considered as perfect. The left panel of Figure 6 demonstrates this approach for the case of an imperfect prediction. We repeat these checks by predicting also the lower halfs of the pre-smoothed functions given their upper halfs which leads in total to $2\,T$ checks. This is a very conservative approach as it ignores the estimation errors with respect to the predictions. For instance, it is quite possible that the imperfect prediction shown in the left panel of Figure 6 can be explained by the estimation errors with respect to the recovered function.

The function-wise pre-smoothing is done using a local linear estimator based on a cross validated smoothing parameter. The Gaussian confidence intervals require an estimate of the unknown variance σ^2_ε. For this we use the nonparametric variance estimator of Gasser, Sroka and Jennen-Steinmetz (1986), say $\hat{\sigma}^2_{\mathrm{GSJ},t}$, and take the average over all function-wise estimators as the final variance estimate, i.e., $\hat{\sigma}^2_\varepsilon = T^{-1}\sum_{t=1}^{T}\hat{\sigma}^2_{\mathrm{GSJ},t}$. We do not apply any bias corrections for constructing the confidence intervals, since the price functions have anyways a very low curvature such that the bias can be considered subordinate.

Using this approach, 22% of the $2\,T$ predictions are considered as imperfect. Under the hypothesis of perfect predictability and the chosen significance level of $\alpha = 5\%$, we expect only 5% of the predictions to be falsely classified as imperfect – ignoring now the conservative aspect of our approach. That is, it seems that the hypothesis of perfect predictability is violated. The histogram in the right panel of Figure 6 clearly shows that most of the imperfect predictions happen when predicting functions over domain values greater than $60,000$ MW of electricity demand. This is not surprising, since at these values of electricity demand so-called scarcity price premia come into play (Burger, Graeber and Schindlmayr, 2008, Ch. 4.3.1). These premia lead to much more complex price formations than it is the case for lower values of electricity demand where electricity prices essentially equal their marginal costs plus some profit surcharge. In fact, if we exclude the price functions that cover electricity demand values of $\geq 63,300$ MW from our checks, then only 5% of the predictions are classified as imperfect which is consistent with our perfect predictability hypothesis.

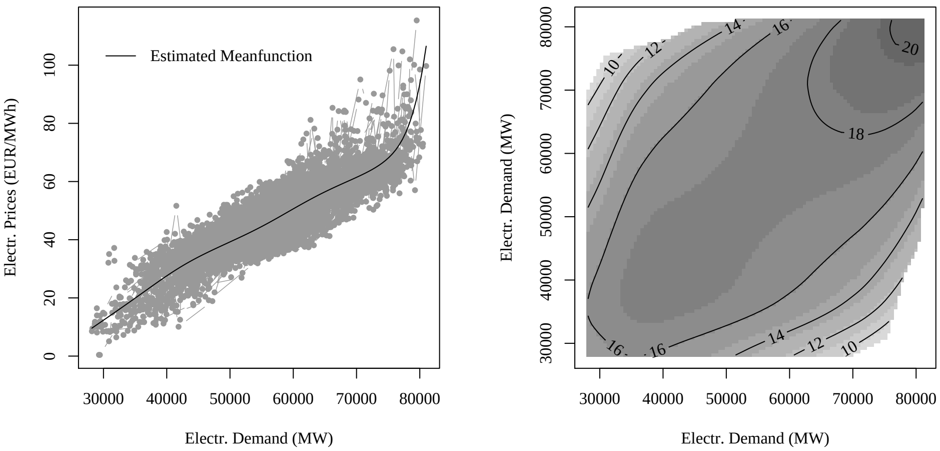

FIG 7. LEFT PANEL: *Estimated mean function plus scatter plot of the data pairs* (Y_{it}, U_{it}). RIGHT PANEL: *Contour plot of the estimated covariance function. The white regions reflect the outer off-diagonal parts which are infeasible to estimate.*

7.3. *Empirical results.* The estimated mean function and the estimated covariance function are shown in Figure 7. 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_{it}, U_{jt}); see also the right panel in Figure 1. For predicting the missing parts of each function X_t we use the repetitive prediction algorithm as described in Section 4. In a first run, we use the information with respect to the actually observed fragments in order to predict the missing fragments as visualized in Figure 3. For the consecutive runs we use the enlarged function (or a part of it) as the new observed fragment in order to predict further missing fragments as visualized in Figure 4. As in our simulation study we use five prediction runs which allows us to recover slightly over 90% of the price functions over the total support from $28,166$ MW to $81,011$ MW of electricity demand. The predicted functions are shown in the left panel of Figure 8.

Our preceding analysis on checking for perfect predictability, suggests that the predictions based on functions that do not cover electricity demand values of $\geq 63,300$ MW are more reliable than those that are based on functions that cover electricity demand values of $\geq 63,300$ MW. While the former set of functions is likely to suffer only from estimation errors, the latter set of functions is likely to suffer additionally from prediction errors. The different sets are shown in the right panel of Figure 8. Though, we emphasize that the predictions are optimal among all linear predictions – no matter whether the prefect predictability condition holds or not. The prediction of negative prices is well justified as the EPEX allows for negative prices since the price reform in 2008 (Nicolosi, 2010). Electricity producers are often willing to sell electricity at negative prices (i.e., to pay for selling), since shutting-off and re-starting their power plants is typically very expensive.

The predicted functions could now be used to re-estimate the covariance function over its total support. Given the re-estimated covariance function, we could then to recover also the rest of the functions over the total support. Though, this is not further pursued here for reasons of space.

8. Conclusion.

We propose a new prediction model that allows to predict the missing fragments of partially observed functional data from the observed fragments. By contrast to the alternative approaches in the literature, we consider the most challenging case of partially observed functional data for which it is impossible to estimate the covariance function over its total support. Our estimation algorithm allows to predict the missing parts of the random functions under this very challenging setup.

We show that our predictor is the best linear predictor – a property that

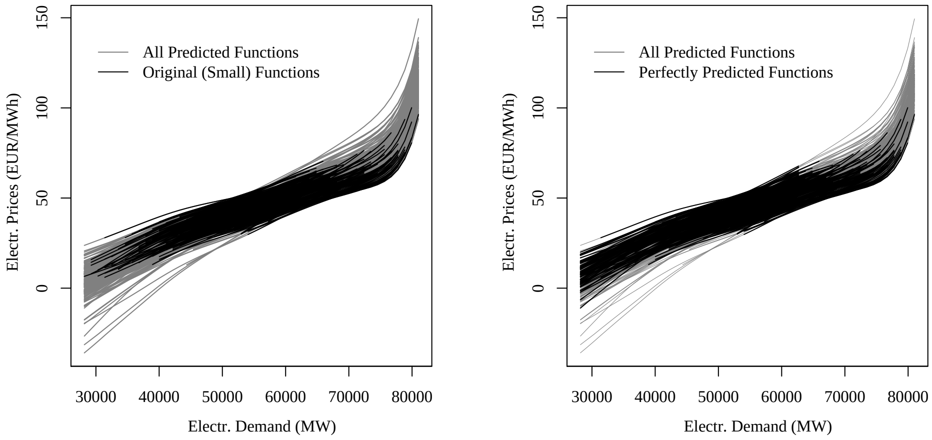

Fig 8. Left Panel: *All predicted functions (gray) and together with their original small fragments (black).* Right Panel: *All predicted functions (gray) and together with the perfectly predicted functions (black).*

is also claimed by the alternative functional predictors in the literature. Though, by contrast to the results in the literature, our optimality result is not based on the usual, but unverifiable, model assumption of orthogonality between the functional predictor and the functional prediction error. We can proof our result without this critical assumption, by making use of the very specific nature of our prediction problem: Namely, to predict the missing fragments of a random functions from its own observed fragments.

In our estimation theory, we focus on the relevant practical situation in which we do not directly observe the fragments, but only their noisy discretization points. This situation involves the use of nonparametric estimation procedures and we derive the uniform rates of consistency for our nonparametric estimation procedures. The finite sample properties are investigated through simulations and a real data application.

References.

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

Burger, M., Graeber, B. and Schindlmayr, G. (2008). *Managing Energy Risk: An Integrated View on Power and Other Energy Markets*, 1. ed. Wiley.

Crambes, C. and Mas, A. (2013). Asymptotics of prediction in functional linear regression with functional outputs. *Bernoulli* **19** 2627–2651.

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

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*, 1. ed. *Springer Series in Statistics.* Springer.

Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

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.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68** 109–126.

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.

He, G., Müller, H. G. and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *Journal of Multivariate Analysis* **85** 54–77.

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.

Hörmann, S. and Kidziński, Ł. (2015). A note on estimation in Hilbertian linear models. *Scandinavian journal of statistics* **42** 43–62.

James, G. M., Hastie, T. J. and Sugar, C. A. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63** 533–550.

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

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*, 2. ed. *Springer Series in Statistics.* Springer.

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

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

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.

Dominik Liebl and Alois Kneip

Statistische Abteilung

University of Bonn

Adenauerallee 24-26

53113 Bonn, Germany

E-mail: dliebl@uni-bonn.de

Supplemental paper for:

On the Optimal Reconstruction of Partially Observed Functional Data

by Dominik Liebl and Alois Kneip

CONTENT

Appendix A of this supplemental material contains the detailed list of assumptions. The proofs of Theorems 2.1, 2.2, 5.1, and 5.2 are given in Appendix B. The main steps in our proofs of the Theorems 5.1 and 5.2 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 A2), consider a different asymptotic setup (see Assumption A4), and exploit the assumption of kernel functions with compact support (see Assumption A7).

APPENDIX A: LIST OF ASSUMPTIONS

A2 (Weak Dependency) Let U_{it} i.i.d. as U_t for all $i \in \{1,\dots,n\}$. Let $(X_t)_t$, $(U_t)_t$, $(A_t)_t$, and $(B_t)_t$ be strictly stationary processes, where $(U_t)_t$, $(A_t)_t$, and $(B_t)_t$ are independent from $(X_t)_t$. Define the following autocovariance functions: $\gamma_h^U = \text{Cov}_h(U_t, U_{t+h})$, $\gamma_h(u,v) = \text{Cov}_h(X_t(u), X_{t+h}(v))$, and $\dot{\gamma}_h((u_1,v_1),(u_2,v_2)) = \text{Cov}_h(\dot{X}_t(u_1,v_1), \dot{X}_{t+h}(u_2,v_2))$ for $h \in \{1,2,\dots\}$, where $\dot{X}_s(u,v) = (X_j(u)-\mu(u))(X_j(v)-\mu(v))$. We assume that there are (generic) constants c and r, with $0 < c < \infty$ and $0 < r < 1$, such that $|\gamma_h^U| \leq cr^h$, $\sup_{(u,v)\in I_0^2} |\gamma_h(u,v)| \leq cr^h$, and $\sup_{(u_1,v_1,u_2,v_2)\in I_0^4} |\dot{\gamma}_h((u_1,v_1),(u_2,v_2))| \leq cr^h$.

A3 (Random Design under Proper Subsets) Given a time point t, U_{it} is a continuous random variable, i.i.d. as $U_t \sim f_{U_t}$, where $f_{U_t}(u)$ is a random probability density function (pdf) defined as

$$f_{U_t}(u) := f_{U_0}(u)\Big(\mathbb{1}(u \in \mathcal{I}_t) / \textstyle\int_{\mathcal{I}_t} f_{U_0}(v)dv\Big), \tag{21}$$

where $\mathbb{1}(.)$ denotes the indicator function, $\mathcal{I}_t$ denotes a proper (random) subset as in Assumption RS, and the pdf $f_{U_0}(u) > 0$ for all $u \in I_0 = [a,b] \subset \mathbb{R}$ and zero else.

For estimating μ (i.e., pooled data): Unconditionally on time points t, the data (Y_{it}, U_{it}) is i.i.d. as $(Y,U) \sim g_{YU}$, where $g_{YU}(y,u) = f_{Y|U}(y|u)f_U(u)$, and where $f_U(u) = \mathbb{E}(f_{U_t}(u))$. It is assumed $f_U(u) > 0$ for all $u \in \mathcal{I}_0$ and zero else, that f'_U is continuous, and that there exist

(generic) constants $c > 0$ such that $\sup_{u\in I_0} |f''_U(u)| < c < \infty$ and $\sup_{(u,y)\in I_0\times\mathbb{R}} |(\partial^2/\partial u^2) g_{YU}(y,u)| \leq c < \infty$.
For estimating γ (i.e., pooled data): Furthermore, $(Y_{it}, Y_{is}, U_{it}, U_{is})$, $t \neq s$, is i.i.d. as $(Y_1, Y_2, U_1, U_2) \sim g_{YYUU}$ and there exist a (generic) constant $c > 0$ such that
$\sup_{(y_1,y_2,u_1,u_2)\in\mathbb{R}^2\times I_0^2} |(\partial^4/(\partial u_1^2 \partial u_2^2)) g_{YYUU}(y_1,y_2,u_1,u_2)| \leq c < \infty$.

A4 (Asymptotic Scenario) $Tn \to \infty$, where $n = n(T) \geq 2$ such that $n(T) \sim T^\theta$ with $0 \leq \theta < \infty$. Here "$a_T \sim b_T$" denotes that two sequences a_T and b_T are asymptotically equivalent up to some positive constant $0 < c < \infty$, i.e., that $\lim_{T\to\infty}(a_T/b_T) = c$.

A5 (Smoothness) For estimating μ: The functions $\mu''(u)$, $\partial\gamma(u,v)/\partial u$, and $\partial\gamma(u,v)/\partial v$ are continuous for all $u, v \in I_0$.
For estimating γ: All second order derivatives of $\gamma(u,v)$ and all first order derivatives of $\dot{\gamma}((u_1,v_1),(u_2,v_2))$ are continuous for all points within its supports I_0^2 and I_0^4.

A6 (Bandwidths) For estimating μ: $h_\mu = h_{Tn,\mu} \to 0$ and $(Tn\, h_\mu)^{-1/2} \to \infty$ as $Tn \to \infty$. For estimating γ: $h_\gamma = h_{TN,\gamma} \to 0$ and $(TN\, h_\gamma)^{-1/2} \to \infty$ as $TN \to \infty$, where $N = n^2 - n$.

A7 (Kernel Function) The kernel function κ is assumed to be a univariate, symmetric, pdf with compact support $\operatorname{supp}(\kappa) = [-1, 1]$, such as, e.g., the univariate Epanechnikov kernel.

A8 (Eigenvalues) For $k \in \{1, 2, \dots\}$, let $\lambda_k = \mathcal{O}(k^{-a})$, with $a > 1$, and $\lambda_k - \lambda_{k+1} \geq$ const. × k^{-a-1}. Correspondingly, for λ_k^{S} and λ_k^{L}, but possibly with different constants and parameters a.

A9 (Additional Regularity Assumption on X_t)

$$\sup_{u\in I_0}\{\mathbb{E}(|X_t(u)|^C)\} < \infty \quad \text{for all} \quad C > 0,$$

$$\sup_{u,v\in I_0}(\mathbb{E}[\{|u-v|^{-\varepsilon}|X_t(u) - X_t(v)|\}^C]) < \infty \quad \text{for some} \quad \varepsilon > 0 \quad \text{and}$$

$$\sup_{k\geq 1} \lambda_k^{-r}\{\mathbb{E}[(\int_a^b (X_t(u) - \mathbb{E}(X_t(u)))\phi_k(u)\,du)^{2r}]\} < \infty$$

for each $r = 1, 2, \dots$.

APPENDIX B: PROOFS

Proof of Theorem 2.1: Note that for any K and every $u \in I_{\mathrm{L}}$, and particularly for every $u \in I_{\mathrm{L}} \setminus I_{\mathrm{S}}$, we have

$$0 \leq \mathbb{E}\left(\left(X_t^{\mathrm{L}}(u) - \sum_{k=1}^K \xi_{tk}^{\mathrm{S}} \tilde{\phi}_k^{\mathrm{L}}(u)\right)^2\right) = \gamma^{\mathrm{L}}(u,u) - \sum_{k=1}^K \lambda_k^{\mathrm{S}} \tilde{\phi}_k^{\mathrm{L}}(u)^2, \tag{22}$$

which implies that $\mathbb{V}(\sum_{k=1}^{K}\xi_{tk}^{\mathrm{S}}\tilde{\phi}_k^{\mathrm{L}}(u))=\sum_{k=1}^{K}\lambda_k^{\mathrm{S}}\tilde{\phi}_k^{\mathrm{L}}(u)^2$ converges to a fixed limit $0\leq\mathbb{V}(\tilde{X}_t^{\mathrm{L}}(u))<\infty$ as $K\to\infty$ for all $u\in I_{\mathrm{L}}$. Moreover, continuity of $\gamma^{\mathrm{L}}(u,v)$ implies continuity of $\mathbb{V}(\tilde{X}_t^{\mathrm{L}}(u))$.

Proof of Theorem 2.2, part (a): For all $v\in I_{\mathrm{S}}$ and $u\in I_{\mathrm{L}}\setminus I_{\mathrm{S}}$ we have that

$$\begin{aligned}\mathbb{E}\left(X_t^{\mathrm{S}}(v)Z_t(u)\right)&=\mathbb{E}\left(X_t^{\mathrm{S}}(v)\left(X_t^{\mathrm{L}}(u)-\tilde{X}_t^{\mathrm{L}}(u)\right)\right)=\\&=\mathbb{E}\left(\sum_{k=1}^{\infty}\xi_{tk}^{\mathrm{S}}\phi_k^{\mathrm{S}}(v)\Big(X_t^{\mathrm{L}}(u)-\sum_{k=1}^{\infty}\xi_{tk}^{\mathrm{S}}\tilde{\phi}_k^{\mathrm{L}}(u)\Big)\right)=\\&=\sum_{k=1}^{\infty}\phi_k^{\mathrm{S}}(v)\Big(\,\mathbb{E}(\xi_{tk}^{\mathrm{S}}X_t^{\mathrm{L}}(u))-\lambda_r\tilde{\phi}_k^{\mathrm{L}}(u)\Big)=0,\end{aligned}$$

since $\mathbb{E}(\xi_{tk}^{\mathrm{S}}X_t^{\mathrm{L}}(u))=\lambda_k^{\mathrm{S}}\,\tilde{\phi}_k^{\mathrm{L}}(u)$. This proves Eq. (5), while Eq. (6) directly follows from the definition of $Z_t(x)$.

Part (b): Note that the Riesz representation theorem implies that

$$\ell_u(X_t^{\mathrm{S}})=\int_{I_{\mathrm{S}}}b(v)X_t^{\mathrm{S}}(v)dv$$

for some $b\in L^2(I_{\mathrm{S}})$. By Eq. (5) and the orthogonality property of the least squares projection we thus obtain

$$\begin{aligned}&\mathbb{E}\left(\left(X_t^{\mathrm{L}}(u)-\ell_u(X_t^{\mathrm{S}})\right)^2\right)=\\&=E\left(\left(\tilde{X}_t^{\mathrm{L}}(u)+Z_t(u)-\int_{I_{\mathrm{S}}}b(v)X_t^{\mathrm{S}}(v)dv\right)^2\right)=\\&=\mathbb{E}\left(\left(\tilde{X}_t^{\mathrm{L}}(u)-\int_{I_{\mathrm{S}}}b(v)X_t^{\mathrm{S}}(v)dv\right)^2\right)+\mathbb{E}(Z_t(u)^2)+\\&+2\left(\mathbb{E}(\tilde{X}_t^{\mathrm{L}}(u)\,Z_t(u))-\int_{I_{\mathrm{S}}}b(v)\,\mathbb{E}\left(X_t^{\mathrm{S}}(v)Z_t(u)\right)dv\right)=\\&=\mathbb{E}\left(\left(\tilde{X}_t^{\mathrm{L}}(u)-\int_{I_{\mathrm{S}}}b(v)X_t^{\mathrm{S}}(v)dv\right)^2\right)+\mathbb{E}(Z_t^2(u))\geq\mathbb{E}(Z_t^2(u)).\end{aligned}$$

Part (c): From basic statistics we know that $\mathbb{V}(Z_t(u)-Z_s(u))=\mathbb{V}(Z_t(u))+\mathbb{V}(Z_s(u))-2\,Cov(Z_t(u),Z_s(u))$ for all $u\in I_{\mathrm{L}}$. Under our stationarity and weak dependency assumptions (Assumption A2) we have that $\mathbb{V}(Z_t(u)-Z_s(u))=2\ \mathbb{V}(Z_t(u))+\mathcal{O}(r^{|t-s|})$, where we use the typical α-mixing covariance inequality; see, e.g., in Fan and Yao (2003) Ch. 2.6.2. Rearranging and using that $\mathbb{E}(Z_t(u)) = \mathbb{E}(Z_s(u)) = 0$ for all $u \in I_{\mathrm{L}}$ and all $t, s \in \{1, \dots, T\}$ yields $\mathbb{V}(Z_t(u)) = \frac{1}{2}\,\mathbb{E}((Z_t(u) - Z_s(u))^2) + \mathcal{O}(r^{|t-s|})$. From result (a) we know that $Z_t(u)$ and $X_t^{\mathrm{S}}(v)$ are orthogonal and therefore uncorrelated for all $u \in I_{\mathrm{L}} \setminus I_0$ and all $v \in I_{\mathrm{S}}$, i.e., $\mathbb{E}\left(X_t^{\mathrm{S}}(v) Z_t(u)\right) = \mathrm{Cov}\left(X_t^{\mathrm{S}}(v), Z_t(u)\right) = 0$. Under the assumption of a Gaussian time series process $(X_t)_{t\in\mathbb{Z}}$, we have then independency between $Z_t(u)$ and $X_t^{\mathrm{S}}(v)$, such that

$$\mathbb{V}(Z_t(u)) = \frac{1}{2}\,\mathbb{E}\left(\mathbb{E}\left((Z_t(u) - Z_s(u))^2\right)|X_t^{\mathrm{S}} = X_s^{\mathrm{S}}\right) + \mathcal{O}(r^{|t-s|}),$$

where $X_t^{\mathrm{S}} = X_s^{\mathrm{S}}$ means that $X_t^{\mathrm{S}}(u) = X_s^{\mathrm{S}}(u)$ for all $u \in I_{\mathrm{S}}$.

□

Proof of Theorem 5.1, part (a): Note that the estimator $\hat{\mu}(u; h_\mu)$ can be written as

$$\hat{\mu}(u; h_\mu) = e_1^\top S_{1,Tn,u}^{-1} S_{2,Tn,u}, \tag{23}$$

with 2×2 matrix

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

and 2×1 vector

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

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

$$\begin{aligned} S_{1,Tn,u} &= \begin{pmatrix} \Psi_{0,Tn}(u;h_\mu) & \Psi_{1,Tn}(u;h_\mu) \\ \Psi_{1,Tn}(u;h_\mu) & \Psi_{2,Tn}(u;h_\mu) \end{pmatrix} \\ &= \begin{pmatrix} f_U(u) & 0 \\ 0 & f_U(u)\nu_2(\kappa) \end{pmatrix} + \mathcal{O}_p^{\mathrm{Unif}}\left(h_\mu^2 + \frac{1}{\sqrt{Tn\,h_\mu}} + \frac{1}{\sqrt{T}}\right) \quad \text{and} \end{aligned} \tag{24}$$

$$S_{2,Tn,u} = \begin{pmatrix} \Psi_{3,Tn}(u;h_\mu) \\ \Psi_{4,Tn}(u;h_\mu) \end{pmatrix} = \begin{pmatrix} \mu(u) f_U(u) \\ 0 \end{pmatrix} + \mathcal{O}_p^{\mathrm{Unif}}\left(h_\mu^2 + \frac{1}{\sqrt{Tn\,h_\mu}} + \frac{1}{\sqrt{T}}\right), \tag{25}$$

where we write $\Psi_{q,Tn}(u;h_\mu)-m_q(u)=\mathcal{O}_p^{\mathrm{Unif}}(\texttt{rate}_n)$ in order to denote that $\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-m_q(u)|=\mathcal{O}_p(\texttt{rate}_n)$. Taking the inverse of (24) gives

$$S_{Tn,u}^{-1}=\begin{pmatrix}1/f_U(u) & 0\\ 0 & 1/(f_U(u)\nu_2(\kappa))\end{pmatrix}+\mathcal{O}_p^{\mathrm{Unif}}\left(h_\mu^2+\frac{1}{\sqrt{Tn\,h_\mu}}+\frac{1}{\sqrt{T}}\right). \tag{26}$$

Plugging (26) and (25) into (23) leads to

$$\sup_{u\in I_0}|\hat\mu(u;h_\mu)-\mu(u)|=\mathcal{O}_p\left(h_\mu^2+\frac{1}{\sqrt{Tn\,h_\mu}}+\frac{1}{\sqrt{T}}\right).$$

□

Proof of Theorem 5.1, part (b): Let us initially consider the infeasible estimator $\hat\gamma_C$ that is based on the infeasible "clean" dependent variables $C_{ijt}=(Y_{it}-\mu(U_{it}))(Y_{jt}-\mu(U_{jt}))$ instead of the estimator $\hat\gamma$ in (14) that is based on the "dirty" dependent variables $\hat C_{ijt}=(Y_{it}-\hat\mu(U_{it}))(Y_{jt}-\hat\mu(U_{jt}))$, which are contaminated through having to estimate the unknown mean function μ. Equivalently to the estimator $\hat\mu$ above, we can write the estimator $\hat\gamma_C$ as

$$\hat\gamma_C(u,v;h_\gamma)=e_1^\top\tilde S_{1,TN,(u,v)}^{-1}\tilde S_{2,TN,(u,v)}, \tag{27}$$

with

$$\tilde S_{1,TN,(u,v)}^{-1}=\begin{pmatrix}\Theta_{0,TN}(u,v;h_\gamma) & \Theta_{1,TN}(u,v;h_\gamma)\\ \Theta_{1,TN}(u,v;h_\gamma) & \Theta_{2,TN}(u,v;h_\gamma)\end{pmatrix}^{-1}$$

$$=\begin{pmatrix}1/f_{UU}(u,v) & 0\\ 0 & 1/f_{UU}(u,v)(\nu_2(\kappa))^2\end{pmatrix}+\mathcal{O}_p^{\mathrm{Unif}}\left(h_\gamma^2+\frac{1}{\sqrt{TN\,h_\gamma^2}}+\frac{1}{\sqrt{T}}\right) \tag{28}$$

and $\tilde S_{2,TN,(u,v)}=$

$$=\begin{pmatrix}\Theta_{3,TN}(u,v;h_\gamma)\\ \Theta_{4,TN}(u,v;h_\gamma)\end{pmatrix}=\begin{pmatrix}\gamma(u,v)f_{UU}(u,v)\\ 0\end{pmatrix}+\mathcal{O}_p^{\mathrm{Unif}}\left(h_\gamma^2+\frac{1}{\sqrt{TN\,h_\gamma^2}}+\frac{1}{\sqrt{T}}\right), \tag{29}$$

where we use the notation and the results from Lemma B.2, and where we write $\Theta_{q,TN}(u,v;h_\gamma)-\eta_q(u,v)=\mathcal{O}_p^{\mathrm{Unif}}(\texttt{rate}_n)$ in order to denote that $\sup_{(u,v)\in I_0^2}|\Theta_{q,TN}(u,v;h_\gamma)-\eta_q(u,v)|=\mathcal{O}_p(\texttt{rate}_n)$.

Plugging (28) and (29) into (27) leads to

$$\sup_{(u,v)\in I_0^2} |\hat{\gamma}_C(u,v;h_\mu) - \gamma(u,v)| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\,h_\gamma^2}} + \frac{1}{\sqrt{T}}\right). \tag{30}$$

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

$$\begin{aligned}\hat{C}_{ijt} = C_{ijt} &+ (Y_{it} - \mu(U_{it}))(\mu(U_{jt}) - \hat{\mu}(U_{jt}))\\ &+ (Y_{jt} - \mu(U_{jt}))(\mu(U_{it}) - \hat{\mu}(U_{it}))\\ &+ (\mu(U_{it}) - \hat{\mu}(U_{it}))(\mu(U_{jt}) - \hat{\mu}(U_{jt})).\end{aligned}$$

Using our finite moment assumptions on Y_{it} (Assumption A1) and our result in Theorem 5.1, part (a), we have that

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

for all $i \neq j \in \{1,\dots,n\}$ and $t \in \{1,\dots,T\}$.

Case of non-parametric rates: As long as n does not diverge too fast with T, i.e., as long as θ in $n \sim T^\theta$, $0 < \theta < 1$, is sufficiently small, the non-parametric rate components, i.e., h_μ^2 and $(Tn\,h_\mu)^{-1/2}$ are of a larger order than the parametric $T^{-1/2}$ component. For the sake of a simple argument, let us focus on the case of optimal bandwidth choices, i.e., $h_\mu \sim (Tn)^{-1/5}$ and $h_\gamma \sim (TN)^{-1/6}$. Then the non-parametric rate components are dominating if $0 < \theta < 1/4$ and we have that $\hat{C}_{ijt} = C_{ijt} + \mathcal{O}_p\left(T^{(-2/5)(1+\theta)}\right)$ and $\sup_{(u,v)\in I_0^2} |\hat{\gamma}_C(u,v;h_\mu) - \gamma(u,v)| = \mathcal{O}_p\left(T^{(-1/3)(1+2\theta)}\right)$. But $(-2/5)(1+\theta) < (-1/3)(1+2\theta) \Leftrightarrow \theta < 1/4$ which implies that the difference between C_{ijt} and $\hat{C}_{ijt}$ is of an order of magnitude smaller than the approximation error between $\hat{\gamma}_C$ and γ. It follows then form standard arguments that $\hat{\gamma}(u,v;h_\mu)$ converges at the same rate as $\hat{\gamma}_C(u,v;h_\mu)$, i.e.,

$$\sup_{(u,v)\in I_0^2} |\hat{\gamma}(u,v;h_\mu) - \gamma(u,v)| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\,h_\gamma^2}} + \frac{1}{\sqrt{T}}\right).$$

Similar but more lengthy arguments apply for the case of non-optimal bandwidth choices and for the case of the parametric $T^{-1/2}$ rate. □

LEMMA B.1. *Define*

$$\Psi_{q,Tn}(u;h_\mu) = \frac{1}{Tnh_\mu}\sum_{it}\kappa\left(\frac{U_{it}-u}{h_\mu}\right)\psi_q\left(U_{it}-u,Y_{it}\right), \tag{31}$$

where

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

Then, under Assumptions A1-A7,

$$\tau_{q,Tn} = \sup_{u\in I_0}\left|\Psi_{q,Tn}(u;h_\mu)-m_q(u)\right| = \mathcal{O}_p\left(h_\mu^2+\frac{1}{\sqrt{Tn\,h_\mu}}+\frac{1}{\sqrt{T}}\right),$$

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

LEMMA B.2. *Define*

$$\Theta_{q,TN}(u,v;h_\gamma) = \frac{1}{TNh_\gamma}\sum_{i\neq j,t}\kappa\left(\frac{U_{it}-u}{h_\gamma}\right)\kappa\left(\frac{U_{jt}-v}{h_\gamma}\right)\vartheta_q\left(U_{it}-u,U_{jt}-u,C_{ijt}\right), \tag{32}$$

where

$$\vartheta_q\left(U_{it}-u,U_{jt}-v,C_{ijt}\right) = \begin{cases} (U_{it}-u)^q\,(U_{jt}-v)^q & \textit{for } q\in\{0,1,2\}\\ C_{ijt} & \textit{for } q=3\\ (U_{it}-u)\,(U_{jt}-v)\,C_{ijt} & \textit{for } q=4. \end{cases}$$

Then, under Assumptions A1-A7,

$$\varrho_{q,Tn} = \sup_{(u,v)\in I_0^2}\left|\Theta_{q,TN}(u,v;h_\gamma)-\eta_q(u,v)\right| = \mathcal{O}_p\left(h_\gamma^2+\frac{1}{\sqrt{TN\,h_\gamma^2}}+\frac{1}{\sqrt{T}}\right),$$

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

Proof of Lemma B.1: Remember that $\mathbb{E}(|X_n|) = \mathcal{O}(\mathtt{rate}_n)$ implies that $X_n = \mathcal{O}_p(\mathtt{rate}_n)$, therefore, we focus in the following on "$\mathbb{E}(|X_n|)$". Adding a zero and applying the triangle inequality yields that $\mathbb{E}(\tau_{q,Tn}) =$

$$\begin{aligned}\mathbb{E}(\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-m_q(u)|) \leq \sup_{u\in I_0}|\,\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))-m_p(u)|+\\ +\,\mathbb{E}(\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))|).\end{aligned} \tag{33}$$

Let us first focus on the second summand in (33). The next steps will make use of the Fourier transformation of the kernel function κ see, e.g., (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 $\mathrm{i} = \sqrt{-1}$. Remember that, by Assumption A7, $\kappa(.)$ 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, Eq. (1.34)) $\mathcal{F}[\kappa(./h_\mu)/h_\mu](x) = \mathcal{F}[\kappa](xh_\mu) = \kappa^{\text{ft}}(xh_\mu)$ which yields

$$\begin{aligned}\kappa\left(s/h_\mu\right)/h_\mu &= \frac{1}{2\pi}\int_{\mathbb{R}}\mathcal{F}[\kappa(./h_\mu)/h_\mu](x)\exp\left(\mathrm{i}xs\right)dx\ \mathbb{1}_{(|s|<h_\mu)}\\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\kappa^{\text{ft}}(xh_\mu)\exp\left(\mathrm{i}xs\right)dx\ \mathbb{1}_{(|s|<h_\mu)}.\end{aligned} \tag{34}$$

Plugging (34) into (31) yields $\Psi_{q,Tn}(u;h_\mu) =$

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

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

$$\mathbb{E}(\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))|) \leq \frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in I_0}\left|\int_{\mathbb{R}}\tilde{\omega}_{q,Tn}(u,x)\cdot\kappa^{\text{ft}}(xh_\mu)dx\right|\right),$$

where

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

Using further that κ^{ft} is symmetric, since κ is symmetric by Assumption A7, and that $\exp\big(\mathrm{i}xU_{it}\big)=\cos\big(xU_{it}\big)+\mathrm{i}\sin\big(xU_{it}\big)$ leads to

$$\frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in I_0}\left|\int_{\mathbb{R}}\tilde{\omega}_{q,Tn}(u,x)\cdot\kappa^{\mathrm{ft}}(xh_\mu)dx\right|\right)=\frac{1}{2\pi}\,\mathbb{E}\left(\sup_{u\in I_0}\left|\int_{\mathbb{R}}\omega_{q,Tn}(u,x)\cdot\kappa^{\mathrm{ft}}(xh_\mu)dx\right|\right),$$

where

$$\begin{aligned}\omega_{q,Tn}(u,x) =&\frac{1}{Tn}\sum_{it}\big[\cos\big(xU_{it}\big)\psi_q\left(U_{it}-u,Y_{it}\right)\, \mathbb{1}_{(|U_{it}-u|<h_\mu)}-\\ &\qquad\qquad \mathbb{E}\left(\cos\big(xU_{it}\big)\psi_q\left(U_{it}-u,Y_{it}\right)\mathbb{1}_{(|U_{it}-u|<h_\mu)}\right)\big],\end{aligned}\tag{35}$$

such that

$$\begin{aligned}&\mathbb{E}(\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))|)\\ &\leq\frac{1}{2\pi}\int_{\mathbb{R}}\mathbb{E}\left(\sup_{u\in I_0}\big|\omega_{q,Tn}(u,x)\big|\right)\cdot\Big|\kappa^{\mathrm{ft}}(xh_\mu)\Big|\,dx\\ &\leq\frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\left(\sup_{u\in I_0}\big|\omega_{q,Tn}(u,x)\big|\right)^2\right)}\cdot\Big|\kappa^{\mathrm{ft}}(xh_\mu)\Big|\,dx\\ &=\frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\sup_{u\in I_0}\,\big(\omega_{q,Tn}(u,x)\big)^2\right)}\cdot\Big|\kappa^{\mathrm{ft}}(xh_\mu)\Big|\,dx.\end{aligned}\tag{36}$$

In order to simplify the notation we will denote

$$W^q_{it}(x,u)=\cos\big(xU_{it}\big)\psi_q\left(U_{it}-u,Y_{it}\right),$$

such that $\mathbb{E}\Big(\sup_{u\in I_0}\,\left(\omega_{q,Tn}(u,x)\right)^2\Big)=$

$$\begin{aligned}\mathbb{E}\Big(\sup_{u\in I_0}\,\Big(&\frac{1}{(Tn)^2}\sum_{it}\left[W^q_{it}(x,u)\mathbb{1}_{(|U_{it}-u|<h_\mu)}-\mathbb{E}(W^q_{it}(x,u)\mathbb{1}_{(|U_{it}-u|<h_\mu)})\right]^2+\\ &\frac{1}{(Tn)^2}\sum_{(i,t)\neq(j,s)}\left[(W^q_{it}(x,u)\mathbb{1}_{(|U_{it}-u|<h_\mu)}-\mathbb{E}(W^q_{it}(x,u)\mathbb{1}_{(|U_{it}-u|<h_\mu)}))\cdot\right.\\ &\qquad\left.\cdot(W^q_{js}(x,u)\mathbb{1}_{(|U_{js}-u|<h_\mu)}-\mathbb{E}(W^q_{js}(x,u)\mathbb{1}_{(|U_{js}-u|<h_\mu)}))\right]\Big)\Big).\end{aligned}$$

As u takes only values within the compact interval $[a,b]$, there exist constants C_1 and C_2 such that, uniformly for all $u \in [a,b]$, $\mathbb{P}(|U_{it}-u| < h_\mu) \leq C_1 h_\mu < \infty$, for all i,t, and $\mathbb{P}(|U_{it}-u| < h_\mu$ AND $|U_{js}-u| < h_\mu) \leq C_2 h_\mu^2 < \infty$, for all $(i,t) \neq (j,s)$. Together with the triangle inequality, this yields that $\mathbb{E}\left(\sup_{u\in I_0}\,(\omega_{q,Tn}(u,x))^2\right) \leq$

$$\frac{C_1 h_\mu}{(Tn)^2}\sum_{it}\mathbb{E}\left(\sup_{u\in I_0}\,[W_{it}^q(x,u)-\mathbb{E}(W_{it}^q(x,u))]^2\right)+$$
$$\frac{C_2 h_\mu^2}{(Tn)^2}\sum_{(i,t)\neq(j,s)}\mathbb{E}\left(\sup_{u\in I_0}\,\left[(W_{it}^q(x,u)-\mathbb{E}(W_{it}^q(x,u)))(W_{js}^q(x,u)-\mathbb{E}(W_{js}^q(x,u)))\right]\right).$$

From our moment assumptions (Assumption A1) and the fact that $I_0 = [a,b]$ is compact we can conclude that there must exist a constant C_3 such that, point-wise for every $x \in \mathbb{R}$,

$$\mathbb{E}\left(\big(\sup_{u\in I_0}\left|W_{it}^q(x,u)-\mathbb{E}(W_{it}^q(x,u))\right|\big)^2\right) \leq C_3 < \infty \tag{37}$$

for all i,t.

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

$$\mathbb{E}\left(\sup_{u\in I_0}\left|W_{it}^q(x,u)-\mathbb{E}(W_{it}^q(x,u))\right|\cdot\sup_{u\in I_0}\left|W_{jt}^q(x,u)-\mathbb{E}(W_{jt}^q(x,u))\right|\right)$$
$$\leq C_4 < \infty \tag{38}$$

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

Between function dependencies: Our weak dependency assumption (Assumption A2) and the fact that I_0 is compact yields that point-wise for every $x \in \mathbb{R}$

$$\mathbb{E}\left(\sup_{u\in I_0}\left|W_{it}^q(x,u)-\mathbb{E}(W_{it}^q(x,u))\right|\cdot\sup_{u\in I_0}\left|W_{js}^q(x,u)-\mathbb{E}(W_{js}^q(x,u))\right|\right)$$
$$\leq c r^{|t-s|} \tag{39}$$

for all i,j and $|t-s| \geq 1$, where $0 < c < \infty$ and $0 < r < 1$.

Eq.s (37), (38), and (39) yield that $\mathbb{E}\left(\sup_{u\in I_0}\,(\omega_{q,Tn}(u,x))^2\right) \leq$

$$\begin{aligned}&\leq \frac{C_1 h_\mu}{(Tn)^2}\sum_{it} C_3 + \frac{C_2 h_\mu^2}{(Tn)^2}\sum_{i\neq j,t} C_4 + \frac{C_2 h_\mu^2}{(Tn)^2}\sum_{i,j,t\neq s} c r^{|t-s|}\\ &= \mathcal{O}\left(\frac{h_\mu}{Tn}+\frac{h_\mu^2(n-1)}{Tn}+\frac{h_\mu^2}{T}\right) = \mathcal{O}\left(\frac{h_\mu}{Tn}+\frac{h_\mu^2}{T}\right),\end{aligned}$$

such that

$$\sqrt{\mathbb{E}\left(\sup_{u\in I_0}\left(\omega_{q,Tn}(u,x)\right)^2\right)}=\mathcal{O}\left(\sqrt{\frac{h_\mu}{Tn}}+\frac{h_\mu}{\sqrt{T}}\right). \tag{40}$$

Plugging (40) into (36) and integration by substitution leads to

$$\begin{aligned}&\mathbb{E}(\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))|)=\\ &\leq\frac{1}{2\pi}\int_{\mathbb{R}}\sqrt{\mathbb{E}\left(\sup_{u\in I_0}\left(\omega_{q,Tn}(u,x)\right)^2\right)}\cdot\left|\kappa^{\mathrm{ft}}(xh_\mu)\right|dx=\mathcal{O}\left(\frac{1}{\sqrt{Tn\,h_\mu}}+\frac{1}{\sqrt{T}}\right).\end{aligned}\tag{41}$$

Let us now focus on the first summand in (33). From standard arguments in nonparametric statistics we know that

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

for each $u\in I_0$ and for all $q\in\{0,\dots,4\}$. Under Assumption A3, the "$\mathcal{O}(h_\mu^2)$" terms become uniformly valid for all $u\in I_0$, since for $q\in\{0,1,2,4\}$, we have then $\sup_{\in I_0}|f''_U(u)|\leq c<\infty$ and additionally for $q\in\{3,4\}$, we have then $\sup_{(u,y)\in I_0\times\mathbb{R}}|(\partial^2/\partial u^2)g_{YU}(y,u)|\leq c<\infty$, where $c>0$ are generic constants (see Assumption A3). We can conclude with respect to the first summand in (33) that

$$\sup_{u\in I_0}|\mathbb{E}(\Psi_{q,Tn}(u;h_\mu))-m_p(u)|=\mathcal{O}(h_\mu^2)\quad\text{for all}\quad q\in\{0,\dots,4\}. \tag{42}$$

Finally, plugging our results (41) and (42) into (33) leads to

$$\tau_{q,Tn}=\sup_{u\in I_0}|\Psi_{q,Tn}(u;h_\mu)-m_q(u)|=\mathcal{O}_p\left(h_\mu^2+\frac{1}{\sqrt{Tn\,h_\mu}}+\frac{1}{\sqrt{T}}\right) \tag{43}$$

for all $q\in\{0,\dots,4\}$. □

The proof of Lemma B.2: Analogously to that of B.1.

The proof of Theorem 5.1, parts (c) and (d) Under the additional regularity Assumption A9 it follows from the eigenvalue and eigenfunction expansions in Hall and Hosseini-Nasab (2006), that with probability 1,

$$\sup_{k\geq 1}|\hat{\lambda}_k-\lambda_k|\leq\hat{\Delta}_{TN}\quad\text{and}$$

$$\sup_{u\in I_0}|\hat{\phi}_k(u)-\phi_k(u)|\leq\sqrt{8}\frac{\hat{\Delta}_{TN}}{\delta_k}\quad\text{for all}\quad 1\leq k\leq\bar{K}_{TN}-1,$$

where $\hat{\Delta}_{TN} = \sqrt{\int_{(u,v)\in I_0^2}(\hat{\gamma}(u,v)-\gamma(u,v))^2 d(u,v)}$, $\delta_k = \min_{1\leq i\leq k}\{\lambda_i - \lambda_{i+1}\}$, and $\bar{K}_{TN} = \inf\{k \geq 1 : \lambda_k - \lambda_{k+1} \leq 2\hat{\Delta}_{TN}\}$. See also Theorem 2 in the in Appendix A.1. of Hall and Hosseini-Nasab (2006) and its discussion.

From part (b) of our Theorem 5.1, it follows that $\hat{\Delta}_{TN} = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\,h_\gamma^2}} + \frac{1}{\sqrt{T}}\right)$ which leads to

$$\sup_{k\geq 1}|\hat{\lambda}_k - \lambda_k| = \mathcal{O}_p\left(h_\gamma^2 + \frac{1}{\sqrt{TN\,h_\gamma^2}} + \frac{1}{\sqrt{T}}\right) \quad \text{and} \tag{44}$$

$$\sup_{u\in I_0}|\hat{\phi}_k(u) - \phi_k(u)| = \mathcal{O}_p\left(\delta_k^{-1}\left(h_\gamma^2 + \frac{1}{\sqrt{TN\,h_\gamma^2}} + \frac{1}{\sqrt{T}}\right)\right), \tag{45}$$

where (45) holds for all $1 \leq k \leq \bar{K}_{TN} - 1$. This concludes our proof of Theorem 5.1, parts (c) and (d).

The proof of Theorem 5.2.
Estimating the pc-scores:

$$\hat{\xi}_{kt}^{\mathrm{S}} = \sum_{i=1}^{n} \hat{\phi}_k^{\mathrm{S}}(U_{it}^{\mathrm{S}})(Y_{it}^{\mathrm{S}} - \hat{\mu}(U_{it}^{\mathrm{S}}; h_\mu))(U_{it}^{\mathrm{S}} - U_{i-1,t}^{\mathrm{S}}), \text{ with } U_{0,t}^{\mathrm{S}} = A_t.$$

Using that $Y_{it}^{\mathrm{S}} = X_t^{\mathrm{S}}(U_{it}^{\mathrm{S}}) + \varepsilon_{it}$, a weak law of large numbers, and results (a) and (d) of Theorem 5.1 lead to

$$\begin{aligned}\hat{\xi}_{kt}^{\mathrm{S}} &= \int_{A_t}^{B_t} \hat{\phi}_k^{\mathrm{S}}(u)(X_t^{\mathrm{S}}(u) - \hat{\mu}(u; h_\mu))\,du + \mathcal{O}_p\left(n^{-1/2}\right)\\ &= \xi_{kt}^{\mathrm{S}} + \mathcal{O}_p\left(n^{-1/2} + r_{Tn}^{\mu} + \frac{r_{TN}^{\gamma}}{\delta_k}\right),\end{aligned} \tag{46}$$

where $r_{Tn}^{\mu} = h_\mu^2 + 1/\sqrt{Tn\,h_\mu} + 1/\sqrt{T}$ and $r_{TN}^{\gamma} = h_\gamma^2 + 1/\sqrt{TN\,h_\gamma^2} + 1/\sqrt{T}$.

Estimating the functional prediction model: Let us denote

$$\tilde{X}_{t,K}^{\mathrm{L}}(u) = \mu(u) + \sum_{k=1}^{K} \frac{\xi_{tk}^{\mathrm{S}}}{\lambda_k^{\mathrm{S}}} \int_{I_{\mathrm{S}}} \phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{L}}(u,v)dv \quad \text{and}$$

$$\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u) = \sum_{k=K+1}^{\infty} \frac{\xi_{tk}^{\mathrm{S}}}{\lambda_k^{\mathrm{S}}} \int_{I_{\mathrm{S}}} \phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{L}}(u,v)dv$$

such that $\tilde{X}_t^{\mathrm{L}}(u) = \tilde{X}_{t,K}^{\mathrm{L}}(u) + \tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)$. This allows us to decompose the prediction error into an estimation error part and an regularization error part:

(47)
$$\sup_{u\in I_0}\left|\tilde{X}_t^{\mathrm{L}}(u) - \hat{\tilde{X}}_{t,K}^{\mathrm{L}}(u)\right| \le \sup_{u\in I_0}\left|\tilde{X}_{t,K}^{\mathrm{L}}(u) - \hat{\tilde{X}}_{t,K}^{\mathrm{L}}(u)\right| + \sup_{u\in I_0}\left|\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)\right|.$$

Let us focus on the estimation error, i.e., the first term on the right hand side of (47). Using the results (a) and (d) of Theorem 5.1 and from (46)

$$\sup_{u\in I_0}\left|\tilde{X}_{t,K}^{\mathrm{L}}(u) - \hat{\tilde{X}}_{t,K}^{\mathrm{L}}(u)\right| = \mathcal{O}_p\left(K\left(n^{-1/2} + r_{Tn}^{\mu}\right) + r_{TN}^{\gamma}\sum_{k=1}^{K}\delta_k^{-1}\right). \tag{48}$$

In order to approximate the regularization error, i.e., the second term on the right hand side of (47), we use that, by the boundedness of γ and of the eigenfunctions, there exists a constant C, $0 < C < \infty$, such that

$$\sup_{u\in I_{\mathrm{L}}}\left|\int_{I_{\mathrm{S}}}\phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{L}}(u,v)dv\right| \le \lambda^{\mathrm{S}}\,C.$$

The later, the triangle inequality, and Jensen's inequality yield that

$$\begin{aligned}\mathbb{E}\left(\sup_{u\in I_0}\left|\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)\right|\right) &= \mathbb{E}\left(\sup_{u\in I_0}\left|\sum_{k=K+1}^{\infty}\frac{\xi_{tk}^{\mathrm{S}}}{\lambda_k^{\mathrm{S}}}\int_{I_{\mathrm{S}}}\phi_k^{\mathrm{S}}(v)\gamma^{\mathrm{L}}(u,v)dv\right|\right)\\ &\le C\sum_{k=K+1}^{\infty}\sqrt{\mathbb{E}\,|\xi_{tk}^{\mathrm{S}}|^2} = \mathcal{O}\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda_k^{\mathrm{S}}}\right).\end{aligned}$$

Therefore,

$$\sup_{u\in I_0}\left|\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)\right| = \mathcal{O}_p\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda_k^{\mathrm{S}}}\right). \tag{49}$$

Under our moment assumption (Assumption A1), we know that $\sum_{k=1}^{\infty}\sqrt{\lambda_k^{\mathrm{S}}} < \infty$, which implies that

$$\begin{aligned}&\sup_{u\in I_0}\left|\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)\right| = \mathcal{O}_p\left(1\right) \quad \text{for fixed } K \text{ and that}\\ &\sup_{u\in I_0}\left|\tilde{X}_{t,K_+}^{c,\mathrm{L}}(u)\right| = o_p\left(1\right) \quad \text{for } K\to\infty.\end{aligned}$$

From (48) and (49), we can conclude that

$$\sup_{u\in I_0}\left|\tilde{X}^{\mathrm{L}}_t(u)-\hat{\tilde{X}}^{\mathrm{L}}_{t,K}(u)\right| =$$

$$\mathcal{O}_p\left(K\,\left(n^{-1/2}+r^{\mu}_{Tn}\right)+r^{\gamma}_{TN}\sum_{k=1}^{K}\delta_k^{-1}\right)+\mathcal{O}_p\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda^{\mathrm{S}}_k}\right). \tag{50}$$

Case-by-case considerations under optimal bandwidth choices: Let us in the following focus on the case of optimal bandwidth choices, i.e., $h_\mu \sim (Tn)^{-1/5}$ and $h_\gamma \sim (TN)^{-1/6}$. Then we have that for $0<\theta<1/4$: $r^{\mu}_{Tn}\sim(Tn)^{-2/5}$ and $r^{\gamma}_{TN}\sim(TN)^{-1/3}$, which implies that $r^{\mu}_{Tn}=o\left(r^{\gamma}_{TN}\right)$, but also that $r^{\gamma}_{TN}=o\left(n^{-1/2}\right)$. These considerations yield to

$$\sup_{u\in I_0}\left|\tilde{X}^{\mathrm{L}}_t(u)-\hat{\tilde{X}}^{\mathrm{L}}_{t,K}(u)\right| =$$

$$=\mathcal{O}_p\left(K\,n^{-1/2}+(TN)^{-1/3}\sum_{k=1}^{K}\delta_k^{-1}\right)+\mathcal{O}_p\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda^{\mathrm{S}}_k}\right). \tag{51}$$

Under optimal bandwidth choices and for $1/4\le\theta<1$ we have that $r^{\mu}_{Tn}\sim r^{\gamma}_{TN}\sim T^{-1/2}$ and that $T^{-1/2}=o\left(n^{-1/2}\right)$. These considerations yield to

$$\sup_{u\in I_0}\left|\tilde{X}^{\mathrm{L}}_t(u)-\hat{\tilde{X}}^{\mathrm{L}}_{t,K}(u)\right| =$$

$$\mathcal{O}_p\left(K\,n^{-1/2}+T^{-1/2}\sum_{k=1}^{K}\delta_k^{-1}\right)+\mathcal{O}_p\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda^{\mathrm{S}}_k}\right). \tag{52}$$

Under optimal bandwidth choices and for $1\le\theta<\infty$ we have that $r^{\mu}_{Tn}\sim r^{\gamma}_{TN}\sim T^{-1/2}$ and that $n^{-1/2}=\mathcal{O}\left(T^{-1/2}\right)$. But note also that $KT^{-1/2}=\mathcal{O}(T^{-1/2}\sum_{k=1}^{K}\delta_k^{-1})$. These considerations yield to

$$\sup_{u\in I_0}\left|\tilde{X}^{\mathrm{L}}_t(u)-\hat{\tilde{X}}^{\mathrm{L}}_{t,K}(u)\right| =$$

$$\mathcal{O}_p\left(T^{-1/2}\sum_{k=1}^{K}\delta_k^{-1}\right)+\mathcal{O}_p\left(\sum_{k=K+1}^{\infty}\sqrt{\lambda^{\mathrm{S}}_k}\right). \tag{53}$$

Deriving explicit rates:
From Assumption A8 we have that $\lambda_k^{\mathrm{S}} = \mathcal{O}(k^{-a})$, with $a > 1$, and $\lambda_k^{\mathrm{S}} - \lambda_{k+1}^{\mathrm{S}} \geq \text{const.} \times k^{-a-1}$ or equivalently, $(\lambda_k^{\mathrm{S}} - \lambda_{k+1}^{\mathrm{S}})^{-1} \leq (\text{const.} \times k^{a+1})$. From the latter it follows that

$$(\lambda_k^{\mathrm{S}} - \lambda_{k+1}^{\mathrm{S}})^{-1} \leq (\text{const.} \times k^{a+1}) \leq (\text{const.} \times K^{a+1}), \text{ for } 1 \leq k \leq K.$$

Observe furthermore, that

$$\delta_k^{\mathrm{S}} = \min_{1\leq i\leq k}\{\lambda_i^{\mathrm{S}} - \lambda_{i+1}^{\mathrm{S}}\} \quad \Leftrightarrow \quad (\delta_k^{\mathrm{S}})^{-1} = \max_{1\leq i\leq k}\{(\lambda_i^{\mathrm{S}} - \lambda_{i+1}^{\mathrm{S}})^{-1}\}.$$

The latter two results yield that $(\delta_k^{\mathrm{S}})^{-1} \leq \text{const.} \times k^{a+1}$, but also that $(\delta_k^{\mathrm{S}})^{-1} \leq \text{const.} \times K^{a+1}$, for $1 \leq k \leq K$. The latter result implies that

$$\sum_{k=1}^{K}(\delta_k^{\mathrm{S}})^{-1} \leq K(\text{const.} \times K^{a+1}) = \text{const.} \times K^{a+2}. \tag{54}$$

The assumption $\lambda_k^{\mathrm{S}} = \mathcal{O}(k^{-a})$ leads to $\sum_{k=K+1}^{\infty}\sqrt{\lambda_k^{\mathrm{S}}} = \mathcal{O}(\sum_{k=K+1}^{\infty} k^{-a/2})$. Using this and the approximation in Eq. (54) we get from Eq.s (51)-(53) our results (a)-(c) of Theorem 5.2. □

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