ECARES

High-Dimensional Functional Factor Models

Marc Hallin
SBS-EM, ECARES, Université libre de Bruxelles

Gilles Nisol
SBS-EM, ECARES, Université libre de Bruxelles
Katholieke Universiteit Leuven

Shahin Tavakoli
Department of Statistics, University of Warwick, UK

June 2019

ECARES working paper 2019-16

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

High-Dimensional Functional Factor Models

Gilles Nisol, Shahin Tavakoli* and Marc Hallin*

ESAT - STADIUS,
Katholieke Universiteit Leuven
3001 Leuven
Belgium
e-mail: gnisol@esat.kuleuven.be

Department of Statistics
University of Warwick
CV4 7AL Coventry
United Kingdom
e-mail: s.tavakoli@warwick.ac.uk

ECARES and Département de Mathématique
Université Libre de Bruxelles CP 114
50, avenue F.D. Roosevelt
1050 Bruxelles
Belgium
e-mail: mhallin@ulb.ac.be

Abstract: . In this paper, we set up the theoretical foundations for a high-dimensional functional factor model approach in the analysis of large panels of functional time series (FTS). We first establish a representation result stating that if the first r eigenvalues of the covariance operator of a cross-section of N FTS are unbounded as N diverges and if the $(r+1)$th one is bounded, then we can represent each FTS as a sum of a common component driven by r factors, common to (almost) all the series, and a weakly cross-correlated idiosyncratic component (all the eigenvalues of the idiosyncratic covariance operator are bounded as $N \to \infty$). Our model and theory are developed in a general Hilbert space setting that allows for panels mixing functional and scalar time series. We then turn to the estimation of the factors, their loadings, and the common components. We derive consistency results in the asymptotic regime where the number N of series and the number T of time observations diverge, thus exemplifying the "blessing of dimensionality" that explains the success of factor models in the context of high-dimensional (scalar) time series. Our results encompass the scalar case, for which they reproduce and extend, under weaker conditions, well-established results (Bai & Ng 2002). We provide numerical illustrations that corroborate the convergence rates predicted by the theory, and provide finer understanding of the interplay between N and T for estimation purposes. We conclude with an empirical illustration on a dataset of intraday S&P100 and Eurostoxx 50 stock returns, along with their scalar overnight returns.

*The author(s) would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Statistical Scalability where work on this paper was undertaken. This work was supported by EPSRC grant no EP/K032208/1.

MSC 2010 subject classifications: Primary 62M10, 62H25; secondary 60G10, 62P05.
Keywords and phrases: Functional time series, High-dimensional time series, Factor model, Panel data, Functional data analysis..

1. Introduction

Throughout the last decades, researchers have been dealing with datasets of increasing size and complexity. In particular, Functional Data Analysis (FDA; see e.g. Ramsay & Silverman 2005, Ferraty & Vieu 2006, Horváth & Kokoszka 2012, Hsing & Eubank 2015, Wang et al. 2015) has received much interest and, in view of its relevance in a number of applications, fast growing popularity. In FDA, the observations are taking values in some functional space, usually some Hilbert space H—often, in practice, the space $L^2([0,1],\mathbb{R})$ of squared integrable functions. When an ordered sequence of functional observations exhibits serial dependence, we enter the realm of *Functional Time Series* (FTS) (Hörmann & Kokoszka 2010, 2012). Many standard univariate and low-dimensional multivariate time-series methods have been adapted to this functional setting, either using a time-domain approach (Kokoszka & Reimherr 2013*a*,*b*, Hörmann et al. 2013, Aue et al. 2014, 2015, Horváth et al. 2014, Aue et al. 2017, Górecki et al. 2018, Bücher et al. 2018, Gao et al. 2018), a frequency domain approach under stationarity assumptions (Panaretos & Tavakoli 2013*a*,*b*, Hörmann et al. 2015, Tavakoli & Panaretos 2016, Hörmann et al. 2018, Rubín & Panaretos 2018, Guo & Qiao 2018) or under local stationarity assumptions (van Delft et al. 2017, van Delft & Eichler 2018, van Delft & Dette 2018, Barigozzi et al. 2019).

Parallel to this development of functional time series analysis, data in high dimensions (e.g. Bühlmann & van de Geer 2011, Fan et al. 2013) have become pervasive in data sciences and related disciplines where, under the name of Big Data, they constitute one of the most active subject of contemporary statistical research.

This contribution stands at the intersection of those two strands of literature, cumulating the challenges of function-valued observations and those of high dimension. Datasets, in this context, consist of large collections of N scalar or functional time series—equivalently, functional time series in high dimension (from fifty, say, to several hundreds)—observed over a period of time T. Typical examples are continuous-time series of concentrations for a large number of pollutants, or/and collected over a large number of sites, daily series of returns observed at high intraday frequency for a large collection of stocks, or intraday energy consumption curves (available, for instance, at `data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households`), to name only a few. Not all component series in the dataset are required to be function-valued, though, and mixed panels of scalar and functional series can be considered as well. In order to model such datasets, we develop a class of *high-dimensional functional factor models* inspired by the factor model approaches developed, mostly, in time series econometrics, which have proven effective, flexible, and quite efficient in the scalar case.

Factor models for FTS are largely unexplored. The only developments in this direction (that we are aware of) are Hays et al. (2012), who consider a Gaussian likelihood approach to functional dynamic factor modelling, and Kokoszka et al. (2015), who consider functional dynamic factor models where the factors are functional; both are limited to *one* FTS, though, whereas our approach is for *large panels* of FTS. More recently, Gao et al. (2018) have used factor models for forecasting panels of FTS, but they use a two-stage approach combining a separate dynamic functional PCA on each FTS in the panel, followed by a combination of separate scalar factor models (one on each PC score). This implicitly assumes that the number of relevant principal components per FTS is the same (which is quite restrictive), and is linked to the number of overall factors. Our approach is mostly motivated by the time series econometrics literature, and differs from these papers because we consider models where the factors are scalar and the loadings are functional; moreover, we do not make Gaussian assumptions. Our approach is principled, we do not impose a model through a two-stage procedure, and do not base our model on a PCA with the same truncation level on each separate FTS.

Early instances of factor model methods for time series can be traced back to the pioneering contributions by Geweke (1977), Sargent & Sims (1977), Chamberlain (1983), and Chamberlain & Rothschild (1983). The factor models considered in Geweke and Sargent and Sims are *exact*, that is, involve mutually orthogonal (all leads, all lags) idiosyncratic components, a most restrictive assumption that cannot be expected to hold in practice. Chamberlain (1983) and Chamberlain & Rothschild (1983) are relaxing this exactness assumption into an assumption of *mildly* cross-correlated idiosyncratics (the so-called "approximate factor models"). Finite-N identifiability is the price to be paid for that relaxation; the resulting model, however, remains asymptotically (as N tend to infinity) identified, which is perfectly in line with the spirit of high-dimensional asymptotics. This idea of an factor models in high dimensions has been picked up and developed, mostly, by Stock & Watson (2002*a*,*b*), Bai & Ng (2002), Bai (2003), Forni et al. (2000), and their many followers; see also Forni et al. (2015) and Forni et al. (2017) for extensions to the so-called *generalized* or *general dynamic factor model*.

Our objective here is to propose a representation theorem (analogue to the classical results of Chamberlain & Rothschild 1983, Chamberlain 1983) linking high-dimensional functional factor models to properties of the eigenvalues of the panel covariance operator (Theorems 2.2 and, drawing inspiration from Stock & Watson (2002*a*,*b*), Bai & Ng (2002) 2.3), to develop the corresponding estimation theory for the unobserved factors, loadings, and common component (see Theorems 3.1, 3.4, and 3.5). This is laying the theoretical foundations for modeling high-dimensional functional time series via factor models. While our contributions are for panels of FTS, our results encompass and extend those of factor models in large dimensions for scalar time series (or "approximate factor models"), for which we reproduce and extend some results under weaker assumptions than available in the literature (Chamberlain & Rothschild 1983, Bai & Ng 2002, Stock & Watson 2002*a*).

The paper is organized as follows. In Section 2, we introduce high-dimensional functional factor models for panels of functional time series (FTS) and show that this class of models can be characterized by conditions on the spectrum of the panel. In Section 3, we introduce an estimator of the factors through an eigendecomposition of the observed panel data, and study the consistency of our estimator. In Section 4, we conduct some numerical experiments, and provide an empirical illustration of our approach in Section 5. We conclude in Section 6 with a discussion. Technical results and all the proofs, as well as additional simulation results, are contained in the Online Supplement.

2. Model and Representation Theorem

Since our goal is to develop a model for panels of time series that could be either functional or scalar, we need to introduce some notation, in particular for vectors or matrices of Hilbert space elements, and their representations as operators. While this could seem a priori tedious, it will actually be very useful later on, as it simplifies the exposition, makes proofs clearer, and allows for weaker assumptions.

2.1. Notation

Throughout, we denote by

$$\mathcal{X}_{N,T} := \{X_{it},\ i = 1, \dots, N,\ t = 1, \dots, T\}$$

an observed $N \times T$ panel (cross-section) of time series, where the random variables X_{it} take values in a separable Hilbert space H_i equipped with the inner product $\langle \cdot, \cdot \rangle_i$. Those series can be of different types. A case of interest is the one for which some series are scalars ($H_i = \mathbb{R}$) and some others are square-integrable functions from $[0, 1]$ to $\mathbb{R}$ ($H_i = L^2([0, 1])$). We tacitly assume that all X_{it}'s are random elements with mean zero and finite second-order moments defined on some common probability space $(\Omega, \mathcal{F}, \mathbb{P})$; we also assume that $\mathcal{X}_{N,T}$ constitutes the finite realization of some a second-order stationary double-indexed process $\mathcal{X} := \{X_{it},\ i \in \mathbb{N},\ t \in \mathbb{Z}\}$.

Define $\boldsymbol{H}_N := H_1 \oplus H_2 \oplus \cdots \oplus H_N$, with typical elements of the form $\boldsymbol{v} := (v_1, v_2, \dots, v_N)'$ or $\boldsymbol{w} := (w_1, w_2, \dots, w_N)'$. The space $\boldsymbol{H}_N$, naturally equipped with the inner product

$$\langle \boldsymbol{v}, \boldsymbol{w} \rangle_{\boldsymbol{H}_N} := \sum_{i=1}^{N} \langle v_i, w_i \rangle_i ,$$

is a Hilbert space. Writing $\langle \cdot, \cdot \rangle$ for $\langle \cdot, \cdot \rangle_{\boldsymbol{H}_N}$ when no confusion is possible, let $\|\cdot\| := \langle \cdot, \cdot \rangle^{1/2}$ be the resulting norm. Write $\mathcal{L}(H_1, H_2)$ for the space of bounded (linear) operators from H_1 to H_2, and use the shorthand notation $\mathcal{L}(H)$ for $\mathcal{L}(H, H)$. Denote the operator norm of $V \in \mathcal{L}(H_1, H_2)$ by

$$|||V|||_\infty := \sup_{x \in H_1, x \neq 0} \|Vx\| / \|x\|,$$

and write V^T for the adjoint of V, which satisfies $\langle Vu_1, u_2 \rangle = \langle u_1, V^\mathsf{T} u_2 \rangle$ for all $u_1 \in H_1, u_2 \in H_2$. In particular, we have (see Hsing & Eubank 2015)

$$|||V|||_\infty = \left|\left|\left| V^\mathsf{T} \right|\right|\right|_\infty = \left|\left|\left| V^\mathsf{T} V \right|\right|\right|_\infty^{1/2}.$$

In order to make our results readable and facilitate proofs, we need to introduce an extension of classical matrix algebra (and linear mappings between Euclidean spaces) to matrix mappings between direct sums of Hilbert spaces (such as $\boldsymbol{H}_N$). For an element $v_i \in H_i$, we write, with slight abuse of notation, $v_i \in \mathcal{L}(\mathbb{R}, H_i)$ for the mapping $v_i : \alpha \mapsto \alpha v_i$ from $\mathbb{R}$ to H_i, and $v_i^\mathsf{T} \in \mathcal{L}(H_i, \mathbb{R})$ for its adjoint, which is defined by

$$H_i \ni f \longmapsto v_i^\mathsf{T} f := v_i^\mathsf{T}(f) := \langle f, v_i \rangle_i \,.$$

Similarly, we denote by $\boldsymbol{v} \in \mathcal{L}(\mathbb{R}, \boldsymbol{H}_N)$ the mapping $a \mapsto a\boldsymbol{v}$ from $\mathbb{R}$ to $\boldsymbol{H}_N$, and by $\boldsymbol{v}^\mathsf{T} = \left(v_1^\mathsf{T}, \ldots, v_N^\mathsf{T}\right) \in \mathcal{L}(\boldsymbol{H}_N, \mathbb{R})$ its adjoint, from $\boldsymbol{H}_N$ to $\mathbb{R}$:

$$\boldsymbol{H}_N \ni \boldsymbol{w} \longmapsto \boldsymbol{v}^\mathsf{T} \boldsymbol{w} := v_1^\mathsf{T} w_1 + \cdots + v_N^\mathsf{T} w_N.$$

Unlike $(\cdot)'$ which denotes *transposition* (that does not change the nature of the elements), $(\cdot)^\mathsf{T}$ refers to *adjunction.* Note, in particular, that $\boldsymbol{v}^\mathsf{T}\boldsymbol{w} = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$ and $\boldsymbol{v}^\mathsf{T}\boldsymbol{v} = \|\boldsymbol{v}\|^2$. Letting $\boldsymbol{v}_j = (v_{1j}, \ldots, v_{Nj})' \in \boldsymbol{H}_N$ with $v_{ij} \in H_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, r$, define the linear mapping

$$\boldsymbol{V} = \begin{pmatrix} v_{11} & \cdots & v_{1r} \\ v_{21} & \cdots & v_{2r} \\ \vdots & \ddots & \vdots \\ v_{N1} & \cdots & v_{Nr} \end{pmatrix} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_r) \in \mathcal{L}(\mathbb{R}^r, \boldsymbol{H}_N)$$

as $(a_1, \ldots, a_r)' \mapsto \boldsymbol{v}_1 a_1 + \cdots + \boldsymbol{v}_r a_r$, with adjoint

$$\boldsymbol{V}^\mathsf{T} := \begin{pmatrix} v_{11}^\mathsf{T} & \cdots & v_{N1}^\mathsf{T} \\ v_{12}^\mathsf{T} & \cdots & v_{N2}^\mathsf{T} \\ \vdots & \ddots & \vdots \\ v_{1r}^\mathsf{T} & \cdots & v_{Nr}^\mathsf{T} \end{pmatrix} \in \mathcal{L}(\boldsymbol{H}_N, \mathbb{R}^r)$$

mapping $\boldsymbol{w}$ to $\boldsymbol{V}^\mathsf{T}\boldsymbol{w} := (\boldsymbol{v}_1^\mathsf{T}\boldsymbol{w}, \ldots, \boldsymbol{v}_r^\mathsf{T}\boldsymbol{w})'$. If $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_T) \in \mathbb{R}^{r \times T}$, then $\boldsymbol{V}\boldsymbol{A}$ should be understood as $(\boldsymbol{V}\boldsymbol{a}_1, \ldots, \boldsymbol{V}\boldsymbol{a}_T) \in \mathcal{L}(\mathbb{R}^T, \boldsymbol{H}_N)$. Similarly, if $\boldsymbol{W}_N = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T) \in \mathcal{L}(\mathbb{R}^T, \boldsymbol{H}_N)$, then

$$\boldsymbol{V}^\mathsf{T}\boldsymbol{W} := (\boldsymbol{V}^\mathsf{T}\boldsymbol{w}_1, \ldots, \boldsymbol{V}^\mathsf{T}\boldsymbol{w}_T) \in \mathcal{L}(\mathbb{R}^T, \mathbb{R}^r).$$

Note that this notation is compatible with the usual matrix multiplication: for instance, $\boldsymbol{V}\boldsymbol{V}^\top = \boldsymbol{v}_1\boldsymbol{v}_1^\top + \cdots + \boldsymbol{v}_r\boldsymbol{v}_r^\top \in \mathcal{L}(\boldsymbol{H}_N)$, and $\boldsymbol{V}^\top\boldsymbol{V}$ is the matrix with (i,j)th entry $\boldsymbol{v}_i^\top\boldsymbol{v}_j = \langle \boldsymbol{v}_j, \boldsymbol{v}_i\rangle \in \mathbb{R}$.

To work with our panel data, we will use the following notation. We let $\boldsymbol{X}^i := (X_{i1}, X_{i2}, \ldots, X_{iT}) \in \mathcal{L}(\mathbb{R}^T, H_i)$, $\boldsymbol{X}_t := (X_{1t}, X_{2t}, \ldots, X_{Nt})' \in \boldsymbol{H}_N$, and $\boldsymbol{X}_{NT} := (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T) \in \mathcal{L}(\mathbb{R}^T, \boldsymbol{H}_N)$. In order to keep the presentation simple, the dependence of $\boldsymbol{X}_t$ on N and the dependence of $\boldsymbol{X}^i$ on T do not explicitly appear in the notation. We denote by $\lambda^X_{N,1}, \lambda^X_{N,2}, \ldots$ the eigenvalues of the covariance of $(X_{1t}, \ldots, X_{Nt})'$, in decreasing order of magnitude; in view of stationarity, these eigenvalues do not depend on t. Finally, denote by $\|.\|^2_{L^2(\Omega)}$ the variance, and by $\langle \cdot, \cdot \rangle_{L^2(\Omega)}$ the covariance, of real-valued random variables. Unless otherwise mentioned, convergence of sequences of random variables is in mean square.

2.2. Model

The basic idea in all factor-model approaches to the analysis of high-dimensional time series consists in decomposing the observation X_{it} into the sum $\chi_{it} + \xi_{it}$ of two unobservable and mutually orthogonal components, the *common* component χ_{it} and the idiosyncratic one ξ_{it}. The various factor models that are found in the literature only differ in the way χ_{it} and ξ_{it} are characterized. The characterization we are adopting here is inspired from Forni & Lippi (2001).

Definition 2.1. *The functional zero-mean second-order stationary process $\mathcal{X} := \{X_{it},\, i \in \mathbb{N};\; t \in \mathbb{Z}\}$* admits a (high-dimensional) functional factor representation with r factors*, or* follows a (high-dimensional) functional factor model with r factors

$$X_{it} = \chi_{it} + \xi_{it} = \boldsymbol{b}^i\boldsymbol{u}_t + \xi_{it}, \qquad i \in \mathbb{N},\, t \in \mathbb{Z} \tag{2.1}$$

(χ_{it} and ξ_{it} unobservable) if there exist $\boldsymbol{b}^i = (b_{i1}, \ldots, b_{ir}) \in \mathcal{L}(\mathbb{R}^r, H_i)$ with $b_{ij} \in H_i$, $i \in \mathbb{N}$, H_i-valued processes $\{\xi_{it};\; t \in \mathbb{Z}\}$, $i \in \mathbb{N}$, and a real r-dimensional second-order stationary process $\{\boldsymbol{u}_t = (u_{1t}, \ldots, u_{rt})';\; t \in \mathbb{Z}\}$, co-stationary with $\mathcal{X}$, such that (2.1) holds with

- *(i)* $\mathbb{E}\,\boldsymbol{u}_t = \boldsymbol{0}$ *and* $\mathbb{E}\,[\boldsymbol{u}_t\boldsymbol{u}_t']$ *positive definite;*
- *(ii)* $\mathbb{E}\,[u_{jt}\xi_{it}] = 0$ *for all* $t \in \mathbb{Z}$*,* $j = 1, \ldots, r$*, and* $i \in \mathbb{N}$*;*
- *(iii) denoting by $\lambda^\xi_{N,j}$ the jth (in decreasing order of magnitude) eigenvalue of the covariance operator of $\boldsymbol{\xi}_t := (\xi_{1t}, \ldots, \xi_{Nt})'$, $\lambda^\xi_1 := \sup_N \lambda^\xi_{N,1} < \infty$;*
- *(iv) denoting by $\lambda^\chi_{N,j}$ the jth (in decreasing order of magnitude) eigenvalue of the covariance operator of $\boldsymbol{\chi}_t := (\chi_{1t}, \ldots, \chi_{Nt})'$, $\lambda^\chi_r := \sup_N \lambda^\chi_{N,r} = \infty$.*

If (ii) is strengthened into

- *(ii)′* $\mathbb{E}\,[u_{jt}\xi_{is}] = 0$ *for all* $t, s \in \mathbb{Z}$*,* $j = 1, \ldots, r$*, and* $i \in \mathbb{N}$*,* *we say that* (2.1) *provides a* (high-dimensional) strong functional factor representation with r factors *of* $\mathcal{X}$. *The* r *scalar random variables* u_{jt} *are called* factors*; the* b_{ij}*'s are the* (functional) factor loading operators*;* χ_{it} *is called the* common component*,* ξ_{it} *the* idiosyncratic one.

This definition calls for some remarks and comments.

(a) In the terminology of Hallin & Lippi (2013) or Forni et al. (2015) and Forni et al. (2017), equation (2.1), where the factors are loaded contemporaneously, is called a *static functional factor representation*, as opposed to the *general dynamic factor representation*, where the b_{ij}'s are linear one-sided square-summable filters of the form $b_{ij}(L) = \sum_{k=0}^{\infty} b_{ijk} L^k$ (L the lag operator), and the u_{jt}'s are mutually orthogonal second-order white noises (the *common shocks*) satisfying (ii)$'$. The strong static r-factor model is a particular case of the general dynamic factor one, with $q \le r$ common shocks. When the idiosyncratic processes $\{\xi_{it};\ t \in \mathbb{Z}\}$ themselves are mutually orthogonal at all leads and lags, static and general dynamic factor models are called *exact*; with this assumption relaxed into (iv) above, they sometimes are called *approximate*. In the sequel, what we call *factor models* all are *approximate static factor models*.

(b) The functional factor representation (2.1) also can be written, with obvious notation $\boldsymbol{\chi}_t$ and $\boldsymbol{\xi}_t$, in vector form

$$\boldsymbol{X}_t = \boldsymbol{\chi}_t + \boldsymbol{\xi}_t = \boldsymbol{B}_N \boldsymbol{u}_t + \boldsymbol{\xi}_t,$$

where the $N \times r$ matrix $\boldsymbol{B}_N$ has i-th row $\boldsymbol{b}^i \in \mathcal{L}(\mathbb{R}^r, H_i)$. It can also be written in matrix form as

$$\boldsymbol{X}_{NT} = \boldsymbol{\chi}_{NT} + \boldsymbol{\xi}_{NT} = \boldsymbol{B}_N \boldsymbol{u} + \boldsymbol{\xi}_{NT},$$

where $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T)$.

(c) Condition (iii) essentially requires that cross-correlations among the components of $\{\boldsymbol{\xi}_t;\ t \in \mathbb{Z}\}$ are not pervasive as $N \to \infty$. A sufficient assumption on $\boldsymbol{\xi}_{NT}$ for condition (iii) to hold is

$$\sum_{j=1}^{\infty} \left|\left|\left| \mathbb{E}\, \xi_{it} \xi_{jt}^{\mathsf{T}} \right|\right|\right|_{\infty} < M < \infty, \quad \forall i = 1, 2, \ldots,$$

see Lemma S2.14 in the Online Supplement.

(d) Condition (iv) requires pervasiveness, as $N \to \infty$, of (instantaneous) correlations among the components of $\{\boldsymbol{\chi}_t;\ t \in \mathbb{Z}\}$; it is equivalent to a condition on the sequence of factor loadings $\boldsymbol{B}_N$, which should be such that factors are loaded again and again as $N \to \infty$. A sufficient condition for this is $\boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{B}_N / N \to \boldsymbol{\Sigma}_B$, where $\boldsymbol{\Sigma}_B$ is positive definite.

(e) It follows from Lemma S2.18 in the Online Supplement that if $\mathcal{X}$ has a (possibly strong) functional factor representation with r factors, then

$\lambda_{r+1}^{\mathcal{X}} < \infty$. This in turn implies that the number r of factors is uniquely defined.

(f) The factor loadings and the factors are only jointly identifiable, since, for any collection of $r \times r$ invertible matrices $\boldsymbol{Q}_t$,

$$\boldsymbol{B}_N \boldsymbol{u}_t = (\boldsymbol{B}_N \boldsymbol{Q}_t^{-1})(\boldsymbol{Q}_t \boldsymbol{u}_t),$$

so that $\boldsymbol{v}_t = \boldsymbol{Q}_t \boldsymbol{u}_t$ provides the same decomposition of $\mathcal{X}$ into common plus idiosyncratic as (2.1).

(g) It is often assumed that $\{\boldsymbol{u}_t;\ t \in \mathbb{Z}\}$ is an r-dimensional Vector Auto-Regressive (VAR) process driven by $q \leq r$ white noises (Amengual & Watson 2007), but this is not required here.

2.3. Representation Theorem

The following results shows that the class of processes $\mathcal{X}$ admitting a functional factor model representation (in the sense of Definition 2.1) can be characterized in terms of the eigenvalues $\lambda_{N,j}^X$ of the covariance operator of the observations $\boldsymbol{X}_t$—while Definition 2.1 involves the eigenvalues $\lambda_{N,j}^\chi$ and $\lambda_{N,j}^\xi$ of the covariance operators of the unobserved common and idiosyncratic components. Moreover, when $\mathcal{X}$ admits a functional factor model representation, its decomposition into a common and an idiosyncratic component is unique.

Let $\lambda_j^X := \lim_{N\to\infty} \lambda_{N,j}^X = \sup_N \lambda_{N,j}^X$: this limit exists, as $\lambda_{N,j}^X$ is monotone increasing with N.

Theorem 2.2. *The process $\mathcal{X}$ admits a (high-dimensional) functional factor model representation with r factors if and only if $\lambda_r^X = \infty$ and $\lambda_{r+1}^X < \infty$.*

The following result tells us that the common component χ_{it} is asymptotically identifiable, and provides its expression in terms of an $L^2(\Omega)$ projection.

Theorem 2.3. *Let $\mathcal{X}$ admit (in the sense of Definition 2.1) the functional factor model representation $X_{it} = \chi_{it} + \xi_{it}$, $i \in \mathbb{N}, t \in \mathbb{Z}$, with r factors. Then (see the Online Supplement, Section S1 for a formal definition of* proj_{H_i}*),*

$$\chi_{it} = \mathrm{proj}_{H_i}(X_{it}|\mathcal{D}_t), \quad \forall i \in \mathbb{N}, t \in \mathbb{Z}$$

where

$$\mathcal{D}_t := \left\{ p \in L^2(\Omega) \,|\, p = \lim_{N\to\infty} \langle \boldsymbol{\alpha}_N, \boldsymbol{X}_t \rangle_{L^2(\Omega)}, \boldsymbol{\alpha}_N \in \boldsymbol{H}_N, \|\boldsymbol{\alpha}_N\| \overset{N\to\infty}{\longrightarrow} 0 \right\} \subset L^2(\Omega)\,;$$

the common and the idiosyncratic parts of the factor model representation thus are unique, and asymptotically identified.

The proofs of Theorems 2.2 and 2.3 are provided in the Online Supplement, Section S2.1; they are inspired from Forni & Lippi (2001)—see also Chamberlain (1983) and Chamberlain & Rothschild (1983). Notice, however, that, unlike these references, our results do not require the minimal eigenvalue of the covariance of $\boldsymbol{X}_t$ to be bounded from below.

3. Estimation

Assuming that a functional factor model with r factors holds for $\mathcal{X}$, we shall estimate the factors $\boldsymbol{u}_t$ using principal component analysis. This is a method often used for estimating factors (Bai & Ng 2002, Fan et al. 2013), but other methods are available as well (Forni & Reichlin 1998, Forni et al. 2000). The idea of this method is to find factor loadings $\boldsymbol{B}$ in $\mathcal{L}(\mathbb{R}^r, \boldsymbol{H}_N)$ and factor scores $\boldsymbol{u} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_T) \in \mathcal{L}(\mathbb{R}^T, \mathbb{R}^r)$ such that

$$P(\boldsymbol{B}, \boldsymbol{u}) := \sum_t \|\boldsymbol{X}_t - \boldsymbol{B}\boldsymbol{u}_t\|^2$$

is minimized. Denoting by $|||\cdot|||_2$ the Hilbert–Schmidt norm (see Section S2.2 in the Online Supplement), we can rewrite this objective function as

$$P(\boldsymbol{B}, \boldsymbol{u}) = |||\boldsymbol{X}_{NT} - \boldsymbol{B}\boldsymbol{u}|||_2^2.$$

Under this form, the solution is clear: by the Eckart–Young–Mirsky Theorem (Hsing & Eubank 2015, Theorem 4.4.7), we know that the objective function is minimized by choosing $\boldsymbol{B}\boldsymbol{u}$ to be equal to $\boldsymbol{B}_*\boldsymbol{u}_*$, the r-term truncation of the singular value decomposition of $\boldsymbol{X}_{NT}$. Let us write the singular value decomposition of $\boldsymbol{X}_{NT}$ as

$$\boldsymbol{X}_{NT} = \sum_{i=1}^{N} \hat{\lambda}_i^{1/2} \hat{\boldsymbol{e}}_i \hat{\boldsymbol{f}}_i^{\mathsf{T}}, \tag{3.1}$$

where $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq 0$, $\hat{\boldsymbol{e}}_i$s (belonging to $\boldsymbol{H}_N$) and $\hat{\boldsymbol{f}}_i$s (belonging to $\mathbb{R}^T$) are rescaled to have norm $\sqrt{T}$. The $\hat{\lambda}_i$s, thus, are rescaled singular values—we show in Lemma S2.11 in the Online Supplement that this rescaling allows $\hat{\lambda}_1 = O_{\mathrm{P}}(1)$. To make the notation simple, the sum is ranging over $i = 1, \ldots, N$: if $N > T$, the last $(N - T)$ $\hat{\lambda}_i$s are set to zero. We now have a multitude of choices for $\boldsymbol{u}_*$, of which we select

$$\tilde{\boldsymbol{u}} := \begin{pmatrix} \hat{\boldsymbol{f}}_1^{\mathsf{T}} \\ \vdots \\ \hat{\boldsymbol{f}}_r^{\mathsf{T}} \end{pmatrix} \in \mathbb{R}^{r \times T}. \tag{3.2}$$

The reason for this choice is the following: $\tilde{\boldsymbol{u}}$ can be obtained by computing the first r eigenvectors of $\boldsymbol{X}_{NT}^{\mathsf{T}}\boldsymbol{X}_{NT}$, and rescaling them by $\sqrt{T}$. Note that computing $\boldsymbol{X}_{NT}^{\mathsf{T}}\boldsymbol{X}_{NT}$ requires computing $O(T^2)$ inner products in $\boldsymbol{H}_N$, and then computing the leading r eigenvectors of a $T \times T$ matrix. Dual to $\tilde{\boldsymbol{u}}$ are the corresponding factor loadings

$$\tilde{\boldsymbol{B}}_N := \left(\hat{\lambda}_1^{1/2}\hat{\boldsymbol{e}}_1, \ldots, \hat{\lambda}_r^{1/2}\hat{\boldsymbol{e}}_r\right) \in \mathcal{L}(\mathbb{R}^r, \boldsymbol{H}_N),$$

for which $\tilde{\boldsymbol{B}}_N\tilde{\boldsymbol{u}} = \boldsymbol{B}_*\boldsymbol{u}_*$. The loadings $\tilde{\boldsymbol{B}}_N$ can be obtained by an eigendecomposition of $\boldsymbol{X}_{NT}\boldsymbol{X}_{NT}^{\mathsf{T}}$. However, this would require an eigendecomposition of an operator in $\mathcal{L}(\boldsymbol{H}_N)$, which could be computationally much more demanding than performing an eigendecomposition of $\boldsymbol{X}_{NT}^\mathsf{T}\boldsymbol{X}_{NT}$ to obtain $\tilde{\boldsymbol{u}}$, and then multiply it with $\boldsymbol{X}_{NT}$ to obtain $\tilde{\boldsymbol{B}}_N$. We also point out that the idealistic approach of using a Karhunen–Loève truncation (or PCA projection) for each X_{it} separately, prior to conducting the global PCA, is *not* a good idea in general, as there is no guarantee that the common component will be picked by the individual Karhunen–Loève truncations, and it might well be that it actually removes all the common component (see Section 4 for examples).

In order be able to estimate the factor scores and loadings, we shall need the following regularity assumptions, which we discuss below. These assumptions, which are adaptations of standard assumptions in scalar factor models (Bai & Ng 2002), imply, in particular, that $\boldsymbol{X}_{NT}$ follows a functional factor model with r factors.

Assumption A. $(\boldsymbol{u}_t)_t$ and $(\boldsymbol{\xi}_t)_t$ are mean zero second-order co-stationary, with $\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{\xi}_t^\mathsf{T}\right] = \mathbf{0}$; the covariance operator $\boldsymbol{\Sigma}_{\boldsymbol{u}} := \mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t^\mathsf{T}\right]$ is $r \times r$ positive definite and $T^{-1}\boldsymbol{u}\boldsymbol{u}^\mathsf{T} \xrightarrow{\mathrm{P}} \boldsymbol{\Sigma}_{\boldsymbol{u}}$ as $T \to \infty$.

Assumption B. $N^{-1}\boldsymbol{B}_N^\mathsf{T}\boldsymbol{B}_N \to \boldsymbol{\Sigma}_B$, as $N \to \infty$, for some $r \times r$ positive-definite matrix $\boldsymbol{\Sigma}_B$.

Assumption C. Let $\nu_N(h) := \mathbb{E}\left[\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_{t-h}/N\right]$.

(C1) There exists a constant M such that for all $N \geq 1$, $\sum_{h\in\mathbb{Z}} |\nu_N(h)| \leq M$, and

(C2) $\left|\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s/N - \nu_N(t-s)\right|$ is $O_\mathrm{P}(N^{-1/2})$ uniformly in $t, s \geq 1$.

Assumption D. There exists $M < \infty$ such that $\|b_{il}\| < M$ for all $i \in \mathbb{N}$ and $l = 1, \dots, r$, and $\sum_{j=1}^{\infty} \left|\!\left|\!\left| \mathbb{E}\, \xi_{it}\xi_{jt}^\mathsf{T} \right|\!\right|\!\right|_\infty < M$ for all $i \in \mathbb{N}$.

Assumption E(α). Letting $C_{N,T} := \min\{\sqrt{N}, \sqrt{T}\}$,

$$\left|\!\left|\!\left|\boldsymbol{u}\boldsymbol{\xi}_{NT}^\mathsf{T}\right|\!\right|\!\right|_2^2 = O_\mathrm{P}\left(NT^2C_{N,T}^{-(1+\alpha)}\right)$$

for some $\alpha \in [0, 1]$.

Assumption A has some basic requirements about the model (factors and idiosyncratics are co-stationary and uncorrelated at lag zero), and the factors. It assumes, in particular, that $|\!|\!|\boldsymbol{u}|\!|\!|_2 = O_\mathrm{P}(\sqrt{T})$. Since $\boldsymbol{u}\boldsymbol{u}^\mathsf{T} = \sum_{t=1}^T \boldsymbol{u}_t\boldsymbol{u}_t^\mathsf{T}$, it also implies a weak law of large numbers for $(\boldsymbol{u}_t\boldsymbol{u}_t^\mathsf{T})_t$, which holds under various dependence assumptions on $(\boldsymbol{u}_t)_t$, see e.g. Brillinger (2001), Bradley (2005), Dedecker et al. (2007).

Assumption B deals with the factor loadings, and implies in particular that $|\!|\!|\boldsymbol{B}_N|\!|\!|_2$ is of order $\sqrt{N}$. Intuitively, it means that the factors are loaded again and again as the cross-section increases.

Assumptions A and B together intuitively mean that (almost) all the common components $\boldsymbol{b}^i\boldsymbol{u}_t$ have dimension r. They could be weakened by assuming that the r largest eigenvalues of $\boldsymbol{B}_N^\mathsf{T}\boldsymbol{B}_N/N$ and $\boldsymbol{u}\boldsymbol{u}^\mathsf{T}/T$ are bounded away from infinity and zero, see e.g. Fan et al. (2013).

Assumption C is an assumption on the idiosyncratic terms: (C1) limits the total variance and lagged cross-covariances of the idiosyncratic component; (C2) imposes a uniform rate of convergence in the law of large numbers for $(\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s/N)_N$. A sufficient condition for this is

There exists $\varepsilon > 0$ and $M < \infty$ such that $\mathbb{E}\left|\sqrt{N}\left(\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s/N - \nu_N(t-s)\right)\right|^\varepsilon < M$, for all $s, t, N \geq 1$.

In particular, (C2) implicitly limits the cross-sectional and lagged correlations of the idiosyncratic components.

Assumption D limits the cross-sectional correlation of the idiosyncratic components, and bounds the norm of the loadings. It implies that $\left|\!\left|\!\left|\boldsymbol{B}_N^\mathsf{T}\boldsymbol{\xi}_{NT}\right|\!\right|\!\right|_2^2$ is $O_\mathrm{P}(NT)$—see Lemma S2.15 in the Online Supplement—and could be replaced by this weaker condition in the proofs of Theorems 3.1, 3.2, 3.4 and 3.5.

Assumption E(α) imposes limits on the lagged cross-correlations between the factors and the idiosyncratics. Notice that Assumptions A and C jointly imply Assumption E(α) for $\alpha = 0$ (see Lemma S2.10 in the Online Supplement), so that $\alpha = 0$ corresponds to the absence of restrictions on these cross-correlations; $\alpha = 1$ is the strongest case of this assumptions, and corresponds to the weakest cross-correlations between factors and idiosyncratics: it is implied by the following stronger (but more easily interpretable) conditions (see Lemma S2.16 in the Online Supplement):

(i) $\mathbb{E}\left[(\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s)u_{lt}u_{ls}\right] = \mathbb{E}\left[\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s\right]\mathbb{E}\left[u_{lt}u_{ls}\right]$ for all $l = 1, \ldots, r$ and all $s, t \in \mathbb{Z}$,

(ii) $\sum_{h\in\mathbb{Z}} |\nu_N(h)| < \infty$.

Notice that Assumption E(α) with $\alpha = 1$ is still less stringent than Assumption D in Bai & Ng (2002).

Note that Assumptions A, B, and D imply that the first r eigenvalues of $\mathrm{cov}(\boldsymbol{X}_t)$ diverge while the $(r+1)$th one remains bounded (Lemma S2.14 in the Online Supplement), hence the common and idiosyncratic components are asymptotically identified (Theorem 2.2).

The first result of this section (Theorem 3.1, see below) tells us, essentially, that $\tilde{\boldsymbol{u}}$ consistently estimates the true factors. Since the true factors are only identified up to an invertible transformation, however, consistency here is about the convergence of the row space spanned by $\tilde{\boldsymbol{u}}$ to the one spanned by $\boldsymbol{u}$. The discrepancy between these row spaces can be measured by

$$\delta_{N,T} := \min_{\mathbf{R}\in\mathbb{R}^{k\times r}} |\!|\!|\tilde{\boldsymbol{u}} - \mathbf{R}\boldsymbol{u}|\!|\!|_2/\sqrt{T},$$

(recall that $|\!|\!|\cdot|\!|\!|_2$ denotes the Hilbert–Schmidt norm: see the Online Supplement, Section S2.2). $\delta_{N,T}$ is the rescaled Hilbert–Schmidt norm of the residual of the least squares fit of the rows of $\tilde{\boldsymbol{u}}$ onto the row space of $\boldsymbol{u}$, and we make explicit its dependence on N, T. The $T^{-1/2}$ rescaling is needed because $|\!|\!|\tilde{\boldsymbol{u}}|\!|\!|_2^2 = rT$—any rescaling of order $T^{-1/2}$ would lead to the same conclusion.

We now can state one of the main results of this section.

Theorem 3.1. *Under Assumptions A, B, C, and D,*

$$\delta_{N,T} = O_\mathrm{P}(C_{N,T}^{-1}),$$

where $C_{N,T} := \min\{\sqrt{N}, \sqrt{T}\}$.

Proof. Define

$$\tilde{\boldsymbol{R}} := \hat{\boldsymbol{\Lambda}}^{-1} \tilde{\boldsymbol{u}} \boldsymbol{u}^{\mathsf{T}} \boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{B}_N / NT, \tag{3.3}$$

where $\hat{\boldsymbol{\Lambda}}$ is the $r \times r$ diagonal matrix with $\hat{\lambda}_i$s in the diagonal, and are defined in (3.1). Let us show that

$$\left|\left|\left| \tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}} \boldsymbol{u} \right|\right|\right|_2 / \sqrt{T} = O_{\mathrm{P}}(C_{N,T}^{-1}).$$

Defining $\hat{\boldsymbol{u}} := \hat{\boldsymbol{\Lambda}} \tilde{\boldsymbol{u}}$ and $\boldsymbol{Q} := \hat{\boldsymbol{\Lambda}} \tilde{\boldsymbol{R}}$, we have

$$\left|\left|\left| \tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}} \boldsymbol{u} \right|\right|\right|_2 \leq \left|\left|\left| \hat{\boldsymbol{\Lambda}}^{-1} \right|\right|\right|_\infty |||\hat{\boldsymbol{u}} - \boldsymbol{Q}\boldsymbol{u}|||_2.$$

By Lemma S2.12 in the Online Supplement, $\left|\left|\left| \hat{\boldsymbol{\Lambda}}^{-1} \right|\right|\right|_\infty = O_{\mathrm{P}}(1)$, and a straightforward calculation yields $\hat{\boldsymbol{u}} = \tilde{\boldsymbol{u}} \boldsymbol{X}_{NT}^{\mathsf{T}} \boldsymbol{X}_{NT} / (NT)$, whereby

$$\hat{\boldsymbol{u}} - \boldsymbol{Q}\boldsymbol{u} = \frac{1}{nT} \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{\xi}_{NT} + \frac{1}{NT} \tilde{\boldsymbol{u}} \boldsymbol{u}^{\mathsf{T}} \boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{\xi}_{NT} + \frac{1}{nT} \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{B}_N \boldsymbol{u}^{\mathsf{T}},$$

and, therefore,

$$\begin{aligned} |||\hat{\boldsymbol{u}} - \boldsymbol{Q}\boldsymbol{u}|||_2 \leq \frac{1}{nT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 + \frac{1}{NT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{u}^{\mathsf{T}} \boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 \\ + \frac{1}{nT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{B}_N \boldsymbol{u}^{\mathsf{T}} \right|\right|\right|_2. \end{aligned}$$

Let us consider each terms separately. For the first term, by Lemma S2.10 in the Online Supplement,

$$\frac{1}{nT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 \leq \frac{1}{nT} |||\tilde{\boldsymbol{u}}|||_\infty \left|\left|\left| \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 = O_{\mathrm{P}}(\sqrt{T} C_{N,T}^{-1}).$$

For the second term, it follows from Assumption D that

$$\frac{1}{NT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{u}^{\mathsf{T}} \boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 \leq \frac{1}{NT} |||\tilde{\boldsymbol{u}}|||_\infty \left|\left|\left| \boldsymbol{u}^{\mathsf{T}} \right|\right|\right|_\infty \left|\left|\left| \boldsymbol{B}_N^{\mathsf{T}} \boldsymbol{\xi}_{NT} \right|\right|\right|_2 = O_{\mathrm{P}}(\sqrt{T/N}).$$

For the third term, still from Assumption D,

$$\frac{1}{nT} \left|\left|\left| \tilde{\boldsymbol{u}} \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{B}_N \boldsymbol{u}^{\mathsf{T}} \right|\right|\right|_2 \leq \frac{1}{nT} |||\tilde{\boldsymbol{u}}|||_\infty \left|\left|\left| \boldsymbol{\xi}_{NT}^{\mathsf{T}} \boldsymbol{B}_N \right|\right|\right|_\infty \left|\left|\left| \boldsymbol{u}^{\mathsf{T}} \right|\right|\right|_2 = O_{\mathrm{P}}(\sqrt{T/N}).$$

Piecing all these together completes the proof. □

This result essentially means that the factors are (asymptotically) consistently estimated. Note in particular that $\delta_{N,T} \equiv \delta_{N,T}(\tilde{\boldsymbol{u}}, \boldsymbol{u})$ is not symmetric in $\tilde{\boldsymbol{u}}, \boldsymbol{u}$, and hence is not a metric. Nevertheless, small values of $\delta_{N,T}$ imply that the row space of the estimated factors is close to the row space of the true factors. By classical least squares theory, we have

$$\delta_{N,T} = \left|\left|\left| (\boldsymbol{I}_T - \boldsymbol{P}_{\boldsymbol{u}}) \tilde{\boldsymbol{u}}^{\mathsf{T}} / \sqrt{T} \right|\right|\right|_2,$$

where $\boldsymbol{P}_{\boldsymbol{u}}$ is the projection onto the column space of $\boldsymbol{u}^\intercal$. This formula will be useful in Section 4.

Under additional constraints on the factor loadings and the factors and adequate additional assumptions, it is possible to show that the estimated factors $\tilde{\boldsymbol{u}}$ converge exactly (up to a sign) to the true factors $\boldsymbol{u}$ (Stock & Watson 2002*a*). For this, we need, for instance,

Assumption F. All the eigenvalues of $\boldsymbol{\Sigma}_B\boldsymbol{\Sigma}_u$ are distinct.

Under Assumptions A, B, and F, for N and T large enough, we can choose the loadings and factors such that $\boldsymbol{u}\boldsymbol{u}^\intercal/T = I_r$, and $\boldsymbol{B}_N^\intercal \boldsymbol{B}_N/N$ is diagonal with distinct positive entries whose gaps remain bounded from below, as $N, T \to \infty$. With this new assumption, we can show that the factors are estimated consistently up to a sign.

Theorem 3.2. *Assume that Assumptions A, B, C, D, and F hold. Assume furthermore that we have transformed the loadings and factors in such a way that, for N and T large enough, $\boldsymbol{u}\boldsymbol{u}^\intercal/T = I_r$ and $\boldsymbol{B}_N^\intercal \boldsymbol{B}_N/N$ is diagonal with distinct decreasing entries. Then, there exists an $r \times r$ diagonal matrix $\boldsymbol{R}_{NT}$ (depending on N, T) with entries ± 1 such that*

$$|||\tilde{\boldsymbol{u}} - \boldsymbol{R}_{NT}\boldsymbol{u}|||_2/\sqrt{T} = O_{\mathrm{P}}(C_{N,T}^{-1}) \quad \text{as } N, T \to \infty.$$

Proof. Notice that, by our assumptions, for N, T large enough,

$$\boldsymbol{\chi}_{NT}^\intercal \boldsymbol{\chi}_{NT}/(NT) = \sum_{k=1}^{r} \lambda_k \boldsymbol{u}_{(k)} \boldsymbol{u}_{(k)}^\intercal \tag{3.4}$$

where the λ_ks are distinct, and $\boldsymbol{u}_{(k)}$ is the kth row of $\boldsymbol{u}$, written as a column. Note that λ_k depends on N, T, but we suppress this dependency in the notation. Notice in particular that given our identification assumptions, (3.4) is in fact a spectral decomposition. We now recall the spectral decomposition $\boldsymbol{X}_{NT}^\intercal \boldsymbol{X}_{NT}/(NT) = \sum_{k\geq 1} \hat{\lambda}_k \hat{\boldsymbol{f}}_k \hat{\boldsymbol{f}}_k^\intercal$. Lemma 4.3 of Bosq (2000) then yields

$$\left\| \hat{\boldsymbol{f}}_k - \operatorname{sign}(\hat{\boldsymbol{f}}_k^\intercal \boldsymbol{u}_{(k)})\boldsymbol{u}_{(k)} \right\| / \sqrt{T} = O_{\mathrm{P}}\left(\left|\left|\left| \boldsymbol{X}_{NT}^\intercal \boldsymbol{X}_{NT} - \boldsymbol{\chi}_{NT}^\intercal \boldsymbol{\chi}_{NT} \right|\right|\right|_\infty /(NT) \right),$$

for $k = 1, \ldots, r$, since the gaps between the $\lambda_1, \ldots, \lambda_r$ remain bounded from below by Assumption F. Now,

$$\left|\left|\left| \boldsymbol{X}_{NT}^\intercal \boldsymbol{X}_{NT} - \boldsymbol{\chi}_{NT}^\intercal \boldsymbol{\chi}_{NT} \right|\right|\right|_\infty \leq \left|\left|\left| \boldsymbol{\xi}_{NT}^\intercal \boldsymbol{\xi}_{NT} \right|\right|\right|_\infty + 2|||\boldsymbol{u}|||_\infty \left|\left|\left| \boldsymbol{B}_N^\intercal \boldsymbol{\xi}_{NT} \right|\right|\right|_\infty,$$

and applying Lemmas S2.10, S2.11 and S2.15 in the Online Supplement, we get

$$N^{-1}T^{-1} \left|\left|\left| \boldsymbol{X}_{NT}^\intercal \boldsymbol{X}_{NT} - \boldsymbol{\chi}_{NT}^\intercal \boldsymbol{\chi}_{NT} \right|\right|\right|_\infty = O_p(C_{N,T}^{-1}).$$

This completes the proof, since the kth row of $\tilde{\boldsymbol{u}}$ is $\hat{\boldsymbol{f}}_k$ for $k = 1, \ldots, r$. □

Notice that we do not assume any particular dependency between N and T in the results above: the order of the estimation error depends only $\min\{n, T\}$. A couple of remarks are in order.

Remark 3.3. (i) As mentioned earlier, Assumptions A, B, and D imply that the common and idiosyncratic part are asymptotically identified, see Lemma S2.14 and Theorem 2.2. The extra assumptions needed for consistent estimation of the factors row space (and for the loadings and common component, see Theorems 3.4 and 3.5 below) are there because the covariance $\mathrm{cov}(\boldsymbol{X}_t)$ is unknown, and its first r eigenvectors must be estimated.

(ii) Notice, in particular, that Theorem 3.1 holds for the case $H_i = \mathbb{R}$ for all i, where it coincides with Theorem 1 of Bai & Ng (2002). However we obtain this result under weaker conditions, as we do not assume $\mathbb{E}\|\boldsymbol{u}_t\|^4 < \infty$ nor $\mathbb{E}\|\xi_{it}\|^8 < M < \infty$ (an assumption that is unlikely to hold in most equity return series). Nor do we assume

$$(NT)^{-1}\sum_{i,j=1}^{N}\sum_{t,s=1}^{T}\left|\left|\left|\mathbb{E}\left[\xi_{it}\xi_{js}^{\mathsf{T}}\right]\right|\right|\right|_{\infty} < M < \infty,$$

and we are weakening their assumption

$$\mathbb{E}\left|\sqrt{N}\left(\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s/N - \nu_N(t-s)\right)\right|^4 < M < \infty,$$

on idiosyncratic cross-covariances into a uniform boundedness in probability assumption on $\sqrt{N}\left(\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s/N - \nu_N(t-s)\right)$. The main tools that allow us to derive results under weaker assumptions are inequalities between Schatten norms (see Section S2.2) of compositions of operators, whereas classical results mainly use the Cauchy–Schwartz inequality.

(iii) Note that we could change the N^{-1} term in Assumption B to be $N^{-\alpha}$, for $\alpha \in (0, 1)$, in which case we have *weak* (or *semi-weak*) factors (Chudik et al. 2011, Lam & Yao 2012, Onatski 2010), which would affect the rate of convergence in Theorems 3.1, 3.2, 3.4, and 3.5; see also Boivin & Ng (2006).

(iv) We do not make any Gaussian assumptions and, unlike Lam & Yao (2012), we do not assume that the idiosyncratic component is white noise.

(v) Bai & Ng (2002) allow for limited correlation between the factors and the idiosyncratic components. This is only an illusory increase of generality, since it transfers to the idiosyncratic part of the impact of the factors on some given cross-sectional unit X_{it} which, consequently, will not benefit fully from the panel-wide contribution to the estimation of the factors.

(vi) The results could be extended to conditionally heteroscedastic common shocks and idiosyncratic components, as frequently assumed in the scalar case (see, e.g., Alessi et al. 2009, Trucios et al. 2019) This, which would come at the cost of additional identifiability constraints, is left for further research.

The next result of this section deals with the consistent estimation of the factor loadings. Define $\bar{\boldsymbol{B}}_N := \tilde{\boldsymbol{B}}_N \hat{\boldsymbol{\Lambda}}^{-1/2}$, which is the same as $\tilde{\boldsymbol{B}}_N$, but with unit norm columns. Similarly to the factors, the factor loadings are only identified up to an invertible transformation, and we therefore measure consistency by quantify the discrepancy between the column space of the estimate $\tilde{\boldsymbol{B}}_N$ and the column space of the true factors $\boldsymbol{B}_N$ by

$$\varepsilon_{N,T} := \min_{\mathbf{R} \in \mathbb{R}^{r \times r}} \left|\left|\left| \bar{\boldsymbol{B}}_N - \boldsymbol{B}_N \mathbf{R} \right|\right|\right|_2 / \sqrt{N}, \tag{3.5}$$

which is the Hilbert–Schmidt norm of the residual of $\bar{\boldsymbol{B}}_N$ projected onto the column space of $\boldsymbol{B}_N$, which depends on both N and T. The $\sqrt{n}$ renormalization is needed as $\left|\left|\left| \bar{\boldsymbol{B}}_N \right|\right|\right|_2^2 = rN$. We then have, for the of factor loadings, the following consistency result.

Theorem 3.4. *Under Assumptions A, B, C, D, and E(α),*

$$\varepsilon_{N,T} = O_{\mathrm{P}}\left(C_{N,T}^{-\frac{1+\alpha}{2}} \right),$$

where $C_{N,T} := \min\{\sqrt{N}, \sqrt{T}\}$.

Proof. We shall show that $\left|\left|\left| \tilde{\boldsymbol{B}}_N - \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1} \right|\right|\right|_2 = O_{\mathrm{P}}\left(\sqrt{N / C_{N,T}^{(1+\alpha)}} \right)$, where $\tilde{\boldsymbol{R}}$ is defined in (3.3), and is invertible by Lemma S2.13; the desired result then follows, since

$$\bar{\boldsymbol{B}}_N - \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1} \hat{\boldsymbol{\Lambda}}^{-1/2} = (\tilde{\boldsymbol{B}}_N - \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1}) \hat{\boldsymbol{\Lambda}}^{-1/2},$$

and $\left|\left|\left| \hat{\boldsymbol{\Lambda}}^{-1/2} \right|\right|\right|_\infty = O_{\mathrm{P}}(1)$ by Lemma S2.12.

First, notice that $\tilde{\boldsymbol{B}}_N = T^{-1} \boldsymbol{X}_{NT} \tilde{\boldsymbol{u}}^\mathsf{T} / T$, so that

$$\begin{aligned}
\tilde{\boldsymbol{B}}_N &= \boldsymbol{B}_N \boldsymbol{u} \tilde{\boldsymbol{u}}^\mathsf{T} / T + \boldsymbol{\xi}_{NT} \tilde{\boldsymbol{u}}^\mathsf{T} / T \\
&= \boldsymbol{B}_N \left(\tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{u}} + \boldsymbol{u} - \tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{u}} \right) \tilde{\boldsymbol{u}}^\mathsf{T} / T + \boldsymbol{\xi}_{NT} \tilde{\boldsymbol{u}}^\mathsf{T} / T \\
&= \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1} + \boldsymbol{B}_N \left(\boldsymbol{u} - \tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{u}} \right) \tilde{\boldsymbol{u}}^\mathsf{T} / T + \boldsymbol{\xi}_{NT} (\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}} \boldsymbol{u})^\mathsf{T} / T + \boldsymbol{\xi}_{NT} \boldsymbol{u}^\mathsf{T} \tilde{\boldsymbol{R}}^\mathsf{T},
\end{aligned}$$

where we have used the fact that $\tilde{\boldsymbol{u}} \tilde{\boldsymbol{u}}^\mathsf{T} / T = I_r$. Hence,

$$\begin{aligned}
\left|\left|\left| \tilde{\boldsymbol{B}}_N - \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1} \right|\right|\right|_2 \leq \frac{1}{T} \Big\{ &|||\boldsymbol{B}_N|||_\infty \left|\left|\left| \boldsymbol{u} - \tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{u}} \right|\right|\right|_2 |||\tilde{\boldsymbol{u}}|||_2 \\
&+ |||\boldsymbol{\xi}_{NT}|||_\infty \left|\left|\left| (\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}} \boldsymbol{u}) \right|\right|\right|_2 + |||\boldsymbol{\xi}_{NT} \boldsymbol{u}^\mathsf{T}|||_2 \left|\left|\left| \tilde{\boldsymbol{R}} \right|\right|\right|_\infty \Big\}.
\end{aligned}$$

By Lemma S2.13 and Theorem 3.1, we have

$$\left|\left|\left| \boldsymbol{u} - \tilde{\boldsymbol{R}}^{-1} \tilde{\boldsymbol{u}} \right|\right|\right|_2 \leq \left|\left|\left| \tilde{\boldsymbol{R}}^{-1} \right|\right|\right|_\infty \left|\left|\left| \tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}} \boldsymbol{u} \right|\right|\right|_2 = O_{\mathrm{P}}(\sqrt{T} / C_{N,T});$$

thus, the first summand is $O_{\mathrm{P}}\left(\sqrt{N/C_{N,T}^2}\right)$. By Lemma S2.10 and Theorem 3.1, the second summand is $O_{\mathrm{P}}\left(\sqrt{N/C_{N,T}^3}\right)$. As for the last summand, it is $O_{\mathrm{P}}\left(\sqrt{N/C_{N,T}^{1+\alpha}}\right)$ by Assumption E(α). This completes the proof. □

The rate of convergence for the loadings thus crucially depends on the value of $\alpha \in [0,1]$ in Assumption E(α). The larger α (i.e., the weaker the cross-correlation between factors and idiosyncratics), the better the rate. Unless $\alpha = 1$, that rate is slower than for the estimation of the factors. As in Theorem 3.2, it could be shown that, under additional identification assumptions, the loadings can be estimated consistently up to a sign. Details are left to the reader.

We can now turn to the estimation of the common component $\boldsymbol{\chi}_{NT}$ itself. Let $\hat{\boldsymbol{\chi}}_{NT} := \tilde{\boldsymbol{B}}_N \tilde{\boldsymbol{u}}$. Using Theorems 3.1 and 3.4, we obtain the following result.

Theorem 3.5. *Under Assumptions A, B, C, D, and E(α),*

$$\frac{1}{\sqrt{NT}} |||\boldsymbol{\chi}_{NT} - \hat{\boldsymbol{\chi}}_{NT}|||_2 = O_{\mathrm{P}}\left(C_{N,T}^{-\frac{1+\alpha}{2}}\right) \qquad \alpha \in [0,1].$$

The $\sqrt{NT}$ renormalization is used because the Hilbert–Schmidt norm of $\boldsymbol{\chi}_{NT}$ is of order $\sqrt{NT}$.

Proof. Recalling the definition (3.3), we have

$$|||\hat{\boldsymbol{\chi}}_{NT} - \boldsymbol{\chi}_{NT}|||_2 \leq \left|\left|\left|\tilde{\boldsymbol{B}}_N - \boldsymbol{B}_N \tilde{\boldsymbol{R}}^{-1}\right|\right|\right|_2 |||\tilde{\boldsymbol{u}}|||_\infty + |||\boldsymbol{B}_N|||_\infty \left|\left|\left|\tilde{\boldsymbol{R}}^{-1}\right|\right|\right|_\infty \left|\left|\left|\tilde{\boldsymbol{u}} - \tilde{\boldsymbol{R}}\boldsymbol{u}\right|\right|\right|_\infty.$$

The desired result follows from applying the results from the proofs of Theorems 3.1 and 3.4, and Lemma S2.13. □

Again, the rate of convergence depends on α, which quantifies the amount of cross-correlation between the factors and the idiosyncratic component.

4. Numerical Experiments

In this section, we assess the finite-sample performance of our estimators of on simulated panels.

Panels of size $(N = 100) \times (T = 200)$ were generated as follows from a functional factor model with three factors. All functional time series in the panel are represented in an orthonormal basis of dimension 7, with basis functions $\varphi_1, \ldots, \varphi_7$; the particular choice of orthonormal functions φ_i has no influence on the results. Each of the three factors is independently generated from a Gaussian AR(1) process with coefficient a_k and variance $1 - a_k^2$, $k = 1, 2, 3$. Those coefficients are picked at random from a uniform on $(-1, 1)$ at the beginning of the simulations and kept fixed across the 500 replications. The a_ks are then rescaled so that the operator norm of the companion matrix of the three-dimensional VAR process $\boldsymbol{u}_t$ is 0.8.

The factor loading coefficients were chosen of the form $\boldsymbol{b}^i = (b_{i1}, b_{i2}, b_{i3}) := (\tilde{b}_{i1}\varphi_1, \tilde{b}_{i2}\varphi_2, \tilde{b}_{i3}\varphi_3)$ with $\tilde{b}_{il} \in \mathbb{R}$—namely, the l-th loading is always aligned with the first three basis functions φ_l, $l = 1, 2, 3$: the $N \times 3$ coefficient matrix $\tilde{B} = (\tilde{b}_{il})$ therefore uniquely defines the N loadings. Those coefficients were generated as follows: first pick a value at random from a uniform over $[0, 1]^{3N}$, then rescale each fixed-i triple to have unit Euclidean norm. This rescaling implies that the total variance of each common component (for each i) is equal to 1; $\tilde{B}$ is kept fixed across replications.

The idiosyncratic components belong to the space spanned by $\varphi_1, \ldots, \varphi_7$; their coefficients $(\langle \xi_{it}, \varphi_j \rangle)_j$ were generated from

$$(\langle \xi_{it}, \varphi_1 \rangle, \ldots, \langle \xi_{it}, \varphi_7 \rangle)' \overset{i.i.d}{\sim} \mathcal{N}(\mathbf{0}, c \cdot \boldsymbol{E} / \operatorname{Tr}(\boldsymbol{E})), \; i = 1, \ldots, N, \; t = 1, \ldots, T.$$

Since the total variance of each common component is one, the constant c is the relative amount of idiosyncratic noise: $c = 1$ means equal common and idiosyncratic variances, while larger values of c make estimation of factors, loadings, and common components more difficult. We considered four Data Generating Processes (DGPs):

`DGP1`: $c = 1$, $\boldsymbol{E} = \operatorname{diag}(1, 2^{-2}, 3^{-2}, \ldots, 7^{-2})$,
`DGP2`: $c = 1$, $\boldsymbol{E} = \operatorname{diag}(7^{-2}, 6^{-2}, \ldots, 1)$,
`DGP3`: $c = 8$, $\boldsymbol{E} = \operatorname{diag}(1, 2^{-2}, 3^{-2}, \ldots, 7^{-2})$,
`DGP4`: $c = 8$, $\boldsymbol{E} = \operatorname{diag}(7^{-2}, 6^{-2}, \ldots, 1)$.

In `DGP1` and `DGP3`, we have chosen to align the largest idiosyncratic variances with the span of the factor loadings (spanned by $\{v_1, v_2, v_3\}$). In this case, $\boldsymbol{X}_{nT}\tilde{\boldsymbol{u}}^\top\tilde{\boldsymbol{u}}$ is picking the three common shocks, but also the idiosyncratic components (which have large variances). On the contrary, in `DGP2, DGP4`, we have chosen the directions of largest idiosyncratic variance to be orthogonal to the span of the factor loadings. We then face two situations: (i) N is small enough (equivalently, the total idiosyncratic variance of any component is big enough) that the first eigenvectors of $\boldsymbol{X}_{NT}^\top\boldsymbol{X}_{NT}$ mainly correspond to idiosyncratic components: the product $\boldsymbol{X}_{nT}\tilde{\boldsymbol{u}}^\top\tilde{\boldsymbol{u}}$ then essentially filters out the common component, and our estimators of the factors and factor loadings are quite poor; (ii) N is big enough that the first eigenvectors of $\boldsymbol{X}_{NT}^\top\boldsymbol{X}_{NT}$ correspond mostly to the common component. In this case, the idiosyncratic component is almost absent in $\boldsymbol{X}_{nT}\tilde{\boldsymbol{u}}^\top\tilde{\boldsymbol{u}}$, and our estimators of the factors and factor loadings are fairly accurate. In particular, while it might seem that `DGP1` and `DGP3` are much more favorable than `DGP2` and `DGP4`, the reality is more subtle, with the latter scenarios being sometimes more favorable, as we will see below.

The variance of the idiosyncratic components (for each i) is equal to 1 for `DGP1` and `DGP2`, and equal to 8 for `DGP3` and `DGP4`; the latter thus are more difficult. In particular, while one might feel that performing a Karhunen–Loève truncation for each separate FTS (each i) is a good idea, this actually performs quite poorly in `DGP4`, where the first (population) eigenfunctions (for each i) are *exactly* orthogonal to the common component.

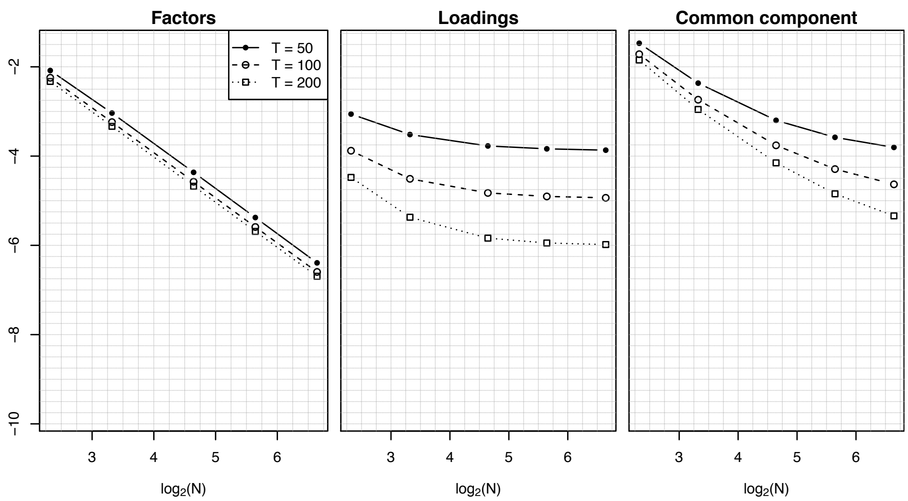
For $N = 10, 25, 50, 100$ and $T = 50, 100, 200$, we have considered the subpanels of the first N and T observation from the "large" 100×200 panel. For each replication and each choice of N and T, we estimated the factors and factor loadings using principal component analysis over the $N \times T$ panel, as explained in Section 3, assuming that the number of factors is known to be three. We have then computed the approximation error $\delta^2_{N,T}$ for the factors, $\varepsilon^2_{N,T}$ for the loadings, and $\phi_{N,T}$ for the common component (see Section 3), with

$$\phi_{N,T} := |||\boldsymbol{\chi}_{NT} - \hat{\boldsymbol{\chi}}_{NT}|||_2^2/(NT). \tag{4.1}$$

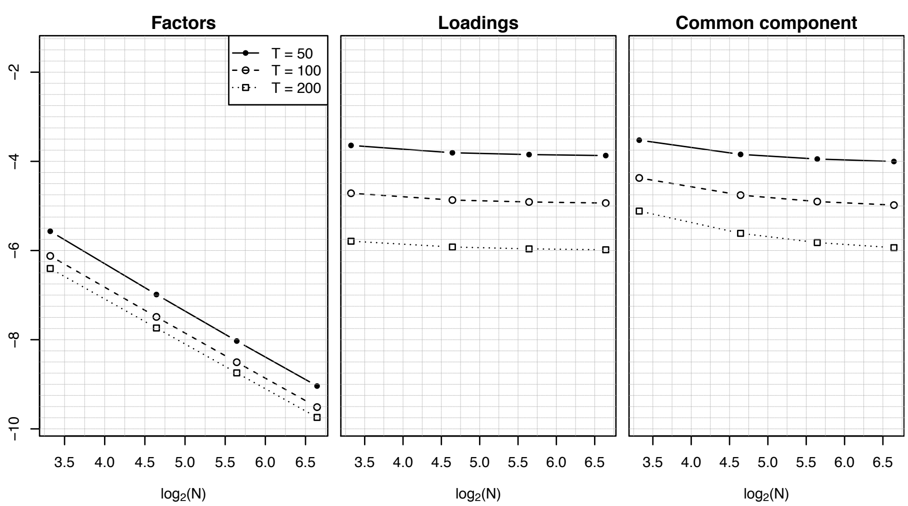
The results, averaged over the 500 replications, are shown in Figures 1a, 1b, and Figures S1a, and S1b in the Online Supplement for `DGP1, DGP2, DGP3`, and `DGP4`, respectively. A careful inspection of these figures allows one to infer whether the asymptotic regime predicted by the theoretical results (see Section 3) has been reached. We will give a detailed description of this for `DGP1`, Figure 1a.

Looking at the left plot in Figure 1a, the local slope α of the curve $\log_2(N) \mapsto \log_2 \delta^2_{N,T}(\tilde{\boldsymbol{u}}, \boldsymbol{u})$ for fixed T tells us that the error rate is N^α, for fixed T. Here, $\alpha \approx -1$; hence, the error rates for the factors is about N^{-1} for each T. For N fixed, the spacings β between $\log_2 \delta^2_{N,T}(\tilde{\boldsymbol{u}}, \boldsymbol{u})$ from $T = 50$ to 100 indicates that the error rate is T^β for N fixed. Since $0 \leq -\beta < 0.25$, the error rate for fixed N is less than $T^{-0.25}$. The simulation results give us insight into which of the terms T^{-1} or N^{-1} is dominant, and for the factors in `DGP1`, the dominant term in N^{-1} for $T \in [50, 200]$. We do expect to see an error rate T^{-1} for large fixed N large, and simulations (with $N = 1000$, not shown here) confirm that this is indeed the case. The middle plot of Figure 1a shows the error rate for the loadings. Since the factors and the idiosyncratic component are independent in our simulations, we expect to have the same $O_{\mathrm{P}}(\max(T^{-1}, N^{-1}))$ error rates as for the factors. For the larger values of N, it is clear that the dominant term is T^{-1}. Smaller values of N actually exhibit a transition from the N^{-1} to the T^{-1} regime: the spacings β between the lines becomes more uniform and close to -1 as N increases, and the slope α decreases in magnitude as N increases, and seems to converge to zero. The right sub-figure shows the error rates for the common component, for which we expect, in this setting, the same $O_{\mathrm{P}}(\max(T^{-1}, N^{-1}))$ error rates as for the loadings. Inspection reveals similar effects as for the factor loadings: for $T = 200$, the error rate is close to N^{-1} for small Ns. For $N = 100$, it is almost T^{-1} for small Ts.

Figure 1b (`DGP2`) can be interpreted in a similar fashion, but we shall not delve into this. Compared to `DGP1`, the errors on the factors are much smaller, those on the loadings are a bit smaller for small N; but the gap dereases with N, and the error on the common component are smaller in general, and much



(a) Simulation scenario DGP1.



(b) Simulation scenario DGP2.

Figure 1: Estimations errors (in $\log_2$ scale) for DGP1 (subfigure (a)) and DGP2 (subfigure (b)). For each subfigure, we have the estimation error for the factors ($\log_2 \delta^2_{N,T}$, left), loadings ($\log_2 \varepsilon^2_{N,T}$, middle), and common component ($\log_2 \phi_{N,T}$, right, $\phi_{N,T}$ defined in (4.1)) as functions of $\log_2 N$. The scales of the vertical axes are the same. Each curve corresponds to one value of $T \in \{50, 100, 200\}$, sampled for $N \in \{10, 25, 50, 100\}$.

smaller for small N. This corroborates the previous comment on the difficulty in assessing a priori which, of `DGP1` or `DGP2`, is more favourable.

Results for `DGP3` and `DGP4` are shown in Figure S1, in the Online Supplement. A comparison between `DGP3, DGP4` and `DGP1, DGP2` is interesting because they only differ by the scale of the idiosyncratic components. We see that the errors are much higher for `DGP3, DGP4` than for `DGP1, DGP2`, as expected. Notice that the dominant term for the factors is no longer of order N^{-1} over all values of N, T: it seems to kick in for $N \in [25, 100]$ in `DGP3, DGP4`, but looks slightly higher thanN^{-1} for $N \in [10, 25]$ and $T \in [100, 200]$ in `DGP4` (noticeably so for $T = 200$). A similar phenomenon occurs for the loadings and common component in `DGP4`, for $N \in [10, 25]$ and $T = 200$. These rates do not contradict the theoretical results of Section 3, which hold for $N, T \to \infty$, so `DGP4`, in particular, indicates that the values of N considered there are too small for the asymptotics to have kicked in, and prompts further theoretical investigations about the estimation error rates in finite samples.

5. Empirical illustration

Our model can be used to tackle a plethora of applications in many different domains. Instances of such applications include joint analysis of fertility (or mortality) curves across different regions or countries, modeling electricity demand curves of households or including yield and financial curves into macroeconomic factor models. In the example developed here, we demonstrate the empirical relevance of our method on financial data. More specifically, we jointly model intraday returns of a large collection of US and European stocks.

In order to model the co-movements of asset returns, the financial literature has been considering factor models for several decades. Factor models are intrinsically related to optimal portfolio allocation through the concept of diversification. Namely, investors try to remove the idiosyncratic risk by appropriately weighting the different assets in their portfolios. An early instance of this is the Capital Asset Pricing Model (CAPM) introduced by Sharpe (1964), Lintner (1965), Treynor (1961). In this model, the returns of the various stocks under study are modelled as linear functions of the market return with different factor loadings, called β's. The Fama-French 3-factor model extends CAPM by adding two extra factors (Fama & French 1996). Ross (1976) developed his *Arbitrage Pricing Theory* by proposing a factor model with unspecified number of factors. Chamberlain & Rothschild (1983) proved their representation theorem in the latter context. None of these models is able to handle intraday returns curves, though, let only mixed with overnight and daily returns. These are precisely the type of data our functional approach is made for.

Our dataset contains the returns for 95 S&P100 stocks and 48 Eurostoxx 50 ones observed from 1st of January 2018 to 12th of July 2018 (list available in Tables S1 and S2 in the Online Supplement; we had to dismiss a few stocks for which the data were not available throughout the observation period). For the US stocks, one-minute frequency prices are available, whereas we only have the opening prices for European stocks. Our dataset thus is a mix of high-frequency series (treated as functional series) and scalar ones.

For the US stocks, we have computed *cumulative intraday returns* (CIDR) as defined by Horváth & Kokoszka (2012). If $p_{id,t}$ is the price of stock i at day d and time t (rescaled between 0 and 1), its CIDR at time t is defined as $\log(p_{id,t}) - \log(p_{id,0})$. We have also computed their overnight returns, that is $\log(p_{id+1,0}) - \log(p_{id,1})$. For the European stocks, we have computed daily returns based on the opening prices, namely $\log(p_{id+1,0}) - \log(p_{id,0})$. For each observation date d, we thus have three categories of series:

(i) the CIDR curves of 95 S&P100 stocks represented in a 7-dimensional B-splines basis,
(ii) the overnight returns of the same 95 S&P100 stocks, and
(iii) the daily returns of 48 Eurostoxx 50 stocks.

Prior to the analysis, all series have been centered about their empirical means. Moreover, we have divided all time series belonging to the same category by a constant so that the average variance within each category is one: the objective is to balance the influence of each category. In order to avoid missing data problems, we have chosen to disregard the days on which the US stock exchange was closed; whenever the US stock exchange was open but the European one was closed, we replaced the missing European price by the previous available value.

In Figure 2, we have plotted the first 10 eigenvalues of the covariance operator for different values of N. Since our panel consists of time series of different natures, we have permuted 50 times every cross section. The curves obtained are then computed as the average values obtained for the first N times series (functional or scale) of the permuted panels. Based on the scales of these curves, we have decided to keep $r = 3$ factors. Esablishing a more rigorous criterion for the identification of the number of factors is left for future research.

We have then computed, for each series i, the percentage A_{ij} of variance explained by the j-th factor. In Figure 3, we have plotted six different figures, arranged in three rows and three columns. The plot at the j-th row and k-th column has A_{ij} on the x-axis and A_{ik} on the y-axis; each point represents one series. We notice that factor 1 represents a high proportion of the variance for most European returns and overnight S&P100 returns, but only a very small percentage of the variance of the intraday CIDRs. Factor 2 explains a large percentage of variance for both overnight and intraday S&P100 returns, but a small fraction of the variance of the European returns. These suggest non-negligible co-movement between these series, which calls for further investigation (which is beyond the scope of this paper). The fractions of variance explained by factor 3 is quite small, and its interpretation is also difficult.

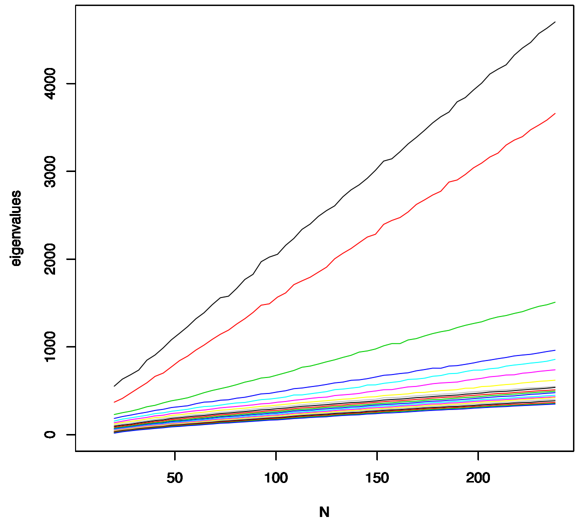


Figure 2: First 10 eigenvalues averaged over 50 permuted cross-sections of varying sizes N.

6. Discussion

We proposed a new paradigm for modelling large panels of functional and scalar time series, based on a new concept of *(high-dimensional) functional factor model*. This model permits to reduce the serial information contained in the panel into a few scalar time series of factors, which encode most of the cross-sectional correlation of the panel. The residual terms of each FTS, uncorrelated with the factors at lag 0, are only mildly cross-correlated (along the cross-section). In particular, this model is weaker than strict factor models (which require mutually strictly uncorrelated idiosyncratic components) or other factor model (such as Lam & Yao 2012, which requires idiosyncratic components to be white noise). We extend to the functional context the classical representation results of Chamberlain (1983), Chamberlain & Rothschild (1983) and propose consistent estimation procedures for the factors, the factor loadings, and the common components, as both the size N of the cross-section and the period T of observation tend to infinity, with no constraints on their relative rates of divergence. Our results also hold for the particular case of scalar panel data, where they reproduce and extend the well-established results of Bai & Ng (2002), but under weaker assumptions. Our proof techniques are therefore of independent interest, in particular since they considerably simplify existing ones, while extending their validity to a functional setting. We then illustrated

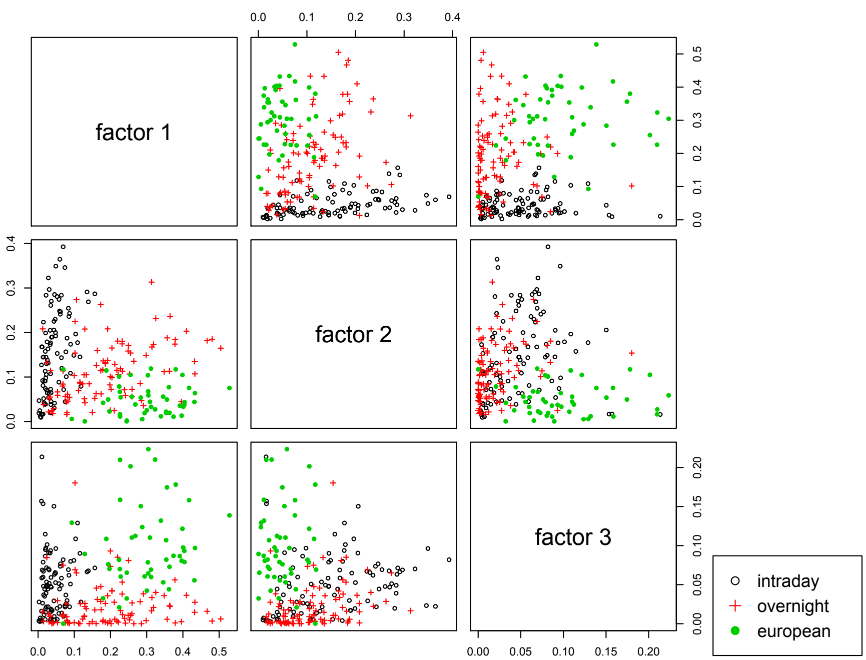


Figure 3: For every subplot (j, k), the variance explained by the j-th factor versus variance explained by the k-th factor is plotted for each 238 series of our empirical illustration (Section 5). This is a pairs plot, whence the symmetry of subplots with respect to the diagonal.

the consistency results by some numerical experiments, which confirm that the rates predicted by the theory are indeed observed empirically, and also provide finer description of the interplay between the size of the cross-section and the length of the (functional) time series. We concluded the paper by providing an empirical illustration of a functional factor model applied to a panel of time series of mixed nature (some functional and some scalar), and showing how one can use the model to assess co-movement between the series.

Extensions of this present work could be in the direction of developing information criteria for identifying the number of factors (Bai & Ng 2002), or extending the theory and methodology of Fan et al. (2013) for dealing with estimation of high-dimensional conditional covariance operator matrices. The factor models presented here have also links with high-dimensional covariance models with very spiked eigenvalues (Cai et al. 2017). Further extensions could be in the direction of generalized dynamic factor models (Forni et al. 2000, Forni & Lippi 2001, Forni et al. 2015, 2017).

Supplementary Material

The R code reproducing the numerical experiments of Section 4, as well as code used for Section 5, can be obtained by contacting the authors by email.

Online Supplement: "High-Dimensional Functional Factor Models" (doi: TYPESETTERS: PLEASE UPDATE THIS FIELD). Contains additional proofs, background results, and additional figures and tables.

References

Alessi, L., Barigozzi, M. & Capasso, M. (2009), 'Estimation and forecasting in large datasets with conditionally heteroskedastic dynamic common factors', *ECB Working Paper No. 1115* .

Amengual, D. & Watson, M. W. (2007), 'Consistent estimation of the number of dynamic factors in a large N and T panel', *J. Bus. Econ. Stat.* **25**(1), 91–96.

Aue, A., Hörmann, S., Horváth, L. & Hušková, M. (2014), 'Dependent functional linear models with applications to monitoring structural change', *Stat. Sinica* **24**(3).

Aue, A., Horváth, L. & Pellatt, D. F. (2017), 'Functional Generalized Autoregressive Conditional Heteroskedasticity', *J. Time Ser. Anal.* **38**(1), 3–21.

Aue, A., Norinho, D. D. & Hörmann, S. (2015), 'On the Prediction of Stationary Functional Time Series', *J. Am. Stat. Assoc.* **110**(509), 378–392. **URL:** *http://dx.doi.org/10.1080/01621459.2014.909317*

Bai, J. (2003), 'Inferential theory for factor models of large dimensions', *Econometrica* **71**(1), 135–171.

Bai, J. & Ng, S. (2002), 'Determining the number of factors in approximate factor models', *Econometrica* **70**(1), 191–221.

Barigozzi, M., Hallin, M. & Soccorsi, S. (2019), 'Time-varying general dynamic factor models and the measurement of financial connectedness', *Available at SSRN 3329445* .

Boivin, J. & Ng, S. (2006), 'Are more data always better for factor analysis?', *J. Econometrics* **132**(1), 169–194.

Bosq, D. (2000), *Linear Processes in Function Spaces*, Springer.

Bradley, R. C. (2005), 'Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions', *Probab. Surv.* **2**(0), 107–144.
URL: *http://projecteuclid.org/euclid.ps/1115386870*

Brillinger, D. R. (2001), *Time Series: Data Analysis and Theory*, Vol. 36 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Data analysis and theory, Reprint of the 1981 edition.
URL: *http://dx.doi.org/10.1137/1.9780898719246*

Bücher, A., Dette, H. & Heinrichs, F. (2018), 'Detecting deviations from second-order stationarity in locally stationary functional time series', *arXiv preprint arXiv:1808.04092* .

Bühlmann, P. & van de Geer, S. (2011), *Statistics for high-dimensional data: Methods, Theory and Applications*, Springer.

Cai, T., Han, X. & Pan, G. (2017), 'Limiting laws for divergent spiked eigenvalues and largest non-spiked eigenvalue of sample covariance matrices', *arXiv preprint arXiv:1711.00217* .

Chamberlain, G. (1983), 'Funds, factors, and diversification in arbitrage pricing models', *Econometrica* **51**, 1281–1304.

Chamberlain, G. & Rothschild, M. (1983), 'Arbitrage, factor structure, and mean-variance analysis on large asset markets', *Econometrica* **51**, 1305–1324.

Chudik, A., Pesaran, M. H. & Tosetti, E. (2011), 'Weak and strong cross-section dependence and estimation of large panels', *Economet. J.* **14**(1).

Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S. & Prieur, C. (2007), *Weak Dependence: With Examples and Applications*, Vol. 190 of *Lecture Notes in Statistics*, Springer.

Fama, E. & French, K. (1996), 'Multifactor explanations of asset pricing anomalies', *J. Financ.* **51**(1), 55–84.

Fan, J., Liao, Y. & Mincheva, M. (2013), 'Large covariance estimation by thresholding principal orthogonal complements', *J. Roy. Stat. Soc. B* **75**(4), 603–680.

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000), 'The Generalized Dynamic-Factor Model: Identification and Estimation', *Rev. Econ. Stat.* **65**(4), 453–554.

Forni, M., Hallin, M., Lippi, M. & Zaffaroni, P. (2015), 'Dynamic factor models with infinite-dimensional factor spaces: One-sided representations', *J. Econo-*

metrics **185**(2), 359–371.

Forni, M., Hallin, M., Lippi, M. & Zaffaroni, P. (2017), 'Dynamic factor models with infinite-dimensional factor space: Asymptotic analysis', *J. Econometrics* **199**(1), 74–92.

Forni, M. & Lippi, M. (2001), 'The generalized dynamic factor model: Representation theory', *Economet. Theor.* **17**(06), 1113–1141.

Forni, M. & Reichlin, L. (1998), 'Let's Get Real: A Factor Analytical Approach to Disaggregated Business Cycle Dynamics', *Rev. Econ. Stud.* **65**(3), 453–473.

Gao, Y., Shang, H. L. & Yang, Y. (2018), 'High-dimensional functional time series forecasting: An application to age-specific mortality rates', *J. Multivariate Anal.* **170**, 232–243.
URL: *http://www.sciencedirect.com/science/article/pii/S0047259X1730739X*

Geweke, J. (1977), 'The Dynamic Factor Analysis of Economic Time-Series Models'.

Górecki, T., Hörmann, S., Horváth, L. & Kokoszka, P. (2018), 'Testing Normality of Functional Time Series', *J. Time Ser. Anal.* .

Guo, S. & Qiao, X. (2018), 'A General Theory for Large-Scale Curve Time Series via Functional Stability Measure', *arXiv preprint arXiv:1812.07619* .

Hallin, M. & Lippi, M. (2013), 'Factor models in high-dimensional time series. a time-domain approach', *Stoch. Proc. Appl.* **123**(7), 2678–2695.

Hays, S., Shen, H. & Huang, J. Z. (2012), 'Functional dynamic factor models with application to yield curve forecasting', *Ann. Appl. Stat.* **6**, 870–894.

Hörmann, S., Horváth, L. & Reeder, R. (2013), 'A functional version of the ARCH model', *Economet. Theor.* **29**, 267–288.

Hörmann, S., Kidziński, L. & Hallin, M. (2015), 'Dynamic functional principal components', *J. Roy. Stat. Soc. B* **77**, 319–348.

Hörmann, S. & Kokoszka, P. (2010), 'Weakly dependent functional data', *Ann. Stat.* **38**, 1845–1884.

Hörmann, S. & Kokoszka, P. (2012), Functional time series, *in* C. R. Rao & T. S. Rao, eds, 'Time Series', Vol. 30 of *Handbook of Statistics*, Elsevier.

Hörmann, S., Kokoszka, P. & Nisol, G. (2018), 'Testing for periodicity in functional time series', *Ann. Stat.* **46**(6A), 2960–2984.

Horváth, L. & Kokoszka, P. (2012), *Inference for Functional Data with Applications*, Springer.

Horváth, L., Rice, G. & Whipple, S. (2014), 'Adaptive bandwidth selection in the long run covariance estimator of functional time series', *Comput. Stat. Data An.* **100**, 676–693.

Hsing, T. & Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley.

Kokoszka, P., Miao, H. & Zhang, X. (2015), 'Functional dynamic factor model for intraday price curves', *J. Financ. Economet.* **13**, 456–477.

Kokoszka, P. & Reimherr, M. (2013*a*), 'Asymptotic normality of the principal components of functional time series', *Stoch. Proc. Appl.* **123**, 1546–1558.

Kokoszka, P. & Reimherr, M. (2013*b*), 'Determining the order of the functional autoregressive model', *J. Time Ser. Anal.* **34**, 116–129.

Lam, C. & Yao, Q. (2012), 'Factor modeling for high-dimensional time series: Inference for the number of factors', *Ann. Stat.* **40**(2), 694–726.

Lintner, J. (1965), 'The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets', *Rev. Econ. Stat.* **47**(1), 13–37.

Onatski, A. (2010), 'Determining the number of factors from empirical distribution of eigenvalues', *Rev. Econ. Stat.* **92**(4), 1004–1016.

Panaretos, V. M. & Tavakoli, S. (2013*a*), 'Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series.', *Stoch. Proc. Appl.* **123**(7), 2779–2807.

Panaretos, V. M. & Tavakoli, S. (2013*b*), 'Fourier analysis of stationary time series in function space.', *Ann. Stat.* **41**(2), 568–603.

Ramsay, J. O. & Silverman, B. W. (2005), *Functional Data Analysis*, Springer Series in Statistics, second edn, Springer, New York.

Ross, S. A. (1976), 'The arbitrage theory of capital asset pricing', *J. Econ. Theory* **13**, 341–360.

Rubín, T. & Panaretos, V. M. (2018), 'Sparsely Observed Functional Time Series: Estimation and Prediction', *arXiv preprint arXiv:1811.06340* .
URL: *http://arxiv.org/abs/1811.06340*

Sargent, T. J. & Sims, C. A. (1977), Business cycle modeling without pretending to have too much a priori economic theory, Working Papers 55, Federal Reserve Bank of Minneapolis.

Sharpe, W. (1964), 'Capital asset prices: a theory of market equilibrium under conditions of risk', *J. Financ.* **19**, 425–442.

Stock, J. H. & Watson, M. W. (2002*a*), 'Forecasting using principal components from a large number of predictors', *J. Am. Stat. Assoc.* **97**(460), 1167–1179.
URL: *http://www.jstor.org/stable/3085839*

Stock, J. H. & Watson, M. W. (2002*b*), 'Macroeconomic forecasting using diffusion indexes', *J. Bus. Econ. Stat.* **20**(2), 147–162.
URL: *https://doi.org/10.1198/073500102317351921*

Tavakoli, S. & Panaretos, V. M. (2016), 'Detecting and localizing differences in functional time series dynamics: a case study in molecular biophysics', *J. Am. Stat. Assoc.* **111**(515), 1020–1035.

Treynor, J. L. (1961), 'Market value, time, and risk', *Unpublished manuscript* .

Trucios, C. C., Mazzeu, J., Hallin, M., Hotta, L. K., Valls Pereira, P. L. & Zevallos, M. (2019), 'Forecasting conditional covariance matrices in high-dimensional time series: a general dynamic factor approach', *ECARES Working Paper 2019-14* .

van Delft, A., Characiejus, V. & Dette, H. (2017), 'A nonparametric test for stationarity in functional time series', *arXiv preprint arXiv:1708.05248* .
URL: *http://arxiv.org/abs/1708.05248*

van Delft, A. & Dette, H. (2018), 'A similarity measure for second order prop-

erties of non-stationary functional time series with applications to clustering and testing', *arXiv preprint arXiv:1810.08292* .
URL: *http://arxiv.org/abs/1810.08292*

van Delft, A. & Eichler, M. (2018), 'Locally stationary functional time series', *Electron. J. Stat.* **12**(1), 107–170.

Wang, J.-L., Chiou, J.-M. & Mueller, H.-G. (2015), 'Review of Functional Data Analysis', *Annu. Rev. Stat. Appl.* **3**, 257–295.

Supplementary material for "High-Dimensional Functional Factor Models"

Gilles Nisol, Shahin Tavakoli* and Marc Hallin*

ESAT - STADIUS,
Katholieke Universiteit Leuven
3001 Leuven
Belgium
e-mail: gnisol@esat.kuleuven.be

Department of Statistics
University of Warwick
CV4 7AL Coventry
United Kingdom
e-mail: s.tavakoli@warwick.ac.uk

ECARES and Département de Mathématique
Université Libre de Bruxelles CP 114
50, avenue F.D. Roosevelt
1050 Bruxelles
Belgium
e-mail: mhallin@ulb.ac.be

All the numberings (Sections, Figures, Tables, Lemmas, etc.) in this document start with the letter "S". All references not starting with "S" refer to the main paper.

S1. Orthogonal Projections

Let H be a separable Hilbert space. For $X, Y : \Omega \to H$, let

$$\|X\|_{L^2(\Omega)} := \sqrt{\langle X, X \rangle_{L^2(\Omega)}},$$

where $\langle X, Y \rangle_{L^2(\Omega)} = \mathbb{E} \langle X, Y \rangle$. Although the notation is similar to that used for the norm and covariance of random variables, it will be clear from the context which norm is being used. Let $L^2_H(\Omega)$ be the space of H-valued random elements $X : \Omega \to H$ with $\mathbb{E} X = 0$ and $\|X\|_{L^2(\Omega)} < \infty$.

*The author(s) would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme Statistical Scalability where work on this paper was undertaken. This work was supported by EPSRC grant no. EP/K032208/1.

Ticker	Company name	Sector	Ticker	Company name	Sector
AAPL	Apple Inc	Technology	HD	Home Depot	Consumer Discretionary
ABT	Abbott Laboratories	Health Care	HON	Honeywell	Industrial
ACN	Accenture	Technology	IBM	International Business Machines	Technology
AGN	Allergan	Health Care	INTC	Intel Corp.	Technology
AIG	American International Group	Financials	JNJ	Johnson & Johnson	Health Care
ALL	Allstate	Financials	JPM	JPMorgan Chase & Co.	Financials
AMGN	Amgen Inc.	Health Care	KHC	Kraft Heinz	Consumer Staples
AMZN	Amazon.com	Consumer Discretionary	KMI	Kinder Morgan	Energy
AXP	American Express	Financials	KO	The Coca-Cola Company	Consumer Staples
BA	Boeing Co.	Industrial	LLY	Eli Lilly and Company	Health Care
BAC	Bank of America Corp	Financials	LMT	Lockheed Martin	Industrial
BIIB	Biogen	Health Care	LOW	Lowe’s	Consumer Discretionary
BK	The Bank of New York Mellon	Financials	MA	MasterCard Inc	Technology
BKNG	Booking Holdings	Consumer Discretionary	MCD	McDonald’s Corp	Consumer Discretionary
BLK	BlackRock Inc	Financials	MDLZ	Mondel¿z International	Consumer Staples
BMY	Bristol-Myers Squibb	Health Care	MDT	Medtronic plc	Health Care
BRKB	Berkshire Hathaway	Financials	MET	MetLife Inc.	Financials
C	Citigroup Inc	Financials	MMM	3M Company	Industrial
CAT	Caterpillar Inc.	Industrial	MO	Altria Group	Consumer Staples
CELG	Celgene Corp	Health Care	MRK	Merck & Co.	Health Care
CHTR	Charter Communications	Communication Services	MS	Morgan Stanley	Financials
CL	Colgate-Palmolive	Consumer Staples	MSFT	Microsoft	Technology
CMCSA	Comcast Corp.	Communication Services	NEE	NextEra Energy	Utilities
COF	Capital One Financial Corp.	Financials	NFLX	Netflix	Communication Services
COP	ConocoPhillips	Energy	NKE	Nike, Inc.	Consumer Discretionary
COST	Costco Wholesale Corp.	Consumer Staples	NVDA	NVIDIA Corp.	Technology
CSCO	Cisco Systems	Technology	ORCL	Oracle Corporation	Technology
CVS	CVS Health	Health Care	OXY	Occidental Petroleum Corp.	Energy
CVX	Chevron Corporation	Energy	PEP	PepsiCo	Consumer Staples
DHR	Danaher Corporation	Health Care	PFE	Pfizer Inc	Health Care
DIS	The Walt Disney Company	Communication Services	PG	Procter & Gamble Co	Consumer Staples
DUK	Duke Energy	Utilities	PM	Philip Morris International	Consumer Staples
DWDP	DowDuPont	Materials	PYPL	PayPal Holdings	Technology
EMR	Emerson Electric Co.	Industrial	QCOM	Qualcomm Inc.	Technology
EXC	Exelon	Utilities	RTN	Raytheon Co.	Industrial
F	Ford Motor Company	Consumer Discretionary	TXN	Texas Instruments	Technology
FB	Facebook	Technology	UNH	UnitedHealth Group	Health Care
FDX	FedEx	Industrial	UNP	Union Pacific Corporation	Industrial
FOX	Fox Corporation B	Communication Services	UPS	United Parcel Service	Industrial
FOXA	Fox Corporation A	Communication Services	USB	U.S. Bancorp	Financials
GD	General Dynamics	Industrial	UTX	United Technologies	Industrial
GE	General Electric	Industrial	V	Visa Inc.	Technology
GILD	Gilead Sciences	Health Care	VZ	Verizon Communications	Communication Services
GM	General Motors	Consumer Discretionary	WBA	Walgreens Boots Alliance	Consumer Staples
GOOG	Alphabet Inc. C	Technology	WFC	Wells Fargo	Financials
GOOGL	Alphabet Inc. A	Technology	WMT	Walmart	Consumer Staples
GS	Goldman Sachs	Financials	XOM	Exxon Mobil Corp.	Energy
HAL	Halliburton Company	Energy			

Table S1: List of US stocks under study.

Ticker	Company name	Sector
ADS.DE	Adidas	Consumer Discretionary
AD.AS	Ahold Delhaize	Consumer Staples
AI.PA	Air Liquide	Materials
AIR.PA	Airbus	Industrial
ALV.DE	Allianz	Financials
ABI.BR	Anheuser-Busch InBev	Consumer Staples
ASML.AS	ASML Holding	Technology
CS.PA	AXA	Financials
BBVA.MC	Banco Bilbao Vizcaya Argentaria	Financials
SAN.MC	Banco Santander	Financials
BAS.DE	BASF	Materials
BAYN.DE	Bayer	Health Care
BMW3.DE	BMW	Consumer Discretionary
BNP.PA	BNP Paribas	Financials
CRG.IR	CRH	Materials
SGO.PA	Compagnie de Saint-Gobain	Materials
DAI.DE	Daimler AG	Consumer Discretionary
DPW.DE	Deutsche Post	Industrial
DTE.DE	Deutsche Telekom	Communication Services
ENEL.MI	Enel	Utilities
ENGI.PA	Engie	Utilities
ENI.MI	Eni	Energy
EOAN.DE	E.ON	Utilities
FRE.DE	Fresenius SE	Health Care
BN.PA	Groupe Danone	Consumer Staples
IBE.MC	Iberdrola	Utilities
ITX.MC	Inditex	Consumer Discretionary
INGA.AS	ING Group NV	Financials
ISP.MI	Intesa Sanpaolo	Financials
OR.PA	L'Oréal	Consumer Staples
MC.PA	LVMH Moët Hennessy Louis Vuitton	Consumer Discretionary
MUV2.DE	Munich Re	Financials
NOKIAsci hub .HE	Nokia	Technology
ORA.PA	Orange S.A.	Communication Services
PHIA.AS	Philips	Health Care
SAF.PA	Safran	Industrial
SAN.PA	Sanofi	Health Care
SAP.DE	SAP SE	Technology
SU.PA	Schneider Electric	Industrial
SIE.DE	Siemens	Industrial
GLE.PA	Société Générale SA	Financials
TEF.MC	Telefónica	Communication Services
FP.PA	TOTAL S.A.	Energy
URW.AS	Unibail-Rodamco	Real Estate
UNA.AS	Unilever	Consumer Staples
DG.PA	Vinci SA	Industrial
VIV.PA	Vivendi	Consumer Discretionary
VOW.DE	Volkswagen Group	Consumer Discretionary

Table S2

List of euro stocks under study.

For any finite-dimensional subspace $\mathcal{U} \subset L^2(\Omega)$, let

$$\operatorname{span}_H(\mathcal{U}) := \left\{ \sum_{j=1}^{m} b_j u_j : b_j \in H, u_j \in \mathcal{U}, m = 1, 2, \dots \right\}. \tag{S1.1}$$

Since $\mathcal{U}$ is finite-dimensional, $\operatorname{span}_H(\mathcal{U}) \subseteq L^2_H(\Omega)$ is a closed subspace. Indeed, let $u_1, \dots, u_r \in \mathcal{U}$ be an orthonormal basis, $r < \infty$, then

$$\|b_1 u_1 + \dots + b_r u_r\|^2_{L^2(\Omega)} = \|b_1\|^2 + \dots + \|b_r\|^2.$$

By the orthogonal decomposition Theorem (Hsing & Eubank 2015, Theorem 2.5.2), for any $X \in L^2_H(\Omega)$, there exists a unique $U[X] \in \operatorname{span}_H(\mathcal{U})$ such that

$$X = U[X] + V[X], \tag{S1.2}$$

where $V[X] = X - U[X] \in \operatorname{span}_H(\mathcal{U})^\perp$; hence $\mathbb{E}[uV[X]] = 0$ for all $u \in \mathcal{U}$ and

$$\mathbb{E}\|X\|^2 = \mathbb{E}\|U[X]\|^2 + \mathbb{E}\|V[X]\|^2. \tag{S1.3}$$

We have the following definition.

Definition S1.1. *Equation* (S1.2) *is called the* orthogonal decomposition *of* X *onto* $\operatorname{span}_H(\mathcal{U})$ *and its orthogonal complement;* $U[X] =: \operatorname{proj}_H(X|\mathcal{U})$ *is the* orthogonal projection *of* X *onto* $\operatorname{span}_H(\mathcal{U})$.

If $u_1, \dots, u_r \in L^2(\Omega)$ form a basis of $\mathcal{U}$, then $\operatorname{proj}_H(X|\mathcal{U}) = \sum_{l=1}^r b_l u_l$ for some unique $b_1, \dots, b_r \in H$. Furthermore, if the u_ls are orthonormal, then $b_l = \mathbb{E}[Xu_l]$, $l = 1, \dots, r$. We shall also use the notation

$$\operatorname{proj}_H(X|\boldsymbol{u}) := \operatorname{proj}_H(X|u_1, \dots, u_r) := \operatorname{proj}_H(X|\mathcal{U}),$$

where $\mathcal{U} = \operatorname{span}(u_1, \dots, u_r)$ and $\boldsymbol{u} = (u_1, \dots, u_r)'$.

S2. Proofs

S2.1. Proofs of Theorems 2.2 and 2.3

We denote by $\boldsymbol{\Sigma}_N$ the covariance operator of $(X_{1t}, \dots, X_{Nt})'$, which does not depend on t by stationarity. Denoting by $\boldsymbol{p}_{N,i} \in \boldsymbol{H}_N$ the ith eigenvector of $\boldsymbol{\Sigma}_N$, we have the following eigendecomposition of the covariance operator,

$$\boldsymbol{\Sigma}_N = \sum_{i=1}^{\infty} \lambda_{N,i} \boldsymbol{p}_{N,i} \boldsymbol{p}_{N,i}^{\mathsf{T}}.$$

(Notice that the notation here differs from that of Section 3, but this is not an issue since we are dealing here with the population level $T = \infty$) The eigenvectors $(\boldsymbol{p}_{N,i})_i \subset \boldsymbol{H}_N$ form an orthonormal basis of the image $\operatorname{Im}(\boldsymbol{\Sigma}_N) \subset \boldsymbol{H}_N$

of $\boldsymbol{\Sigma}_N$. We can extend the set of eigenvectors of $\boldsymbol{\Sigma}_N$ to form an orthonormal basis of $\boldsymbol{H}_N$. With a slight abuse of notation, we shall denote this basis by $(\boldsymbol{p}_{N,i})_i$, possibly reordering the eigenvalues (and having eigenvalues equal to zero): we might no longer have non-increasing eigenvalues, but we can enforce $\lambda^x_{N,1} \geq \cdots \geq \lambda^x_{N,r+1}$ and $\lambda^x_{N,r+1} \geq \lambda^x_{N,r+1+j}, \forall j \geq 1$. Define

$$\boldsymbol{P}_N := \begin{pmatrix} \boldsymbol{p}_{N,1} & \cdots & \boldsymbol{p}_{N,r} \end{pmatrix} \in \mathcal{L}(\mathbb{R}^r, \boldsymbol{H}_N)$$

and

$$\boldsymbol{Q}_N := \begin{pmatrix} \boldsymbol{p}_{N,r+1} & \boldsymbol{p}_{N,r+2} & \cdots \end{pmatrix} \in \mathcal{L}(\ell_2, \boldsymbol{H}_N)$$

where $\ell_2 := \left\{(\alpha_1, \alpha_2, \ldots) : \alpha_i \in \mathbb{R}, \sum_i \alpha_i^2 < +\infty\right\}$. Denote by $\boldsymbol{\Lambda}_N \in \mathcal{L}(\mathbb{R}^r)$ and $\boldsymbol{\Phi}_N \in \mathcal{L}(\ell_2)$ the diagonal matrices with diagonal elements $(\lambda_{N,1}, \ldots, \lambda_{N,r})$ and $(\lambda_{N,r+1}, \lambda_{N,r+2}, \ldots)$, respectively. Then,

$$\boldsymbol{\Sigma}_N = \boldsymbol{P}_N \boldsymbol{\Lambda}_N \boldsymbol{P}_N^\mathsf{T} + \boldsymbol{Q}_N \boldsymbol{\Phi}_N \boldsymbol{Q}_N^\mathsf{T} \quad \text{and} \quad \boldsymbol{P}_N \boldsymbol{P}_N^\mathsf{T} + \boldsymbol{Q}_N \boldsymbol{Q}_N^\mathsf{T} = \boldsymbol{I}_N, \tag{S2.1}$$

where $\boldsymbol{I}_N$ is the identity operator on $\boldsymbol{H}_N$.

The analysis we are going to perform is for fixed t, letting $N \to \infty$. We therefore omit the index t, unless needed, and write $\boldsymbol{X}_N$ for $(X_{1t}, \ldots, X_{Nt})'$. Let

$$\boldsymbol{\psi}^N := \boldsymbol{\Lambda}_N^{-1/2} \boldsymbol{P}_N^\mathsf{T} \boldsymbol{X}_N = (\lambda_{N,1}^{-1/2} \boldsymbol{p}_{N,1}^\mathsf{T} \boldsymbol{X}_N, \ldots, \lambda_{N,r}^{-1/2} \boldsymbol{p}_{N,r}^\mathsf{T} \boldsymbol{X}_N)'. \tag{S2.2}$$

Notice that, by Lemma S2.18, $\boldsymbol{\psi}^N$ is well defined for N large enough since $\lambda_{N,r}$ tends to infinity. Using (S2.1), we get

$$\boldsymbol{X}_N = \boldsymbol{P}_N \boldsymbol{P}_N^\mathsf{T} \boldsymbol{X}_N + \boldsymbol{Q}_N \boldsymbol{Q}_N^\mathsf{T} \boldsymbol{X}_N = \boldsymbol{P}_N \boldsymbol{\Lambda}_N^{1/2} \boldsymbol{\psi}^N + \boldsymbol{Q}_N \boldsymbol{Q}_N^\mathsf{T} \boldsymbol{X}_N, \tag{S2.3}$$

where the two summands are uncorrelated since $\boldsymbol{P}_N^\mathsf{T} \boldsymbol{\Sigma}_N \boldsymbol{Q}_N = 0 \in \mathcal{L}(\ell_2, \boldsymbol{H}_N)$, the zero operator: (S2.3) is in fact the orthogonal decomposition of $\boldsymbol{X}_N$ into $\text{span}_{\boldsymbol{H}_N}(\boldsymbol{\psi}^N)$ and its orthogonal complement, defined in Appendix S1. For $m < n$, let us define products such as $\boldsymbol{P}_M^\mathsf{T} \boldsymbol{P}_N$ by extending the smaller matrix by adding zeros. For instance,

$$\boldsymbol{P}_M^\mathsf{T} \boldsymbol{P}_N := \begin{pmatrix} \boldsymbol{p}_{M,1}^\mathsf{T} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{p}_{M,r}^\mathsf{T} & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{p}_{N,1} & \cdots & \boldsymbol{p}_{N,r} \end{pmatrix},$$

where we have added $N - M$ columns of zeros to $\boldsymbol{P}_M^\mathsf{T}$. Let $\mathcal{O}(r)$ be the set of $r \times r$ orthogonal matrices, i.e. matrices $\boldsymbol{C}$ such that $\boldsymbol{C}\boldsymbol{C}^\mathsf{T} = \boldsymbol{C}^\mathsf{T}\boldsymbol{C} = \boldsymbol{I}_r$, the $r \times r$ identity matrix. For $\boldsymbol{C} \in \mathcal{O}(r)$, left-multiplying both sides of (S2.3) by $\boldsymbol{C}\boldsymbol{\Lambda}_M^{-1/2} \boldsymbol{P}_M^\mathsf{T}$, with $m < n$, yields

$$\begin{aligned} \boldsymbol{C}\boldsymbol{\psi}^M &= \boldsymbol{C}\boldsymbol{\Lambda}_M^{-1/2} \boldsymbol{P}_M^\mathsf{T} \boldsymbol{P}_N \boldsymbol{\Lambda}_N^{1/2} \boldsymbol{\psi}^N + \boldsymbol{C}\boldsymbol{\Lambda}_M^{-1/2} \boldsymbol{P}_M^\mathsf{T} \boldsymbol{Q}_N \boldsymbol{Q}_N^\mathsf{T} \boldsymbol{X}_N \\ &= \boldsymbol{D}\boldsymbol{\psi}^N + \boldsymbol{R}\boldsymbol{X}_N, \end{aligned} \tag{S2.4}$$

which is the orthogonal decomposition of $\boldsymbol{C}\boldsymbol{\psi}^M$ onto the span of $\boldsymbol{\psi}^N$ and its orthogonal complement, since the two summands are uncorrelated. Notice that $\boldsymbol{D} = \boldsymbol{D}[\boldsymbol{C}, M, N]$, and similarly $\boldsymbol{R} = \boldsymbol{R}[\boldsymbol{C}, M, N]$, where we use square brackets to denote dependence on variables. The following Lemma gives a bound on the residual term in (S2.4). Write $\boldsymbol{A} \succeq \boldsymbol{B}$ for $\boldsymbol{A} - \boldsymbol{B}$ non-negative definite.

Lemma S2.1. *The largest eigenvalue of the covariance operator of the residual $\boldsymbol{R}[\boldsymbol{C}, M, N]\boldsymbol{X}_N$ in (S2.4) is bounded by $\lambda_{N,r+1}/\lambda_{M,r}$.*

Proof. We know that $\boldsymbol{I}_N \succeq \boldsymbol{Q}_N \boldsymbol{Q}_N^\top$ and $\lambda_{N,r+1}\boldsymbol{Q}_N\boldsymbol{Q}_N^\top \succeq \boldsymbol{Q}_N \boldsymbol{\Phi}_N \boldsymbol{Q}_N^\top$. Hence, $\lambda_{N,r+1}\boldsymbol{I}_N \succeq \boldsymbol{Q}_N \boldsymbol{\Phi}_N \boldsymbol{Q}_N^\top$. Multiplying to the left by $\boldsymbol{C}\boldsymbol{\Lambda}_M^{-1/2}\boldsymbol{P}_M^\top$ and to the right by its adjoint, and using the fact that $\boldsymbol{\Phi}_N = \boldsymbol{Q}_N^\top \boldsymbol{\Sigma}_N \boldsymbol{Q}_N$, we get

$$\lambda_{N,r+1}\boldsymbol{C}\boldsymbol{\Lambda}_M^{-1}\boldsymbol{C}^\top \succeq \boldsymbol{C}\boldsymbol{\Lambda}_M^{-1/2}\boldsymbol{P}_M^\top \boldsymbol{Q}_N \boldsymbol{Q}_N^\top \boldsymbol{\Sigma}_N \boldsymbol{Q}_N \boldsymbol{Q}_N^\top \boldsymbol{P}_M \boldsymbol{\Lambda}_M^{-1/2}\boldsymbol{C}^\top = \boldsymbol{R}\boldsymbol{\Sigma}_N\boldsymbol{R}^\top.$$

Lemma S2.18 completes the proof, since the largest eigenvalue on the left-hand side is $\lambda_{N,r+1}/\lambda_{M,r}$. □

Let us now take covariances on both sides of (S2.4). We get

$$\boldsymbol{I}_r = \boldsymbol{D}\boldsymbol{D}^\top + \boldsymbol{R}\boldsymbol{\Sigma}_N\boldsymbol{R}^\top.$$

Denoting by δ_i the ith largest eigenvalue of $\boldsymbol{D}\boldsymbol{D}^\top$, we have

$$1 - \frac{\lambda_{N,r+1}}{\lambda_{M,r}} \leq \delta_i \leq 1, \tag{S2.5}$$

by Lemma S2.1 and Lemma S2.18. Thus, for $N > M \geq M^*$, all δ_is are strictly positive and, since $\lambda_{r+1} < \infty$ and $\lambda_r = \infty$, δ_i can be made arbitrarily close to one by choosing M^* large enough. Denoting by $\boldsymbol{U}\boldsymbol{\Delta}^{1/2}\boldsymbol{V}^\top$ the singular decomposition of $\boldsymbol{D}$, where $\boldsymbol{\Delta}$ is the diagonal matrix of $\boldsymbol{D}\boldsymbol{D}^\top$'s eigenvalues $\delta_1 \geq \delta_2 \geq \cdots \geq \delta_r$, define

$$\boldsymbol{F} := \boldsymbol{F}[\boldsymbol{D}] = \boldsymbol{F}[\boldsymbol{C}, M, N] = \boldsymbol{U}\boldsymbol{V}^\top, \qquad \boldsymbol{D} = \boldsymbol{U}\boldsymbol{\Delta}^{1/2}\boldsymbol{V}^\top. \tag{S2.6}$$

Notice that (S2.5) implies that $\boldsymbol{F}$ is well-defined for M, N large enough, and that $\boldsymbol{F} \in \mathcal{O}(r)$. The following Lemma shows that $\boldsymbol{C}\boldsymbol{\psi}^M$ is well approximated by $\boldsymbol{F}\boldsymbol{\psi}^N$.

Lemma S2.2. *For every $\varepsilon > 0$, there exists an M_ε such that, for all $N > M \geq M_\varepsilon$, $\boldsymbol{F} = \boldsymbol{F}[\boldsymbol{C}, M, N]$ is well defined, and the largest eigenvalue of the covariance of*

$$\boldsymbol{C}\boldsymbol{\psi}^M - \boldsymbol{F}\boldsymbol{\psi}^N = \boldsymbol{C}\boldsymbol{\psi}^M - \boldsymbol{F}[\boldsymbol{C}, M, N]\boldsymbol{\psi}^N$$

is smaller than ε for all $N > M \geq M_\varepsilon$.

Proof. First notice that it suffices to take $M_\varepsilon > M^*$ for $\boldsymbol{F}$ to be well defined. We have

$$\boldsymbol{C}\boldsymbol{\psi}^M - \boldsymbol{F}\boldsymbol{\psi}^M = \boldsymbol{R}\boldsymbol{X}_N + (\boldsymbol{D} - \boldsymbol{F})\boldsymbol{\psi}^N,$$

and since the two summands on the right-hand side are uncorrelated, the covariance of the sum is the sum of their covariances. Denoting by $\boldsymbol{S}$ the covariance of the left-hand side, and by $|||\boldsymbol{S}|||_\infty$ the operator norm of $\boldsymbol{S}$, and noting that $\boldsymbol{D} - \boldsymbol{F} = \boldsymbol{U}(\boldsymbol{\Delta}^{1/2} - \boldsymbol{I}_r)\boldsymbol{V}^\top$, we get

$$|||\boldsymbol{S}|||_\infty \leq \left|\left|\left|\boldsymbol{R}\boldsymbol{\Sigma}_N\boldsymbol{R}^\top\right|\right|\right|_\infty + \left|\left|\left|\boldsymbol{U}(\boldsymbol{\Delta}^{1/2} - \boldsymbol{I}_r)\boldsymbol{V}^\top\boldsymbol{V}(\boldsymbol{\Delta}^{1/2} - \boldsymbol{I}_r)\boldsymbol{U}^\top\right|\right|\right|_\infty \tag{S2.7}$$

$$= \left|\left|\left|\boldsymbol{R}\boldsymbol{\Sigma}_N\boldsymbol{R}^\top\right|\right|\right|_\infty + \left|\left|\left|\boldsymbol{\Delta}^{1/2} - \boldsymbol{I}_r\right|\right|\right|_\infty^2, \tag{S2.8}$$

since $\boldsymbol{U}$ and $\boldsymbol{V}$ are unitary matrices. The first summand of (S2.8) can be made smaller than $\varepsilon/2$ for m large enough (by Lemma S2.1), and the second summand can be made smaller than $\varepsilon/2$ using (S2.5). ☐

A careful inspection of the proofs of these results shows that they hold for all values of t, i.e., writing $\boldsymbol{\psi}^{n,t} = \boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{P}_N^\top\boldsymbol{X}_N$, the result of Lemma S2.2 holds for the difference $\boldsymbol{D}\boldsymbol{\psi}^{m,t} - \boldsymbol{F}\boldsymbol{\psi}^{n,t}$, with a value of M_ε that does not depend on t. The following results provides a construction of the process $\boldsymbol{u}_t$.

Proposition S2.3. *There exists an r-dimensional second-order stationary process $\boldsymbol{u}_t = (u_{1t}, \ldots, u_{rt})'$ such that*

- *(i)* $u_{it} \in \mathcal{D}_t$ *for* $i = 1, \ldots, r$ *and all* $t \in \mathbb{Z}$,
- *(ii)* $\mathbb{E}\,\boldsymbol{u}_t\boldsymbol{u}_t^\top = \boldsymbol{I}_r$, $\boldsymbol{u}_t$ *is second-order stationary, and* $\boldsymbol{u}_t$ *and* $\boldsymbol{X}_N$ *are second-order co-stationary.*

Proof. Recall that M_ε is defined in Lemma S2.2. The idea of the proof is that $\boldsymbol{\psi}^{mt}$ is converging after suitable rotation.

Step 1: Let $s_1 = M_{1/2^2}$, $\boldsymbol{F}_1 = \boldsymbol{I}_r$, and $\boldsymbol{u}^{1,t} = \boldsymbol{F}_1\boldsymbol{\psi}^{s_1,t}$.
Step 2: Let $s_2 = \max\{s_1, M_{1/2^4}\}$, let $\boldsymbol{F}_2 = \boldsymbol{F}[\boldsymbol{F}_1, s_1, s_2]$, and let $\boldsymbol{u}^{2,t} = \boldsymbol{F}_2\boldsymbol{\psi}^{s_2,t}$.

⋮

Step k+1: Let $s_{k+1} = \max\{s_k, M_{1/s^{2(k+1)}}\}$, $\boldsymbol{F}_{k+1} = \boldsymbol{F}[\boldsymbol{F}_k, s_k, s_{k+1}]$, and $\boldsymbol{u}^{k+1,t} = \boldsymbol{F}_{k+1}\boldsymbol{\psi}^{s_{k+1},t}$.

⋮

Denoting by $u_j^{k,t}$ the jth coordinate of $\boldsymbol{u}^{k,t}$, we have

$$\left\|u_j^{k,t} - u_j^{k+1,t}\right\|_{L^2(\Omega)}^2 = \left\|\boldsymbol{F}_k\boldsymbol{\psi}^{s_k,t} - \boldsymbol{F}[\boldsymbol{F}_k, s_k, s_{k+1}]\boldsymbol{\psi}^{s_{k+1},t}\right\|_{L^2(\Omega)}^2 \leq \frac{1}{2^{2k}},$$

and thus

$$\left\|u_j^{k,t} - u_j^{k+h,t}\right\|_{L^2(\Omega)} \leq \sum_{l=1}^{h}\left\|u_j^{k+l-1,t} - u_j^{k+l,t}\right\|_{L^2(\Omega)} \leq \frac{1}{2^{k-1}}.$$

Therefore $(u_j^{k,t})_{k\geq 1}$ is a Cauchy sequence and converges in $L^2(\Omega)$ to some limit u_{jt}, $j = 1, \ldots, r$.

Notice that $\mathbb{E}\, \boldsymbol{u}^{k,t}(\boldsymbol{u}^{k,t})^{\mathsf{T}} = \boldsymbol{I}_r$ for each k since $\boldsymbol{F}_k \in \mathcal{O}(r)$, so that

$$\mathbb{E}\, \boldsymbol{u}_t \boldsymbol{u}_t^{\mathsf{T}} = \lim_{k\to\infty} \mathbb{E}\, \boldsymbol{u}^{k,t}(\boldsymbol{u}^{k,t})^{\mathsf{T}} = \lim_{k\to\infty} \boldsymbol{I}_r = \boldsymbol{I}_r.$$

Furthermore, $\mathbb{E}\, \boldsymbol{u}_t \boldsymbol{u}_{t+h}^{\mathsf{T}}$ is well-defined (and finite) for every $h \in \mathbb{Z}$, and

$$\mathbb{E}\, \boldsymbol{u}_t \boldsymbol{u}_{t+h}^{\mathsf{T}} = \lim_{k\to\infty} \mathbb{E}\, \boldsymbol{u}^{k,t}(\boldsymbol{u}^{k,t+h})^{\mathsf{T}} = \lim_{k\to\infty} \boldsymbol{F}_k \mathbb{E}\left[\boldsymbol{\psi}^{s_k,t}(\boldsymbol{\psi}^{s_k,t+h})^{\mathsf{T}}\right] \boldsymbol{F}_k^{\mathsf{T}}.$$

The term inside the limit is independent of t (since $\boldsymbol{X}_N$ is second-order stationary), and hence $\mathbb{E}\, \boldsymbol{u}_t \boldsymbol{u}_{t+h}^{\mathsf{T}}$ does not depend on t, and $(\boldsymbol{u}_t)_{t\in\mathbb{Z}}$ is second-order stationary. Furthermore,

$$\mathbb{E}\, X_{it} \boldsymbol{u}_{t+s}^{\mathsf{T}} = \lim_{k\to\infty} \mathbb{E}\left[X_{it}(\boldsymbol{u}^{k,t+s})^{\mathsf{T}}\right] = \lim_{k\to\infty} \mathbb{E}\left[X_{it} \boldsymbol{X}_{s_k,t+s}^{\mathsf{T}}\right] \boldsymbol{P}_{s_k}^{\mathsf{T}} \boldsymbol{\Lambda}_{s_k}^{-1} \boldsymbol{F}_k^{\mathsf{T}},$$

and since the term inside the limit does not depend on t, it follows that $\boldsymbol{u}_t$ is co-stationary with $\boldsymbol{X}_N$, for all N.

Let us now show that $u_{jt} \in \mathcal{D}_t$. Recall that

$$\boldsymbol{u}^{k,t} = \boldsymbol{F}_k \boldsymbol{\psi}^{s_k,t} = \boldsymbol{F}_k \boldsymbol{\Lambda}_{s_k}^{-1/2} \boldsymbol{P}_{s_k}^{\mathsf{T}} \boldsymbol{X}_N,$$

and let us write $\boldsymbol{G}_k = \boldsymbol{F}_k \boldsymbol{\Lambda}_{s_k}^{-1/2} \boldsymbol{P}_{s_k}^{\mathsf{T}}$. Notice that $u_j^{k,t} = \text{row}_j(\boldsymbol{G}_k) \boldsymbol{X}_{s_k,t}$, where $\text{row}_j(\boldsymbol{G})$ denotes the jth row of $\boldsymbol{G}$ and $\text{row}_j(\boldsymbol{G}_k)^{\mathsf{T}} \in \boldsymbol{H}_{s_k}$; hence, its squared norm is equal to the jth diagonal entry of $\boldsymbol{G}_k \boldsymbol{G}_k^{\mathsf{T}}$, which itself is bounded by

$$\left|\!\left|\!\left| \boldsymbol{G}_k \boldsymbol{G}_k^{\mathsf{T}} \right|\!\right|\!\right|_\infty = \left|\!\left|\!\left| \boldsymbol{F}_k \boldsymbol{\Lambda}_{s_k}^{-1} \boldsymbol{F}_k \right|\!\right|\!\right|_\infty \leq \left|\!\left|\!\left| \boldsymbol{\Lambda}_{s_k}^{-1} \right|\!\right|\!\right|_\infty = \lambda_{s_k,r}^{-1}.$$

Since $\lim_{k\to\infty} \lambda_{s_k,r}^{-1} = 0$, $u_{jt} \in \mathcal{D}_t$. □

We now know that each space $\mathcal{D}_t$ has dimension at least r. The following results tells us that this dimension is exactly r.

Lemma S2.4. *The dimension of $\mathcal{D}_t$ is r, and $\{u_{1t}, \dots, u_{rt}\}$ is an orthonormal basis for it.*

Proof. We already know that the dimension of $\mathcal{D}_t$ is at least r, and that $u_{1t}, \dots, u_{rt} \in \mathcal{D}_t$ are orthonormal. We only need to show that the dimension of $\mathcal{D}_t$ is less than or equal to r to finish the proof. First of all, let us drop the index t to simplify notation. Assume that $\mathcal{D}$ has dimension larger than r. Hence there exists $d_1, \dots, d_{r+1} \in \mathcal{D}$ orthonormal, with $d_j = \lim_{N\to\infty} d_{jN}$ in $L^2(\Omega)$, where $d_{jN} = \boldsymbol{v}_{jN}^{\mathsf{T}} \boldsymbol{X}_N$, and $\|\boldsymbol{v}_{jN}\|^2 = \boldsymbol{v}_{jN}^{\mathsf{T}} \boldsymbol{v}_{jN} \to 0$ as $N \to \infty$.

Let $\boldsymbol{A}^{(N)}$ be the $(r+1) \times (r+1)$ matrix with (i,j)th coordinate $A_{ij}^{(N)} = \mathbb{E}\,[d_{iN} d_{jN}]$. On the one hand, $\boldsymbol{A}^{(N)} \to \boldsymbol{I}_{r+1}$. On the other hand,

$$A_{ij}^{(N)} = \boldsymbol{v}_{iN}^{\mathsf{T}} \boldsymbol{\Sigma}_N \boldsymbol{v}_{jN} = \boldsymbol{v}_{iN}^{\mathsf{T}} \boldsymbol{P}_N \boldsymbol{\Lambda}_N \boldsymbol{P}_N^{\mathsf{T}} \boldsymbol{v}_{jN} + \boldsymbol{v}_{iN}^{\mathsf{T}} \boldsymbol{Q}_N \boldsymbol{\Phi}_N \boldsymbol{Q}_N^{\mathsf{T}} \boldsymbol{v}_{jN},$$

and, from the Cauchy–Schwarz inequality,

$$|\boldsymbol{v}_{iN}^{\mathsf{T}} \boldsymbol{Q}_N \boldsymbol{\Phi}_N \boldsymbol{Q}_N^{\mathsf{T}} \boldsymbol{v}_{jN}| \leq |||\boldsymbol{Q}_N|||_\infty^2 |||\boldsymbol{\Phi}_N|||_\infty \|\boldsymbol{v}_{jN}\| \|\boldsymbol{v}_{iN}\|$$

$$\leq \lambda_{N,r+1}\|\boldsymbol{v}_{jN}\|\|\boldsymbol{v}_{iN}\| \to 0,$$

as $N \to \infty$. Therefore, the limit of $\boldsymbol{A}^{(N)}$ is the same as the limit of $\boldsymbol{B}^{(N)}$, whose (i,j)th entry is $B_{ij}^{(N)} = \boldsymbol{v}_{iN}^{\mathsf{T}}\boldsymbol{P}_N\boldsymbol{\Lambda}_N\boldsymbol{P}_N^{\mathsf{T}}\boldsymbol{v}_{jN}$. But this is impossible since $\boldsymbol{B}^{(N)}$ is of rank at most r for all N. Therefore, the dimension of $\mathcal{D}$ is at most r. □

Consider the orthogonal decomposition

$$X_{it} = \gamma_{it} + \delta_{it}, \quad \text{with} \quad \gamma_{it} = \text{proj}_{H_i}(X_{it}|\mathcal{D}_t) \quad \text{and} \quad \delta_{it} = X_{it} - \gamma_{it},$$

of X_{it} into its projection onto $\text{span}_{H_i}(\mathcal{D}_t)$ and its orthogonal complement. Here, $\text{proj}_{H_i}(\cdot|\mathcal{D}_t)$ denotes the orthogonal projection onto $\text{span}_{H_i}(\mathcal{D}_t)$—see Appendix S1 for definitions. Since $\gamma_{it} \in \text{span}_{H_i}(\mathcal{D}_t)$, we can write it as a linear combination

$$\gamma_{it} = b_{i1}u_{1t} + \cdots + b_{ir}u_{rt},$$

(with coefficients in H_i) of $u_{1t}, \ldots, u_{rt}$, where $b_{ij} = \mathbb{E}\,\gamma_{it}u_{jt} = \mathbb{E}\,X_{it}u_{jt}$ does not depend on t in view of the co-stationarity of $\boldsymbol{u}_t$ and $\boldsymbol{X}_N$.

The only technical result needed before being able to prove Theorem 2.2 is that $\boldsymbol{\xi}$ is idiosyncratic, that is, $\lambda_1^{\xi} < \infty$. The rest of this section is devoted to the derivation of this result.

Although $\boldsymbol{\psi}^N$ does not necessarily converge, we know intuitively that the projection onto the entries of $\boldsymbol{\psi}^N$ should somehow converge. The following notion and result formalises this.

Definition S2.5. *Let $(\boldsymbol{v}_N)_N$ be an r-dimensional process with mean zero and $\mathbb{E}\,\boldsymbol{v}_N\boldsymbol{v}_N^{\mathsf{T}} = \boldsymbol{I}_r$. Consider the orthogonal decomposition*

$$\boldsymbol{v}_M = \boldsymbol{A}^{MN}\boldsymbol{v}_N + \boldsymbol{\rho}^{MN},$$

and let $\text{cov}(\boldsymbol{\rho}^{MN})$ be the covariance matrix of $\boldsymbol{\rho}^{MN}$. We say that $(\boldsymbol{v}_N)_N$ generates a Cauchy sequence of subspaces *if for all $\varepsilon > 0$, there is an $M_\varepsilon \geq 1$ such that for all N and $M > M_\varepsilon$, $\text{Tr}[\text{cov}(\boldsymbol{\rho}^{MN})] < \varepsilon$.*

Lemma S2.6. *Let $Y \in L_H^2(\Omega)$. If $(\boldsymbol{v}_N)_N$ generates a Cauchy sequence of subspaces, and $Y_N = \text{proj}_H(Y|\boldsymbol{v}_N)$, then $(Y_N)_N$ converges in $L_H^2(\Omega)$.*

Proof. Let $Y = Y_N + r_N = \boldsymbol{b}^N\boldsymbol{v}_N + r_N$ and $Y_M + r_M = \boldsymbol{b}^M\boldsymbol{v}_M + r_M$ be orthogonal decompositions, with $\boldsymbol{b}^k = (b_{k1}, \ldots, b_{kr})$, $b_{ki} \in H$, $k = N, M$. We therefore get

$$Y_N - Y_M = \boldsymbol{b}^N\boldsymbol{v}_N - \boldsymbol{b}^M\boldsymbol{v}_M = r_M - r_N.$$

The squared norm of the left-hand side can be written as the inner product between the middle and right expressions. Namely,

$$\begin{aligned}\|Y_N - Y_M\|_{L^2(\Omega)}^2 &= \left\langle \boldsymbol{b}^N\boldsymbol{v}_N - \boldsymbol{b}^M\boldsymbol{v}_M, r_M - r_N\right\rangle_{L^2(\Omega)} \\ &= \left\langle \boldsymbol{b}^N\boldsymbol{v}_N, r_M\right\rangle_{L^2(\Omega)} + \left\langle \boldsymbol{b}^M\boldsymbol{v}_M, r_N\right\rangle_{L^2(\Omega)}\end{aligned}$$

$$= S_1^{NM} + S_2^{MN},$$

where the cross-terms are zero by orthogonality. Since $\boldsymbol{v}_N = \boldsymbol{A}^{NM}\boldsymbol{v}_M + \boldsymbol{\rho}^{NM}$, and since $\boldsymbol{v}_M$ is uncorrelated with r_M,

$$S_1^{NM} = \left\langle \boldsymbol{b}^N \boldsymbol{\rho}^{NM}, r_M \right\rangle_{L^2(\Omega)},$$

and the Cauchy–Schwarz inequality, along with simple matrix algebra yields

$$\left|S_1^{NM}\right|^2 \leq \operatorname{Tr}\left[(\boldsymbol{b}^N)^\mathsf{T}\boldsymbol{b}^N\right] \|r_M\|_{L^2(\Omega)} \operatorname{Tr}\left[\operatorname{cov}(\boldsymbol{\rho}^{NM})\right]. \tag{S2.9}$$

Notice that $\|Y\|^2_{L^2(\Omega)} = \operatorname{Tr}[(\boldsymbol{b}^N)^\mathsf{T}\boldsymbol{b}^N] + \|r_N\|^2_{L^2(\Omega)} = \operatorname{Tr}[(\boldsymbol{b}^M)^\mathsf{T}\boldsymbol{b}^M] + \|r_M\|^2_{L^2(\Omega)}$. Therefore, the first two terms of the right-hand side in (S2.9) are bounded and, since $\boldsymbol{v}_N$ generates a Cauchy sequence of subspaces, $|S_1^{NM}|$ can be made arbitrarily small for large N, M. A similar argument holds for $|S_2^{MN}|$, and therefore $(Y_N)_N \subset L^2_H(\Omega)$ is a Cauchy sequence, and thus converges. □

We now show that $\boldsymbol{\psi}^N$, defined in (S2.2), generates a Cauchy sequence of subspaces.

Lemma S2.7. *$(\boldsymbol{\psi}^N)_N$ generates a Cauchy sequence of subspaces.*

Proof. For $N > M$, we already have the orthogonal decomposition

$$\boldsymbol{\psi}^M = \boldsymbol{D}\boldsymbol{\psi}^N + \boldsymbol{\rho}^{MN}, \tag{S2.10}$$

with $\boldsymbol{D} = \boldsymbol{\Lambda}_M^{-1/2}\boldsymbol{P}_M^\mathsf{T}\boldsymbol{P}_N\boldsymbol{\Lambda}_N^{1/2}$. Lemma S2.1 gives $\operatorname{Tr}(\operatorname{cov}(\boldsymbol{\rho}^{MN})) \leq r\lambda_{N,r+1}/\lambda_{M,r}$. We now need to show that the residual of the projection of $\boldsymbol{\psi}^N$ onto $\boldsymbol{\psi}^M$ is also small. The projection of $\boldsymbol{\psi}^N$ onto $\boldsymbol{\psi}^M$ is $\mathbb{E}\left[\boldsymbol{\psi}^N(\boldsymbol{\psi}^M)^\mathsf{T}\right]\boldsymbol{\psi}^M$. Expanding the expectation, we get

$$\begin{aligned}
\mathbb{E}\left[\boldsymbol{\psi}^N(\boldsymbol{\psi}^M)^\mathsf{T}\right] &= \mathbb{E}\left[\boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{P}_N^\mathsf{T}\boldsymbol{X}_N\boldsymbol{X}_M^\mathsf{T}\boldsymbol{P}_M\boldsymbol{\Lambda}_M^{-1/2}\right] \\
&= \boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{P}_N^\mathsf{T}\boldsymbol{\Sigma}_N\boldsymbol{P}_M\boldsymbol{\Lambda}_M^{-1/2} \\
&= \boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{P}_N^\mathsf{T}\left(\boldsymbol{P}_N\boldsymbol{\Lambda}_N\boldsymbol{P}_N^\mathsf{T} + \boldsymbol{Q}_N\boldsymbol{\Phi}_N\boldsymbol{Q}_N^\mathsf{T}\right)\boldsymbol{P}_M\boldsymbol{\Lambda}_M^{-1/2} \\
&= \boldsymbol{\Lambda}_N^{-1/2}\boldsymbol{\Lambda}_N\boldsymbol{P}_N^\mathsf{T}\boldsymbol{P}_M\boldsymbol{\Lambda}_M^{-1/2} = \boldsymbol{D}^\mathsf{T},
\end{aligned}$$

where the second equality comes from the fact that we expand the smaller matrix $\boldsymbol{P}_M$ with rows of zeros, and the third equality comes from (S2.1). We therefore have $\boldsymbol{\psi}^N = \boldsymbol{D}^\mathsf{T}\boldsymbol{\psi}^M + \boldsymbol{\rho}^{NM}$. Taking covariances, we get

$$\boldsymbol{I}_r = \boldsymbol{D}^\mathsf{T}\boldsymbol{D} + \operatorname{cov}(\boldsymbol{\rho}^{NM}) = \boldsymbol{D}\boldsymbol{D}^\mathsf{T} + \operatorname{cov}(\boldsymbol{\rho}^{MN})$$

where the second equality follows from (S2.10). Taking traces yields

$$\operatorname{Tr}(\operatorname{cov}(\boldsymbol{\rho}^{MN})) = \operatorname{Tr}(\operatorname{cov}(\boldsymbol{\rho}^{NM})),$$

which completes the proof. □

We now know that $\boldsymbol{\psi}^N = (\psi_1^N, \ldots, \psi_r^N)'$ generates a Cauchy sequence of subspaces. Let us show that the projection onto $\text{span}_H(\psi_1^N, \ldots, \psi_r^N)$ converges to the projection onto $\text{span}_H(\mathcal{D})$ (we are dropping the index t for ease of notation).

Lemma S2.8. *For each $i \geq 1$, writing X_i for X_{it},*

$$\lim_{N\to\infty} \text{proj}_{H_i}(X_i|\boldsymbol{\psi}^N) = \text{proj}_{H_i}(X_i|\mathcal{D}).$$

Proof. Let

$$\gamma_i^N = \text{proj}_{H_i}(X_i|\boldsymbol{\psi}^N), \qquad \delta_i^N = X_i - \gamma_i^N. \tag{S2.11}$$

We know by Lemmas S2.6 and S2.7 that

$$\gamma_i^N \to \gamma_i^* \quad \text{and} \quad \delta_i^N \to \delta_i^*, \qquad \text{as } N \to \infty. \tag{S2.12}$$

Let us show that $\gamma_i^* \in \text{span}_{H_i}(\mathcal{D}_t)$. The orthogonal decomposition of γ_i^* into its projection onto $\text{span}_{H_i}(\mathcal{D}_t)$ and its orthogonal complement is

$$\gamma_i^* = \mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right]\boldsymbol{u} + r_i$$

and, by orthogonality,

$$\|\gamma_i^*\|^2_{L^2(\Omega)} = \text{Tr}\left(\mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right](\mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right])^\mathsf{T}\right) + \|r_i\|^2_{L^2(\Omega)}. \tag{S2.13}$$

We also know by (S2.3) that $\gamma_i^N = \text{row}_i(\boldsymbol{P}_N)\boldsymbol{\Lambda}_N^{1/2}\boldsymbol{\psi}^N$, and therefore

$$\|\gamma_i^*\|_{L^2(\Omega)} = \lim_{N\to\infty} \text{Tr}(\text{cov}(\gamma_i^N)) = \lim_{N\to\infty} \text{Tr}(\text{row}_i(\boldsymbol{P}_N)\boldsymbol{\Lambda}_N \, \text{row}_i(\boldsymbol{P}_N)^\mathsf{T})$$

where we notice that $\text{row}_i(\boldsymbol{P}_N) : \mathbb{R}^r \to H_i$.

Recall from Proposition S2.3 that $\boldsymbol{u} = \lim_{N\to\infty} \boldsymbol{F}_{s_N}\boldsymbol{\Lambda}_{s_N}^{-1/2}\boldsymbol{P}_{s_N}^\mathsf{T}\boldsymbol{X}_{s_N}$. This implies that $\mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right] = \lim_{N\to\infty} \text{row}_i(\boldsymbol{P}_{s_N})\boldsymbol{\Lambda}_{s_N}^{1/2}\boldsymbol{F}_{s_N}^\mathsf{T}$, which in turn implies that

$$\text{Tr}\left(\mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right](\mathbb{E}\left[\gamma_i^* \boldsymbol{u}^\mathsf{T}\right])^\mathsf{T}\right) = \lim_{N\to\infty} \text{Tr}(\text{row}_i(\boldsymbol{P}_{s_N})\boldsymbol{\Lambda}_{s_N} \, \text{row}_i(\boldsymbol{P}_{s_N})^\mathsf{T}),$$

and therefore $\|r_i\|_{L^2(\Omega)} = 0$ by (S2.13), and $\gamma_i^* \in \text{span}_{H_i}(\mathcal{D}_t)$.

Finally, let us show that δ_i^* is orthogonal to $\mathcal{D}$. Writing u_j instead of u_{jt},

$$\begin{aligned}
\mathbb{E}\left[\delta_i^* u_j\right] &= \lim_{N\to\infty} \mathbb{E}\left[\delta_i^{s_N} u_j^{s_N}\right] \\
&= \lim_N \mathbb{E}\left[\delta_i^{s_N} \, \text{row}_j(\boldsymbol{F}_N)\boldsymbol{\psi}^{s_N}\right] \\
&= \lim_N \mathbb{E}\left[\delta_i^{s_N}(\boldsymbol{\psi}^{s_N})^\mathsf{T}\right] \text{row}_j(\boldsymbol{F}_N)^\mathsf{T} = 0,
\end{aligned}$$

since δ_i^k is orthogonal to $\boldsymbol{\psi}^k$ for all k. The result follows. □

Recall the definition of δ_i^* in (S2.12). We can now show that the largest eigenvalue $\lambda_{N,1}^{\delta^*}$ of the covariance of $\boldsymbol{\delta}_N^* = (\delta_1^*, \ldots, \delta_N^*)'$ is bounded.

Lemma S2.9. *$\boldsymbol{\delta}^*$ is idiosyncratic, i.e.* $\sup_{N\geq 1}\lambda^{\delta^*}_{N,1} < \infty$.

Proof. Let $\boldsymbol{\Sigma}^{\delta^N}_M$ be the covariance of $(\delta^N_1,\dots,\delta^N_M)$, $N\geq M$. Since δ^N_i converges to δ^*_i in $L^2(\Omega)$, $\boldsymbol{\Sigma}^{\delta^N}_M$ converges to $\boldsymbol{\Sigma}^{\delta^*}_M$. This implies that $\lambda^{\delta^N}_{M,1}$ converges to $\lambda^{\delta^*}_{M,1}$ as $N\to\infty$, since $\left|\lambda^{\delta^N}_{M,1}-\lambda^{\delta^*}_{M,1}\right| \leq \left|\left|\left|\boldsymbol{\Sigma}^{\delta^N}_M-\boldsymbol{\Sigma}^{\delta^*}_M\right|\right|\right|_\infty$ (Hsing & Eubank 2015). Since $\boldsymbol{\Sigma}^{\delta^N}_M$ is a compression of $\boldsymbol{\Sigma}^{\delta^N}_N$, we have

$$\lambda^{\delta^N}_{M,1}\leq \lambda^{\delta^N}_{N,1} = \lambda^x_{N,r+1},$$

where we have used the fact that, by definition, $\lambda^{\delta^N}_{N,1}=\lambda^x_{N,r+1}$. Taking the limit as $N\to\infty$, we get

$$\lambda^{\delta^*}_{M,1}\leq \lambda^x_{r+1}<\infty,$$

and, since this holds true for each m, it follows that $\lambda^{\delta^*}_1\leq\lambda^x_{r+1}<\infty$. □

Proof of Theorem 2.2. We have already shown the "only if" part. Let us assume $\lambda^x_r=\infty$, $\lambda^x_{r+1}<\infty$. Then we know that X_{it} has the representation

$$X_{it}=\gamma_{it}+\delta_{it},\quad \text{with}\quad \gamma_{it}=\mathrm{proj}_{H_i}(X_{it}|\mathcal{D}_t),\quad\text{and}\quad \delta_{it}=X_{it}-\gamma_{it}.$$

We know that $\gamma_{it}=b_{i1}u_{1t}+\cdots+b_{ir}u_{rt}$ is co-stationary with $\mathcal{X}$ since $\mathcal{D}_t$ is obtained as an $L^2(\Omega)$ limit of projections of X_t. It follows from Lemma S2.9 that $\lambda^\delta_{N,1}\leq\lambda^x_{r+1}$; using Lemma S2.18, we get $\lambda^\chi_{N,r}\geq\lambda^x_{N,r}-\lambda^\delta_{N,1}$, and thus $\lambda^\chi_{N,r}\to\infty$ as $N\to\infty$. □

Proof of Theorem 2.3 . Assume that $p\in\mathcal{D}_t$, so that $p=\lim_N\langle\boldsymbol{a}_N,\boldsymbol{X}_N\rangle$ for $\boldsymbol{a}_N\in\boldsymbol{H}_N$ with $\|a_N\|\to 0$. Since $\lambda^\xi_1<\infty$, the non-correlation of χ and ξ yields $p=\lim_N\langle\boldsymbol{a}_N,\boldsymbol{\chi}_N\rangle$, which implies that $p\in\mathrm{span}(\boldsymbol{v}_t)$ for $\boldsymbol{v}_t=(v_{1t},\dots,v_{rt})$, where $\boldsymbol{\chi}_N:=(\chi_{1t},\dots,\chi_{Nt})'$. Therefore,

$$\mathrm{span}(\boldsymbol{v}_t)\supset\mathrm{span}(\mathcal{D}_t)=\mathrm{span}(\boldsymbol{u}_t),$$

where $\boldsymbol{u}_t$ is constructed in Proposition S2.3. But $\mathrm{span}(\boldsymbol{v}_t)$ and $\mathrm{span}(\boldsymbol{u}_t)$ both have dimension r, so they are equal, and therefore

$$\chi_{it}=\mathrm{proj}_{H_i}(X_{it}|\,\mathrm{span}(\boldsymbol{v}_t))=\mathrm{proj}_{H_i}(X_{it}|\mathcal{D}_t).$$

If $X_{it}=\gamma_{it}+\delta_{it}$ is another functional factor representation with r factors, then we have

$$\gamma_{it}=\mathrm{proj}_{H_i}(X_{it}|\mathcal{D}_t)=\chi_{it}\quad\text{and}\quad\delta_{it}=X_{it}-\gamma_{it}=X_{it}-\chi_{it}=\xi_{it},$$

which shows the uniqueness of the decomposition. □

S2.2. Technical Results for Section 3

Let us first recall some basic definitions and properties of classes of operators on separable Hilbert spaces (Weidmann 1980, Chapter 7). Let H_1, H_2 be separable (real) Hilbert spaces. Denote by $\mathcal{S}_\infty(H_1, H_2)$ the space of compact (linear) operators from H_1 to H_2. The space $\mathcal{S}_\infty(H_1, H_2)$ is a subspace of $\mathcal{L}(H_1, H_2)$ and consists of all the operators $A \in \mathcal{L}(H_1, H_2)$ that admit a singular value decomposition

$$A = \sum_{j\geq 1} s_j[A] u_j v_j^\mathsf{T},$$

where $(s_j[A])_j \subset [0, \infty)$ are the singular values of A, ordered in decreasing order, satisfying $\lim_{j\to\infty} s_j[A] = 0$, $(u_j)_j \subset H_1$ and $(v_j)_j \subset H_2$ are orthonormal vectors. An operator $A \in \mathcal{S}_\infty(H_1, H_2)$ satisfying $|||A|||_1 := \sum_j s_j[A] < \infty$ is called a *trace-class* operator, and the subspace of trace-class operator is denoted by $\mathcal{S}_1(H_1, H_2)$. We have that $|||A|||_\infty \leq |||A|||_1 = |||A^\mathsf{T}|||_1$ and if $C \in \mathcal{L}(H_2, H)$, then $|||CA|||_1 \leq |||C|||_\infty |||A|||_1$. An operator $A \in \mathcal{S}_\infty(H_1, H_2)$ satisfying $|||A|||_2 := \sqrt{\sum_j (s_j[A])^2} < \infty$ is called *Hilbert–Schmidt*, and the subspace of Hilbert–Schmidt operators is denoted by $\mathcal{S}_2(H_1, H_2)$. We have that $|||A|||_\infty \leq |||A|||_2 = |||A^\mathsf{T}|||_2$ and if $C \in \mathcal{L}(H_2, H)$, then $|||CA|||_2 \leq |||C|||_\infty |||A|||_2$. Furthermore, if $B \in \mathcal{S}_2(H_2, H)$ then $|||BA|||_1 \leq |||B|||_2 |||A|||_2$, and if $A \in \mathcal{S}_1(H, H_1)$ then $|||A|||_2 \leq |||A|||_1$. We shall use the shorthand notation $\mathcal{S}_1(H)$ for $\mathcal{S}_1(H, H)$, and similarly for $\mathcal{S}_2(H)$. If $A \in \mathcal{S}_1(H)$, then we define its *trace* by

$$\mathrm{Tr}(A) = \sum_{i\geq 1} \langle Ae_i, e_i \rangle,$$

where $(e_i) \subset H$ is a complete orthonormal sequence (COS). The sum does not depend on the choice of the COS, and $|\mathrm{Tr}(A)| \leq |||A|||_1$. Furthermore, if A is symmetric positive semi-definite (i.e. $\langle Au, u\rangle \geq 0, \forall u \in H$), then $\mathrm{Tr}(A) = |||A|||_1$. If $A \in \mathcal{L}(H_1, H_2)$ and $B \in \mathcal{L}(H_2, H_1)$ and either $|||A|||_1 < \infty$ or $|||A|||_2 + |||B|||_2 < \infty$, we have $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. The spaces $\mathcal{S}_\infty(H)$, $\mathcal{S}_2(H)$, and $\mathcal{S}_1(H)$ are also called *Schatten spaces*.

Recall that $C_{N,T} := \min\{\sqrt{N}, \sqrt{T}\}$,

Lemma S2.10. *Under Assumptions C,*

$$|||\boldsymbol{\xi}_{NT}^\mathsf{T}\boldsymbol{\xi}_{NT}|||_2 = O_\mathrm{P}(NT/C_{N,T}). \tag{S2.14}$$

In particular, $|||\boldsymbol{\xi}_{NT}|||_\infty = O_\mathrm{P}(\sqrt{NT/C_{N,T}})$.

Proof. We have

$$|||(NT)^{-1}\boldsymbol{\xi}_{NT}^\mathsf{T}\boldsymbol{\xi}_{NT}|||_2^2 = \sum_{t,s=1}^{T} (\boldsymbol{\xi}_t^\mathsf{T}\boldsymbol{\xi}_s/N)^2/T^2 \leq 2T^{-2}\sum_{t,s}(\nu_N(t-s)^2 + \eta_{st}^2),$$

where $\eta_{st} := N^{-1}\boldsymbol{\xi}_t^\top \boldsymbol{\xi}_s - \nu_N(t-s)$. First, by Assumption C,

$$T^{-2}\sum_{t,s} \nu_N(t-s)^2 = O(T^{-1}) \quad \text{for all } T \geq 1.$$

Second, by Assumption C, $\eta_{st}^2 = O_{\mathrm{P}}(N^{-1})$ uniformly in t, s, and therefore

$$T^{-2}\sum_{t,s} \eta_{st}^2 = O_{\mathrm{P}}(N^{-1}),$$

which entails (S2.14). The second statement of the Lemma then follows since $|||\boldsymbol{\xi}_{NT}|||_\infty^2 = \left|\left|\left|\boldsymbol{\xi}_{NT}^\top \boldsymbol{\xi}_{NT}\right|\right|\right|_\infty$. □

Lemma S2.11. *Under Assumptions A, B, and C,*

$$\hat{\lambda}_1 = O_{\mathrm{P}}(1) \qquad \textit{and} \qquad |||\boldsymbol{X}_{NT}|||_\infty = O_{\mathrm{P}}(\sqrt{NT}).$$

In particular, $|||\hat{\boldsymbol{u}}|||_2 = O_{\mathrm{P}}(\sqrt{T})$, *where* $\hat{\boldsymbol{u}}$ *is defined in the proof of Theorem 3.1.*

Proof. We have, by definition of $\hat{\lambda}_1$, and using Assumptions A, B, C and Lemma S2.10,

$$\begin{aligned}
\hat{\lambda}_1 &= \left|\left|\left|\boldsymbol{X}_{NT}^\top \boldsymbol{X}_{NT}/(NT)\right|\right|\right|_\infty \leq (NT)^{-1}|||\boldsymbol{X}_{NT}|||_\infty^2 \\
&\leq 2(NT)^{-1}(|||\boldsymbol{B}_N\boldsymbol{u}|||_\infty^2 + |||\boldsymbol{\xi}_{NT}|||_\infty^2) \\
&\leq 2(NT)^{-1}(|||\boldsymbol{B}_N|||_\infty^2 |||\boldsymbol{u}|||_\infty^2 + |||\boldsymbol{\xi}_{NT}|||_\infty^2) \\
&\leq 2(NT)^{-1}(O(N)O_{\mathrm{P}}(T) + O_{\mathrm{P}}(NTC_{N,T}^{-1})) = O_{\mathrm{P}}(1).
\end{aligned}$$

The last statement of the Lemma follows from the fact that

$$T^{-1}|||\hat{\boldsymbol{u}}|||_2^2 = \hat{\lambda}_1^2 + \cdots + \hat{\lambda}_k^2 \leq k\hat{\lambda}_1^2 = O_{\mathrm{P}}(1).$$ □

For a sequence of random variables $Y_N > 0$ and a sequence of constants $a_N > 0$, we write $Y_N = \Omega_p(a_N)$ if and only if $Y_N^{-1} = O_{\mathrm{P}}(a_N^{-1})$.

Lemma S2.12. *Under Assumptions A, B, C, and D,* $\hat{\lambda}_r = \Omega_p(1)$.

Proof. Write $\lambda_k[A]$ for the k-th largest eigenvalue of a self-adjoint operator A. By definition,

$$\begin{aligned}
\hat{\lambda}_r &:= \lambda_r\left[\boldsymbol{X}_{NT}^\top \boldsymbol{X}_{NT}/(NT)\right] \\
&= \lambda_r\left[\boldsymbol{u}^\top \boldsymbol{B}_N^\top \boldsymbol{B}_N \boldsymbol{u}/(NT) + (NT)^{-1}(\boldsymbol{u}^\top \boldsymbol{B}_N^\top \boldsymbol{\xi}_{NT} + \boldsymbol{\xi}_{NT}^\top \boldsymbol{B}_N \boldsymbol{u} + \boldsymbol{\xi}_{NT}^\top \boldsymbol{\xi}_{NT})\right].
\end{aligned}$$

Since the operator norm of second summand is $O_{\mathrm{P}}(1)$ under Assumptions A, C, and D (see Lemma S2.10), we have, by Lemma S2.17,

$$\left|\hat{\lambda}_r - \lambda_r[\boldsymbol{u}^\top \boldsymbol{B}_N^\top \boldsymbol{B}_N \boldsymbol{u}/(NT)]\right| = O_{\mathrm{P}}(1).$$

We therefore just need to show that $\lambda_r[\boldsymbol{u}^\top \boldsymbol{B}_N^\top \boldsymbol{B}_N \boldsymbol{u}/(NT)] = \Omega_p(1)$. Using the Courant–Fischer–Weyl minimax characterization of eigenvalues (Hsing & Eubank 2015), we get that

$$\lambda_r[\boldsymbol{u}^\top \boldsymbol{B}_N^\top \boldsymbol{B}_N \boldsymbol{u}/(NT)] \geq \lambda_r[\boldsymbol{B}_N^\top \boldsymbol{B}_N/N] \cdot \lambda_r[\boldsymbol{u}^\top \boldsymbol{u}/T].$$

Now, by Assumption B, $\lambda_r[\boldsymbol{B}_N^\mathsf{T}\boldsymbol{B}_N/N] = \Omega_p(1)$, and by Assumption A,

$$\lambda_r[\boldsymbol{u}^\mathsf{T}\boldsymbol{u}/T] = \lambda_r[\boldsymbol{u}\boldsymbol{u}^\mathsf{T}/T] = \Omega_p(1).$$

The result follows. □

Lemma S2.13. *Recalling the definition of $\tilde{\boldsymbol{R}}$ in (3.3), denote by $s_j[A]$ the jth largest singular value of a matrix A. Under Assumptions A, B, C, and D,*

$$s_1[\tilde{\boldsymbol{R}}] = O_\mathrm{P}(1) \quad \textit{and} \quad s_r[\tilde{\boldsymbol{R}}] = \Omega_p(1)$$

In other words, $\tilde{\boldsymbol{R}}$ has a bounded norm, is invertible, and its inverse has a bounded norm.

Proof. The first statement follows directly from Lemma S2.12. For the second statement, using the Courant–Fischer–Weyl minimax characterization of singular values (Hsing & Eubank 2015), we obtain

$$s_r\left[\tilde{\boldsymbol{R}}\boldsymbol{u}/\sqrt{T}\right] \leq s_1\left[\boldsymbol{u}/\sqrt{T}\right] s_r[\tilde{\boldsymbol{R}}] = \left(s_1\left[\boldsymbol{u}\boldsymbol{u}^\mathsf{T}/\sqrt{T}\right]\right)^{1/2} s_r[\tilde{\boldsymbol{R}}].$$

Hence, given that $s_1[\boldsymbol{u}\boldsymbol{u}^\mathsf{T}/T] = O_\mathrm{P}(1)$, by Assumption A,

$$s_r[\tilde{\boldsymbol{R}}] \geq \left(s_1\left[\boldsymbol{u}\boldsymbol{u}^\mathsf{T}/T\right]\right)^{-1/2} s_r\left[\tilde{\boldsymbol{R}}\boldsymbol{u}/\sqrt{T}\right] = \Omega_p(1) s_r\left[\tilde{\boldsymbol{R}}\boldsymbol{u}/\sqrt{T}\right],$$

and by Theorem 3.1 and Lemma S2.17,

$$s_r\left[\tilde{\boldsymbol{R}}\boldsymbol{u}/\sqrt{T}\right] = s_r\left[\tilde{\boldsymbol{u}}/\sqrt{T}\right] + O_\mathrm{P}(1) = 1 + O_\mathrm{P}(1).$$

Therefore, by Lemma S2.12, $s_r\left[\boldsymbol{Q}_k\boldsymbol{u}/\sqrt{T}\right] = \Omega_p(1)$, which completes the proof. □

S2.3. Background Results and Technical Lemmas

Lemma S2.14. *Assume that $\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t^\mathsf{T}\right]$ is positive definite, and that Assumption B holds. If $\mathbb{E}\,\boldsymbol{u}_t = \boldsymbol{0}$, $\mathbb{E}\,\boldsymbol{\xi}_t = \boldsymbol{0}$, and $\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{\xi}_t^\mathsf{T}\right] = \boldsymbol{0}$,*

$$\lambda_r\left[\,\mathbb{E}\,\boldsymbol{X}_t\boldsymbol{X}_t^\mathsf{T}\right] = \Omega(N).$$

If, in addition, $\sum_{j=1}^\infty \left|\!\left|\!\left|\,\mathbb{E}\,\xi_{it}\xi_{jt}^\mathsf{T}\right|\!\right|\!\right|_\infty < M < \infty$ for all i, then

$$\lambda_{r+1}\left[\,\mathbb{E}\,\boldsymbol{X}_t\boldsymbol{X}_t^\mathsf{T}\right] = O(1).$$

In other words, the r-th largest eigenvalue of the covariance of $\boldsymbol{X}_t$ diverges and the $(r+1)$-th largest eigenvalue remains bounded as $N \to \infty$, which implies that the covariance of $\boldsymbol{X}_t$ satisfies the condition of Theorem 2.2.

Proof. Assuming $\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{\xi}_t^\top\right]=0$, we get $\Sigma := \boldsymbol{B}_N\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t^\top\right]\boldsymbol{B}_N + \mathbb{E}\left[\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top\right]$. Using Lemma S2.18, we get

$$\lambda_r[\Sigma] \geq \lambda_r\left[\boldsymbol{B}_N\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t^\top\right]\boldsymbol{B}_N^\top\right] \geq \lambda_r\left[\mathbb{E}\left[\boldsymbol{u}_t\boldsymbol{u}_t^\top\right]\right]\lambda_r[\boldsymbol{B}_N\boldsymbol{B}_N^\top],$$

where the second inequality comes from the Weyl–Fischer characterization of eigenvalues (Hsing & Eubank 2015). By assumption, the first term is bounded away from zero. For the second term, we have

$$\lambda_r[\boldsymbol{B}_N\boldsymbol{B}_N^\top] = \lambda_r[\boldsymbol{B}_N^\top\boldsymbol{B}_N] = \Omega(N),$$

by Assumption B. For the second statement, using Lemma S2.18 we get

$$\lambda_{r+1}[\,\mathbb{E}\,\boldsymbol{X}_t\boldsymbol{X}_t^\top] \leq \lambda_{r+1}[\boldsymbol{B}_N(\,\mathbb{E}\,\boldsymbol{u}_t\boldsymbol{u}_t^\top)\boldsymbol{B}_N^\top] + \lambda_1[\,\mathbb{E}\,\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top] \leq \lambda_1[\,\mathbb{E}\,\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top]$$

since $\boldsymbol{B}_N(\,\mathbb{E}\,\boldsymbol{u}_t\boldsymbol{u}_t^\top)\boldsymbol{B}_N^\top$ has rank at most r. Now the (i,j)-th entry of $\mathbb{E}\,\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top$ is $\mathbb{E}\,\xi_{it}\xi_{jt}^\top$. We want to show that $\left|\!\left|\!\left|\,\mathbb{E}\,\boldsymbol{\xi}_t\boldsymbol{\xi}_t^\top\right|\!\right|\!\right|_\infty = O(1)$. We will show that for any norm $\|\cdot\|_*$ on the operators $\mathcal{L}(\boldsymbol{H}_N)$ that is a matrix norm, that is, satisfies, for all $A,B\in\mathcal{L}(\boldsymbol{H}_N)$ and $\gamma\in\mathbb{R}$,

(i) $\|A\|_* \geq 0$ with equality if and only if $A=0$,
(ii) $\|\gamma A\|_* \leq |\gamma|\,\|A\|_*$,
(iii) $\|A+B\|_* \leq \|A\|_* + \|B\|_*$,
(iv) $\|AB\|_* \leq \|A\|_*\|B\|_*$,

we have $\lambda_1[A] \leq \|A\|_*$ if A is compact and self-adjoint. Indeed, let $Ax=\gamma x$ for some non-zero $x\in\boldsymbol{H}_N$ and $\gamma>0$. Then $Axx^\top=\gamma xx^\top$, and thus

$$|\gamma|\left\|xx^\top\right\|_* = \left\|\gamma xx^\top\right\|_* = \left\|Axx^\top\right\|_* \leq \|A\|_*\left\|xx^\top\right\|_*.$$

Simplifying by $\left\|xx^\top\right\|_*$ yields $|\gamma|\leq\|A\|_*$. To complete the proof, we still need that $\|A\|_* := \max_i\sum_{j=1}^N |\!|\!|a_{ij}|\!|\!|_\infty$ (where $(A)_{ij}=a_{ij}\in\mathcal{L}(H_j,H_i)$) is a matrix norm. This, however, is straightforward and details are omitted. □

Lemma S2.15. *Under Assumption D, there exists $M_1<\infty$ such that*

$$(NT)^{-1}\,\mathbb{E}\left|\!\left|\!\left|\boldsymbol{B}_N^\top\boldsymbol{\xi}_{NT}\right|\!\right|\!\right|_2^2 \leq M_1, \quad \textit{for all } N,T\geq 1.$$

In particular, $\left|\!\left|\!\left|\boldsymbol{B}_N^\top\boldsymbol{\xi}_{NT}\right|\!\right|\!\right|_2 = O_{\mathrm{P}}(\sqrt{NT})$.

Proof. We have $\left|\!\left|\!\left|\boldsymbol{B}_N^\top\boldsymbol{\xi}_{NT}\right|\!\right|\!\right|_2^2 = \sum_{s=1}^T\left\|\boldsymbol{B}_N^\top\boldsymbol{\xi}_s\right\|^2$. Since $\boldsymbol{B}_N^\top\boldsymbol{\xi}_s = \sum_{i=1}^N \boldsymbol{b}^{i\top}\xi_{is}$,

$$\begin{aligned}N^{-1}\,\mathbb{E}\left\|\boldsymbol{B}_N^\top\boldsymbol{\xi}_s\right\|^2 &= N^{-1}\sum_{i,j=1}^N\mathbb{E}\left[\xi_{is}^\top\boldsymbol{b}^i\boldsymbol{b}^{j\top}\xi_{js}\right]\\ &= N^{-1}\sum_{i,j=1}^N\mathbb{E}\left[\mathrm{Tr}(\xi_{is}^\top\boldsymbol{b}^i\boldsymbol{b}^{j\top}\xi_{js})\right]\end{aligned}$$

$$= N^{-1} \sum_{i,j=1}^{N} \operatorname{Tr}(\boldsymbol{b}^i \boldsymbol{b}^{j\mathsf{T}} \mathbb{E}\left[\xi_{js}\xi_{is}^{\mathsf{T}}\right])$$
$$\leq N^{-1} \sum_{i,j=1}^{N} |||\boldsymbol{b}^i \boldsymbol{b}^{j\mathsf{T}}|||_1 |||\mathbb{E}\left[\xi_{js}\xi_{is}^{\mathsf{T}}\right]|||_\infty$$
$$\leq (\max_{i,j} |||\boldsymbol{b}^i \boldsymbol{b}^{j\mathsf{T}}|||_1) N^{-1} \sum_{i,j=1}^{N} |||\mathbb{E}\left[\xi_{js}\xi_{is}^{\mathsf{T}}\right]|||_\infty$$
$$\leq (\max_{i,j} |||\boldsymbol{b}^i|||_2 |||\boldsymbol{b}^j|||_2) N^{-1} \sum_{i,j=1}^{N} |||\mathbb{E}\left[\xi_{js}\xi_{is}^{\mathsf{T}}\right]|||_\infty \leq rM^3,$$

where we have used Hölder's inequality for operators. The claim follows directly since the bound is independent of s. □

Lemma S2.16. *Assume that* $\mathbb{E}\left[(\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s)u_{lt}u_{ls}\right] = \mathbb{E}\left[\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s\right]\mathbb{E}\left[u_{lt}u_{ls}\right]$ *for all* $l = 1, \ldots, r$ *and* $s, t \in \mathbb{Z}$ *and that* $\sum_{t\in\mathbb{Z}} |\nu_N(t)| < M < \infty$. *Then Assumption E(α) holds with* $\alpha = 1$.

Proof. We have $|||\boldsymbol{u}\boldsymbol{\xi}_{NT}^{\mathsf{T}}|||_2^2 = \operatorname{Tr}\left[\boldsymbol{u}\boldsymbol{\xi}_{NT}^{\mathsf{T}}\boldsymbol{\xi}_{NT}\boldsymbol{u}^{\mathsf{T}}\right] = \sum_{l=1}^{r}\sum_{s,t=1}^{T} u_{lt}u_{ls}(\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s)$, and thus

$$\mathbb{E}\,|||\boldsymbol{u}\boldsymbol{\xi}_{NT}^{\mathsf{T}}|||_2^2 = \sum_{l=1}^{r}\sum_{s,t=1}^{T} \mathbb{E}\left[u_{lt}u_{ls}\right]\mathbb{E}\left[\boldsymbol{\xi}_t^{\mathsf{T}}\boldsymbol{\xi}_s\right],$$
$$\leq n \left(\max_l \mathbb{E}\, u_{lt}^2\right) \sum_{s,t=1}^{T} |\nu_N(t-s)|$$
$$= O(NT).$$

Hence, $|||\boldsymbol{u}\boldsymbol{\xi}_{NT}^{\mathsf{T}}|||_2^2 = O_{\mathrm{P}}(NT) \leq O_{\mathrm{P}}(NTC_{N,T}^{-2})$. □

The following Lemma tell us that the singular values of compact operators are stable under compact perturbations.

Lemma S2.17. *(Weidmann 1980, Chapter 7) Let* $A, B : H_1 \to H_2$ *be compact operators between two separable Hilbert spaces* H_1 *and* H_2*, with the singular value decompositions*

$$A = \sum_{j\geq 1} s_j[A]u_j v_j^{\mathsf{T}}, \quad \text{and} \quad B = \sum_{j\geq 1} s_j[B]w_j z_j^{\mathsf{T}},$$

where $(s_j[A])_j$ *are the singular values of* A*, arranged in decreasing order, and* $(s_j[B])_j$ *are the singular values of* B *arranged in decreasing order. Then*

$$|s_j[A] - s_j[B]| \leq |||A - B|||_\infty, \quad \forall j \geq 1.$$

Lemma S2.18. *Let $D, E \in \mathcal{S}_\infty(H)$ be symmetric positive semi-definite operators on a separable Hilbert space H, and let $\lambda_s[C]$ denote the s-th largest eigenvalue of an operator $C \in \mathcal{S}_\infty(H)$.*

(i) Letting $F = D + E$, we have, for all $i \geq 1$,

$$\lambda_i[F] \leq \min(\lambda_1[D] + \lambda_i[E], \lambda_i[D] + \lambda_1[E]) \text{ and } \max(\lambda_i[D], \lambda_i[E]) \leq \lambda_i[F].$$

(ii) Let G be a compression of D, meaning that $G = PDP$ for some orthogonal projection operator $P \in \mathcal{L}(H)$ ($P^2 = P = P^{\mathsf{T}}$). Then

$$\lambda_i[G] \leq \lambda_i[D] \quad \text{for all } i \geq 1.$$

Proof. This is a straightforward consequence of the Courant–Fischer–Weyl minimax characterization of eigenvalues of compact operators, see, e.g. Hsing & Eubank (2015). □

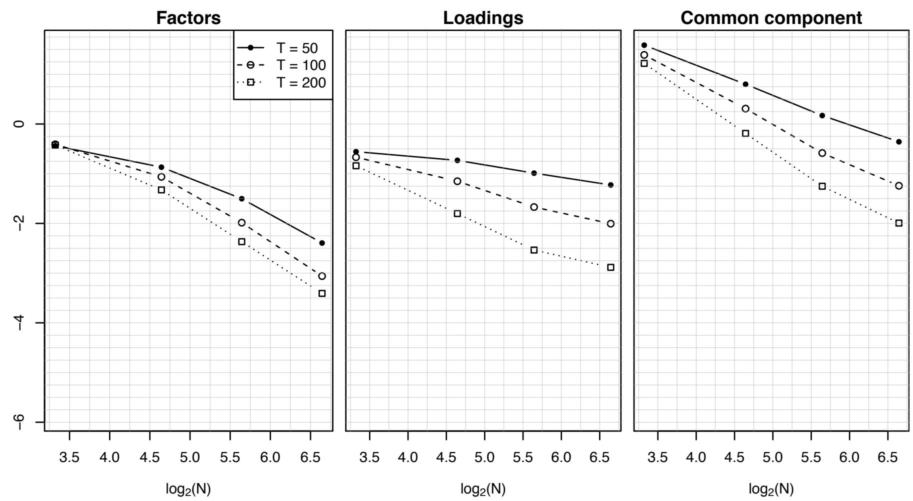
S3. Additional Simulations

Figure S1 shows the simulation results for `DGP3, DGP4`, described in Section 4.

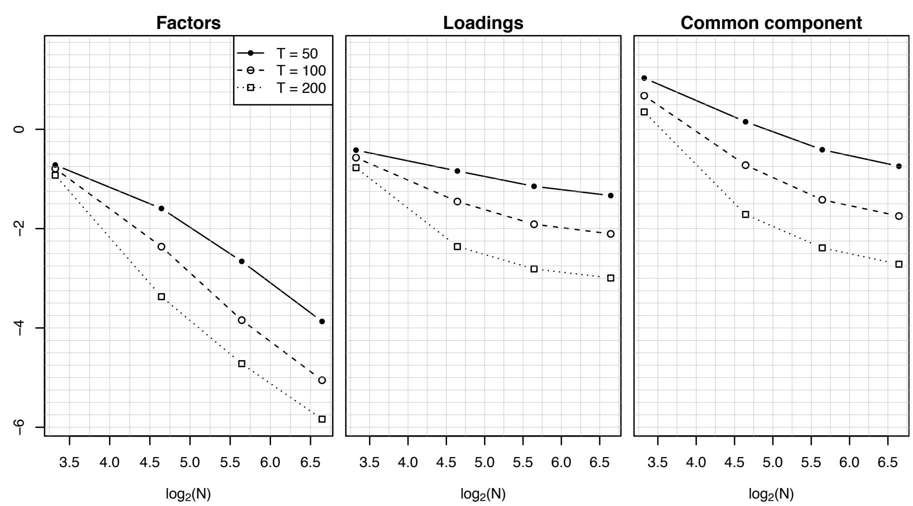
References

Hsing, T. & Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Wiley.

Weidmann, J. (1980), *Linear operators in Hilbert spaces*, Vol. 68 of *Graduate Texts in Mathematics*, Springer-Verlag, New York-Berlin. Translated from the German by Joseph Szücs.



(a) Simulation scenario `DGP3`.



(b) Simulation scenario `DGP4`.

Figure S1: Estimations errors (in $\log_2$ scale) for `DGP3` (subfigure (a)) and `DGP4` (subfigure (b)). For each subfigure, we have the estimation error for the factors ($\log_2 \delta^2_{N,T}$, left), loadings ($\log_2 \varepsilon^2_{N,T}$, middle), and common component ($\log_2 \phi_{N,T}$, right, $\phi_{N,T}$ defined in (4.1)) as functions of $\log_2 N$. The scales of the vertical axes are the same. Each curve corresponds to one value of $T \in \{50, 100, 200\}$, sampled for $N \in \{10, 25, 50, 100\}$.