Statistica Sinica **28** (2018), 2521-2539
doi:https://doi.org/10.5705/ss.202016.0536

# METHODOLOGY AND CONVERGENCE RATES FOR FUNCTIONAL TIME SERIES REGRESSION*

Tung Pham and Victor M. Panaretos

*University of Melbourne and Ecole Polytechnique Fédérale de Lausanne*

*Dedication:* Peter Hall was a mentor to the first author and role model for both authors, and this paper is dedicated to his memory.

*Abstract:* The functional linear model extends the notion of linear regression to the case where the response and covariates are iid elements of an infinite-dimensional Hilbert space. The unknown to be estimated is a Hilbert-Schmidt operator, whose inverse is by definition unbounded, rendering the problem of inference ill-posed. In this paper, we consider the more general context where the sample of response/covariate pairs forms a weakly dependent stationary process in the respective product Hilbert space: simply stated, the case where we have a regression between functional time series. We consider a general framework of potentially nonlinear processes, expoiting recent advances in the spectral analysis of functional time series. This allows us to quantify the inherent ill-posedness, and to motivate a Tikhonov regularisation technique in the frequency domain. Our main result is the rate of convergence for the corresponding estimators of the regression coefficients, the latter forming a summable sequence in the space of Hilbert-Schmidt operators. In a sense, our main result can be seen as a generalisation of the classical functional linear model rates to the case of time series, and rests only upon Brillinger-type mixing conditions. It is seen that, just as the covariance operator eigenstructure plays a central role in the independent case, so does the spectral density operator's eigenstructure in the dependent case. While the analysis becomes considerably more involved in the dependent case, the rates are strikingly comparable to those of the i.i.d. case, but at the expense of an additional factor caused by the necessity to estimate the spectral density operator at a nonparametric rate, as opposed to the parametric rate for covariance operator estimation.

*Key words and phrases:* Frequency analysis, functional linear model, spectral density operator, system identification, Tikhonov regularisation

## 1. Introduction

Functional regression generalises the classical linear model of multivariate statistics to the case where the parameter, the response, and the error com-

*Research supported by an ERC Starting Grant Award to Victor M. Panaretos
*Part of this work was carried out while the first author was with the EPFL.

ponents reside in general separable Hilbert spaces, while the design matrix is replaced by a linear operator between these spaces (Grenander (1981)). The most studied case is that where the covariate lies in space $(L^2[0,1], \langle\cdot,\cdot\rangle, \|\cdot\|)$ of square integrable real functions on the unit interval (Horváth and Kokoszka (2012), Hsing and Eubank (2015), Ramsay and Silverman (2005)). Here one has independent random elements $X, \epsilon \in L^2[0,1]$, and a bounded linear operator $\mathcal{B} : L^2[0,1] \to \mathcal{H}$ mapping into a separable Hilbert space $\mathcal{H}$, yielding the regression model

$$Y = \mathcal{B}X + \varepsilon.$$

The random elements $X$ and $Y$ are assumed observable, but $\varepsilon$ is unobservable and $\mathcal{B}$ is unknown and to be estimated from i.i.d. replicates $\{(X_n, Y_n)\}_{n\in\mathbb{N}}$ of $(X, Y)$. The most studied case is the so-called scalar-on-function regression where $\mathcal{H} = \mathbb{R}$, so $\mathcal{B}$ reduces to a bounded linear functional $\mathcal{B}f = \langle f, \beta\rangle$, and the function $\beta$ is the parameter of interest. More general is the case in which $\mathcal{H} = L^2[0,1]$, and the operator $\mathcal{B}$ is an integral operator with kernel $\beta \in L^2([0,1]^2)$,

$$\mathcal{B}f = \int_0^1 \beta(\sigma,\tau)f(\tau)d\tau, \quad \forall\, f \in L^2[0,1].$$

In either of these cases, writing down the normal equations reveals an ill-posed inverse problem: the equations involve the application of the inverse of the trace-class covariance operator $\mathcal{R}$ of the random element $X$. Worse still, the operator $\mathcal{R}$ is unknown, and needs to be replaced by its empirical version. Consequently, the statistical methodology for functional regression must involve some means of regularisation, the most popular being PCA regression (or spectral truncation), where one replaces the empirical covariance $\hat{\mathcal{R}}$ by its best rank $K$ approximation in nuclear norm, for some regularisation parameter $K$ (that is of course allowed to grow with $n$; see, e.g. Ramsay and Silverman (2005, Chap. 10); Ferraty and Vieu (2000); Cuevas, Febrero and Fraiman (2002); Cardot and Sarda (2006)).

In a landmark contribution on the functional linear model, Hall and Horowitz (2007) demonstrated that, while the PCA estimator can achieve minimax rates (in probability) in some cases, the ridge estimator (corresponding to Tikhonov regularisation, and adding a multiple of the identity to the empirical covariance) can have important advantages. Theoretically, the Tikhonov estimator can achieve the minimax mean square error (MSE) rate, whereas the truncated PCA estimators would need to undergo a nonlinear modification to achieve similar MSE rates (see, e.g. Hall and Hosseini-Nasab (2006, Thm. 5, Appendix A.2), and the remarks following Hall and Horowitz (2007, Thm. 1). Practically, Hall and Horowitz (2007) showed that the Tikhonov estimator enjoyed better stability properties and was robust to eigenvalue ties. The results of Hall and Horowitz (2007) apply to the scalar-on-function case, and extensions thereof have recently been considered in the function-on-function case (Imaizumi and Kato (2016)).

In this paper, we attack the problem of extending the Tikhonov-based methodology and rates of convergence of Hall and Horowitz (2007) to the case of the function-on-function regression of time series (which can also be seen as a functional linear system identification problem). Here, the observed covariates $\{X_t\}_{t\in\mathbb{Z}}$ are no longer i.i.d., but constitute a stationary process in $L^2[0,1]$. The resulting response process $\{Y_t\}_{t\in\mathbb{Z}}$ is then also a stationary process, linearly coupled to the $X_t$ and $\epsilon_t$ via a sequence of operators $\{\mathcal{B}_t\}_{t\in\mathbb{Z}}$,

$$Y_t = \sum_s \mathcal{B}_{t-s} X_s + \epsilon_t, \quad t \in \mathbb{Z}.$$

Of interest is the estimation of the operators (or *filter*) $\{\mathcal{B}_t\}$, on the basis of the observation of a finite stretch of pairs $\{(X_t, Y_t)\}_{t=0}^{T-1}$. This case is considerably more challenging than the i.i.d. function-on-function case. The reason is that beyond the intrinsic covariation of each regressor function $X_t$, encapsulated in the covariance $\mathcal{R}$, one needs to account for the temporal covariation between lagged regressor functions $X_t$ and $X_{t+s}$. These too contribute to the ill-posedness of the problem, which is now doubly ill-posed: one needs to solve an operator deconvolution problem, where the "Fourier division" step is replaced with the solution of an integral equation. To account for these two layers of ill-posedness, one needs to consider the frequency domain framework (Panaretos and Tavakoli (2013a,b)), and it turns out that the operator that needs to be inverted as part of the normal equations is now the spectral density operator of the process $\{X_t\}$,

$$\mathcal{F}_\omega^{XX} = \frac{1}{2\pi} \sum_t e^{-\mathbf{i}t\omega} \mathcal{R}_t^X,$$

the Fourier transform of the lag $t$ autocovariance operators $\mathcal{R}_t$ of $\{X_t\}$.

Just as estimation in the i.i.d case is based on the spectral truncation or the ridge regularisation of the covariance operator, estimation in the time series case can be based on the spectral truncation or ridge regularisation of the spectral density operator (achieved by *harmonic* or *dynamic* PCA, see Panaretos and Tavakoli (2013a) and Hörmann, Kidziński and Hallin (2015)). The spectral truncation approach was recently considered and studied by Hörmann, Kidziński and Kokoszka (2015), and indeed this appears to be the first contribution to the theory of time series regression without any structural assumptions past weak dependence (to be contrasted to the functional regression of *linear processes*, which are much better understood, see Bosq (2012)). Hörmann, Kidziński and Kokoszka (2015) show that by truncating the spectral density operator at a certain rate, one can obtain consistent estimators of the operators $\{\mathcal{B}_t\}$ under weak dependence conditions. An elegant aspect of their approach is that the "correct" truncation rate can in principle be deduced from the data. Still, convergence rates have to date not been established.

Inspired by the work of Hall and Horowitz (2007), we set forth to establish such convergence rates. In view of the technical difficulties of PCA regression in the i.i.d. case, it seems unlikely that MSE error rates would be attainable for the truncated harmonic PCA estimator without some nonlinear modification – after all, the i.i.d. setup is a special case of the time series setup, and so any difficulties encountered in the former apply to the latter, too. This motivates us to introduce a different regularisation method than that of Hörmann, Kidziński and Kokoszka (2015), adopting the Tikhonov perspective. In this framework, we establish the rate of convergence under Brillinger-type weak dependence conditions (Brillinger (2001)), and mild ill-posedness assumptions formulated in direct analogy to the assumptions of Hall and Horowitz (2007) (and of Imaizumi and Kato (2016)). The convergence rate turns out to be the same as in the i.i.d. case, except for the presence of a bandwidth factor that results from the fact that one needs to estimate the spectral density operator by smoothing the periodogram operator; unless one *knows* the processes to actually be uncorrelated in $t \in \mathbb{Z}$, this is a term that cannot be escaped.

The paper is organised as follows. Section 2 establishes notational conventions and analytic notions employed throughout Section 3 then briefly reviews the framework of functional time series, including the key objects of frequency domain functional time series used in the sequel. Functional time series regression and its diagonalisation are considered in Sections 4 and 5. This motivates the methodological contribution of the paper, the Fourier-Tikhonov estimator, presented in Section 6 and discussed in detail in comparison to PCA-based methodology. Our central result is given in Section 7.7, is the MSE rate of convergence of the Fourier-Tikhonov estimator.

## 2. Basic Definitions and Notation

We work in the usual context of functional data analysis, that assumes that each datum arises as the realisation of a random element of the separable Hilbert space $L^2([0,1])$ of square integrable real functions on $[0,1]$. The latter is equipped with the standard inner product and norm

$$\langle f, g\rangle = \int_0^1 f(\tau)g(\tau)d\tau, \quad \|f\|^2 = \int_0^1 f^2(\tau)d\tau = \langle f, f\rangle.$$

Given a linear operator $\mathcal{B} : L^2([0,1]) \to L^2([0,1])$, we denote its adjoint by $\mathcal{B}^*$, its generalised inverse by $\mathcal{B}^\dagger$, and its inverse by $\mathcal{B}^{-1}$, if well defined. The Schatten-$\infty$ norm (operator norm), Schatten-2 norm (Hilbert-Schmidt norm), and Schatten-1 norm (nuclear norm) are respectively, denoted by

$$|||\mathcal{B}|||_\infty = \sup_{\|h\|=1} \|\mathcal{B}h\|, \quad |||\mathcal{B}|||_2 = \sqrt{\mathrm{trace}\,(\mathcal{B}^*\mathcal{B})}, \quad |||\mathcal{B}|||_1 = \mathrm{trace}\left(\sqrt{\mathcal{B}^*\mathcal{B}}\right).$$

Occasionally, we abuse notation and apply a Schatten norm to the kernel of the corresponding integral operator, in which case it should be understood that the norm applies to the induced operator. For example, if $f \in L^2([0,1]^2)$, we may write $|||f|||_1$ to denote the Schatten-1 norm of the operator $g \mapsto \int_0^1 f(s,t)g(t)dt$.

The identity operator is denoted by $\mathcal{I}$. For a pair of elements $f, g \in L^2[0,1]$, we define the tensor product (operator) as $f \otimes g : L^2[0,1] \to L^2[0,1]$

$$(f \otimes g)u = \langle g, u\rangle f, \quad u \in \mathcal{H}.$$

We make use of the same notation for tensor products of operators: if $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{G}$ are operators $L^2[0,1] \to L^2[0,1]$, we write

$$(\mathcal{A} \otimes \mathcal{B})\,\mathcal{G} = \mathrm{trace}\,(\mathcal{B}^*\mathcal{G})\,\mathcal{A}.$$

Finally, we use $Conv_C\big(L^2([0,1]^2,\mathbb{C}) \times L^2([0,1]^2,\mathbb{C})\big)$ to denote the set of finite convex combinations of elements of the form $f \times g$ with $f, g \in L^2([0,1]^2,\mathbb{C})$ whose induced operators have Schattern 1-norm uniformly bounded by a constant $C$. A generic element of $Conv_C\big(L^2([0,1]^2,\mathbb{C}) \times L^2([0,1]^2,\mathbb{C})\big)$ is denoted by $\vartheta_1 \odot \vartheta_2$, understood as implying that this element can be written in the form

$$\vartheta_1(a_1,a_2) \odot \vartheta_2(a_3,a_4) = \sum_{j=1}^{J} \pi_j f_j(a_1,a_2)h_j(a_3,a_4),$$

for a probability measure $\{\pi_j\}_{j=1}^J$ and functions $f_j, h_j \in L^2([0,1]^2,\mathbb{C})$ such that the Schatten-1 norms $\{|||f_j|||_1, |||h_j|||_1\}_{j=1}^J$ of the operators $L^2([0,1],\mathbb{C}) \to L^2([0,1],\mathbb{C})$ with kernels $f_j$ and $h_j$ are all bounded by $C$. This notation is used frequently to abbreviate cumbersome terms in Taylor expansions involving linear combinations of products of kernels.

## 3. Functional Time Series Background

A *functional time series* is a sequence of random elements $\{X_t\}$ of $L^2[0,1]$, indexed by $t \in \mathbb{Z}$ (interpreted as time). The argument of each function $X_t$ is denoted by, $\tau \in [0,1]$,

$$X_t(\tau) : [0,1] \to \mathbb{R}, \quad \text{for } t \in \mathbb{Z}.$$

We consider only *strictly stationary* time series: given any finite index set $I \subset \mathbb{Z}$, and any $s \in \mathbb{Z}$, it holds that

$$\{X_t\}_{t\in I} \overset{d}{=} \{X_{t+s}\}_{t\in I}.$$

The mean function and lag $t$ covariance kernel of $\{X_t\}$ are given by,

$$\begin{aligned} \mu^X(\tau) &= \mathbb{E}\{X_t(\tau)\}, \\ r_t^X(\tau,\sigma) &= \mathbb{E}\left[\left\{X_{t+s}(\tau) - \mu^X(\tau)\right\}\left\{X_s(\sigma) - \mu^X(\sigma)\right\}\right], \; t,s \in \mathbb{Z}, \end{aligned}$$

and are well-defined for almost all $\tau \in [0,1]$ and $(\tau,\sigma) \in [0,1]^2$, respectively, when $\mathbb{E}\|X_0\|^2 < \infty$. The lag $t$ covariance operator $\mathcal{R}_t^X : L^2[0,1] \to L^2[0,1]$ is then defined by the action

$$\mathcal{R}_t^X h = \mathbb{E}\left\{\left(X_{t+s} - \mu^X\right) \otimes \left(X_s - \mu^X\right) h\right\} = \operatorname{cov}\left(\langle X_0, h\rangle, X_t\right), \quad h \in L^2[0,1],$$

and is a nuclear integral operator with integral kernel $r_t^X$. Assuming that the sequence $\mathcal{R}_t^X$ is nuclear-summable,

$$\sum_t \left|\!\left|\!\left|\mathcal{R}_t^X\right|\!\right|\!\right|_1 < \infty,$$

we can define the spectral density operator $\mathcal{F}_\omega^X$ at frequency $\omega \in [-\pi,\pi]$ as

$$\mathcal{F}_\omega^{XX} = \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega}\mathcal{R}_t^X.$$

where $\mathbf{i}^2 = -1$. This is a nuclear, self-adjoint, and non-negative operator with integral kernel

$$f_\omega^{XX}(\tau,\sigma) = \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega} r_t^X(\tau,\sigma).$$

The corresponding spectral decompositions are

$$\mathcal{F}_\omega^{XX} = \sum_{i=1}^{\infty} \lambda_i^\omega \varphi_i^\omega \otimes \overline{\varphi_i^\omega} \quad \text{and} \quad f_\omega^{XX}(\tau,\sigma) = \sum_{i=1}^{\infty} \lambda_i^\omega \varphi_i^\omega(\tau)\overline{\varphi_i^\omega(\sigma)},$$

with $\{\lambda_i^\omega\}$ the sequence of non-negative eigenvalues, and $\{\varphi_i^\omega\}$ the corresponding orthonormal eigenfunctions.

Given a second functional time series $\{Y_t\}$ satisfying the same (corresponding) assumptions, we can define the lag $t$ cross-covariance kernel as

$$r_t^{YX}(\tau,\sigma) = \mathbb{E}\left[\left\{X_{t+s}(\tau) - \mu^X(\tau)\right\}\left\{Y_s(\sigma) - \mu^Y(\sigma)\right\}\right],\ \tau,\sigma \in [0,1]\ \&\ t,s \in \mathbb{Z},$$

which, in turn, induces the lag $t$ cross-covariance operator $\mathcal{R}_t^{YX} : L^2(0,1) \to L^2[0,1]$ by

$$\mathcal{R}_t^{YX} h = \mathbb{E}\left\{\left(X_{t+s} - \mu^X\right) \otimes \left(Y_s - \mu^Y\right)\right\} h = \operatorname{cov}\left(\langle Y_0, h\rangle, X_t\right), \quad h \in L^2(0,1).$$

The cross-spectral density operator $\mathcal{F}_\omega^{YX}$ at frequency $\omega \in [-\pi,\pi]$ is then defined as

$$\mathcal{F}_\omega^{YX} = \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega}\mathcal{R}_t^{YX}$$

with associated integral kernel

$$f_\omega^{YX}(\tau,\sigma) = \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega} r_t^{YX}(\tau,\sigma).$$

These can be expanded in the basis of eigenfunctions of $\mathcal{F}_\omega^{XX}$, yielding the Fourier representations

$$\mathcal{F}_\omega^{YX} = \sum_{i,j=1}^{\infty} a_{ij}^\omega \varphi_i^\omega \otimes \overline{\varphi_j^\omega} \quad \& \quad f_\omega^{YX}(\tau,\sigma) = \sum_{i,j=1}^{\infty} a_{ij}^\omega \varphi_i^\omega(\tau)\overline{\varphi_j^\omega(\sigma)}.$$

Finally, we consider *cumulant kernels* (and corresponding operators) as a means of quantifying the strength of temporal dependence in $\{X_t\}$ via Brillinger mixing conditions. Given any $(\tau_1,\dots\tau_k) \in [0,1]^k$, we define the order-$k$ cumulant kernel of $\{X_t\}$ as

$$\operatorname{cum}\left(X_{t_1}(\tau_1),\dots,X_{t_k}(\tau_k)\right) = \sum_{\nu=(\nu_1,\dots,\nu_p)} (-1)^{p-1}(p-1)!\prod_{l=1}^{p}\mathbb{E}\left\{\prod_{j\in\nu_l} X_{t_j}(\tau_j)\right\},$$

with summation being over unordered partitions $\nu = (\nu_1,\dots,\nu_p)$ of $\{1,\dots,k\}$. The kernel exists almost everywhere on $[0,1]^k$ provided $\mathbb{E}\|X_0\|^k < \infty$. A cumulant kernel of of order $2k$ gives rise to a corresponding $2k$*-th order cumulant operator* $\mathcal{R}_{t_1,\dots,t_{2k-1}} : L^2([0,1]^k,\mathbb{R}) \to L^2([0,1]^k,\mathbb{R})$, defined by right integration. More generally, any $g \in L^2([0,1]^{2k},\mathbb{R})$ induces a corresponding operator $\mathcal{G}$ on $L^2([0,1]^k$, defined as

$$\mathcal{G}h(\tau_1,\dots,\tau_k) = \int_{[0,1]^k} g(\tau_1,\dots,\tau_{2k}) \times h(\tau_1,\dots,\tau_k)d\tau_1\cdots d\tau_k,$$

provided the integral is well-defined.

## 4. Functional Time Series Regression

In the context of a functional time series regression, we consider a collection of *covariates* $\{X_t\}$ and associated *responses* $\{Y_t\}$, each comprising a strictly stationary time series of random elements in $L^2[0,1]$. A functional linear model for the pair $(X_t, Y_t)$ stipulates that the two time series are defined on the same probability space and are *approximately linearly coupled.* That is, there exists a sequence of Hilbert-Schmidt operators $\{\mathcal{B}_t\}$ with integral kernels $\{b_t\}$,

$$\mathcal{B}_t : L^2 \to L^2, \quad b_t(\sigma,\tau) : [0,1]^2 \to \mathbb{R}, \quad (\mathcal{B}_t f)(\tau) = \int_0^1 b_t(\sigma,\tau) f(\sigma) d\sigma, \quad f \in L^2,$$

and a collection of centred i.i.d. *perturbations* in $L^2$, $\{\epsilon_t\}_{t\in\mathbb{Z}}$, such that

$$Y_t = \sum_s \mathcal{B}_{t-s} X_s + \epsilon_t, \quad t \in \mathbb{Z}. \tag{4.1}$$

Notice that the temporal convolution is the only possible linear coupling, if both $X_t$ and $Y_t$ are to be stationary.

**Assumption 1** (Moment and Dependence Assumptions)**.** *In the context of model* (4.1)*, we assume,*

(A1) *the* filter $\{\mathcal{B}_t\}$ *is* Hilbert-Schmidt *summable,*

$$\sum_t |||\mathcal{B}_t|||_2 = \sum_t \left( \int_0^1 \int_0^1 \left| b_t(\tau,\sigma) \right|^2 d\sigma d\tau \right)^{1/2} < \infty.$$

(A2) *the i.i.d. perturbation process* $\{\epsilon_t\}$ *is independent of the covariate process* $\{X_t\}$*, and*

$$\mathbb{E}\|X_t\|^2 + \mathbb{E}\|\epsilon_t\|^2 < \infty, \quad \mathbb{E}(X_t) = \mathbb{E}(\epsilon_t) = 0,$$

(A3) *the covariance operators* $\{\mathcal{R}_t^X\}_{t\in\mathbb{Z}}$ *are* nuclear *summable,*

$$\sum_t |||\mathcal{R}_t^X|||_1 < \infty.$$

Whenever Assumptions (A1)–(A3) are satisfied, it holds that $\{Y_t\}$ is also second order (Bosq (2012)), and possesses nuclear-summable covariance operators $\{\mathcal{R}_t^Y\}_{t\in\mathbb{Z}}$,

$$\mathbb{E}\|Y_t\|^2 < \infty \quad \& \quad \sum_t |||\mathcal{R}_t^Y|||_1 < \infty.$$

The statistical task at hand is to estimate the unknown sequence of operators (or *filter*) $\{\mathcal{B}_t\}_{t\in\mathbb{Z}}$ on the basis of the observation of a finite stretch of $\{(X_t, Y_t); t = 0, \ldots, T-1\}$. As usual, the $\epsilon_t$ are unobservable.

## 5. Diagonalising the Problem

As with iid functional regression, the key to constructing estimators is to establish a connection between the cross-covariance of $\{X_t\}$ with $\{Y_t\}$, and the sequence of operators $\{\mathcal{B}_t\}$. The next lemma does precisely that.

**Proposition 1.** *In the notation of Section* 2, *and under Assumptions* 2, *it holds that*

$$\mathcal{R}_t^{YX} = \sum_u \mathcal{B}_{t-u}\mathcal{R}_u^X, \quad \sum_t |||\mathcal{R}_t^{YX}|||_1 < \infty, \quad \mathcal{F}_\omega^{YX} = \mathcal{Q}_\omega \mathcal{F}_\omega^{XX},$$

*where* $\mathcal{Q}_\omega$ *is the linear operator with kernel*

$$f_\omega^B(\tau,\sigma) = \sum_t e^{-\mathbf{i}\omega t} b_t(\tau,\sigma),$$

*and satisfies*

$$\int_{-\pi}^{\pi} |||\mathcal{Q}_\omega|||_2^2 d\omega = \sum_t |||\mathcal{B}_t|||_2^2 < \infty.$$

In passing, we note that $\mathcal{Q}_\omega$ and $f_\omega^B(\tau,\sigma)$ also admit a Fourier representations with respect to the eigenfunctions $\{\varphi_n^\omega\}$ of $\mathcal{F}_\omega^{XX}$, and these are denoted as

$$\mathcal{Q}_\omega = \sum_{i,j=1}^{\infty} b_{ij}^\omega \varphi_i^\omega \otimes \overline{\varphi_j^\omega} \quad \text{and} \quad f_\omega^B(\tau,\sigma) = \sum_{i,j=1}^{\infty} b_{ij}^\omega \varphi_i^\omega(\varsigma)\overline{\varphi_j^\omega(\tau)}.$$

*Proof of Proposition* 1. Given any $f \in L^2[0,1]$, we have

$$\left\{\left(\sum_u \mathcal{B}_{t-u}X_u\right) \otimes X_0\right\} f = \langle X_0, f\rangle \sum_u \mathcal{B}_{t-u}X_u = \sum_u \mathcal{B}_{t-u}\langle X_0, f\rangle X_u.$$

As a result, it holds that

$$\mathbb{E}\left(\sum_u \mathcal{B}_{t-u}\langle X_0, f\rangle X_u\right) = \sum_u \mathcal{B}_{t-u}\mathbb{E}\left(\langle X_0, f\rangle X_u\right)$$

using Fubini's theorem and the fact that

$$\begin{aligned}
\sum_u \mathbb{E}\|\mathcal{B}_{t-u}\langle X_0, f\rangle X_u\| &\le \sum_u \mathbb{E}\left(|||\mathcal{B}_{t-u}|||_2 |||X_u \otimes X_0|||_2 \|f\|\right) \\
&\le \|f\| \sum_u |||\mathcal{B}_{t-u}|||_2 \mathbb{E}\left[\|X_u\|\|X_0\|\right] \\
&\le \|f\|^2 \sqrt{\mathbb{E}[\|X_u\|^2]\mathbb{E}[\|X_0\|^2]} \sum_u |||\mathcal{B}_{t-u}|||_2 \\
&< \infty.
\end{aligned}$$

Consequently, since $\{\epsilon_t\}$ is uncorrelated with $\{X_t\}$, we have that

$$\mathcal{R}_t^{YX} f = \mathbb{E}\left(\sum_u \mathcal{B}_{t-u}\langle X_0, f\rangle X_u\right) = \sum_u \mathcal{B}_{t-u}\mathbb{E}\left(\langle X_0, f\rangle X_u\right) = \sum_u \mathcal{B}_{t-u}\mathcal{R}_u f$$

which proves the first part of the proposition, since $f \in L^2[0,1]$ was arbitrary. In order to show that $\mathcal{R}_t^{YX}$ is nuclear-summable, we use Hölder's inequality for Schatten norms, and Tonelli's theorem to write

$$\begin{aligned}\sum_t \|\mathcal{R}_t^{YX}\|_1 &\leq \sum_t \sum_u \|\mathcal{B}_{t-u}\mathcal{R}_u^X\|_1 \\ &\leq \sum_u \sum_t \|\mathcal{B}_{t-u}\|_2 \|\mathcal{R}_u^X\|_2 \\ &\leq \sum_u \|\mathcal{B}_{t-u}\|_2 \sum_t \|\mathcal{R}_u^X\|_2 \\ &< \infty.\end{aligned}$$

It follows that the Fourier transform $\mathcal{F}_\omega^{YX}$ of $\mathcal{R}_t^{YX}$ is well-defined. Following the standard manipulations leading to the convolution formula, we have

$$\begin{aligned}\mathcal{F}_\omega^{YX} &= \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega}\mathcal{R}_t^{YX} \\ &= \frac{1}{2\pi}\sum_t e^{-\mathbf{i}t\omega}\sum_u \mathcal{B}_{t-u}\mathcal{R}_u \\ &= \frac{1}{2\pi}\sum_t\sum_u e^{-\mathbf{i}(t-u)\omega}\mathcal{B}_{t-u}e^{-\mathbf{i}u\omega}\mathcal{R}_u \\ &= \sum_t e^{-\mathbf{i}(t-u)\omega}\mathcal{B}_{t-u}\frac{1}{2\pi}\sum_u e^{-\mathbf{i}u\omega}\mathcal{R}_u \\ &= \mathcal{Q}_\omega\mathcal{F}_\omega^{XX}.\end{aligned}$$

Here, we have made use of Fubini's theorem, noting that

$$\sum_u\sum_t \|e^{-\mathbf{i}t\omega}\mathcal{B}_{t-u}\mathcal{R}_u^X\|_1 = \sum_u\sum_t \|\mathcal{B}_{t-u}\mathcal{R}_u^X\|_1 < \infty.$$

When $\mathcal{F}_\omega^{XX}$ is strictly positive uniformly over $\omega$ (so that its range is $L^2([0,1], \mathbb{C})$ itself), then the proposition implies that

$$\mathcal{Q}_\omega = \mathcal{F}_\omega^{YX}\left(\mathcal{F}_\omega^{XX}\right)^{-1} = \left(\sum_{i,j} a_{ij}^\omega \varphi_i^\omega \otimes \overline{\varphi_j^\omega}\right)\left(\sum_j \frac{1}{\lambda_j^\omega}\varphi_j^\omega \otimes \overline{\varphi_j^\omega}\right) = \sum_j\sum_i \frac{a_{ij}^\omega}{\lambda_j^\omega}\varphi_i^\omega \otimes \overline{\varphi_j^\omega}. \tag{5.2}$$

It follows that the operator $\mathcal{B}_t$ can be deduced by inverse Fourier transforming,

$$\mathcal{B}_t = \int_{-\pi}^{\pi} \mathcal{Q}_\omega \exp\left(-\mathbf{i}t\omega\right)d\omega.$$

This allows us to formulate an estimation strategy in the Fourier domain, as described in the next section.

## 6. Methodology for Estimation

The results in the previous Section suggest the following estimation strategy, when we have a finite stretch $\{(X_t, Y_t)\}_{t=0}^{T-1}$ of length $T$ of the coupled series at our disposal.

1. Estimate $\mathcal{F}_\omega^{XX}$ and $\mathcal{F}_\omega^{YX}$ nonparametrically, say by $\widehat{\mathcal{F}}_{\omega,T}^{XX}$ and $\widehat{\mathcal{F}}_{\omega,T}^{YX}$. This can be done using the approach introduced in Panaretos and Tavakoli (2013b), and described in more detail in Section 6.1.
2. Construct a regularised estimator of $(\mathcal{F}_\omega^{XX})^{-1}$ based on $\widehat{\mathcal{F}}_{\omega,T}^{XX}$. Regularisation is necessary, as the operator $\widehat{\mathcal{F}}_{\omega,T}^{XX}$ will be of finite rank and its maximal eigenvalue diverge as $T$ grows. We consider this problem in Section 6.2.

Once these steps have been completed, one can plug the corresponding estimators into 5.2 to obtain the regularised estimator of $\mathcal{Q}_\omega$, and consequently of $\mathcal{B}_t$. This is defined in Section 6.5.

### 6.1. Estimation of $\mathcal{F}_\omega^{XX}$ and $\mathcal{F}_\omega^{YX}$

Following Panaretos and Tavakoli (2013b), let $W(x) : \mathbb{R} \to (0, \infty)$ be a positive real function such that

1. $W$ is of bounded variation.
2. $W(x) = 0$ if $|x| \geq 1$.
3. $\int_{-\infty}^{\infty} W(x)dx = 1$,
4. $\int_{-\infty}^{\infty} W(x)^2 dx < \infty$.

Define a kernel of bandwidth $B_T$ as

$$W^{(T)}(x) = \frac{1}{B_T} \sum_{k\in\mathbb{Z}} W\left(\frac{x + 2k\pi}{B_T}\right).$$

We use this kernel in order to construct estimators in the frequency domain. Specifically, defining the discrete Fourier transforms of the two time series as

$$\widetilde{X}_{\omega,T} = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X_t \exp(-\mathbf{i}\omega t) \quad \& \quad \widetilde{Y}_{\omega,T} = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} Y_t \exp(-\mathbf{i}\omega t),$$

the periodogram operator of $\{X_t\}$ at frequency $\omega$ (and its corresponding kernel) are given by the empirical covariance (and its corresponding kernel) of the discrete Fourier transform at frequency $\omega$,

$$\mathcal{P}_{\omega,T}^{XX} = \widetilde{X}_{\omega,T} \otimes \overline{\left(\widetilde{X}_{\omega,T}\right)} \quad \& \quad p_\omega^{(T)}(\tau, \sigma) = \widetilde{X}_{\omega,T}(\tau) \otimes \overline{\left\{\widetilde{X}_{\omega,T}(\sigma)\right\}}.$$

Similarly, the empirical cross-covariance of the discrete Fourier transforms of $X$ and $Y$ yields the cross-periodogram operator,

$$\mathcal{P}^{YX}_{\omega,T} = \widetilde{Y}_{\omega,T} \otimes \overline{\left(\widetilde{X}_{\omega,T}\right)}.$$

These can be smoothed using $W^{(T)}$, in order to yield the estimators of the spectral density operator of $X$ (and spectral density kernel), and of the cross-spectral density operator of $(X,Y)$,

$$\widehat{\mathcal{F}}^{XX}_{\omega,T} = \frac{1}{T}\sum_{s=0}^{T-1} W^{(T)}(\omega - \nu_s)\mathcal{P}^{XX}_{\omega,T} \quad \text{and}$$

$$f^{(T)}_{\omega}(\tau,\sigma) = \frac{1}{T}\sum_{s=0}^{T-1} W^{(T)}(\omega - \nu_s) p^{(T)}_{\nu_s}(\tau,\sigma), \tag{6.3}$$

$$\widehat{\mathcal{F}}^{YX}_{\omega,T} = \frac{1}{T}\sum_{s=0}^{T-1} W^{(T)}(\omega - \nu_s)\mathcal{P}^{YX}_{\nu_s,T}, \tag{6.4}$$

where

$$\nu_s = \frac{2\pi s}{T}, \quad s = 0,1,\ldots,T-1.$$

### 6.2. Regularised estimation of $(\mathcal{F}^{XX}_{\omega})^{-1}$

Once we have the estimators $\widehat{\mathcal{F}}^{XX}_{\omega,T}$ and $\widehat{\mathcal{F}}^{XX}_{\omega,T}$, a naive approach to estimating $\mathcal{Q}_\omega$ is to use the estimator

$$\widehat{\mathcal{F}}^{YX}_{\omega,T}\left(\widehat{\mathcal{F}}^{XX}_{\omega,T}\right)^{\dagger},$$

where $\left(\widehat{\mathcal{F}}^{XX}_{\omega,T}\right)^{\dagger}$ is the pseudoinverse of $\widehat{\mathcal{F}}^{XX}_{\omega,T}$. However, as can clearly be seen using the spectral decompositions

$$\widehat{\mathcal{F}}^{XX}_{\omega,T} = \sum_{n=1}^{T} \widehat{\lambda}^{\omega}_n \widehat{\varphi}_n(\omega) \otimes \widehat{\varphi}_n(\omega), \quad \left(\widehat{\mathcal{F}}^{XX}_{\omega,T}\right)^{\dagger} = \sum_{n=1}^{T} (\widehat{\lambda}^{\omega}_n)^{-1} \widehat{\varphi}_n(\omega) \otimes \widehat{\varphi}_n(\omega),$$

the eigenvalues of $(\widehat{\mathcal{F}}^{XX}_{\omega,T})^{\dagger}$ do not remain bounded as $T$ diverges when $\widehat{\mathcal{F}}^{XX}_{\omega,T}$ is consistent for $\mathcal{F}^{XX}_{\omega}$ (the latter being nuclear).

This effect is generally *not* annihilated by the application of the integral operator $\widehat{\mathcal{F}}^{YX}_{\omega,T}$ from the left, when forming the naive estimator $\widehat{\mathcal{F}}^{YX}_{\omega,T}(\widehat{\mathcal{F}}^{XX}_{\omega,T})^{\dagger}$. The problem is that the spectrum of $\widehat{\mathcal{F}}^{YX}_{\omega,T}$ depends on $\mathcal{Q}_\omega$, which a priori has no structural relationship with $\mathcal{F}^{XX}_{\omega}$. Said differently, if $\widehat{\mathcal{F}}^{YX}_{\omega,T}$ is expanded in the tensor product basis given by the eigenfunctions of $\mathcal{F}^{XX}_{\omega}$ (extended to a complete system),

$$\widehat{\mathcal{F}}^{YX}_{\omega,T} = \sum_{n,m} \widehat{a}^{\omega}_{n,m} \widehat{\varphi}_n(\omega) \otimes \widehat{\varphi}_m(\omega),$$

there is no reason to expect that the resulting basis coefficients $\{\widehat{a}^{\omega}_{n,m}\}$ will decay sufficiently fast for the products $\widehat{a}^{\omega}_{n,m}(\widehat{\lambda}^{\omega}_n)^{-1}$ to remain bounded in $n$ as $T$ grows. Thus, it is necessary to use some form of regularisation. Two classical strategies are the following.

(i) *Spectral truncation.* Here, one replaces the generalised inverse $(\widehat{\mathcal{F}}^{XX}_{\omega,T})^{\dagger}$ by

$$\sum_{n=1}^{K(T)} (\widehat{\lambda}^{\omega}_n)^{-1} \widehat{\varphi}_n(\omega) \otimes \widehat{\varphi}_n(\omega),$$

where $K(T) < T$ grows sufficiently slowly in order to control the terms $a^{\omega}_{n,m}(\widehat{\lambda}^{\omega}_n)^{-1}$.

(ii) *Tikhonov regularisation.* Here, one replaces the generalised inverse $(\widehat{\mathcal{F}}^{XX}_{\omega,T})^{\dagger}$ by a ridge-regularised inverse

$$\left[\widehat{\mathcal{F}}^{XX}_{\omega,T} + \zeta_T \mathcal{I}\right]^{-1} = \sum_{n=1}^{\infty} (\zeta_T + \widehat{\lambda}^{\omega}_n)^{-1} \widehat{\varphi}_n(\omega) \otimes \widehat{\varphi}_n(\omega),$$

where $\zeta_T$ decays to zero sufficiently slowly in order to control the behaviour of the terms $a^{\omega}_{n,m}(\zeta_T + \widehat{\lambda}^{\omega}_n)^{-1}$.

The first approach (spectral truncation) is essentially the approach described by Hörmann, Kidziński and Kokoszka (2015, Equation 3.4). It can be seen as the extension of functional PCA regression (e.g. Hall and Horowitz (2007), Imaizumi and Kato (2016)) to the case of functional time series. Hörmann, Kidziński and Kokoszka (2015) choose the value $K(T)$ to be dependent on the rate of decay of $\sup_{\omega} \left|\!\left|\!\left| \widehat{\mathcal{F}}^{XX}_{\omega} - \mathcal{F}^{XX}_{\omega} \right|\!\right|\!\right|_{\infty}$ and $\sup_{\omega} \left|\!\left|\!\left| \widehat{\mathcal{F}}^{YX}_{\omega} - \mathcal{F}^{YX}_{\omega} \right|\!\right|\!\right|_{\infty}$ (assumed known), in a way that guarantees consistency of the estimator eventually constructed. In principle, one could be more ambitious and use a a frequency-dependent truncation level $K(T,\omega)$, but it seems unlikely to have detailed enough information on the decay rates $\left|\!\left|\!\left| \widehat{\mathcal{F}}^{XX}_{\omega} - \mathcal{F}^{XX}_{\omega} \right|\!\right|\!\right|_{\infty}$ and $\left|\!\left|\!\left| \widehat{\mathcal{F}}^{YX}_{\omega} - \mathcal{F}^{YX}_{\omega} \right|\!\right|\!\right|_{\infty}$ at each frequency $\omega$.

Though spectral truncation is a very popular technique in the i.i.d. case, it poses some challenges both in terms of theoretical study, as well as practical performance, which might be exacerbated in the dependent case:

To this date, and to the best of our knowledge, there are no results concerning the MSE convergence rates for the spectral truncation estimator, even in the i.i.d. case. Hall and Horowitz (2007) (and Imaizumi and Kato (2016)) establish rates of convergence for small ball probabilities, but not for the MSE itself. Hall and Horowitz (2007) explain that to upgrade to MSE results, the spectral truncation estimator needs to be modified to a non-linear truncated version (see the discussion after Hall and Horowitz (2007, Thm. 1)) and also Hall and Hosseini-Nasab (2006, Thm. 5, Appendix A.2)). It thus seems that in the more challenging weakly dependent case, spectral truncation may not be the most fruitful avenue to obtain MSE convergence rates.

In practical terms, a challenge that spectral truncation encounters is in the case where eigenvalues are nearly tied, because the chosen subspace $\{\varphi_n(\omega)\}_{j=1}^{K(T)}$ makes no reference to the quantity of interest $\mathcal{Q}_\omega$. Specifically, $\mathcal{Q}_\omega$ might be well expressed in some but not other eigenfunctions of $\widehat{\mathcal{F}}_\omega^{XX}$, and this irrespectively of the size of the corresponding eigenvalues (according to which the truncation is performed). Thus, if the eigenvalues $\{\widehat{\lambda}_{K\pm j}^\omega\}_{j=1}^{J(\omega)}$ of $\widehat{\mathcal{F}}_\omega^{XX}$ are nearly tied, the sample variability of the estimator increases, if this is well expressed in some but not all of the eigenfunctions of order $\{K \pm j; j = 1, \ldots, J(\omega)\}$. Intuitively, a certain term that is highly correlated with $\mathcal{Q}_\omega$ may come in or be left out of the truncation simply because of sample variability, leading to variance inflation. This phenomenon was doscumented by Hall and Horowitz (2007) in the standard functional linear model, and can be a serious issue in the time series case, since we are considering a whole range frequencies, and thus of approximate eigenvalue ties $\{\widehat{\lambda}_{K\pm j}^\omega : j = 1, \ldots, J(\omega); \omega \in [-\pi, \pi]\}$.

Hall and Horowitz (2007) introduced and studied Tikhonov regularisation as an alternative that circumvents these issues. Indeed, they were able to deduce convergence rates for the MSE of the Tikhonov estimator, as opposed to the small ball probability rates for spectral truncation. For these two reasons, we follow the Tikhonov approach here, defining $\left[\widehat{\mathcal{F}}_{\omega,T}^{XX} + \zeta_T \mathcal{I}\right]^{-1}$ to be the (regularised) estimator of $\left[\mathcal{F}_\omega^{XX}\right]^{-1}$. We put all the elements together in the next section, to define our estimator.

### 6.3. The smoothed Fourier-Tikhonov estimator of $\{\mathcal{B}_t\}$

Let $B_T > 0$ be a bandwidth and $\zeta_T$ a Tikhonov parameter. The (smoothed) Fourier-Tikhonov estimator of $\{\mathcal{B}_t\}_t$ is defined to be

$$\widehat{\mathcal{B}}_t = \int_{-\pi}^{\pi} \widehat{\mathcal{Q}}_{\omega,T} \exp(-\mathbf{i}t\omega) d\omega, \tag{6.5}$$

where

$$\widehat{\mathcal{Q}}_{\omega,T} = \widehat{\mathcal{F}}_{\omega,T}^{YX} \big(\widehat{\mathcal{F}}_{\omega,T}^{XX} + \zeta_T \mathcal{I}\big)^{-1} \tag{6.6}$$

is the estimator of $\mathcal{Q}_\omega$. Here $\widehat{\mathcal{F}}_{\omega,T}^{YX}$ and $\widehat{\mathcal{F}}_{\omega,T}^{XX}$ are the smoothed periodogram and smoothed cross-periodogram estimators defined in Section 6.1 (see 6.3 and 6.4).

In terms of series representations, $\widehat{\mathcal{Q}}_{\omega,T}$ and $\mathcal{Q}_\omega$ can be comparatively written as

$$\begin{aligned}\widehat{\mathcal{Q}}_{\omega,T} &= \widehat{\mathcal{F}}^{YX}_{\omega,T}\big(\widehat{\mathcal{F}}^{XX}_{\omega,T} + \zeta_T \mathcal{I}\big)^{-1}\\ &= \left(\sum_{i,j} \hat{a}^{\omega}_{ij} \hat{\varphi}^{\omega}_i \otimes \overline{\hat{\varphi}^{\omega}_j}\right)\left(\sum_j \frac{1}{\hat{\lambda}^{\omega}_j + \zeta_T} \hat{\varphi}^{\omega}_j \otimes \overline{\hat{\varphi}^{\omega}_j}\right)\\ &= \sum_j \sum_i \frac{\hat{a}^{\omega}_{ij}}{\hat{\lambda}^{\omega}_j + \zeta_T} \hat{\varphi}^{\omega}_i \otimes \overline{\hat{\varphi}^{\omega}_j},\\ \mathcal{Q}_\omega &= \mathcal{F}^{YX}_{\omega}\big(\mathcal{F}^{XX}_{\omega}\big)^{-1}\\ &= \left(\sum_{i,j} a^{\omega}_{ij} \varphi^{\omega}_i \otimes \overline{\varphi^{\omega}_j}\right)\left(\sum_j \frac{1}{\lambda^{\omega}_j} \varphi^{\omega}_j \otimes \overline{\varphi^{\omega}_j}\right)\\ &= \sum_j \sum_i \frac{a^{\omega}_{ij}}{\lambda^{\omega}_j} \varphi^{\omega}_i \otimes \overline{\varphi^{\omega}_j},\end{aligned}$$

for $\{\hat{\varphi}^{\omega}_j, \hat{\lambda}^{\omega}_j\}$ the eigenfunctions/eigenvalues of $\widehat{\mathcal{F}}^{XX}_{\omega,T}$, and $\{\hat{a}^{\omega}_{ij}\}$ the Fourier coefficients of $\widehat{\mathcal{F}}^{YX}_{\omega,T}$ with respect to the orthonormal system $\{\hat{\varphi}^{\omega}_j\}$.

The asymptotic performance of our estimator, and its dependence on the choice of $\zeta_T$ is investigated in the next Section.

## 7. Rate of Convergence

In this section, we state our main result of this paper, concerning the rate of convergence of the MSE of the Smoothed Fourier-Tikhonov Estimator (6.5). One can establish consistency (without a rate of convergence) by letting $B_T \to 0$ and $TB_T \to \infty$ as $T \to \infty$, provided that the decay rate of $\zeta_T$ is taken to be a suitable function of $\sup_\omega \left\|\widehat{\mathcal{F}}^{XX}_{\omega} - \mathcal{F}^{XX}_{\omega}\right\|_\infty$ and $\sup_\omega \left\|\widehat{\mathcal{F}}^{YX}_{\omega} - \mathcal{F}^{YX}_{\omega}\right\|_\infty$. This follows similar steps as Hörmann, Kidziński and Kokoszka (2015), but adapted to the case of Tikhonov regularisation, and does *not* require any structural assumptions on the rate of decay of $\{\lambda^{\omega}_n\}$ or indeed on the spectra of $\{\mathcal{B}_t\}$, just as the results of Hörmann, Kidziński and Kokoszka (2015) did not either.

We would like to be able to make more refined statements, and, in particular, to establish convergence rates in the form of a rate of decay for the mean square error

$$\mathbb{E}\left(\sum_t \left\|\mathcal{B}_t - \widehat{\mathcal{B}}_t\right\|_2^2\right) = \mathbb{E}\left(\int_0^{2\pi} \left\|\mathcal{Q}_\omega - \widehat{\mathcal{Q}}_{\omega,T}\right\|_2^2\right) d\omega,$$

where equality follows from Parseval's relation. Such rates necessarily depend on the decay rate of $\{\lambda^{\omega}_n\}$, and on the spectra of $\{\mathcal{B}_t\}$. Our goal is thus to establish a convergence rate that links these behaviours, and illustrates their interplay with the tuning parameters $B_T$ and $\zeta_T$.

We work in the so-called *mildly ill-posed* setting, where the spectra involved exhibit a polynomial decay. Specifically, $\mathcal{F}_\omega^{YX}$, $\mathcal{F}_\omega^{XX}$ and $\mathcal{F}^B$ have integral kernels admitting series representations

$$f_\omega^{XX}(\tau,\sigma) = \sum_{i=1}^{\infty} \lambda_i^\omega \varphi_i^\omega(\tau)\overline{\varphi_i^\omega(\sigma)},$$
$$f_\omega^{YX}(\tau,\sigma) = \sum_{i,j=1}^{\infty} a_{ij}^\omega \varphi_i^\omega(\tau)\overline{\varphi_j^\omega(\sigma)},$$
$$f_\omega^{B}(\tau,\sigma) = \sum_{i,j=1}^{\infty} b_{ij}^\omega \varphi_i^\omega(\tau)\overline{\varphi_j^\omega(\sigma)}.$$

Further assumptions are collected here.

**Assumption 2** (Ill-Posedness, Spectral Smoothness, and Weak Dependence)**.** *In the context of model* (4.1)*, we assume the following.*

(B1) *For all $j$ and $\omega$ it holds that*

$$\lambda_j^\omega \asymp Cj^{-\alpha}, \quad \sum_i \left|b_{ij}^\omega\right| \le Cj^{-\beta}.$$

*with $\alpha > 1$, $\beta > 1/2$, and $\alpha < \beta + 1/2$ .*

(B2) *Whenever $\varphi_i^\omega$ is an eigenfunction of $\mathcal{F}_\omega^{XX}$, then so is its complex conjugate $\overline{\varphi_i^\omega}$, so*

$$\left\{\varphi_i^\omega : i = 1, 2, \dots\right\} = \left\{\overline{\varphi_i^\omega} : i = 1, 2, \dots\right\}.$$

*Therefore, for each $i$, there exists unquely an index $i'$ such that $\left\langle \varphi_i^\omega, \varphi_{i'}^\omega \right\rangle = 1$ and $\left\langle \varphi_i^\omega, \varphi_j^\omega \right\rangle = 0$ when $j \neq i'$.*

(B3) *The kernel $W$ is uniformly bounded, compactly supported and, even on $[-1,1]$, $\int_{\mathbb{R}} W(\alpha)d\alpha = 1$; there exists a positive integer $p$ such that $B_T^{p+1} < T^{-1}$ and for $j \le p-1$.*

$$\int_{\mathbb{R}} W(\alpha)\alpha^j d\alpha = 0.$$

(B4)

$$\sum_{t\in\mathbb{Z}} |t|^{p+5} \left\|\left|\mathcal{R}_t\right|\right\|_1 < \infty$$
$$\sum_{t\in\mathbb{Z}} |t|^{p+5} \left\|\left|\mathcal{B}_t\right|\right\|_1 < \infty.$$

(B5) *The functions* $r_t^X$ *and* $b_t$ *are continuous for all* $t \in \mathbb{Z}$ *with respect to* $\tau, \sigma \in [0,1]$, *and*

$$\sum_{t\in\mathbb{Z}} |t|^{p+2} \big\| r_t^X(\tau,\sigma) \big\|_\infty < \infty$$
$$\sum_{t\in\mathbb{Z}} |t|^{p+2} \big\| b_t(\tau,\sigma) \big\|_\infty < \infty.$$

(B6) *There exists constant* $C < \infty$ *such that*

$$\sum_{t_1,t_2,t_3\in\mathbb{Z}} \left\| cum(X_{t_1}, X_{t_2}, X_{t_3}, X_0) \right\|_1 < C.$$

Condition (B1) is the direct extension of the mild ill-posedness conditions of Hall and Horowitz (2007) to the time series context (see Hall and Horowitz (2007, Sec. 3) for a detailed discussion; Imaizumi and Kato (2016, Sec. 3.1) also introduce these conditions in the function-on-function regression case). Condition (B2) has the set of eigenfunctions is closed under conjugation. We do not need to make any assumption on the separation of the eigenvalues, since Tikhonov regularisation is immune to eigenvalue ties. The conditions in (B5) can be seen as weak dependence conditions that suffice for the existence of Taylor expansions of sufficiently high order of the spectral density operator and the Fourier transform of the filter with respect to the frequency argument. Conditions (B4) and (B6) are also weak dependence conditions of Brillinger-type, that are sufficient for the establishment of convergence rates of the spectral density estimator to its estimand (as in Panaretos and Tavakoli (2013b)). Finally, (B3) is a standard higher order kernel assumption that is often encountered in density estimation and deconvolution.

**Theorem 1** (Rate of Convergence)**.** *Let* $\{\widehat{\mathcal{B}}_t\}$ *be the Fourier-Tikhonov estimator* 6.5 *of the coefficients* $\{\mathcal{B}_t\}$ *in the functional time series regression model* 4.1 *satisfying Assumptions* 2. *Then, under conditions* (B1)–(B5)*, there exists a sequence of events* $G_T$ *such that* $\mathbb{P}\big[G_T\big] \to 1$, *and*

$$\mathbb{E}\left( \sum_t \left\| \mathcal{B}_t - \widehat{\mathcal{B}}_t \right\|_2^2 ; G_T \right) = \frac{1}{B_T} O\Big(T^{-(2\beta-1)/(\alpha+2\beta)}\Big), \tag{7.7}$$

*provided the Tikhonov parameter satisfies* $\zeta_T = T^{-\alpha/(\alpha+2\beta)}$ *and the bandwidth*

---

Note, however, that we do not need to make any assumption on the separation of the eigenvalues, since Tikhonov regularisation is immune to eigenvalue ties.

*satisfies* $B_T = T^{-\gamma}$, *with* $\gamma$ *such that*

$$\frac{\alpha - 1}{\alpha + 2\beta} < \gamma < \frac{2\beta - \alpha}{\alpha + 2\beta}.$$

**Remark 1.** Assuming that $(\alpha - 1)/(\alpha + 2\beta) < \gamma < (2\beta - \alpha)/(\alpha + 2\beta)$ is compatible with assumption (B1) since $\alpha < \beta + 1/2$.

If we compare the rate (7.7) with the one obtained by Hall and Horowitz (2007) in the i.i.d case, we see that they are identical except for the presence of the $B_T^{-1}$ factor in our case. Intuitively, this is the price we have to pay for the fact that the estimation of the spectral density operator is intrinsically harder than the estimation of a covariance operator: for densely observed functional data, a covariance operator can be estimated at a parametric rate (Hall, Müller and Wang (2006)), but the spectral density operator can only be estimated at nonparametric rates (Panaretos and Tavakoli (2013b)).

The proof of Theorem 1 is quite lengthy and technical, and is constructed via a series of intermediate results in the Supplementary Material, available online.

### Supplementary Materials

The supplement contains the proof Theorem 1, including several auxiliary results that are required for this proof, but stated separately for tidiness.

### Acknowledgment

We thank the Editor and two referees for their constructive feedback and their useful comments.

## References

Bosq, D. (2012). *Linear Processes in Function Spaces: Theory and Applications* 149, Springer Science & Business Media.

Brillinger, D. R. (2001). *Time Series: Data Analysis and Theory* **36** Siam.

Cardot, H. and Sarda, P. (2006). Linear regression models for functional data. In *The Art of Semiparametrics, Contributions to Statistics* 49–66. Physica-Verlag HD.

Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression: The case of fixed design and functional response. *Canadian Journal of Statistics* **30**, 285–300.

Ferraty, F. and Vieu, P. (2000). Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* **330**, 139–142.

Grenander, U. (1981). *Abstract Inference*. Wiley, New York.

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

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*, 1493–1517.

Hörmann, S., Kidziński, Ł. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **77**, 319–348.

Hörmann, S., Kidziński, Ł. and Kokoszka, P. (2015). Estimation in functional lagged regression. *Journal of Time Series Analysis* **36**, 541–561.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications.* Springer Series in Statistics. Springer, New York.

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.* Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester.

Imaizumi, M. and Kato, K. (2016). Pca-based estimation for functional linear regression with functional responses. *arXiv preprint arXiv: 1609.00286.*

Panaretos, V. M. and Tavakoli, S. (2013a). Cramér–karhunen–loève representation and harmonic principal component analysis of functional time series. *Stochastic Processes and Their Applications* **123**, 2779–2807.

Panaretos, V. M. and Tavakoli, S. (2013b). Fourier analysis of stationary time series in function space. *The Annals of Statistics* **41**, 568–603.

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

Tung Pham, School of Mathematics and Statistics Universiry of Melbourne, Victoria 3010, Australia.

E-mail: pham.t@unimelb.edu.au

Victor M. Panaretos, Institute of Mathematics, Ecole Polytechnique Fédérale de Lausanne, EPFL Station 8 , Lausanne 1015, Switzerland.

E-mail: victor.panaretos@epfl.ch

(Received December 2016; accepted July 2017)