Optimal Dimension Reduction for High-dimensional and Functional Time Series

Marc Hallin,
ECARES and Département de Mathématique, Université libre de Bruxelles

Siegfried Hörmann,
ECARES and Département de Mathématique, Université libre de Bruxelles
and Graz University of Technology

Marco Lippi,
Einaudi Institute for Economics and Finance

November 2017

ECARES working paper 2017-39

ECARES
ULB - CP 114/04
50, F.D. Roosevelt Ave., B-1050 Brussels BELGIUM
www.ecares.org

Optimal Dimension Reduction for High-Dimensional and Functional Time Series

Marc Hallin[a,b], Siegfried Hörmann[a,b,c,*], and Marco Lippi[d]

[a] *ECARES, Université libre de Bruxelles, Belgium*
[b] *Département de Mathématique, Université libre de Bruxelles, Belgium*
[c] *Institute for Statistics, Graz University of Technology, Austria*
[d] *Einaudi Institute for Economics and Finance, Rome, Italy*

Abstract

Dimension reduction techniques are at the core of the statistical analysis of high-dimensional and functional observations. Whether the data are vector- or function-valued, principal component techniques, in this context, play a central role. The success of principal components in the dimension reduction problem is explained by the fact that, for any $K \leq p$, the K first coefficients in the expansion of a p-dimensional random vector $\mathbf{X}$ in terms of its principal components is providing the best linear K-dimensional summary of $\mathbf{X}$ in the mean square sense. The same property holds true for a random function and its functional principal component expansion. This optimality feature, however, no longer holds true in a time series context: principal components and functional principal components, when the observations are serially dependent, are losing their optimal dimension reduction property to the so-called *dynamic principal components* introduced by Brillinger in 1981 in the vector case and, in the functional case, their functional extension proposed by Hörmann, Kidziński and Hallin in 2015.

Keywords. Dimension reduction, time series, principal components, functional principal components, dynamic principal components, Karhunen-Loève expansion.

*Research partially supported by the Communauté francaise de Belgique, Actions de Recherche Concertées, Projets Consolidation 2016-2021, and Interuniversity Attraction Poles Programme (IAP-network P7/06) of the Belgian Science Policy Office.

1 Introduction.

1.1 Dimension reduction, principal components, and time series.

Dimension reduction techniques are at the core of the statistical analysis of high-dimensional observations. This is the case for vector-valued observations; it is the case, too, for functional observations, which are intrinsically infinite-dimensional.

In the vector case, the objective is to provide a "good", or even "optimal", K-dimensional summary $\mathbf{Y} := (Y_1, \ldots, Y_K)'$ of a p-dimensional random vector $\mathbf{X} = (X_1, \ldots, X_p)'$ (p "large", $1 \leq K < p$ "small"), that is, a representation of the form $\mathbf{X}^{(K)} := \sum_{k=1}^{K} Y_k \mathbf{e}_k$, where $\mathbf{e}_1, \ldots, \mathbf{e}_K$ is a (deterministic) orthonormal K-tuple of vectors in $\mathbb{R}^p$, and $\mathbf{Y}$ a K-dimensional random vector defined on the same probability space as $\mathbf{X}$. The quality of this dimension reduction is measured by the mean square error $\mathrm{E}\big[\, \big\|\mathbf{X} - \mathbf{X}^{(K)}\big\|^2 \,\big]$ which, for given K, is to be minimized with respect to $\mathbf{e}_1, \ldots, \mathbf{e}_K$ and $(Y_1, \ldots, Y_K)$.

In the functional case, X denotes a random function $u \mapsto X(u)$, with domain $u \in [0,\, 1]$; we throughout assume that X belongs to the space $L^2([0,1])$ of square-integrable functions mapping $[0,\, 1]$ to $\mathbb{R}$. Here again, the objective is to approximate X (which is intrinsically infinite-dimensional) by means of a K-dimensional summary $\mathbf{Y} := (Y_1, \ldots, Y_K)'$, via a representation of the form $X^{(K)} := \sum_{k=1}^{K} Y_k e_k$, where $\mathbf{Y}$ again is a real-valued random vector with given dimension K, while $e_1, \ldots, e_K$ now is some (deterministic) K-tuple of elements in $L^2([0,1])$. The quality of the dimension reduction is still measured by the mean square error, here $\mathrm{E} \int_0^1 \big[X(u) - X^{(K)}(u)\big]^2 \,\mathrm{d}u$ with $X^{(K)}(u) := \sum_{k=1}^{K} Y_k e_k(u)$, is to be minimized, for given K, with respect to $e_1, \ldots, e_K$ and $(Y_1, \ldots, Y_K)$.

Whether the data are vector-valued or functional, the most popular dimension reduction technique is principal component analysis, which is daily practice, and constitutes the basic ingredient of most high-dimensional data analysis techniques. The objective of this paper is to explain the success of this practice, and to question it in the time series context. Actually, we show that principal components dimension reduction techniques, at least in their classical, *static* form, are inadequate in the presence of serially correlated data, and propose an alternative. The proposed alternative, based on a *dynamic* extension of the classical concept, is shown to recover all the properties that make the success of principal components as a dimension reduction tool in the absence of serial dependence.

1.2 Outline of the paper.

The paper is organized as follows. Section 2 is presenting well known results on principal components-based dimension reduction techniques, emphasizing their optimality features in the vector (Section 2.1) and functional (Section 2.2) cases, respectively. In Section 3, we first explain why those properties are lost in the time series context. We then present (Section 3.1) Brillinger's concept of *dynamic principal components*, and show how dynamic principal components, in the vector-valued time series case, recover the optimality properties (as described in Section 2.1) of traditional principal components, to which they reduce for serially uncorrelated processes. In Section 3.2, we explain how the concept of *functional dynamic principal components* introduced in Hörmann et al. (2015) extends this to the functional case. Section 4 illustrates via simulations the benefits one can expect, in dimension reduction, from considering dynamic principal components instead of the daily-practice traditional (static) ones. Section 5 concludes and gives some perspectives for future research.

2 Principal components and functional principal components.

2.1 Principal Components in $\mathbb{R}^p$

Let the p-dimensional random vector $\mathbf{X}$ have mean $\boldsymbol{\mu} := \mathrm{E}\mathbf{X} = \mathbf{0}$ and covariance $\mathbf{C} := \mathrm{Cov}(\mathbf{X}) := \mathrm{E}\mathbf{X}\mathbf{X}'$, with eigenvectors $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$ and eigenvalues $\lambda_1 > \ldots > \lambda_p$ (for convenience, we assume that those eigenvalues are distinct, an assumption that easily can be relaxed in the sequel for the $p-K$ smallest ones). The spectral decomposition of $\mathbf{X}$ yields

$$\mathbf{X} = \sum_{k=1}^{p} Y_k^* \boldsymbol{\phi}_k := \sum_{k=1}^{p} (\boldsymbol{\phi}_k' \mathbf{X}) \boldsymbol{\phi}_k. \tag{2.1}$$

The scalar random variable $Y_k^* := \boldsymbol{\phi}_k' \mathbf{X} = \langle \mathbf{X}, \boldsymbol{\phi}_k \rangle$ is known as $\mathbf{X}$'s kth *principal component* or kth *principal component score*, and, subjected to being uncorrelated with $Y_1^*, \ldots, Y_{k-1}^*$, has maximal variance among all normed linear combinations of $X_1, \ldots, X_p$. The covariance matrix of $Y_1^*, \ldots, Y_p^*$ is easily seen to be diagonal, with diagonal elements $\lambda_1, \ldots, \lambda_p$.

The success of principal components as a dimension reduction technique is explained by two important properties. The first one is an optimality property: the expansion (2.1) is such that, for any $1 \leq K \leq p$ and any

orthonormal basis $\mathbf{e}_1, \ldots, \mathbf{e}_p$ of $\mathbb{R}^p$,

$$\begin{aligned} \mathrm{E}\Big[\Big\|\mathbf{X} - \sum_{k=1}^{K} Y_k^* \boldsymbol{\phi}_k\Big\|^2\Big] &= \mathrm{E}\Big[\Big\|\mathbf{X} - \sum_{k=1}^{K} \langle \mathbf{X},\, \boldsymbol{\phi}_k\rangle \boldsymbol{\phi}_k\Big\|^2\Big] \quad \Big(= \sum_{k=K+1}^{p} \lambda_k\Big) \\ &\leq \mathrm{E}\Big[\Big\|\mathbf{X} - \sum_{k=1}^{K} \langle \mathbf{X},\, \mathbf{e}_k\rangle \mathbf{e}_k\Big\|^2\Big]. \end{aligned} \tag{2.2}$$

For any dimension K between 1 and p, thus, the K first principal components $(Y_1^*, \ldots, Y_K^*)$ provide the "best" (linear) reduction of $\mathbf{X}$ to dimension K in the mean square error sense. This optimality property readily follows from the fact that Y_1 is the normed linear combination of $X_1, \ldots, X_p$ with highest variance, etc. It actually constitutes the discrete version of a more general optimality property of *Karhunen-Loève expansions* (see Section 2.2).

A second and no less important property of principal components is that they are mutually uncorrelated (their covariance matrix is diagonal), which greatly simplifies their statistical analysis and neatly decomposes their respective contributions to the approximation error $\sum_{k=K+1}^{p} \lambda_k$. A natural measure of performance of the principal component-based reduction of $\mathbf{X}$ to dimension K is the *relative mean square approximation error*

$$\mathrm{RMSE}_{\mathrm{stat}}(K) := \sum_{k=K+1}^{\infty} \lambda_k \Big/ \sum_{k=1}^{\infty} \lambda_k = 1 - \frac{\sum_{k=1}^{K} \lambda_k}{E\|X\|^2}, \quad K = 1, \ldots, p. \tag{2.3}$$

2.2 Principal components in $L^2([0,1])$.

The definition and properties of principal components in $\mathbb{R}^p$ readily extend to the functional context where the variable X under study takes values in the Hilbert space $L^2 = L^2([0,1])$ equipped with inner product

$$\langle x, y\rangle := \int_0^1 x(u)y(u)\mathrm{d}u$$

and norm $\|x\| := \langle x, x\rangle^{1/2}$.

Assume, without loss of generality, that X has mean function

$$\mathrm{E}X = \mu : \; u \mapsto \mu(u) = \mathrm{E}X(u) = 0, \qquad u \in [0,1],$$

and denote by

$$\big\{C(u,v) := \mathrm{Cov}\big(X(u), X(v)\big) : \; (u,v) \in [0,1]^2\big\}$$

its covariance kernel. Instead of matrices, we are dealing, in this functional context, with covariance operators: the covariance operator C is mapping the function $f \in L^2$ to the function $C(f) \in L^2$, with

$$C(f)(u) := \int_0^1 C(u,v)f(v)\mathrm{d}v, \qquad u \in [0,1].$$

Functional principal components rely on the fact that C is a *compact* (more precisely, a *trace class*) operator, so that, for any $f \in L^2$, the image $C(f)$ of f admits an eigendecomposition

$$C(f) = \sum_{k=1}^{\infty} \lambda_k \langle f, \phi_k \rangle \phi_k,$$

where $\phi_k \in L^2$ is the kth eigenfunction of C, associated with eigenvalue λ_k (eigenvalues, as usual, are assumed to be distinct and ordered by decreasing magnitude): $\|\phi_k\| = 1$, $\langle \phi_k, \phi_\ell \rangle = 0$ for $k \neq \ell$, and $\sum_{k=1}^{\infty} \lambda_k < \infty$. The sequence $\phi_1, \phi_2, \ldots$ thus constitutes an orthonormal basis for image space of C, yielding for $X(u)$ the representation, called *Karhunen-Loève expansion*,

$$X(u) = \sum_{k=1}^{\infty} \langle X, \phi_k \rangle \phi_k(u) =: \sum_{k=1}^{\infty} Y_k^* \phi_k(u) \quad u \in [0,1]. \tag{2.4}$$

The scalar random variable $Y_k^* := \langle X, \phi_k \rangle$ is called X's kth *functional principal component* or *functional principal component score*. Those functional principal components have diagonal covariance matrix, with diagonal elements $\mathrm{Var}(Y_k^*) = \lambda_k$. As $k \to \infty$, λ_k tends to 0, so that $Y_k^* \to 0$ in quadratic mean.

Expansions of the form (2.4) are named after Karhunen (1946, 1947) and Loève (1955), who independently introduced them in a more general framework; they enjoy, *mutatis mutandis*, the same optimality properties as their random vector counterpart (2.1), namely, for any $1 \leq K \in \mathbb{N}$ and any orthonormal sequence $e_1, e_2, \ldots$ in $L^2[0,1]$,

$$\mathrm{E} \int_0^1 \Big(X(u) - \sum_{k=1}^{K} Y_k^* \phi_k(u) \Big)^2 \mathrm{d}u \leq \mathrm{E} \int_0^1 \Big(X(u) - \sum_{k=1}^{K} Y_k e_k(u) \Big)^2 \mathrm{d}u, \tag{2.5}$$

with $Y_k = \langle X_k, e_k \rangle$, or, in more compact form,

$$\mathrm{E} \Big\| X - \sum_{k=1}^{K} Y_k^* \phi_k \Big\|^2 \leq \mathrm{E} \Big\| X - \sum_{k=1}^{K} Y_k e_k \Big\|^2; \tag{2.6}$$

call this the *Karhunen-Loève optimality property*. Hence, just as in the vector case, the K first functional principal components $(Y_1^*, \ldots, Y_K^*)$ provide the "best reduction to dimension K" of X in the mean square error sense; moreover, they are mutually orthogonal. A simple consequence of this orthogonality is that the mean square error associated with $\sum_{k=1}^{K} Y_k^* \phi_k$ as an approximation of X is

$$\mathrm{E}\Big\| X - \sum_{k=1}^{K} Y_k^* \phi_k \Big\|^2 = \sum_{k=K+1}^{\infty} \lambda_k. \tag{2.7}$$

Here also, it is natural to use the *relative mean square approximation error*

$$\mathrm{RMSE}_{\mathrm{stat}}(K) := \sum_{k=K+1}^{\infty} \lambda_k \Big/ \sum_{k=1}^{\infty} \lambda_k = 1 - \frac{\sum_{k=1}^{K} \lambda_k}{E\|X\|^2}, \quad K = 1, \ldots, p \tag{2.8}$$

as a scale-free measure of how well we can represent the original function X by a K-tuple of real numbers.

3 Principal components for time series.

Let us now assume that $\{\mathbf{X}_t |\ t \in \mathbb{Z}\}$ is a multivariate second-order stationary p-dimensional vector process. The process

$$(Y_{1t}^*, \ldots, Y_{Kt}^*) := (\langle \mathbf{X}_t, \boldsymbol{\phi}_1 \rangle, \ldots, \langle \mathbf{X}_t, \boldsymbol{\phi}_K \rangle), \quad t \in \mathbb{Z}$$

of the K first principal component scores (a K-dimensional vector-valued stochastic process), being based on the eigenvectors $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ of the covariance matrix $\mathbf{C} := \mathrm{E}[\mathbf{X}_t \mathbf{X}_t']$, only exploits instantaneous covariances, disregarding the information contained in $\mathbf{X}_t$'s leads and lags. Those principal components thus only provide the "best static" K-dimensional linear approximation of the process under study, that is, they restrict to dimension reductions involving instantaneous linear combinations of the $\mathbf{X}_t$'s only. In a time series context, as soon as $\{\mathbf{X}_t\}$ is not second-order white noise, linear combinations involving the past and future values of $\mathbf{X}_t$ contain essential information, and are likely to do a much better job.

Similarly, the K-dimensional vector-valued stochastic process

$$(Y_{1t}^*, \ldots, Y_{Kt}^*) := (\langle X_t, \phi_1 \rangle, \ldots, \langle X_t, \phi_K \rangle), \quad t \in \mathbb{Z}$$

of the K first principal components of the functional second-order stationary process $\{X_t |\ t \in \mathbb{Z}\}$ is based on the eigenfunctions ϕ_k of the covariance

operator C, hence only exploits information contained in the instantaneous covariances $C(u,v) := \mathrm{E}[X_t(u)X_t(v)]$.

As a consequence, the optimality property that makes the success of (functional) principal components in a sample of i.i.d. (or, at least, uncorrelated) observations of a random vector $\mathbf{X}$ (a random function X) only holds if one restricts to suboptimal *static* dimension reduction. Moreover, unless $\{\mathbf{X}_t\}$ (respectively, $\{X_t\}$) is an uncorrelated process, the Y^*_{kt}'s are mutually orthogonal "at lag zero" only: $Y^*_{k_1 t}$ and $Y^*_{k_2 t}$ are uncorrelated for $k_1 \neq k_2$, but $Y^*_{k_1 t}$ and $Y^*_{k_2, t-\ell}$, in general, are (cross-)correlated for $\ell \neq 0$, even when $k_1 \neq k_2$.

Whether classical or functional, thus, principal components, in a time series context, no longer enjoy the two properties that make them an efficient statistical tool in the i.i.d. (uncorrelated) case.

3.1 Brillinger's dynamic principal components.

Can we adapt the traditional, *static*, concepts of principal components to the dynamic context in order to preserve, in the time-series context, their optimal dimension reduction and mutual orthogonality properties? An obvious challenger of the static approach is the concept of *dynamic principal component*, proposed by Brillinger (1981) in a pathbreaking but all-too-often overlooked contribution, for p-dimensional processes admitting a spectral density (a spectral density matrix).

Denote by $\{\mathbf{X}_t := (X_{1t}, \ldots, X_{pt})' \; t \in \mathbb{Z}\}$ a second-order p-dimensional stationary vector process, with values in $\mathbb{R}^p$, mean $\boldsymbol{\mu} = \mathbf{0}$, and cross-covariance matrices $\boldsymbol{\Gamma}_k := \mathrm{E}\left[\mathbf{X}_t \mathbf{X}'_{t-k}\right]$, $k \in \mathbb{Z}$. Intuitively, informative linear approximations of $\mathbf{X}_t$ should fully exploit the cross-covariance structure—the collection of *all* $\boldsymbol{\Gamma}_k$'s rather than $\mathbf{C} = \boldsymbol{\Gamma}_0$ only. Hence, they should involve all present, but also past and future values of the $\mathbf{X}_t$'s. This requires looking for the normed linear combinations of present, past and future observations maximizing (subject to orthogonality constraints) the variance, etc.—dynamic linear combinations (that is, *filters*), not just static ones.

This is what Brillinger's *dynamic principal components* are achieving. Dynamic principal components are based on a factorization of spectral density matrices instead of the factorization of the (instantaneous) covariance matrices $\mathbf{C} = \boldsymbol{\Gamma}_0$. Therefore, let us make the additional (but relatively mild) assumption that the spectral measure of $\{\mathbf{X}_t, t \in \mathbb{Z}\}$ is absolutely continuous with respect to the Lebesgue measure on $[-\pi, \pi]$, that is, $\{\mathbf{X}_t\}$ has a $p \times p$

spectral density matrix

$$\mathbf{\Sigma}(\theta) := \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \mathbf{\Gamma}_\ell e^{-\imath\ell\theta},$$

with entries $(\sigma_{ij}(\theta))$, $\theta \in [-\pi, \pi]$ ($\imath$ stands for the imaginary root of -1).

The matrix $\mathbf{\Sigma}(\theta)$ is Hermitian and positive semidefinite. Its eigenvalues $\lambda_1(\theta) > \lambda_2(\theta) > \ldots > \lambda_p(\theta) > 0$ thus are real; call them the *dynamic eigenvalues* of $\{\mathbf{X}_t\}$ (again, for simplicity, we assume they are all distinct for any frequency θ). Writing $\overline{\mathbf{\Sigma}}(\theta)$ for $\mathbf{\Sigma}(\theta)$'s complex conjugate, note that, for all θ, $\overline{\mathbf{\Sigma}}(\theta) = \mathbf{\Sigma}'(\theta) = \mathbf{\Sigma}(-\theta)$. Denote by $\boldsymbol{\varphi}'_k(\theta)$ a row eigenvector associated with $\lambda_k(\theta)$, and call it a *dynamic eigenvector* of $\{\mathbf{X}_t\}$. Then, $\boldsymbol{\varphi}_k(\theta)$ is a column eigenvector of $\mathbf{\Sigma}(-\theta) = \overline{\mathbf{\Sigma}}(\theta) = \mathbf{\Sigma}'(\theta)$, still associated with eigenvalue $\lambda_k(\theta)$, that is,

$$\boldsymbol{\varphi}'_k(\theta)\mathbf{\Sigma}(\theta) = \lambda_k(\theta)\boldsymbol{\varphi}'_k(\theta) \quad \mathbf{\Sigma}(-\theta)\boldsymbol{\varphi}_k(\theta) = \lambda_k(\theta)\boldsymbol{\varphi}_k(\theta).$$

For all θ, the $p \times p$ matrix of eigenvectors $\boldsymbol{\varphi}(\theta) := \big(\boldsymbol{\varphi}_1(\theta) \ldots \boldsymbol{\varphi}_p(\theta)\big)$ is unitary; namely,

$$\boldsymbol{\varphi}(\theta)\boldsymbol{\varphi}^\dagger(\theta) = \mathbf{I} = \boldsymbol{\varphi}^\dagger(\theta)\boldsymbol{\varphi}(\theta),$$

hence

$$\overline{\boldsymbol{\varphi}}(\theta)\boldsymbol{\varphi}'(\theta) = \mathbf{I} = \boldsymbol{\varphi}'(\theta)\overline{\boldsymbol{\varphi}}(\theta)$$

(where $\boldsymbol{\varphi}^\dagger(\theta)$ and $\overline{\boldsymbol{\varphi}}(\theta)$ stand for the adjoint of $\boldsymbol{\varphi}(\theta)$ and its complex conjugate, respectively). Since $\mathbf{\Sigma}(-\theta) = \overline{\mathbf{\Sigma}}(\theta)$, we can impose $\boldsymbol{\varphi}(-\theta) = \overline{\boldsymbol{\varphi}}(\theta)$. Also note that $\boldsymbol{\varphi}'(\theta)\mathbf{\Sigma}(\theta)\overline{\boldsymbol{\varphi}}(\theta) = \mathbf{\Lambda}(\theta) =: \mathrm{diag}(\lambda_1(\theta), \ldots, \lambda_p(\theta))$ and that we have $\overline{\boldsymbol{\varphi}}(\theta)\mathbf{\Lambda}(\theta)\boldsymbol{\varphi}'(\theta) = \mathbf{\Sigma}(\theta)$.

Before introducing Brillinger's concept of dynamic principal components, let us recall some classical facts on the relation between matrices in the spectral domain and filters in the time domain. Any $(p \times m)$ matrix or vector $\mathbf{M}(\theta)$ with square-integrable θ-measurable elements defined over $[-\pi, \pi]$ can be expanded (componentwise) into a Fourier series

$$\mathbf{M}(\theta) = \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \Big[\int_{-\pi}^{\pi} \mathbf{M}(\theta) e^{\imath\ell\theta} \mathrm{d}\theta \Big] e^{-\imath\ell\theta}$$

where the right-hand side converges in quadratic mean. That expansion creates a correspondence between the square-integrable matrix-valued function $\mathbf{M}(\theta)$ and the square-summable filter

$$\underline{\mathbf{M}}(L) := \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \Big[\int_{-\pi}^{\pi} \mathbf{M}(\theta) e^{\imath\ell\theta} \, \mathrm{d}\theta \Big] L^\ell$$

(L, as usual, stands for the *lag operator*; note that $\mathbf{M}(\theta) = \underline{\mathbf{M}}(e^{-\imath\theta})$).

The $(p \times m)$-dimensional matrix $\mathbf{M}(\theta)$ and the filter $\underline{\mathbf{M}}(L)$ then are strongly connected by the fact that the p-dimensional process $\{\mathbf{X}_t\}$ has spectral density matrix $\mathbf{\Sigma}(\theta)$ iff the m-variate stochastic process $\{\underline{\mathbf{M}}'(L)\mathbf{X}_t\}$ has spectral density matrix

$$\mathbf{M}'(\theta)\mathbf{\Sigma}(\theta)\overline{\mathbf{M}}(\theta) \tag{3.1}$$

where $\overline{\mathbf{M}}$ is the complex conjugate of $\mathbf{M}$.

The $p \times 1$ dynamic eigenvectors $\boldsymbol{\varphi}_k(\theta)$, in particular, can be expanded into

$$\boldsymbol{\varphi}_k(\theta) = \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty}\Big[\int_{-\pi}^{\pi}\boldsymbol{\varphi}_k(\theta)e^{\imath\ell\theta}\mathrm{d}\theta\Big]e^{-\imath\ell\theta} =: \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty}\boldsymbol{\psi}_{k\ell}e^{-\imath\ell\theta},$$

defining $p \times 1$ square-summable filters of the form

$$\underline{\boldsymbol{\varphi}}_k(L) = \frac{1}{2\pi}\sum_{\ell=-\infty}^{\infty}\Big[\int_{-\pi}^{\pi}\boldsymbol{\varphi}_k(\theta)e^{\imath\ell\theta}\mathrm{d}\theta\Big]L^{\ell} =: \sum_{\ell=-\infty}^{\infty}\boldsymbol{\psi}_{k\ell}L^{\ell}$$

where $\boldsymbol{\psi}_{k\ell} := \frac{1}{2\pi}\int_{-\pi}^{\pi}\boldsymbol{\varphi}_k(\theta)e^{\imath\ell\theta}\mathrm{d}\theta$ is a p-dimensional vector of Fourier coefficients. Note that $\boldsymbol{\varphi}(-\theta) = \overline{\boldsymbol{\varphi}}(\theta)$ implies that the $\boldsymbol{\psi}_{k\ell}$'s, hence the filters $\underline{\boldsymbol{\varphi}}_k(L)$, are real.

It follows from (3.1) that the p-tuple

$$\big\{\mathbf{Y}_t^*|\ t \in \mathbb{Z}\big\} := \big\{\underline{\boldsymbol{\varphi}}'(L)\mathbf{X}_t|\ t \in \mathbb{Z}\big\}$$

of (real-valued) univariate processes $\{Y_{kt}^*|\ t \in \mathbb{Z}\}$, where

$$Y_{kt}^* := \underline{\boldsymbol{\varphi}}_k'(L)\mathbf{X}_t = \sum_{\ell=-\infty}^{\infty}\boldsymbol{\psi}_{k\ell}'\mathbf{X}_{t-\ell} = \sum_{\ell=-\infty}^{\infty}\langle\mathbf{X}_{t-\ell},\,\boldsymbol{\psi}_{k\ell}\rangle \tag{3.2}$$

has diagonal spectral density

$$\boldsymbol{\varphi}'(\theta)\mathbf{\Sigma}(\theta)\overline{\boldsymbol{\varphi}}(\theta) = \mathbf{\Lambda}(\theta), \tag{3.3}$$

with diagonal elements $\boldsymbol{\varphi}_k'(\theta)\mathbf{\Sigma}(\theta)\overline{\boldsymbol{\varphi}}_k(\theta) = \lambda_k(\theta)$. Hence, the univariate processes $\{Y_{kt}^*\}$'s are mutually orthogonal at all leads and lags, with variance

$$\mathrm{Var}(Y_{kt}^*) = \lambda_k := \int_{-\pi}^{\pi}\lambda_k(\theta)\mathrm{d}\theta.$$

The univariate process $\{Y^*_{kt} |\ t \in \mathbb{Z}\}$ is called $\{\mathbf{X}_t\}$*'s* k*th dynamic principal component process* or k*th dynamic principal component score process* $(k = 1, \ldots, p)$.

The properties of $\{\mathbf{X}_t\}$'s dynamic principal components extend to the time-series context the standard properties of traditional principal components associated with the eigenvalues and eigenvectors of $\{\mathbf{X}_t\}$'s covariance matrix. In particular, the variance λ_k of Y_{kt} is such that

$$\lambda_k = \begin{cases} \max_{\left\{\sum_{i=1}^p \sum_{\ell=-\infty}^{\infty} a_{i\ell k}^2 = 1\right\}} \mathrm{Var}\left(\sum_{i=1}^p \sum_{\ell=-\infty}^{\infty} a_{i\ell k} X_{i,t-\ell,k}\right), \ k = 1 \\ \max_{\left\{\sum_{i=1}^p \sum_{\ell=-\infty}^{\infty} a_{i\ell k}^2 = 1\right\}} \mathrm{Var}\left(\sum_{i=1}^p \sum_{\ell=-\infty}^{\infty} a_{i\ell k} X_{i,t-\ell,k}\right), \ k = 2, \ldots, p \\ \quad \text{subject to } Y_{kt} := \sum_{i=1}^p \sum_{\ell=-\infty}^{\infty} a_{i\ell k} X_{i,t-\ell,k} \\ \quad \text{orthogonal to } Y_{1s}, \ldots, Y_{k-1,s} \quad \text{for all } s, t \in \mathbb{Z}. \end{cases} \tag{3.4}$$

This is a direct consequence of (3.1) and the classical properties of eigenvalues.

Because $\boldsymbol{\varphi}(\theta)$ is unitary, $\left(\boldsymbol{\varphi}'(\theta)\right)^{-1} = \overline{\boldsymbol{\varphi}}(\theta)$. Therefore,

$$\underline{\overline{\boldsymbol{\varphi}}}(L)\mathbf{Y}_t = \mathbf{X}_t \tag{3.5}$$

since, from (3.1), it has spectral density

$$\overline{\boldsymbol{\varphi}}(\theta)\boldsymbol{\Lambda}(\theta)\boldsymbol{\varphi}'(\theta) = \overline{\boldsymbol{\varphi}}(\theta)\boldsymbol{\varphi}'(\theta)\boldsymbol{\Sigma}(\theta)\overline{\boldsymbol{\varphi}}(\theta)\boldsymbol{\varphi}'(\theta) = \boldsymbol{\Sigma}(\theta).$$

Developing $\underline{\overline{\boldsymbol{\varphi}}}(L)Y_t$ and taking into account the fact that $\boldsymbol{\psi}_{k\ell} = \boldsymbol{\psi}_{k,-\ell}$, (3.5) yields an expansion of $\mathbf{X}_t$ in terms of the Fourier coefficients $\boldsymbol{\psi}_{k\ell}$ of its dynamic eigenvectors $\boldsymbol{\varphi}_k(\theta)$:

$$\begin{aligned} \mathbf{X}_t = \underline{\overline{\boldsymbol{\varphi}}}(L)\mathbf{Y}^*_t &= \sum_{k=1}^{p} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t+\ell}\, \boldsymbol{\psi}_{k\ell} \\ &= \sum_{k=1}^{p} \sum_{\ell=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle \mathbf{X}_{t+\ell-m},\, \boldsymbol{\psi}_{k\ell} \rangle\, \boldsymbol{\psi}_{k\ell}. \end{aligned} \tag{3.6}$$

That expansion of $\mathbf{X}_t$ is to be compared with (2.1), based on the eigenvectors of its covariance matrix $\mathbf{C} = \boldsymbol{\Gamma}_0$.

It readily follows from (3.4) that the truncation $\sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t+\ell}\, \boldsymbol{\psi}_{k\ell}$ of the dynamic expansion (3.6) provides, for any $1 \leq K \leq p$, the "best" reduction of $\{\mathbf{X}_t\}$ to a K-tuple of linear combinations of its present, past, and future values. More precisely, for any sequence

$$\boldsymbol{\varepsilon}_{k\ell}, \quad k = 1, \ldots p, \ \ell \in \mathbb{Z}$$

of p-dimensional vectors such that $\sum_{\ell=-\infty}^{\infty} \|\boldsymbol{\varepsilon}_{k\ell}\| < \infty$, letting

$$Y_{kt} := \sum_{\ell=-\infty}^{\infty} \langle \mathbf{X}_{t-\ell}, \boldsymbol{\varepsilon}_{k\ell} \rangle,$$

we have, for all $K = 1, \ldots, p$,

$$\mathrm{E}\Big\| X_t - \sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t+\ell} \boldsymbol{\psi}_{k\ell} \Big\|^2 \leq \mathrm{E}\Big\| X_t - \sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y_{k,t+\ell} \boldsymbol{\varepsilon}_{k\ell} \Big\|^2.$$

This inequality holds, in particular, for the "static" expansion (2.1), where

$$\boldsymbol{\varepsilon}_{k\ell} = \mathbf{e}_k \quad \text{for } \ell = 0, \qquad \boldsymbol{\varepsilon}_{k\ell} = \boldsymbol{\varepsilon}_{k\ell} = \mathbf{0} \quad \text{else,}$$

with strict inequality as soon as the $\mathbf{X}_t$'s exhibit serial correlation, thus confirming the suboptimality of static principal components in a time series context.

As for mutual orthogonality, noncorrelation of $Y^*_{k_1 t}$ and $Y^*_{k_2 s}$ for $k_1 \neq k_2$, all s and t, readily follows from (3.3).

The dimension reduction based on dynamic principal components thus recovers, in the context of vector-valued time series, the desired properties of standard principal components-based dimension reduction for i.i.d. random vectors. Parallel to (2.3), the quality of that dimension reduction can be measured via the *relative mean square approximation error*

$$\mathrm{RMSE}_{\mathrm{dyn}}(K) := \sum_{k=K+1}^{p} \int_{-\pi}^{\pi} \lambda_k(\theta) d\theta \Big/ \sum_{k=1}^{p} \int_{-\pi}^{\pi} \lambda_k(\theta) d\theta, \quad K = 1, \ldots, p. \tag{3.7}$$

3.2 Functional dynamic principal components.

Brillinger's concept was limited to the case of vector-valued time series, but a functional version of the same has been proposed recently by Hörmann et al. (2015). Such extension requires a neat treatment of spectral analysis for functional time series, a theory that was not available until the contribution of Panaretos and Tavakoli (2013), where we refer to for details.

Let $\{X_t\}$ be an $L^2([0,1])$-valued functional process as in Section 2.2. Provided that $\sum_{\ell=-\infty}^{\infty} \|\mathrm{Cov}(X_t, X_{t-\ell})\|_{\mathcal{S}} < \infty$, where $\| \, . \, \|_{\mathcal{S}}$ stands for the Hilbert-Schmidt norm, Panaretos and Tavakoli (2013) show that

$$\mathcal{F}^X_\theta := \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(X_t, X_{t-\ell}) e^{-\imath \ell \theta}$$

exists, and is also a Hilbert-Schmidt operator: call it the *spectral density operator* of the functional process $\{X_t\}$.

In Section 2.2, dimension reduction was based on "static" functional transformations Ψ_0, say, mapping the L^2-valued variable X_t to an $\mathbb{R}^K$-valued one

$$\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Kt})' := \Psi_0(X_t)$$

(K here is an arbitrary integer). In a dynamic context, we rather need linear *functional filters*, mapping an L^2-valued functional process to an $\mathbb{R}^K$-valued variable, of the form

$$\begin{aligned} (\ \ldots, X_{t-1}, X_t, X_{t+1}, \ldots\ ) \mapsto \mathbf{Y}_t &= (Y_{1t}, \ldots, Y_{Kt})' \\ &:= \sum_{\ell=-\infty}^{\infty} \Psi_\ell(X_{t-\ell}), \end{aligned} \tag{3.8}$$

with $\Psi_\ell : L^2 \to \mathbb{R}^K$.

In order to qualify as a "good" dimension reduction, those Ψ_ℓ's should be such that Y_{kt} and $Y_{k't'}$ be uncorrelated, unless $k = k'$, for all t and t' (orthogonality at all leads and lags—autocorrelations are allowed, but cross-correlations are ruled out). This holds if and only if the (traditional) spectral density matrix

$$\mathbf{\Sigma}^{\mathbf{Y}}(\theta) := \frac{1}{2\pi} \sum_{\ell=-\infty}^{\infty} \mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-\ell}) e^{-\imath\ell\theta}$$

of the K-dimensional vector process $\{\mathbf{Y}_t\}$ is diagonal for (almost) all θ in $[-\pi, \pi]$. Accordingly, let us investigate the relation between the spectral density matrix $\mathbf{\Sigma}^{\mathbf{Y}}(\theta)$ and the spectral density operator $\mathcal{F}_\theta^X$.

For fixed θ, the spectral density operator $\mathcal{F}_\theta^X$ (for every θ, a non-negative self-adjoint Hilbert-Schmidt operator) has similar properties as a covariance operator: for any $f \in L^2$, the image $\mathcal{F}_\theta^X(f)$ of f admits the eigendecomposition

$$\mathcal{F}_\theta^X(f) = \sum_{k=1}^{\infty} \lambda_k(\theta) \langle f, \varphi_k(\theta) \rangle \varphi_k(\theta)$$

where $\lambda_k(\theta)$ and $\varphi_k(\theta)$ are $\mathcal{F}_\theta^X$'s eigenvalues (in descending order of magnitude) and eigenfunctions, respectively.

The relation between functional filters and frequency-indexed operators is similar to, but more delicate than, the relation between matrix filters and frequency-indexed matrices. Since each Ψ_ℓ in the functional filter (3.8) is linear, $\Psi_\ell(f)$ admits, for some K-tuple $\big(\psi_{1\ell}, \ldots, \psi_{K\ell}\big)$ of $L^2([0,1])$ elements, the representation *(Riesz representation)*

$$\Psi_\ell(f) = (\langle f, \psi_{1\ell} \rangle, \ldots, \langle f, \psi_{K\ell} \rangle)', \qquad f \in L^2.$$

Let

$$\psi_k^*(\theta) := \sum_{\ell=-\infty}^{\infty} \psi_{k\ell} e^{\imath\ell\theta} \quad k = 1, \ldots, K$$

(the $\psi_{k\ell}$'s thus are the Fourier coefficients of the ψ_k^*'s). Then, the following relation holds between the spectral density operator $\mathcal{F}_\theta^X$ and the spectral density matrix $\mathbf{\Sigma}^{\mathbf{Y}}(\theta)$ of the filtered p-dimensional real process $\{\mathbf{Y}_t\}$ in (3.8):

$$\mathbf{\Sigma}^{\mathbf{Y}}(\theta) = \begin{pmatrix} \langle \mathcal{F}_\theta^X(\psi_1^*(\theta)), \psi_1^*(\theta) \rangle & \cdots & \langle \mathcal{F}_\theta^X(\psi_K^*(\theta)), \psi_1^*(\theta) \rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{F}_\theta^X(\psi_1^*(\theta)), \psi_K^*(\theta) \rangle & \cdots & \langle \mathcal{F}_\theta^X(\psi_K^*(\theta)), \psi_K^*(\theta) \rangle \end{pmatrix}$$

(we refer to Hörmann et al. (2015) for details). Let us choose the Ψ_ℓ's (that is, the functional filters (3.8)) in such a way that

$$\psi_k^*(\theta) = \sum_{\ell=-\infty}^{\infty} \psi_{k\ell} e^{\imath\ell\theta} = \varphi_k(\theta), \quad k = 1, \ldots, K,$$

that is, choose as $\psi_{k\ell}$'s the Fourier coefficients

$$\psi_{k\ell}^* = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_k(\theta) e^{-\imath\ell\theta} d\theta, \quad \ell \in \mathbb{Z}$$

of $\mathcal{F}_\theta^X$'s kth eigenfunction $\varphi_k(\theta)$ (note again that the $\psi_{k\ell}^*$'s, just as the $\psi_k^*(\theta)$'s and $\varphi_k(\theta)$'s, are L^2 functions, viz. $\psi_{k\ell}^*(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_k(\theta)(u) e^{-\imath\ell\theta} d\theta$). We then obtain

$$Y_{kt}^* := \sum_{\ell=-\infty}^{\infty} \langle X_{t-\ell}, \psi_{k\ell}^* \rangle, \quad t \in \mathbb{Z},\ k = 1, \ldots, K.$$

Call the univariate real-valued process $\{Y_{kt}^*\}$ the *kth dynamic functional principal component (score)* of $\{X_t\}$ (formally, the definition thus is the same as in the $\mathbb{R}^p$-valued case).

The following elementary properties readily follow from the definition:

(i) the series defining Y_{kt}^* is mean square convergent; Y_{kt}^* is real, and has variance $\int_{-\pi}^{\pi} \lambda_k(\theta) d\theta$;

(ii) for $k \neq k'$, the principal components Y_{kt} and $Y_{k's}$ are uncorrelated for all s, t;

(iii) if the X_t's are serially uncorrelated, the functional dynamic principal components coincide with the static ones.

The original process $\{X_t\}$ then admits the *dynamic expansion* (a functional version of the dynamic expansion previously obtained in $\mathbb{R}^p$)

$$X_t = \sum_{k=1}^{\infty} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t+\ell}\, \psi^*_{k\ell}, \quad t \in \mathbb{Z}$$

with $Y^*_{kt} := \sum_{\ell=-\infty}^{\infty} \langle X_{t-\ell}, \psi^*_{k\ell} \rangle$, and that expansion enjoys the Karhunen-Loève optimality property: for any L^2-valued sequence $\psi_{k\ell}$, $k,\ \ell \in \mathbb{Z}$ such that $\sum_{\ell=-\infty}^{\infty} \|\psi_{k\ell}\| < \infty$, and any collection Y_{kt}, $k,\ \ell \in \mathbb{Z}$ of real-valued random variables,

$$\mathrm{E}\Big\| X_t - \sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t-\ell}\ \psi^*_{k\ell} \Big\|^2 \leq \mathrm{E}\Big\| X_t - \sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y_{k,t-\ell}\, \psi_{k\ell} \Big\|^2$$

for all $K \in \mathbb{N}$. In analogy to (2.7), we have

$$\mathrm{E}\Big\| X_t - \sum_{k=1}^{K} \sum_{\ell=-\infty}^{\infty} Y^*_{k,t-\ell}\ \psi^*_{k\ell} \Big\|^2 = \sum_{k=K+1}^{\infty} \int_{-\pi}^{\pi} \lambda_m(\theta) d\theta; \tag{3.9}$$

here again, a convenient measure of perfomance is the *relative mean square approximation error*

$$\mathrm{RMSE}_{\mathrm{dyn}}(K) := \sum_{k=K+1}^{\infty} \int_{-\pi}^{\pi} \lambda_k(\theta) d\theta \Big/ \sum_{k=1}^{\infty} \int_{-\pi}^{\pi} \lambda_k(\theta) d\theta. \tag{3.10}$$

4 Numerical illustration.

In this section, we are providing a very brief numerical illustration of the potential benefits of dynamic principal components, as opposed to the usual static ones, in dimension reduction for functional time series. To this end, based on a very simple functional AR(1) model, we generated synthetic functional data and compared the performances of static and dynamic dimension reduction in terms of their relative respective mean square approximation errors $\mathrm{RMSE}_{\mathrm{stat}}(K)$ and $\mathrm{RMSE}_{\mathrm{dyn}}(K)$; see (2.8) and (3.10).

Consider the functional AR(1) model

$$X_t(u) = \int_0^1 \psi(u,v) X_{t-1}(v) dv + Z_t(u), \tag{4.1}$$

where $\psi(u,v)$ is some square-integrable kernel function and Z_t is i.i.d. functional noise. To guarantee a stationary solution we consider operators ψ

with $\kappa^2 := \int_0^1 \int_0^1 \psi^2(u,v)dudv < 1$. Let (e_j), $j \in \mathbb{N}$ denote some orthonormal basis of $L^2([0,1])$, yielding, for some possibly infinite d, the expansions (equality/convergence in the mean square sense)

$$\psi(u,v) \;=\; \sum_{i=1}^{d}\sum_{j=1}^{d} \psi_{ij} e_i(u) e_j(v),$$

$$Z_t(u) = \sum_{i=1}^{d} z_{ti} e_i(u), \quad \text{and} \quad X_t(u) = \sum_{i=1}^{d} x_{ti} e_i(u).$$

Taking, for lags $1 \le k \le d$, inner products $\langle X_t, e_k \rangle$ on both sides of (4.1), we obtain a VAR(1) representation of the functional AR(1) process, of the form

$$\boldsymbol{X}_t = \boldsymbol{\Psi} \boldsymbol{X}_{t-1} + \boldsymbol{Z}_t, \tag{4.2}$$

where

$$\boldsymbol{\Psi} = \Big((\psi_{ij} \colon 1 \le i,j \le d)\Big), \quad \boldsymbol{Z}_t = (z_{t1}, \ldots, z_{td})' \quad \text{and} \quad \boldsymbol{X}_t = (x_{t1}, \ldots, x_{td})'.$$

In theory, that VAR model is infinite-dimensional, but in practice we work with finite values of d.

What is special about this VAR model (4.2) is that the compactness of ψ and the covariance operator of Z_t implies that the components z_{tk} and $\psi_{k\ell}$ converge to zero as k and ℓ tend to infinity. For example, note that $\mathrm{E}\|Z_t\|^2 = \mathrm{tr}(\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \big(\Sigma_{ij}\big)$ is the covariance matrix of $\boldsymbol{Z}_t$. Consequently, $\sum_{i=1}^{d} \Sigma_{ii} < \infty$, even when d is infinite. Accordingly, the data-generating process of our simulation setup was set as

$$\boldsymbol{Z}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d(\boldsymbol{0}, \boldsymbol{\Sigma}), \quad \text{with} \quad \boldsymbol{\Sigma} = \Big((\Sigma_{ij})\Big) \quad \text{and} \quad \Sigma_{ij} = (i^2 + j^2)^{-0.6};$$

$$\boldsymbol{\Psi} = \frac{\kappa\ \mathbf{v}\mathbf{v}'}{\sqrt{\mathbf{v}'\mathbf{v}}}, \text{ with } \mathbf{v} = \big(v_i\big),\ v_i \overset{\text{ind}}{\sim} \mathcal{N}(0, e^{-i}) \text{ and } \kappa = 0.5,\ 0.7,\ 0.9.$$

For numerical implementation, we used the function `fts.rar()` of the R-package `freqdom.fda` (see Hörmann and Kidziński (2017)), which requires the data to lie in the finite-dimensional subspace of $L^2([0,1])$ spanned by the first d Fourier basis functions, namely $e_1(u) = 1$, $e_2(u) = \sqrt{2}\cos(2\pi u)$, $e_3(u) = \sqrt{2}\sin(2\pi u)$, etc. Here we chose $d = 15$. The results remain similar for larger d. We considered three values for the dependence coefficient: $\kappa \in \{0.5, 0.7, 0.9\}$; the larger κ, the stronger the auto- and cross-correlations. The differences between $\mathrm{RMSE}_{\mathrm{dyn}}(K)$ and $\mathrm{RMSE}_{\mathrm{stat}}(K)$ are maximal for $K = 1$, and decrease with K. However, $\mathrm{RMSE}_{\mathrm{dyn}}(K)$ appears to be less dispersed than $\mathrm{RMSE}_{\mathrm{stat}}(K)$, which is another highly non negligible advantage of the dynamic approach over the static one.

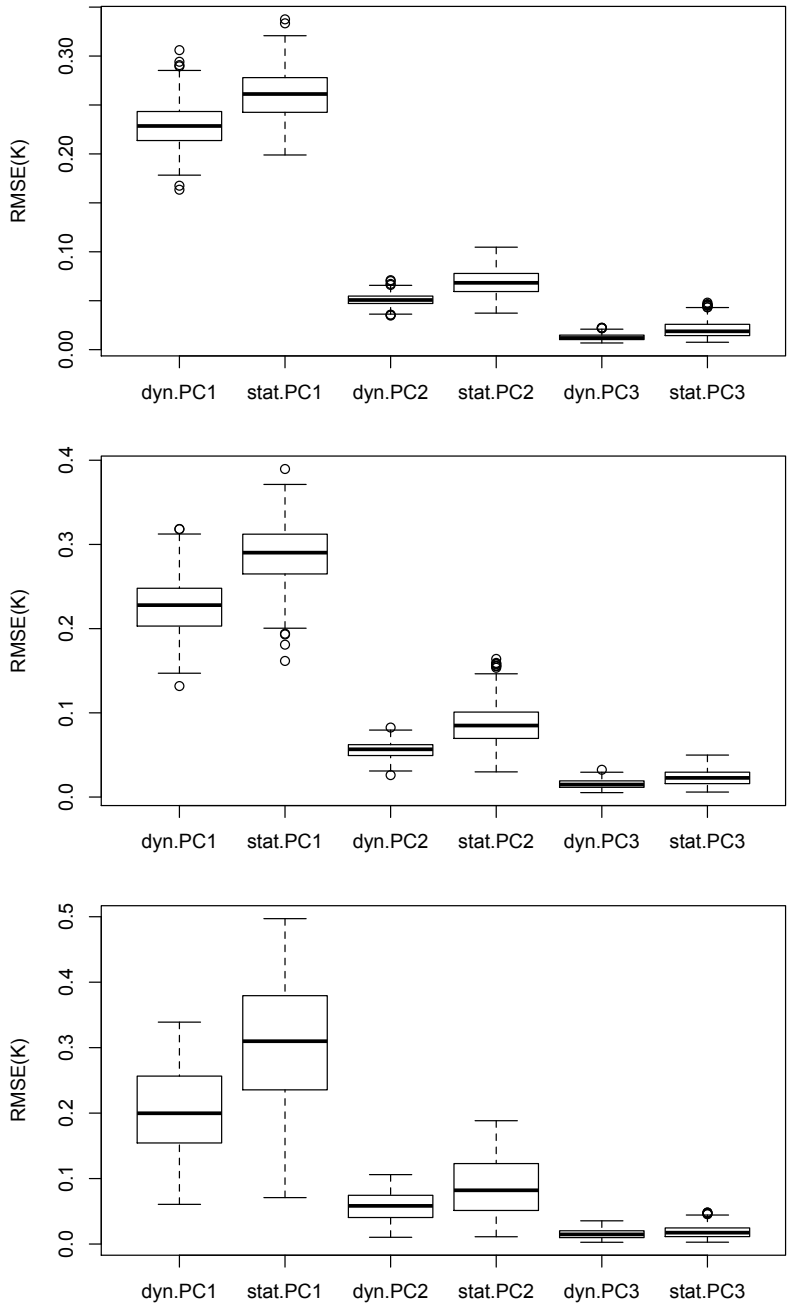

Figure 1: Boxplots of $\mathrm{RMSE}_{\mathrm{stat}}(K)$ and $\mathrm{RMSE}_{\mathrm{dyn}}(K)$ for $K = 1, 2, 3$ principal components, based on 500 iterations, $n = 250$ and $\kappa = 0.5$ (top), $\kappa = 0.7$ (middle) and $\kappa = 0.9$ (bottom).

5 Conclusion

Static principal components and functional principal components, which are everyday practice in a large number of applications, are inadequate and suboptimal in the context of dimension reduction for time series. The resulting expansion indeed does not enjoy, in the presence of serial correlation, the optimality property that makes their success in the i.i.d. case. Moreover, the principal components themselves are no longer mutually uncorrelated, hence do not lead to a neat interpretation of their respective contributions to the total variance. *Dynamic* principal components and functional *dynamic* principal components, by taking into account serial dependence, quite on the contrary recover both properties. Consistent empirical versions are computationally feasible, and simulations, as well as real data examples, are quite encouraging.

On the other hand, a negative feature of dynamic principal components and their functional counterparts is that the filters defining the Y_{kt}^*'s are two-sided, so that, when replacing those filters with consistent estimators (as in Forni et al. 2000), the quality of the resulting dimension reduction deteriorates at the end of the observation period. A solution to that problem, yielding one-sided filtering only, has been provided by Forni and Lippi (2011), with improvements in Forni et al. (2015, 2017). A more direct approach has been proposed recently by Peña and Yohai (2016) and Peña et al. (2017), but its functional counterpart and exact properties in the context of dimension reduction require some further investigation.

Finally, dynamic principal components are a basic ingredient in the General Dynamic Factor methods used (see Forni et al. 2000, 2005, 2015, 2017; Hallin and Lippi 2013) in the analysis of large panels of time series data; whether this can be extended to functional or partly functional panels is the subject of ongoing research.

References

[1] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000) The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics* **82**, 540–554.

[2] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2015) Dynamic factor models with infinite-dimensional factor space: one-sided representations. *Journal of Econometrics* **185**, 359–37.

[3] Forni, M., M. Hallin, M. Lippi, and P. Zaffaroni (2017) Dynamic factor models with infinite-dimensional factor space: asymptotic analysis, *Journal of Econometrics* **199**, 74–92.

[4] Forni, M. and M. Lippi (2011) The generalized dynamic factor model: one-sided representation results. *Journal of Econometrics* **163**, 23–28.

[5] Hallin, M. and M. Lippi (2013) Factor models in high-dimensional time series: a time-domain approach, *Stochastic Processes and their Applications* **123**, 2678-2695.

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

[7] Hörmann, S. and Ł. Kidziński (2017). `freqdom.fda`: Functional Time Series: Dynamic Functional Principal Components; available at `https://CRAN.R-project.org/package=freqdom.fda`.

[8] Karhunen, K. (1946). Zur Spektraltheorie Stochastischer Prozesse. *Annales Academiæ Scientiarum Fennicæ Ser.* A.I. *Mathematica–Physica* **34**, 1–7.

[9] Karhunen, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Annales Academiæ Scientiarum Fennicæ Ser.* A.I. *Mathematica–Physica* **37**, 1–79.

[10] Loève, M. (1955). *Probability Theory*. Princeton, N.J.: VanNostrand.

[11] Peña, D. and V.J. Yohai (2016) Generalized dynamic principal components, *Journal of the American Statistical Association*, available online at `http://dx.doi.org/10.1080/01621459.2015.1072542`.

[12] Peña, D., E. Smucler, and V.J. Yohai (2017) Forecasting multiple time series with one-sided dynamic principal components; available at `arXiv:1708.04705v1`.