Taylor & Francis
Taylor & Francis Group

Nonparametric estimation of functional dynamic factor model

Israel Martínez-Hernández ⬤[a,c], Jesús Gonzalo ⬤[b] and Graciela González-Farías[c]

[a]Department of Mathematics and Statistics, Lancaster University, Lancaster, UK; [b]Universidad Carlos III de Madrid, Getafe, Spain; [c]Centro de Investigación en Matemáticas, A.C. CIMAT, Guanajuato, Mexico

ABSTRACT

Data can be assumed to be continuous functions defined on an infinite-dimensional space for many phenomena. However, the infinite-dimensional data might be driven by a small number of latent variables. Hence, factor models are relevant for functional data. In this paper, we study functional factor models for time-dependent functional data. We propose nonparametric estimators under stationary and nonstationary processes. We obtain estimators that consider the time-dependence property. Specifically, we use the information contained in the covariances at different lags. We show that the proposed estimators are consistent. Through Monte Carlo simulations, we find that our methodology outperforms estimators based on functional principal components. We also apply our methodology to monthly yield curves. In general, the suitable integration of time-dependent information improves the estimation of the latent factors.

1. Introduction

Functional data analysis (FDA) has attracted interest in recent years in different areas in statistics, where data can be collected almost continuously, e.g. in finance, economics, climatology, medicine, and engineering. FDA is a new methodology and an interesting approach to deal with large-scale, high-dimensional and high-frequency data. Unlike the multivariate approach, which depends on the points at which data are taken, FDA can extract additional information about the continuous underlying stochastic process since each curve is treated as a unit (Ramsay and Silverman 2005). In many real applications, functional data, $\{X_n(s); s \in D, n \in \mathbb{Z}\}$, are time-dependent, e.g. yield curves (Diebold and Li 2006), mortality curves (Hyndman and Ullah 2007), electricity consumption curves (Liebl 2013), and intraday price curves (Kokoszka, Miao, and Zhang 2015). When the functional data are time-dependent, they are called functional time series (see Horváth and Kokoszka 2012, for a survey on functional time series). A way to construct a functional time series is to partition a continuous-time stochastic process $\{\mathcal{Y}_t, t \in \mathbb{R}\}$ into consecutive segments of length δ, that is, $\{X_n(s) = \mathcal{Y}_s \mathbb{1}\{s \in [n\delta, (n+1)\delta)\}, n \in \mathbb{Z}\}$, where δ depends on the dataset application (daily, monthly, annual, etc.). Some functional time series are

CONTACT Israel Martínez-Hernández ✉ i.martinezhernandez@lancaster.ac.uk, israel.m.hdz@gmail.com

naturally modelled by using factor models, e.g. yield curves (Diebold and Li 2006), electricity consumption curves (Liebl 2013), and intraday price curves (Kokoszka et al. 2015). In this paper, we consider factor models for functional time series. We propose a new methodology to obtain an estimator of the functional dynamic factor (FDF) model.

Factor models represent a large number of dependent variables from a dataset in terms of a small number of latent variables. When the data are time-dependent, attention is focussed on dynamic factor models. A dynamic factor model can explain a large fraction of the variance in many macroeconomic series (Giannone, Reichlin, and Sala 2005), and it is also consistent with broad applications to various phenomena; see Stock and Watson (1988), Bai and Ng (2002), Bai (2003), Diebold and Li (2006), Friguet, Kloareg, and Causeur (2009), Härdle and Trück (2010) and Desai and Storey (2012). Dynamic factor models have been studied for multivariate time series in both stationary and nonstationary cases. Bai and Ng (2002), Bai (2003), Forni, Hallin, Lippi, and Reichlin (2005), and Lam, Yao, and Bathia (2011) considered the stationary case. Stock and Watson (1988) studied factors in a cointegrated time series, Gonzalo and Granger (1995) proposed a method to estimate the factors in a cointegrated time series, Bai and Ng (2004) studied the factor structure of large dimensional panels in nonstationarity data, and Peña and Poncela (2006) presented a procedure to obtain dynamic factor models for a vector of time series. Here, we also study both cases of functional time series, stationary and nonstationary.

Specifically, we assumed that the functional time series $\{X_n\}$ is driven by K latent factor loadings curves $\{F_1(s), \ldots, F_K(s)\}$ and K latent factor time series $\{\beta_{n,k}\}$, i.e. $X_n(s) = \sum_{k=1}^{K} \beta_{n,k} F_k(s) + \varepsilon_n(s)$. This FDF model was studied previously by other authors. Hays, Shen, and Huang (2012) assumed that $\{X_n(s); s \in D\}$, with $D \subset \mathbb{R}$, is observed in a sample of discrete points, $s \in \{s_1, \ldots, s_m\}$, and that the factors, $\{\beta_{n,k}\}$, follow an AR(p) model. In that paper, the factor loading $F_k(s_j)$ and the components of the AR(p) model are estimated jointly via maximum likelihood and using the EM algorithm. However, the EM algorithm becomes increasingly complicated when the sample size grows or when the number of point observations of each curve grows. Here, we assume that X_n is given in functional form instead of discrete observations of each, and we propose a nonparametric estimator for the latent variables $\{\beta_{n,k}\}$ and F_k. Liebl (2013) used the FDF model to forecast electricity spot prices, where the factor loading curves, F_k, are defined as eigenfunctions of the covariance operator of X_n, and the factor process $\{\beta_{n,k}\}$ is defined as the corresponding score. One disadvantage is that functional principal component analysis (PCA) operates in a static way. That is, if the functional data $\{X_n\}$ are time-dependent, then the dynamics are not accurately represented by the principal components, as noted in Hörmann, Kidziński, and Hallin (2015). In Jungbacker, Koopman, and van der Wel (2014), the factor loading curves F_k are proposed as cubic splines and rely on hypothesis tests to select the number of knots and their locations. This is inefficient when the sample size increases. Kokoszka et al. (2015) assumed that the factor loading curves are known and depend on time. Then, they propose the use of least square estimators to obtain the factors.

In this paper, we propose a nonparametric estimator of the FDF model, considering time-dependent functional data that are either stationary or nonstationary. The factor processes $\{\beta_{n,k}\}$ are assumed to be scalar processes, and they described the dynamics of the data (by dynamics, we mean the dependence structure over time). The factor loadings $\{F_1, \ldots, F_K\}$ are assumed to be continuous functions. The subspace generated by the proposed estimators for the factor loading curves represents the dynamics of the functional

time series. Thus, our interest is in estimating the trajectory of the factor processes without assuming any model in such a way that each trajectory accurately represents the dependency over time. To take into account the temporal dependency, we consider a specific long-run covariance operator.

The remainder of our paper is organised as follows: In Section 2, we introduce mathematical concepts for functional time series. In Section 3, we describe the methodology to obtain the estimators in both cases: stationary and nonstationary models. Additionally, we present algorithms and examples to illustrate the methodology. In Section 4, we study the properties and consistency of the proposed estimators. In Section 5, we evaluate the performance of the proposed estimators under different simulation settings, comparing our results with functional PCA estimators. In Section 6, we apply our methodology to the yield curves. Finally, Section 7 presents a discussion. The proofs are provided in the Appendix.

2. Preliminaries

To describe our methodology, we first introduce some concepts for functional time series. Let $\mathcal{H} = L^2(D)$ be the separable real Hilbert space of square integrable functions defined on compact subset $D \subset \mathbb{R}$, with inner product $\langle f, g \rangle = \int_D f(x)g(x)\, dx$, and the corresponding norm denoted by $\|\cdot\|_{\mathcal{H}}$. A functional random variable is defined as a random variable in $\mathcal{H}$, i.e. $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathcal{H}$. Let $L^p_{\mathcal{H}} = \{X : \Omega \to \mathcal{H} : \mathbb{E}(\|X\|^p) < \infty\}$ be the set of random variables in $\mathcal{H}$ with finite moments of order p. The expected value of $X \in L^1_{\mathcal{H}}$ is defined as a unique element of $\mathcal{H}$, denoted by μ, such that $\mathbb{E}\langle X, y \rangle = \langle \mu, y \rangle$, for all $y \in \mathcal{H}$. In the rest of the paper, we write $\mathbb{E}(X)$ instead of μ to refer to the expected value of X.

A functional time series is a sequence of functional random variables $\{X_n, -\infty < n < \infty\}$ in $\mathcal{H}$. The covariance operator at lag h of $\{X_n\} \in L^2_{\mathcal{H}}$ is defined by $\Gamma_{X_n, X_{n+h}}(z) = \mathbb{E}[\langle X_n - \mathbb{E}(X_n), z \rangle \{X_{n+h} - \mathbb{E}(X_{n+h})\}]$ for all $z \in \mathcal{H}$. This covariance operator can be written as $\Gamma_{X_n, X_{n+h}}(z)(s) = \int \gamma_{n, n+h}(t, s) z(t)\, dt$, where $\gamma_{n, n+h}(t, s) = \text{Cov}(X_n(t), X_{n+h}(s))$ is called the kernel of $\Gamma_{X_n, X_{n+h}}$. A functional time series $\{X_n\} \in L^2_{\mathcal{H}}$ is said to be stationary if (i) $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ and (ii) $\Gamma_{X_{n+h}, X_{m+h}} = \Gamma_{X_n, X_m}$ for all n, m, and h. In this case, we use the notation Γ_h instead of $\Gamma_{X_n, X_{n+h}}$. If $\{X_n\}$ is a stationary functional time series with $\mathbb{E}(X_n) = 0$ and $\Gamma_h = 0$ for all $h \neq 0$, then it is called *functional white noise* and *strong functional white noise* if it is a sequence of i.i.d. functional random variables. In the rest of the paper, we refer to strong functional white noise as an i.i.d. sequence in $L^2_{\mathcal{H}}$. The reader can consult (Bosq 2000; Horváth and Kokoszka 2012) for more details on functional time series.

Let $\{X_n, \ n \in \mathbb{Z}\}$ be a stationary functional time series, and the long-run covariance operator Γ of X_n is defined as

$$\Gamma(z)(s) = \int_D c(t, s) z(t)\, dt, \quad z \in \mathcal{H} \tag{1}$$

where the corresponding kernel is $c(t, s) = \sum_{j=-\infty}^{\infty} \gamma_j(t, s)$, and $\gamma_j(t, s) = \text{Cov}(X_0(t), X_j(s))$. We should note that the assumption of stationarity on $\{X_n\}$ does not guarantee the existence of Γ. For that, we need an additional weak dependence condition stated in Section 4.

Let $\mathcal{B}_{\mathcal{H}} = \{A : \mathcal{H} \to \mathcal{H}; A \text{ is continuous and linear}\}$ be the set of all continuous linear operators from $\mathcal{H}$ to $\mathcal{H}$, with an operator norm denoted by $\| \cdot \|_{\mathcal{B}_{\mathcal{H}}}$. Let $A \in \mathcal{B}_{\mathcal{H}}$; an eigenfunction v of A is defined as a nonzero element of $\mathcal{H}$ such that $A(v) = \alpha v$, where $\alpha \neq 0$ is a scalar number and is the associated eigenvalue.

Let $\{\varepsilon_n\}$ be an i.i.d. sequence in $L^2_{\mathcal{H}}$, and let $\{A_n\} \in \mathcal{B}_{\mathcal{H}}$. A functional linear process $\{Y_n, n \in \mathbb{Z}\}$ with innovations $\{\varepsilon_n\}$ is defined as

$$Y_n(s) = \sum_{j=0}^{\infty} A_j(\varepsilon_{n-j})(s), \quad s \in D.$$

If the functional linear process $\{Y_n\}$ is stationary, and if its covariance long-run covariance operator exists, then this long-run covariance operator can be written as (see Appendix)

$$\Gamma = A\Gamma_{\varepsilon_0}A^*, \tag{2}$$

where Γ_{ε_0} is the covariance operator of ε_0, $A = \sum_{j=0}^{\infty} A_j$ and A^* is the adjoint operator of A.

Now, we introduce the functional $I(1)$ process. We use the approach proposed by Beare, Seo, and Seo (2017), where the notion of cointegration for multivariate time series is extended to an inifinite-dimensional space. Let $\{X_n\} \in L^2_{\mathcal{H}}$ be a functional time series such that the first difference $\Delta X_n = X_n - X_{n-1}$ admits the representation $\Delta X_n = \sum_{j=0}^{\infty} \Phi_j(\varepsilon_{n-j})$, where $\{\varepsilon_n\}$ is an i.i.d. sequence in $L^2_{\mathcal{H}}$, and coefficient operators satisfying $\sum_{j=0}^{\infty} j\|\Phi_j\|_{\mathcal{B}_{\mathcal{H}}} < \infty$. Then, X_n can be written as $X_n = X_0 + \Phi(\sum_{j=1}^{n} \varepsilon_j) + \eta_n$, where $\Phi = \sum_{j=0}^{\infty} \Phi_j$, and $\{\eta_n\}$ is a stationary functional time series. The functional time series $\{X_n\}$ is called an $I(1)$ functional process if and only if the long-run covariance operator of $\{\Delta X_n\}$ is different from zero.

3. Methodology

In this paper, we assume that the functional time series $\{X_n\}$ are given in the functional form. In a real application, functional data are observed on a grid of points, and thus, the continuous curve should be estimated (see Ramsay and Silverman 2005, chap. 3–7).

3.1. Model setting

Assume that we observe N functional data $\{X_1, \ldots, X_N\}$. We assume that the functional data follow the model

$$X_n(s) = Y_n(s) + \varepsilon_n(s),$$

$$Y_n(s) = \sum_{k=1}^{K} \beta_{n,k} F_k(s) \tag{3}$$

where F_k, $k = 1, \ldots, K$, are factor loading curves, $\{\beta_{n,k}\}$ are scalar factor time series, K is the number of factors, and $\{\varepsilon_n\}$ is a sequence of centred, independent and identically distributed innovations in $L^2_{\mathcal{H}}$ with covariance operator Γ_ε. We refer to model (3) as the

functional dynamic factor model or FDF model. The factors $\{\beta_{n,k}\}$ are assumed to be time-dependent, and then $\{X_n\}$ is time-dependent. The K factor processes drive the dynamics of the functional data $\{X_n\}$. Factors and factor loadings are latent functions and latent random variables, respectively.

For model identifiability, we assume orthonormality of the factor loading curves, that is,

$$\int_D F_i(s)F_j(s)\,\mathrm{d}s = \mathbb{1}(j = i), \tag{4}$$

for $i, j = 1, \ldots, K$, where $\mathbb{1}(j = i)$ takes the value 1 if $i = j$ and zero otherwise. In some scenarios, the factor processes $\{\beta_{n,k}\}$ can be known, and then, condition (4) can be omitted (Kokoszka, Miao, Reimherr, and Taoufik 2018).

Since factors and factor loadings are unobserved, F_k and $\{\beta_{n,k}\}$ are not uniquely determined in model (3), even with the constraint (4). However, the linear space $\mathcal{H}_F :=$ span$\{F_1, \ldots, F_K\}$ generated by the factor loadings, called the factor loading space, is uniquely defined. Thus, any orthonormal rotation of the orthonormal basis system $\{F_1, \ldots, F_K\}$ can be a solution to model (3) as well. Therefore, our goal is to estimate $\mathcal{H}_F$. To do so, we assume the following conditions.

Assumption 3.1: The functional white noise $\{\varepsilon_n\}$ is uncorrelated with the functional process $\{Y_n\}$, that is, $\Gamma_{Y_n,\varepsilon_{n+h}} = 0, h \in \mathbb{Z}$.

Assumption 3.2: There exists $i, j \in \{1, \ldots, K\}$ such that $\mathrm{Cov}(\beta_{n,i}, \beta_{n+h,j}) \neq 0$, for some $h > 0$.

Assumptions 3.1 is the usual conditions assumed for the FDF model. Assumption 3.2 requires time-dependent functional data $\{X_n\}$ and ensures $\widehat{\mathcal{H}}_F$ not being an empty set. To motivate our methodology, let us assume that the functional data $\{X_n\}$ are stationary and follow model (3) (we will use similar ideas for the nonstationary case in Section 3.3). Thus, under Assumption 3.1, the covariance operator of X_n satisfies

$$\Gamma_h = \sum_{i,j}^{K} \mathrm{Cov}(\beta_{n,i}, \beta_{n+h,j})F_i \otimes F_j + \mathbb{1}(h = 0)\Gamma_\varepsilon, \quad h \in \mathbb{Z},$$

where $\otimes$ denotes the tensor product. Therefore, $\Gamma_h(v) = 0$ for any $v \in \mathcal{H}_F^{\perp}$, with $h \neq 0$. Moreover, if $\ker(\Gamma_0) = \{z \in \mathcal{H} : \Gamma_0(z) = 0\} = \{0\}$ and Γ_0^{-1} is invertible, then $\ker(\Gamma_h\Gamma_0^{-1}) = \{\mathrm{range}(\Gamma_h^*)\}^{\perp} = \mathcal{H}_F^{\perp}$. Hence, $\mathcal{H}_F$ is the orthogonal complement of the linear space spanned by the eigenfunctions of $\Gamma_h\Gamma_0^{-1}$ (or Γ_h) corresponding to the zero eigenvalues, with $h \neq 0$. Thus, our methodology uses the operator $\Gamma_h\Gamma_0^{-1}$, and we consider the summation over all possible lags $h \neq 0$, that is,

$$\Lambda := \sum_{h \in \mathbb{Z}, h \neq 0} \Gamma_h\Gamma_0^{-1} = (\Gamma - \Gamma_0)\Gamma_0^{-1}, \tag{5}$$

where Γ is the long-run covariance of X_n defined in (1). We notice that, under Assumption 3.2, range$(\Lambda) \neq \emptyset$.

We observe that $\Lambda(v) = 0$ for any $v \in \mathcal{H}_F^\perp$ as well. Consequently, we propose to estimate the space $\mathcal{H}_F$ using the eigenfunctions of the operator Λ corresponding to nonzero eigenvalues. Additionally, these eigenfunctions are defined as estimators of the factor loadings $F_k, k = 1, \ldots, K$.

In general, Γ_0^{-1} is unbounded (Bosq 2000; Martínez-Hernández, Genton, and González-Farías 2019). The method typically employed to address this problem is to truncate the spectral representation of Γ_0 and then compute the inverse from this representation. We adopt this method, and we describe it in Section 3.2.

Remark 3.1: The consideration of the term Γ_0^{-1} in (5) is equivalent to assuming that the data have identity as the covariance operator. The latter case requires transforming the data X_n to $\Gamma_0^{-1/2}(X_n)$. Thus, $\Gamma_h \Gamma_0^{-1}$ is the covariance operator at lag h of the transformed data.

3.2. Estimation for stationary FDF model

Without loss of generality, we assume that $\mathbb{E}(X_n) = 0$. To estimate $\mathcal{H}_F$, we compute the eigenfunctions of $\widehat{\Lambda}$, where $\widehat{\Lambda}$ is the estimator of Λ. To obtain $\widehat{\Lambda}$, we use a smooth periodogram-type estimator (Rice and Shang 2017). Let $\widehat{\gamma}_h(t,s)$ be the kernel estimator of Γ_h defined as

$$
\widehat{\gamma}_h(t,s) = \begin{cases} \dfrac{1}{N} \sum_{i=1}^{N-h} X_i(t) X_{i+h}(s), & h \geq 0 \\ \dfrac{1}{N} \sum_{i=1-h}^{N} X_i(t) X_{i+h}(s), & h < 0 \end{cases}
$$

Now, we define $\widehat{c}(t,s) - \widehat{\gamma}_0(t,s)$ as

$$
\widehat{c}_{b,-0}(t,s) = \sum_{|h| \leq b,\, h \neq 0} \chi\left(\frac{h}{b}\right) \widehat{\gamma}_h(t,s),
$$

where χ is a continuous, symmetric weight function and satisfies $\chi(0) = 1$, $\chi(u) = 0$ if $|u| > c$ for some $c > 0$. In this paper, we use $\chi(h/b) = 1 - |h|/b$, where b is a bandwidth parameter. To obtain a consistent estimator, the bandwidth must satisfy $b = b(N) \to \infty$ as $N \to \infty$ and $b(N)/N = o(1)$ (Horváth, Kokoszka, and Reeder 2013). However, in practice, the selection of b should be done carefully, since this can affect the performance of the estimator in finite samples. Here, we select b similarly as described in Rice and Shang (2017), that is, minimising the mean-squared error $\mathbb{E}\|\widehat{c}_b - c\|^2$, where $\|\cdot\|$ denotes the usual norm in $L^2([0,1]^2)$. We use the notation $\widehat{c}_{-0}(t,s)$ to denote $\widehat{c}_{b,-0}(t,s)$ with the optimal value of the bandwidth b.

Then, the estimator of the operator $\Gamma - \Gamma_0$ is defined by $\widehat{\Gamma - \Gamma_0}(z)(s) = \int \widehat{c}_{-0}(t,s) z(t)\, dt$, $z \in \mathcal{H}$. To obtain a bounded estimator of Γ_0^{-1}, we only consider the first p eigenfunctions corresponding to the largest eigenvalues of Γ_0. Thus, the inverse $\widehat{\Gamma}_0^{-1}$ is approximated by $\widehat{\Gamma}_0^{-1} = \sum_{j=1}^{p} \widehat{\lambda}_j^{-1} \widehat{v}_j \otimes \widehat{v}_j$, where $\widehat{v}_j, j = 1, \ldots, p$, are such that $\widehat{\Gamma}_0(v_j) = \widehat{\lambda}_j \widehat{v}_j$. We select the parameter $p = p(N)$ using the scree plot.

Finally, we obtain an estimator of Λ as $\widehat{\Lambda} = \widehat{\Gamma - \Gamma_0}\widehat{\Gamma}_0^{-1}$, and consecutively, we estimate the eigenfunctions $\{\widehat{\zeta}_1, \ldots, \widehat{\zeta}_{k_0}\}$ of $\widehat{\Lambda}$ corresponding to the first k_0 largest nonzero

eigenvalues $\{\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_{k_0}\}$, with k_0 a positive number. That is, for each $i = 1, \ldots, k_0$, $\widehat{\zeta}_i$ satisfies

$$\widehat{\Lambda}\widehat{\zeta}_i = \widehat{\alpha}_i\widehat{\zeta}_i.$$

As mentioned above, we have that K is the number of nonzero eigenvalues of Λ. In practice, the number of eigenfunctions of $\widehat{\Lambda}$ with nonzero eigenvalues is not exactly K, since the zero-eigenvalues of $\widehat{\Lambda}$ are unlikely to be zero exactly. Here, we estimate the number of factors similarly as in Lam and Yao (2012), using a ratio-based estimator. Explicitly, we define $\widehat{K}$ as

$$\widehat{K} = \arg \min_{1 \leq i \leq k_0} \widehat{\alpha}_{i+1}/\widehat{\alpha}_i, \tag{6}$$

where k_0 should be large enough that $K < k_0$. Although we do not study the theoretical properties of $\widehat{K}$, thorough empirical investigations suggest it works well in practice.

Once we have estimated the eigenfunctions of Λ and the number of factors, we define $\widehat{\mathcal{H}}_F = \mathrm{span}\{\widehat{\zeta}_1, \ldots, \widehat{\zeta}_{\widehat{K}}\}$, and $\widehat{F}_i = \widehat{\zeta}_i$, $i = 1, \ldots, \widehat{K}$. The estimated trajectories of the factor processes are obtained as

$$\{\widehat{\beta}_{n,k}\}_{n=1}^N = \{\langle X_n, \widehat{F}_k \rangle\}_{n=1}^N, \quad k = 1, \ldots, \widehat{K}.$$

Algorithm 1 presents a summary of the steps to obtain the estimators of model (3) under the stationary assumption.

Algorithm 1 Estimators for stationary FDF model

1: Fix k_0 with k_0 large enough.
2: Compute the estimator $\widehat{\Lambda}$.
3: Compute the eigenfunctions $\widehat{\zeta}_1, \ldots, \widehat{\zeta}_{k_0}$ of $\widehat{\Lambda}$, corresponding to the first k_0 largest eigenvalues.
4: Obtain $\widehat{K}$ as in (6).
5: For $k = 1, \ldots, \widehat{K}$, estimate the factor loadings and the trajectories of the factors as follows: $\widehat{F}_k = \widehat{\zeta}_k$, and $\{\widehat{\beta}_{n,k}, \, n \geq 1\} = \{\langle X_n, \widehat{\zeta}_k \rangle, \, n \geq 1\}$.

Example 3.1: We simulate $\{X_n\}$ from an FDF model with $K = 2$ and sample size $N = 200$. The factors are AR(1) processes with coefficients $a_{1,1} = 0.6$ and $a_{2,1} = -0.6$. The factor loadings are defined by $F_1(s) = \sin(2\pi s)$ and $F_2(s) = \cos(2\pi s)$ with $s \in [0, 1]$. Figure 1 shows the estimators obtained from Algorithm 1. The left panel shows the factor loadings and their corresponding estimators, while the centre and right panels show the estimated trajectory for the factors. We observe that $\widehat{F}_i$ is close to the original factor loadings and that the estimated trajectories of the factor processes successfully approximately the original factor processes. Thus, the proposed estimators have good performance in this example. A more exhaustive simulation study is performed in Section 5.

3.3. Estimation for the nonstationary FDF model

In real data applications, the factor processes can be nonstationary. Here, we assume that at least one of the factor processes $\{\beta_{n,k}\}$, $k = 1, \ldots, K$, is an $I(1)$ scalar process. Let

Figure 1. Estimators of $F_1(s)$ and $F_2(s)$ for the stationary FDF model with $K = 2$ and the estimated trajectories for the factor time series $\{\beta_{n,1}\}$ and $\{\beta_{n,2}\}$. The sample size is $N = 200$.

$r \in \{1, \ldots, K\}$ be the number of factor processes that are $I(1)$ processes. We consider two cases: (i) all factor processes are $I(1)$ processes, i.e. $r = K$, and (ii) there are r $I(1)$ factor processes, with $r < K$, and the remaining $K-r$ factor processes are stationary processes. Thus, the factor loading space can be written as $\mathcal{H}_F = \mathcal{H}_F^S \oplus \mathcal{H}_F^N$, where $\mathcal{H}_F^S$ is the linear space generated by the stationary factor processes and $\mathcal{H}_F^N$ is the linear space generated by the nonstationary factor processes. We have that if $z \in \mathcal{H}_F^S$, then the scalar time series $\{\langle X_n, z \rangle\}_n$ is stationary, and if $z \in \mathcal{H}_F^N$, the scalar time series $\{\langle X_n, z \rangle\}_n$ is nonstationary. Case (ii) is related to the concept of cointegrating in a Hilbert space (Beare et al. 2017). Here, we describe estimators for $\mathcal{H}_F^S$ and $\mathcal{H}_F^N$.

Let $\Lambda_{\Delta X}$ be the operator defined in (5) for the functional process ΔX_n. In the following, we describe the two cases for the factor processes.

Case (i): In this case, $\mathcal{H}_F = \mathcal{H}_F^N$, and ΔX_n is a stationary FDF model. We estimate the space $\mathcal{H}_F^N$ as the linear space generated by the eigenfunctions of the operator $\Lambda_{\Delta X}$, where $\Lambda_{\Delta X}$ is estimated as described in Section 3.2 using $\{\Delta X_n, n \geq 1\}$ instead of $\{X_n, n \geq 1\}$. The number of factors is estimated as in (6) with the corresponding eigenvalues of $\widehat{\Lambda}_{\Delta X}$. Finally, we define $\widehat{F}_k = \widehat{\xi}_k$, $k = 1, \ldots, \widehat{K}$, where $\widehat{\xi}_k$ are the eigenfunctions of $\widehat{\Lambda}_{\Delta X}$. This approach guarantees that $\{\widehat{\beta}_{n,k}, \; n \geq 0\} = \{\langle X_n, \widehat{\xi}_k \rangle, \; n \geq 0\}$ is an $I(1)$ process for $k = 1, \ldots, \widehat{K}$ (Proposition 4.3).

Case (ii): We have that $\mathcal{H}_F = \mathcal{H}_F^S \oplus \mathcal{H}_F^N$ with $\mathcal{H}_F^S \neq \emptyset$. First, we estimate the space $\mathcal{H}_F^N$; then, we subtract the estimated space $\widehat{\mathcal{H}}_F^N$ from the entire space $\mathcal{H}$. Then, we estimate $\mathcal{H}_F^S$. We propose to estimate $\mathcal{H}_F^N$ as the linear space generated by the eigenfunctions of $\Lambda_{\Delta X}$, and then we define $\widehat{F}_k = \widehat{\xi}_k$, where $\widehat{\xi}_k$ is the eigenfunction of $\widehat{\Lambda}_{\Delta X}$, for $k = 1, \ldots, \widehat{r}$. Here, we obtain $\widehat{r}$ using criteria (6), with $\widehat{\alpha}_i's$ being eigenvalues of $\widehat{\Lambda}_{\Delta X}$. Then, we obtain the estimated trajectories of the factor processes as $\{\widehat{\beta}_{n,k}, \; n \geq 0\} = \{\langle X_n, \widehat{\xi}_k \rangle, \; n \geq 0\}$, for $k = 1, \ldots, \widehat{r}$. Given $\{\widehat{F}_1, \ldots, \widehat{F}_{\widehat{r}}\}$, we estimate $\mathcal{H}_F^S$ as follows. We define a new functional time series as

$$Z_n(s) = X_n(s) - \sum_{k=1}^{\widehat{r}} \widehat{\beta}_{n,k} \widehat{F}_k(s). \tag{7}$$

Let Λ_Z be the corresponding operator defined in (5) for the functional process $\{Z_n\}$. The functional process $\{Z_n\}$ should be a stationary functional process. Thus, we estimate $\mathcal{H}_F^S$ similarly as in Section 3.2, with $\widehat{K-r}$ as in (6) using the corresponding eigenvalues of $\widehat{\Lambda}_Z$. Let $\{\widehat{\upsilon}_1, \ldots, \widehat{\upsilon}_{\widehat{K-r}}\}$ be the eigenfunctions of $\widehat{\Lambda}_Z$ corresponding to the ith largest

eigenvalues. Then, we define $\widehat{F}_{\widehat{r}+1} = \widehat{\upsilon}_1, \ldots, \widehat{F}_{\widehat{K}} = \widehat{\upsilon}_{\widehat{K-r}}$, with $\widehat{K} = \widehat{r} + \widehat{K-r}$, as the estimators of the factor loadings corresponding to the stationary dynamic and $\{\widehat{\beta}_{n,k}, \ n \geq 0\} = \{\langle X_n, \widehat{\upsilon}_{k-r}\rangle, \ n \geq 0\}$, for $k = r + 1, \ldots, K$, as the estimated trajectories of the $I(0)$ factor processes.

In real applications, we do not know whether $r < K$ or $r = K$. To overcome this problem, we propose to apply a test for independence for the functional process $\{Z_n, n \geq 1\}$, such as a test proposed in Horváth, Hušková, and Rice (2013). If the independency hypothesis is not rejected, then we conclude that $r = K$. Otherwise, $r < K$, and we can proceed to estimate the corresponding stationary loading space $\mathcal{H}_F^S$.

In Algorithm 2, we present a summary of the steps to obtain the estimators of model (3) with r nonstationary factor processes and $K-r$ stationary factor processes.

Algorithm 2 Estimators for nonstationary FDF model

1: Apply Algorithm 1 by considering $\Lambda_{\Delta X}$, and define $\widehat{r} = \widehat{K}_{\Delta X}$, where $\widehat{K}_{\Delta X}$ is the number of factors estimated with eigenvalues of $\widehat{\Lambda}_{\Delta X}$.

2: Save the estimated loading factors and factor processes, $\{\widehat{F}_k\}$ and $\{\widehat{\beta}_{n,k}\}$, $k = 1, \ldots, \widehat{r}$.

3: Obtain the functional process $\{Z_n\}$ as defined in (7).

4: Apply a test of independence for the functional process $\{Z_n\}$.

5: If the test of the independency hypothesis is not rejected, then stop.

6: Else: Compute $\widehat{\Lambda}_Z$ the estimator of the operator Λ_Z corresponding to the functional process (7).

7: Apply Algorithm (1) to $\widehat{\Lambda}_Z$, and define $\widehat{K-r} = \widehat{K}_Z$, where $\widehat{K}_Z$ is the number of factors estimated with eigenvalues of $\widehat{\Lambda}_Z$.

8: Finally, obtain the complete estimators of loading factors and factor processes, $\{\widehat{F}_1, \ldots, \widehat{F}_{\widehat{r}}, \widehat{F}_{\widehat{r}+1}, \ldots, \widehat{F}_{\widehat{K}}\}$ and $\{\widehat{\beta}_{n,1}, \ldots, \widehat{\beta}_{n,\widehat{r}}, \widehat{\beta}_{n,\widehat{r}+1}, \ldots, \widehat{\beta}_{n,\widehat{K}}\}$, where $\widehat{K} = \widehat{r} + \widehat{K-r}$.

Example 3.2: We simulate $\{X_n\}$ from the FDF model with $K = 2$, $r = 1$, and sample size $N = 200$, where $F_1(s) = \sin(2\pi s)$, $F_2(s) = \cos(2\pi s)$, $\Delta\beta_{n,1} = 0.7\Delta\beta_{n-1,1} + u_{n,1}$, and $\beta_{n,2} = 0.7\beta_{n-1,2} + u_{n,2}$, with $s \in [0, 1]$. Figure 2 shows the estimators obtained from Algorithm 2. The left panel shows the factor loadings and their corresponding estimators, while the centre and the right panels show the trajectory estimated for the factors. Similar to Example 3.1, we observe that the proposed estimators have good performance in this nonstationary example. A more exhaustive simulation study is performed in Section 5.

Remark 3.2: As we mentioned before, the estimators of the FDF model are not unique, but the subspace $\mathcal{H}_F$ generated by the factor loadings is unique. Here, we defined $\widehat{F}_k$ as the eigenfunctions of the corresponding long-run covariance operator. However, users can choose an appropriate orthonormal rotation to obtain factors that allow a meaningful-interpretation. For example, one possible rotation is the well-known VARIMAX rotation.

The dynamic modelling for the FDF model is attained by modelling the factor processes $\{\widehat{\beta}_{n,k}\}$ and using the relationship $\widehat{X}_n(s) = \sum_{k=1}^{\widehat{K}} \widehat{\beta}_{n,k}\widehat{F}_k(s)$.

In general, the operator Λ defined in (5) represents the loading space and the dynamic of the functional time series. Moreover, if the functional time series is an $I(1)$ functional

Figure 2. Estimators of $F_1(s)$ and $F_2(s)$ for the nonstationary FDF model and the estimated trajectory for the factor time series $\{\beta_{n,1}\}$ and $\{\beta_{n,2}\}$. In this example, $K = 2$, $r = 1$, and the sample size is $N = 200$.

process, then the operator Λ classifies the space of common trends and the cointegrating space. We conclude that the dynamics of the FDF model over time are accurately described by using the space generated by this operator.

4. Estimation properties

In this section, we study the properties of the estimators described in Section 3. For this, we assume that X_n has a functional linear representation in $L^2_{\mathcal{H}}$. Without loss of generality, we assume that the curves are defined on the unit interval $D = [0, 1]$ and $\mathbb{E}(X_n) = 0$.

Assumption 4.1: The functional time series $\{X_n\}$ has a linear representation $X_n = \sum_{j=0}^{\infty} \Psi_j(\epsilon_n)$, with $\Psi_j \in \mathcal{B}_{\mathcal{H}}$, and $\sum_{m=1}^{\infty} \sum_{j=m}^{\infty} \|\Psi_j\| < \infty$.

Assumption 4.2: $\int \mathbb{E}\{X_0^2\}(s)\,ds < \infty$ and $\lim_{m\to\infty} m(\mathbb{E}[\int \{X_n(s) - X_{n,m}(s)\}^2\,ds])^{1/2} = 0$, where $X_{n,m} = \sum_{j=0}^{m-1} \Psi_j(\epsilon_n) + \sum_{j=m}^{\infty} \Psi_j(\epsilon_n^{(m)})$, and for each m, $\{\epsilon_k^{(m)}\}$ is an independent copy of $\{\epsilon_k\}$.

Assumption 4.1 does not impose any restrictions on the model (3). For example, if the factors processes are (scalar) linear processes, $\beta_{n,k} = \sum_{j=0}^{\infty} \theta_{j,k} u_{n-j,k}$ with $\{u_n\}$ an i.i.d. sequence, then $X_n(s) = \sum_{k=1}^{K} \sum_{j=0}^{\infty} \theta_{j,k} u_{n-j,k} F_k(s) + \varepsilon_n(s)$. This latter expression can be rewritten as $X_n(s) = \sum_{j=0}^{\infty} \Psi_j(\epsilon_n)(s)$, with $\epsilon_n := \sum_{k=1}^{K} u_{n,k} F_k + \varepsilon_n(s)$ and $\Psi_j := \sum_{k=1}^{K} \theta_{j,k} F_k \otimes F_k$. In this case, the condition $\sum_{m=1}^{\infty} \sum_{j=m}^{\infty} \|\Psi_j\| < \infty$ is equivalent to $\sum_{j=0}^{\infty} j|\theta_{j,k}| < \infty$. This latter condition is a common assumption on scalar time series in order to obtain basic results.

Proposition 4.1: *Let $\{X_n, \ n \in \mathbb{Z}\}$ be a functional time series following model* (3)*, and let $\widehat{\Lambda}$ be the estimator of Λ described in Section 3, with kernels $\widehat{g}(s,t)$ and $g(s,t)$, respectively. Under Assumptions 3.1, 3.2, 4.1, and 4.2, $\int\int \{\widehat{g}(s,t) - g(s,t)\}^2\,ds\,dt \xrightarrow{\mathbb{P}} 0$, as $N \to \infty$ and $p \to \infty$.*

Proof: See Appendix. ∎

Assumption 4.1 implies that $\{X_n\}$ is a stationary functional time series. Thus, Proposition 4.1 corresponds to the stationary case of model (3).

As we noted above, the factors $\{\beta_{n,k}\}$ and factor loadings F_k are not uniquely determined in the FDF model. However, the factor loading space is uniquely determined. We only study the theoretical properties of the factor loading space under the assumption that K is known. Here, we say that a subspace $\mathcal{H}_n$ converges to a subspace $\mathcal{H}_0$ if $\|\Pi_n - \Pi_0\| \to 0$, where Π_n and Π_0 are the orthonormal projectors on $\mathcal{H}_n$ and $\mathcal{H}_0$, respectively, and $\|\cdot\|$ an operator norm. We use the notation $\mathcal{H}_n \to \mathcal{H}_0$ to indicate this convergence of a subspace.

Corollary 4.2: *Let $\{X_n,\ n \in \mathbb{Z}\}$ be a functional time series following model* (3) *with K known. Suppose that Γ_0 is invertible. Then, under Assumptions* 3.1, 3.2, 4.1 *and* 4.2,

$$\widehat{\mathcal{H}}_F \xrightarrow{\mathbb{P}} \mathcal{H}_F.$$

To study the nonstationary case, we assume that ΔX_n admits a functional linear representation as well.

Assumption 4.3: The functional time series $\{X_n,\ n \in \mathbb{Z}\} \in L^2_{\mathcal{H}}$ is an $I(1)$ functional process so that $\{\Delta X_n,\ n \geq 1\}$ is a stationary process and admits the representation

$$\Delta X_n = \sum_{j=0}^{\infty} \Phi_j(\epsilon_{n-j}), \quad n \geq 1, \tag{8}$$

with $\Phi_j \in \mathcal{B}_{\mathcal{H}}$, $\Phi_{j_0} \neq 0$ for some $j_0 > 0$, and the covariance operator of ϵ_0, Γ_{ϵ_0}, is positive definite. Furthermore, Assumptions 4.1 and 4.2 are satisfied on the functional process $\{\Delta X_n,\ n \geq 1\}$.

Assumption 4.3 does not impose any restrictions on the model (3). If any factor process $\{\beta_{n,k},\ \in \mathbb{Z}\}$ is an $I(1)$ process, then the functional time series that follows model (3) can always be written as (8), with Φ_j compact and self-adjoint operators. The condition $\Phi_{j_0} \neq 0$ for some $j_0 > 0$ guarantees that ΔX_n is not a functional white noise. Consequently, X_n admits a solution of the form $X_n = X_0 + \Phi(\sum_{i=1}^{n} \varepsilon_n) + \nu_n$, where ν_n is a stationary functional time series, $\{\varepsilon_n\}$ is an i.i.d. sequence, and $\Phi = \sum_{j \geq 0} \Phi_j$.

From Proposition 4.1, we have that $\widehat{\Lambda}_{\Delta X}$ is a consistent estimator of $\Lambda_{\Delta X}$.

Proposition 4.3: *Let $\{X_n,\ n \in \mathbb{Z}\}$ be a nonstationary functional time series such that it follows model* (3). *Then, under Assumptions* 3.1 *and* 4.3 *we have that $\mathcal{H}_F^N = \{\ker(\Lambda_{\Delta X})\}^{\perp}$, and $\mathcal{H}_F^S = \ker(\Lambda_{\Delta X})$.*

Proof: See Appendix. ∎

Proposition 4.3 shows that the $I(1)$ dynamics of the factors are recovered in the subspace generated by nonzero elements of Im $\Lambda_{\Delta X}$.

Corollary 4.4: *Let Γ_0 be invertible, and K and r known. If the assumptions in Proposition* 4.3 *hold, then $\widehat{\mathcal{H}}_F^S \xrightarrow{\mathbb{P}} \mathcal{H}_F^S$ and $\widehat{\mathcal{H}}_F^N \xrightarrow{\mathbb{P}} \mathcal{H}_F^N$.*

Corollary 4.5: *Let $\{\widehat{\beta}_{n,k}\}$ be the estimated trajectory obtained from Algorithm* 2 *with r and K known. Then, if the assumptions in Proposition* 4.3 *hold, we have that, for $k = 1, \ldots, r$,*

$\{\widehat{\beta}_{n,k}, n \in \mathbb{Z}\}$ *is an* $I(1)$ *process, and if* $r < K$, *for* $k = r+1, \ldots, K$, $\{\widehat{\beta}_{n,k}, n \in \mathbb{Z}\}$ *is an* $I(0)$ *process.*

Corollaries 4.4 and 4.5 show that the loading space is consistently estimated for both stationary and nonstationary components.

5. Simulation study

We study the performance of our proposed estimators for the FDF model. We compare our results with estimators obtained from functional PCA. To obtain the functional PCA estimators, we replace $\widehat{F}_k$ with $\widehat{F}_k^{\mathrm{PCA}} = \widehat{\varsigma}_k$ in Algorithm 1, where $\widehat{\varsigma}_k$ denotes the kth eigenfunction of $\widehat{\Gamma}_0$ corresponding to the kth largest eigenvalue (Liebl 2013). Similarly, in Algorithm 2, we consider the covariance operators at lag zero of $\{\Delta X_n\}$ and $\{Z_n\}$ instead of $\widehat{\Lambda}_{\Delta X}$ and $\widehat{\Lambda}_Z$. The PCA estimators are known to perform well when observations are uncorrelated.

5.1. Simulation setting

We simulate $\{X_n(s); \ s \in [0,1], \ n = 1 \ldots, N\}$ from the FDF model with four different models defined as follows:

(1) Model 1: $K = 1$, $F_1(s) = \sin(2\pi s)$, and $\beta_{n,1} = 0.7\beta_{n-1} + u_{n,1}$.
(2) Model 2: $K = 2$, $F_1(s) = \sin(2\pi s)$, $F_2(s) = \cos(2\pi s)$, $\beta_{n,1} = 0.8\beta_{n-1,1} + u_{n,1}$, and $\beta_{n,2} = -0.5\beta_{n-1,2} + u_{n,2}$.
(3) Model 3: $K = 1$, $r = K$, $F_1(s) = \sin(2\pi s)$, and $\Delta\beta_{n,1} = 0.5\beta_{n-1,1} + u_{n,1}$.
(4) Model 4: $K = 2$, $r = 1$, $F_1(s) = \sin(2\pi s)$, $F_2(s) = \cos(2\pi s)$, $\Delta\beta_{n,1} = 0.7\beta_{n-1,1} + u_{n,1}$, and $\beta_{n,2} = 0.5\beta_{n-1,2} + u_{n,2}$.

The processes $\{u_{n,i}\}$ are scalar white noises. In all cases, the functional white noise $\varepsilon_n(s)$ in model (3) is simulated as $\varepsilon_n(s) = W_n(s)$, where $W_n(s)$ is a Brownian motion in $[0,1]$. Models (1) and (2) represent stationary FDF models, while Models (3) and (4) represent nonstationary FDF models.

We evaluate the performance by computing the Integrated Squared Error (ISE) value defined as:

$$\mathrm{ISE}_F = \int_0^1 \left\{ F(s) - \widehat{F}(s) \right\}^2 \mathrm{d}s.$$

We consider sample sizes $N = 200, 300, 500$, and 1000. We simulate 1000 replications for each model. For each replication, we estimate the factor loadings and factor processes. Then, we compute the error value ISE_F, both for our proposed estimators and for PCA estimators.

Additionally, we report the estimated factor number $\widehat{K}$ for each simulation using (6).

5.2. Simulation results

We denote by FDF the simulation results when considering our nonparametric estimators. First, we describe the results for ISE_F values. In doing so, we assume that the true number
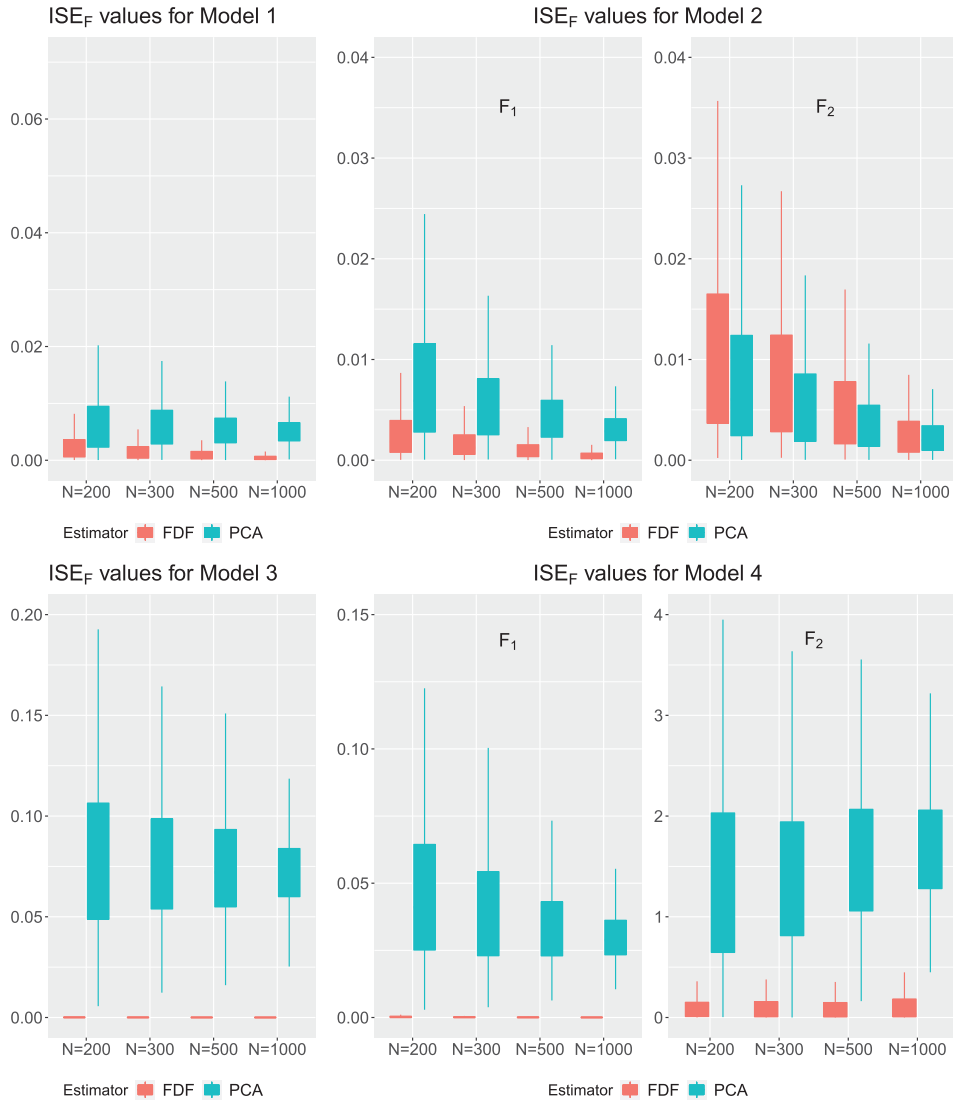
Figure 3. Boxplots of ISE_F values for each simulation. In all cases, our proposed estimators (in red) outperform the PCA estimators. Although for F_2 in Model (2), the PCA estimator performs as well as our estimator.

of factors is known. Figure 3 shows boxplots of the results. Each plot presents the ISE_F values for a specific F_i with different sample sizes. For Models (1) and (3), we have only one loading factor, and for Models (2) and (4), we have two loading factors, F_1 and F_2.

For Model (1) (Figure 3 top left), we observe that our proposed estimator is highly accurate and presents lower error values than the PCA estimator. These error values decrease when the sample size increases. For Model (2), we observe that for the first loading factor F_1 (Figure 3 top centre), the results are similar to the results of Model (1). Our method outperforms the PCA estimator. For F_2 (Figure 3 top right), we see that the PCA method performs
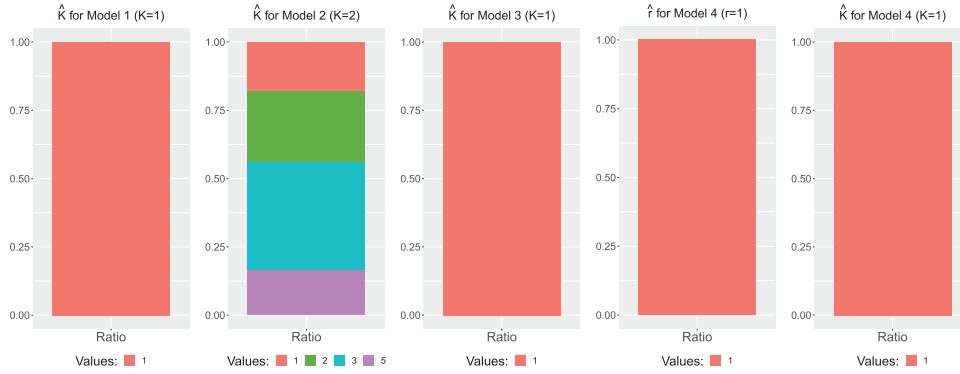
Figure 4. Values and proportions of the number of factors estimated for sample size $N = 300$, and 1000 replications. The ratio method provides a good estimator for K.

as well as our method. Thus, we conclude that, in general, our method outperforms the PCA estimator for the stationary Models (1) and (2).

Now, we analyse the nonstationary models (Models (3) and (4)). We observe that in all cases, our method has the lowest error values. The ISE_F values for our method remain as accurate as in the stationary cases. However, the ISE_F values for the PCA method become significantly larger. We conclude that our proposed method performs well in all simulated cases and outperforms the PCA method.

Finally, we show the results corresponding to the estimation of the number of factors. For each replication, we estimate the number of factors using (6). We only present the results from sample size $N = 300$, since the results from the sample sizes $N = 200, 500$, and 1000 are similar. Figure 4 shows the estimated values $\widehat{K}$ over the 1000 replications. Each bar uses colours to represent the proportion of replications.

For Model (1), we observe that with the ratio method, we obtain $\widehat{K} = 1$ for all replications. For this Model (1), we conclude that the ratio method correctly estimates the value of K. For Model (2), $\widehat{K}$ takes values in $\{1, 2, 3, 5\}$, resulting in $\widehat{K} = 2$ and $\widehat{K} = 3$ with a large proportion. For Model (3), the ratio method successfully estimates the value of K. For Model (4), we need to estimate r and $K-r$, where r is the number of nonstationary factor processes and $K-r$ is the number of stationary factor processes. In this case, we observe that the ratio method provides the correct values of r and $K-r$.

In general, the ratio method correctly estimates the number of factors with the exception of Model (2).

We conclude that our proposed methodology performs well and is superior to the PCA method under time dependence.

6. Data application

We fit the FDF model with our proposed nonparametric estimators to analyse the Federal Reserve interest rates. Then, we study the estimated factor loadings and the trajectories of the factor processes.
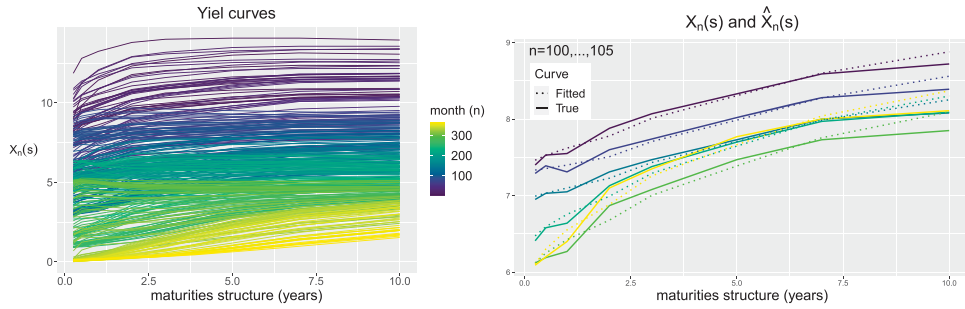
Figure 5. Right: Monthly curves of yield data from June 1982 to May 2012. Left: six consecutive yield curves from the dataset (solid curves) and the corresponding five yield curves obtained from the FDF model with the estimators (dashed curves), that is, $\widehat{X}_n(s) = \sum_{k=1}^{3} \widehat{F}_k(s)\widehat{\beta}_{n,k}$.
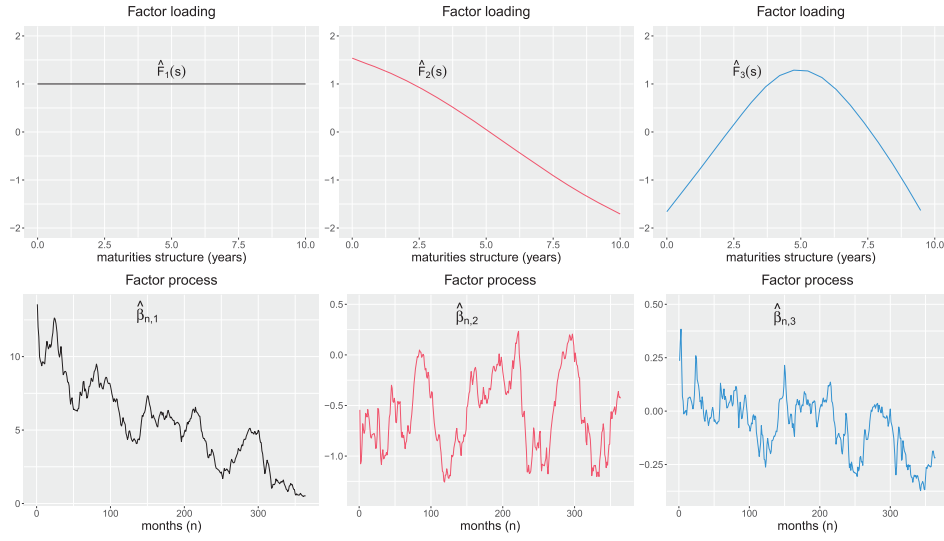
6.1. Yield curve

The interest rate data from the Federal Reserve are available in the R package *YieldCurve*. The dataset represents monthly yield data from June 1982 to May 2012 at different maturities: 3 months, 6 months, 1 year, 2 years, 3 years, 5 years, 7 years and 10 years (see Figure 5). Yield curves are important in economics and finance and can help to determine the current and future position of the economy. The Nelson-Siegel parametrization is commonly used to describe yield curves (see Nelson and Siegel 1987; Diebold and Li 2006; Koopman, Mallee, and der Wel 2010). Our proposal is a nonparametric estimator and can be considered as alternative modelling to the Nelson-Siegel modelling. To justify the functional approach, we assume that the dataset is continuous on maturities. That is, a curve $X_n(s)$ represents interest rates in month n with time to maturity s. To estimate the continuous curves, we fit 15 cubic B-spline basis functions for each monthly observation, i.e. $X_n(s) = \sum_{r=1}^{15} b_{r,n}\phi_r(s)$, where $\{\phi_1, \ldots, \phi_{15}\}$ is the B-spline basis function, with $n = 1, \ldots, 366$.

Yield curves were analysed with a nonparametric approach and with a functional approach. Hays et al. (2012) used a functional approach combined with an EM algorithm to jointly estimate the factor loading curves and the factors. Their approach is difficult to apply if more data are observed at different maturities and almost impossible to apply if the functional data are observed in a dense set. In contrast, our estimators are easy to implement and can be applied to either dense or sparse observations and for stationary and nonstationary functional time series.

We are interested in studying the factor loading curves and the factor processes that drive the interest rates at different maturities by taking into account the dependence structure of the functional time series. To infer the stationarity of the functional time series, we apply a test proposed by Horváth, Kokoszka, and Rice (2014). The p-value of the test, with 15 basis functions, is equal to 0.001, and the smaller the p-value, the more evidence there is against stationarity. Therefore, we follow Algorithm 1 with three factors, and $r = 1$.

Figure 6 shows the results. In the first row, we plot the estimator of factor loading curves, $\widehat{F}_1, \widehat{F}_2$, and $\widehat{F}_3$, and in the second row, we plot the trajectories of the factor processes estimated, $\{\widehat{\beta}_{n,k}, 1 \le n \le 366\}$. Since the factor loading curves are time-invariant,

Figure 6. Estimator of factor loading curves, $\widehat{F}_1, \widehat{F}_2$, and $\widehat{F}_3$, and the corresponding trajectories of the factor processes estimated, $\{\widehat{\beta}_{n,k}, 1 \leq n \leq 366\}$.

they represent the common properties of the yield curves, while the factor processes represent the dynamics over time. Our estimators agree with the level, slope and curvature functions of the Nelson-Siegel parameterisation described by Diebold and Li (2006), which is interesting since we do not assume any specific model (see Appendix for a plot of Nelson-Siegel curves). Therefore, the first factor loading curve F_1 is the level, and it is associated with the long-term factor; these dynamics are described by $\{\beta_{n,1}\}$. The factor loading F_2 is the slope, and it is associated with the short-term factor that is represented by the process $\{\beta_{n,2}\}$. Finally, the factor loading F_3 is the curvature that corresponds to the medium-term factor, and the factor $\{\beta_{n,3}\}$ describes such dynamics.

Figure 5 (right) shows six consecutive yield curves from the dataset with the corresponding six yield curves obtained from the FDF model with the estimators, that is, $\widehat{X}_n(s) = \sum_{k=1}^{3} \widehat{F}_k(s)\widehat{\beta}_{n,k}$. These six curves correspond to the months of October, November, December, January, February, and March of 1990 and 1991, respectively. We observe that the yield curves seem to be accurately represented by the FDF model.

7. Discussion

We propose new nonparametric estimators of the functional dynamic factor model, taking into account the time dependence of the functional data. We use the long-run covariance operator to define a subspace of the continuous functions where the dynamics of the functional data are properly described by the factor processes. We have showed that the proposed estimators of the factor loading curves represent a subspace where factor processes describe the dependence of the functional time series, under both the stationarity and nonstationarity assumptions. We compared our proposed estimators with eigenfunctions of the covariance operator. From the simulation study, we conclude that our proposed estimators have better performance than PCA-based estimators.

From a mathematical point of view, factor loading curves can be considered part of basis functions for the Hilbert space. In principle, factor loading curves could be, for example, eigenfunctions, but eigenvalues do not take into account the time-dependent structure and therefore do not represent the dynamics over time. The ideas developed here can be extended to multivariate functional time series for studying the common factors among the different functional time series and the factors of each functional time series that are not shared. The extended model might be relevant in many applications, such as multi-economy yield curves.

Acknowledgments

Disclosure statement

Funding

ORCID

Israel Martínez-Hernández 🔍 http://orcid.org/0000-0002-4122-2529
Jesús Gonzalo 🔍 http://orcid.org/0000-0002-4309-5025

References

Bai, J. (2003), 'Inferential Theory for Factor Models of Large Dimensions', *Econometrica*, 71, 135–171.
Bai, J., and Ng, S. (2002), 'Determining the Number of Factors in Approximate Factor Models', *Econometrica*, 70, 191–221.
Bai, J., and Ng, S. (2004), 'A PANIC Attack on Unit Roots and Cointegration', *Econometrica*, 72, 1127–1177.
Beare, B.K., Seo, J., and Seo, W. -K. (2017), 'Cointegrated Linear Processes in Hilbert Space', *Journal of Time Series Analysis*, 38, 1010–1027.
Bosq, D. (2000), *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lecture Notes in Statistics*, New York: Springer-Verlag.
Desai, K.H., and Storey, J.D. (2012), 'Cross-dimensional Inference of Dependent High-dimensional Data', *Journal of the American Statistical Association*, 107, 135–151.
Diebold, F.X., and Li, C. (2006), 'Forecasting the Term Structure of Government Bond Yields', *Journal of Econometrics*, 130, 337–364.
Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2005), 'The Generalized Dynamic Factor Model: One-sided Estimation and Forecasting', *Journal of the American Statistical Association*, 100, 830–840.
Friguet, C., Kloareg, M., and Causeur, D. (2009), 'A Factor Model Approach to Multiple Testing Under Dependence', *Journal of the American Statistical Association*, 104, 1406–1415.
Giannone, D., Reichlin, L., and Sala, L. (2005), 'Monetary Policy in Real Time', In *NBER Macroeconomics Annual 2004, Volume 19*, Cambridge, MA: MIT Press, pp. 161–224.

Gonzalo, J., and Granger, C. (1995), 'Estimation of Common Long-memory Components in Coin-tegrated Systems', *Journal of Business & Economic Statistics*, 13, 27–35.

Härdle, W.K., and Trück, S. (2010), 'The Dynamics of Hourly Electricity Prices', SFB 649 Discussion Paper 2010-013.

Hays, S., Shen, H., and Huang, J.Z. (2012), 'Functional Dynamic Factor Models with Application to Yield Curve Forecasting', *The Annals of Applied Statistics*, 6, 870–894.

Hörmann, S., Kidziński, L. u., and Hallin, M. (2015), 'Dynamic Functional Principal Components', *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 77, 319–348.

Horváth, L., Hušková, M., and Rice, G. (2013), 'Test of Independence for Functional Data', *Journal of Multivariate Analysis*, 117, 100–119.

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

Horváth, L., Kokoszka, P., and Reeder, R. (2013), 'Estimation of the Mean of Functional Time Series and a Two-sample Problem', *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 75, 103–122.

Horváth, L., Kokoszka, P., and Rice, G. (2014), 'Testing Stationarity of Functional Time Series', *Journal of Econometrics*, 179, 66–82.

Hyndman, R.J., and Ullah, M.S. (2007), 'Robust Forecasting of Mortality and Fertility Rates: a Functional Data Approach', *Computational Statistics & Data Analysis*, 51, 4942–4956.

Jungbacker, B., Koopman, S.J., and van der Wel, M. (2014), 'Smooth Dynamic Factor Analysis with Application to the US Term Structure of Interest Rates', *Journal of Applied Econometrics*, 29, 65–90.

Kokoszka, P., Miao, H., Reimherr, M., and Taoufik, B. (2018), 'Dynamic Functional Regression with Application to the Cross-section of Returns', *Journal of Financial Econometrics*, 16, 461–485.

Kokoszka, P., Miao, H., and Zhang, X. (2015), 'Functional Dynamic Factor Model for Intraday Price Curves', *Journal of Financial Econometrics*, 13, 456–477.

Koopman, S.J., Mallee, M.I.P., and der Wel, M.V. (2010), 'Analyzing the Term Structure of Interest Rates Using the Dynamic Nelson-Siegel Model with Time-varying Parameters', *Journal of Business & Economic Statistics*, 28, 329–343.

Lam, C., and Yao, Q. (2012), 'Factor Modeling for High-dimensional Time Series: Inference for the Number of Factors', *The Annals of Statistics*, 40, 694–726.

Lam, C., Yao, Q., and Bathia, N. (2011), 'Estimation of Latent Factors for High-dimensional Time Series', *Biometrika*, 98, 901–918.

Liebl, D. (2013), 'Modeling and Forecasting Electricity Spot Prices: a Functional Data Perspective', *The Annals of Applied Statistics*, 7, 1562–1592.

Martínez-Hernández, I., Genton, M.G., and González-Farías, G. (2019), 'Robust Depth-based Estimation of the Functional Autoregressive Model', *Computational Statistics & Data Analysis*, 131, 66–79.

Nelson, C., and Siegel, A.F. (1987), 'Parsimonious Modeling of Yield Curves', *The Journal of Business*, 60, 473–89.

Peña, D., and Poncela, P. (2006), 'Nonstationary Dynamic Factor Analysis', *Journal of Statistical Planning and Inference*, 136, 1237–1257.

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

Rice, G., and Shang, H.L. (2017), 'A Plug-in Bandwidth Selection Procedure for Long-run Covariance Estimation with Stationary Functional Time Series', *Journal of Time Series Analysis*, 38, 591–609.

Stock, J., and Watson, M. (1988), 'Testing for Common Trends', *Journal of the American Statistical Association*, 83, 1097–1107.

Appendices

Appendix 1. Proofs

Derivation of equation (2): Let $Y_n(s) = \sum_{j=0}^{\infty} A_j(\varepsilon_{n-j})(s)$ be a stationary functional linear process, with $\sum_{j=0}^{\infty} \|A_j\| < \infty$. Let $A = \sum_{j=0}^{\infty} A_j$. First, we observe that for all $z \in \mathcal{H}$, the covariance operator at lag h holds

$$\langle \Gamma_h(z), z \rangle = \mathbb{E}(\langle Y_0, z \rangle \langle Y_h, z \rangle) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbb{E}\{\langle A_i(\varepsilon_{-i}), z \rangle \langle A_j(\varepsilon_{h-j}), z \rangle\}$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \mathbb{E}\{\langle \varepsilon_{-i}, A_i^*(z) \rangle \langle \varepsilon_{h-j}, A_j^*(z) \rangle\}.$$

We note that $\mathbb{E}\{\langle \varepsilon_{-i}, A_i^*(z) \rangle \langle \varepsilon_{h-j}, A_j^*(z) \rangle\} = \langle \Gamma_{\varepsilon_{-i}, \varepsilon_{h-j}} A_i^*(z), A_j^*(z) \rangle$. Since $\{\varepsilon_n\}$ are i.i.d., we have that $\Gamma_{\varepsilon_i, \varepsilon_j} = 0$ if $i \neq j$. Thus, $\langle \Gamma_{\varepsilon_{-i}, \varepsilon_{h-j}}(A_i^*(z)), A_j^*(z) \rangle = \langle \Gamma_{\varepsilon_0}(A_i^*(z)), A_{i+h}^*(z) \rangle = \langle A_{i+h} \Gamma_{\varepsilon_0} A_i^*(z), (z) \rangle$. Substituting this into the above expression, we have

$$\Gamma_h = \sum_{i=0}^{\infty} A_{i+h} \Gamma_{\varepsilon_0} A_i^*.$$

Now, we compute the long-run covariance operator Λ. For all $z \in \mathcal{H}$, we have

$$\langle \Lambda(z), z \rangle = \sum_{h=-\infty}^{\infty} \langle \Gamma_h(z), z \rangle = \sum_{h=-\infty}^{\infty} \sum_{i=0}^{\infty} \langle A_{i+h} \Gamma_{\varepsilon_0} A_i^*(z), z \rangle$$

$$= \sum_{i=0}^{\infty} \left\langle \Gamma_{\varepsilon_0} A_i^*(z), \sum_{h=-\infty}^{\infty} A_{i+h}^*(z) \right\rangle,$$

since $A_j = 0$ for all $j < 0$, and changing the variable h to $k = h + i$, we obtain that $\sum_{h=-\infty}^{\infty} A_{i+h}^*(z) = \sum_{k=0}^{\infty} A_k^*(z)$. That is, $\langle \Lambda(z), z \rangle = \langle \Gamma_{\varepsilon_0} A^*(z), A^*(z) \rangle = \langle A \Gamma_{\varepsilon_0} A^*(z), z \rangle$. This implies that the long-run covariance operator is $\Lambda = A \Gamma_{\varepsilon_0} A^*$. ∎

All operators involved in the proof are integral operators, i.e. each operator is represented by a kernel function in $L^2([0, 1] \times [0, 1])$. Thus, the operator norm is defined as the usual norm in $L^2([0, 1] \times [0, 1])$, e.g. $\|\Lambda\|^2 = \iint c_{-0}^2(s, t) \, ds \, dt$.

Proof of Proposition 4.1: Let us first observe that under Assumptions 4.1 and 4.2, the estimator $\widehat{\Gamma}$ is a consistent estimator of Γ. Explicitly, we have that $\iint \{\widehat{c}(s, t) - c(s, t)\}^2 \, ds \, dt \xrightarrow{\mathbb{P}} 0$. This is because with Assumptions 4.1 and 4.2 the functional time series $\{X_n\}$ is an L^2 m-approximable (and hence sttaionary). The reader is referred to Horváth et al. (2013) for the details of this proof.

Similarly, we have that $\iint [\widehat{\gamma_0}(s, t) - \mathbb{E}\{X_1(s)X_1(t)\}]^2 \, ds \, dt \xrightarrow{\mathbb{P}} 0$. That is, $\|\widehat{\Gamma}_0 - \Gamma_0\| \xrightarrow{\mathbb{P}} 0$. Thus,

$$\iint \{\widehat{c}_{-0}(s, t) - c_{-0}(s, t)\}^2 \, ds \, dt = \iint \{\widehat{c}(s, t) - \widehat{\gamma_0}(s, t) - c(s, t) + \widehat{\gamma_0}(s, t)\}^2 \, ds \, dt$$

$$\leq 2 \iint \{\widehat{c}(s, t) - c(s, t)\}^2 \, ds \, dt + 2 \iint \{\widehat{\gamma_0}(s, t) - \gamma_0(s, t)\}^2 \, ds \, dt$$

$$= o_{\mathbb{P}}(1) + o_{\mathbb{P}}(1).$$

Therefore, $\|\widehat{\Gamma - \Gamma_0} - (\Gamma - \Gamma_0)\| \xrightarrow{\mathbb{P}} 0$.

We now turn to the Γ_0^{-1} operator. We observe that Γ_0^{-1} is an integral operator with kernel $\widehat{k}^{-1}(s, t) = \sum_{j=1}^{p} \widehat{\lambda}_j^{-1} \widehat{v}_j(s) \widehat{v}_j(t)$.

For $g, f \in L^2([0,1] \times [0,1])$, we use the notation $(gf)(s,t) := g(s,t)f(s,t)$. Then, we observe that

$$\iint \{(\widehat{c}_{-0}\widehat{k}^{-1})(s,t) - (c_{-0}k^{-1})(s,t)\}^2 \, ds \, dt$$

$$= \iint \left[\{(\widehat{c}_{-0} - c_0)\widehat{k}^{-1}\}(s,t) + \{c_{-0}(\widehat{k}^{-1} - k^{-1})\}(s,t)\right]^2 \, ds \, dt,$$

with $k^{-1}(s,t)$ being the kernel with the true values λ_j and v_j. Thus,

$$\iint \{(\widehat{c}_{-0}\widehat{k}^{-1})(s,t) - (c_{-0}k^{-1})(s,t)\}^2 \, ds \, dt \le 2 \iint \{(\widehat{c}_{-0} - c_0)^2 (\widehat{k}^{-1})^2\}(s,t) \, ds \, dt$$

$$+ 2 \iint \{c_{-0}^2 (\widehat{k}^{-1} - k^{-1})^2\}(s,t) \, ds \, dt.$$

The first component on the right-hand side of the above inequality is bounded by $\iint (\widehat{c}_{-0} - c_0)^2(s,t) \, ds \, dt \iint (\widehat{k}^{-1})^2(s,t) \, ds \, dt$. If Γ_0 is invertible or if $\mathcal{H}$ is finite dimensional, then we have that $\widehat{\Gamma}_0^{-1}$ is bounded, and as a consequence $a_1 := \iint (\widehat{k}^{-1})^2(s,t) \, ds \, dt < \infty$. Then,

$$2 \iint \{(\widehat{c}_{-0} - c_0)^2 (\widehat{k}^{-1})^2\}(s,t) \, ds \, dt \le 2a_1 \iint (\widehat{c}_{-0} - c_0)^2(s,t) \, ds \, dt$$

$$= 2a_1 o_{\mathbb{P}}(1).$$

Similarly, for the second component we have that

$$2 \iint \{c_{-0}^2 (\widehat{k}^{-1} - k^{-1})^2\}(s,t) \, ds \, dt \le 2a_2 \iint (\widehat{k}^{-1} - k^{-1})^2(s,t) \, ds \, dt$$

$$\le 2a_1^2 a_2 \iint (\widehat{k} - k)^2(s,t) \, ds \, dt$$

$$= O(1)o_{\mathbb{P}}(1),$$

with $a_2 := \iint c_{-0}^2(s,t) \, ds \, dt < \infty$. Therefore, we conclude that

$$\iint \{(\widehat{c}_{-0}\widehat{k}^{-1})(s,t) - (c_{-0}k^{-1})(s,t)\}^2 \, ds \, dt \overset{\mathbb{P}}{\to} 0,$$

that is, $\|\widehat{\Lambda} - \Lambda\| \overset{\mathbb{P}}{\to} 0$.

In the case where $\mathcal{H}$ is infinite dimension, one can obtain similar result following similar ideas as in Bosq (2000, chap. 8.3). We omit the proof of this case. ∎

Proof of Corollary 4.2: Let $\widehat{\Pi}_F$ and Π_F be the orthonormal projectors on $\widehat{\mathcal{H}}_F$ and $\mathcal{H}_F$, respectively. Then, for any $z \in \mathcal{H}$ with $\|z\|_{\mathcal{H}} < \infty$, we have that

$$\|\widehat{\Pi}_F(z) - \Pi_F(z)\| \le \sum_{k=1}^{K} \|\langle z, \widehat{F}_k \rangle \widehat{F}_k - \langle z, F_k \rangle F_k\|$$

$$\le a_3 a_4 \sum_{k=1}^{K} \|\widehat{F}_k - F_k\|$$

$$\le a_3 a_4 \sum_{k=1}^{K} \frac{2\sqrt{2}}{\tilde{\alpha}_k} \|\widehat{\Lambda} - \Lambda\|,$$

with $a_3 = \max\{\langle z, \widehat{F}_k \rangle : k = 1, \ldots, K\}$, $a_4 = \max\{\langle z, F_k \rangle : k = 1, \ldots, K\}$, $\tilde{\alpha}_1 = \alpha_1 - \alpha_2$, and $\tilde{\alpha}_j = \min\{\alpha_{j-1} - \alpha_j, \alpha_j - \alpha_{j+1}\}$, $j = 2, \ldots, K$. Since $\alpha_1 > \alpha_2 > \ldots > \alpha_k > 0$, we obtain $\|\widehat{\Pi}_F(z) -$

$\Pi_F(z)\| \leq a_3 a_4 \frac{K2\sqrt{2}}{\tilde{\alpha}} \|\widehat{\Lambda} - \Lambda\|$ with $\tilde{\alpha} = \min\{\tilde{\alpha}_1, \ldots, \tilde{\alpha}_K\}$. Thus, $\|\widehat{\Pi}_F(z) - \Pi_F(z)\| \leq O(1)o_{\mathbb{P}}(1)$, and then $\|\widehat{\Pi}_F - \Pi_F\| \xrightarrow{\mathbb{P}} 0$. ∎

Proof of Proposition 4.3: We need only consider the case in which r is $0 < r < K$. We have that the long-run covariance operator $\Lambda_{\Delta X}$ of the functional process ΔX_n is $\Lambda_{\Delta X} = \Phi \Gamma_{\epsilon_0} \Phi$, where $\Phi = \sum_{j \geq 0} \Phi_j$, with Φ_j compact and self-adjoint operators. Thus, under Assumption 4.3, X_n can be written as

$$X_n = X_0 + \Phi\left(\sum_{i \leq n} \epsilon_i\right) + \nu_n,$$

where ν_n is a stationary functional time series. Let ξ_k be an eigenfunction of $\Lambda_{\Delta X}$; then, $\xi_k \notin \ker(\Phi)$. From this, we can show that $\mathcal{H}_\lambda^N = \{\ker(\Lambda_{\Delta X})\}^\perp$ and $\mathcal{H}_\lambda^S = \ker(\Lambda_{\Delta X})$ (see Beare et al. 2017 for more details). ∎

Proof of Corollary 4.4: The proof is similar in spirit to the proof of Corollary 4.2. ∎

Proof of Corollary 4.5: For $\xi_k \neq 0$, an eigenfunction of $\Lambda_{\Delta X}$, we have that

$$\langle X_n, \xi_k \rangle = \langle X_0, \xi_k \rangle + \left\langle \Phi\left(\sum_{i \leq n} \epsilon_i\right), \xi_k \right\rangle + \langle \nu_n, \xi_k \rangle = X_0^k + \sum_{i \leq n} \epsilon_i^k + \nu_n^k,$$

where $X_0^k = \langle X_0, \xi_k \rangle$, $\epsilon_i^k = \langle \epsilon_i, \Phi \xi_k \rangle$, and $\nu_n^k = \langle \nu_n, \xi_k \rangle$. That is, $\{\langle X_n, \xi_k \rangle\}$ has a random walk component $\sum_{i \leq n} \epsilon_i^k$ since $\Phi(\xi_k) \neq 0$. Thus, for each eigenfunction $\xi_k \neq 0$ of $\Lambda_{\Delta X}$, the process $\{\langle X_n, \xi_k \rangle\} = \{\beta_{n,k}\}$ is an $I(1)$ process, and replacing ξ_k with their corresponding estimator $\widehat{\xi}_k$, we obtain that $\{\widehat{\beta}_{n,k}\}$ is also an $I(1)$ process. In contrast, if $\nu \in \ker(\Lambda_{\Delta X})$, then $\{\langle X_n, \nu \rangle\} = \langle X_0, \nu \rangle + \{\langle \nu_n, \nu \rangle\}$ is stationary. This completes the proof. ∎

Appendix 2. Comparison of factors

Here we present the Nelson-Siegel curves corresponding to the three factors (see, e.g. Diebold and Li 2006). Also, we present our estimators of the Yield curve data (Section 6).

The Nelson-Siegel curves are defined as

$$NF_1(s) = 1, \quad NF_2(s) = \frac{1 - \exp(-\lambda_n s)}{\lambda_n s}, \quad \text{and} \quad NF_3(s) = \frac{1 - \exp(-\lambda_n s)}{\lambda_n s} - \exp(-\lambda_n s).$$

Figure A1 presents the three Nelson-Siegel curves (second row) and our estimators (first row). In the Nelson-Siegel curves, we fix the parameter $\lambda_n = 0.9$.
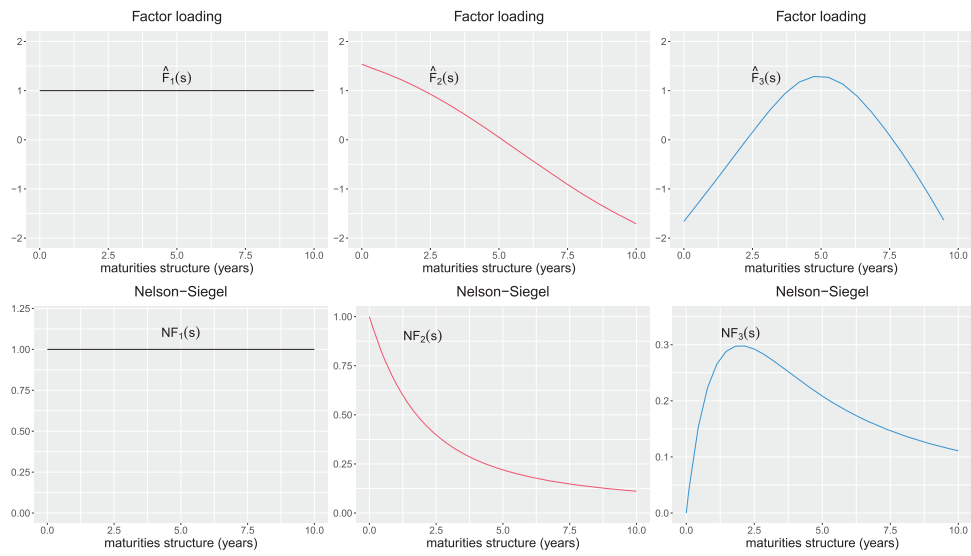
Figure A1. Our estimator of factor loading curves, $\widehat{F}_1$, $\widehat{F}_2$, and $\widehat{F}_3$, and the Nelson-Siegel curves NF_1, NF_2, and NF_3. We fix $\lambda_n = 0.9$.