

Michael Eichler, Giovanni Motta,
Rainer von Sachs

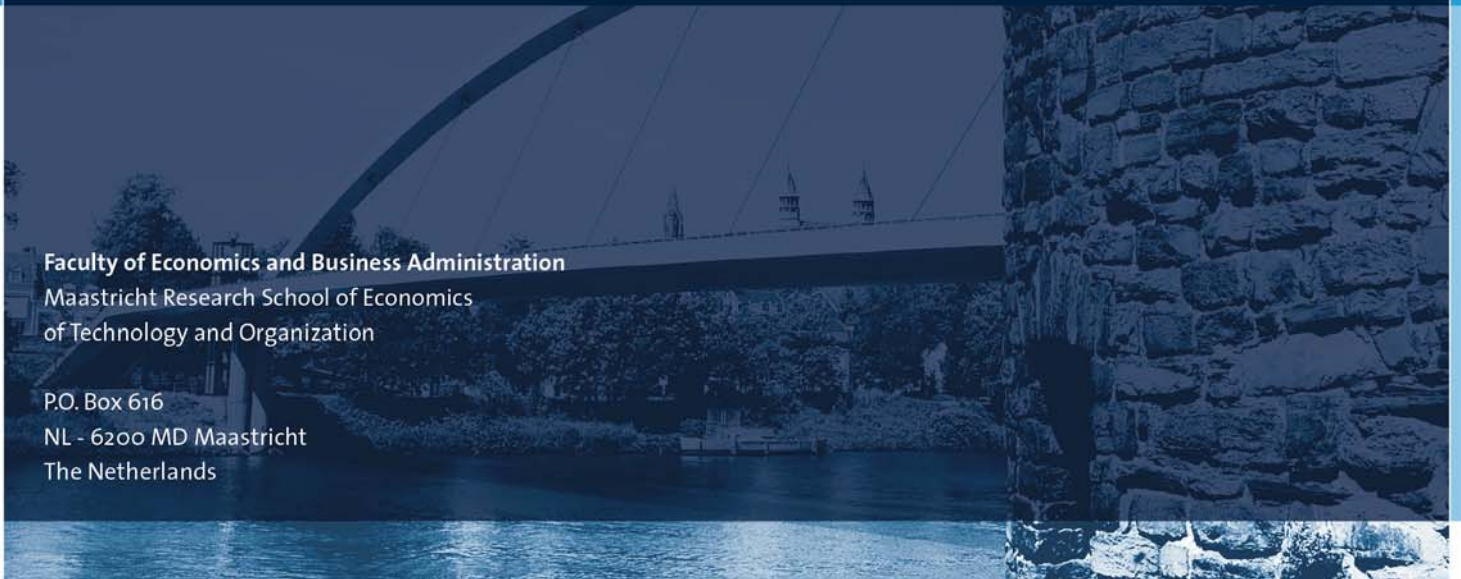
**Fitting dynamic factor models to
non-stationary time series**

RM/09/002

METEOR

Faculty of Economics and Business Administration
Maastricht Research School of Economics
of Technology and Organization

P.O. Box 616
NL - 6200 MD Maastricht
The Netherlands



FITTING DYNAMIC FACTOR MODELS TO NON-STATIONARY TIME SERIES

MICHAEL EICHLER, GIOVANNI MOTTA*, AND RAINER VON SACHS*

University of Maastricht and Université catholique de Louvain

24 December 2008

ABSTRACT. Factor modelling of a large time series panel has widely proven useful to reduce its cross-sectional dimensionality. This is done by explaining common co-movements in the panel through the existence of a small number of common components, up to some idiosyncratic behaviour of each individual series. To capture serial correlation in the common components, a dynamic structure is used as in traditional (uni- or multivariate) time series analysis of second order structure, i.e. allowing for infinite-length filtering of the factors via dynamic loadings. In this paper, motivated from economic data observed over long time periods which show smooth transitions over time in their covariance structure, we allow the dynamic structure of the factor model to be non-stationary over time, by proposing a deterministic time variation of its loadings. In this respect we generalise existing recent work on static factor models with time-varying loadings as well as the classical, i.e. stationary, dynamic approximate factor model. Motivated from the stationary case, we estimate the common components of our dynamic factor model by the eigenvectors of a consistent estimator of the now time-varying spectral density matrix of the underlying data-generating process. This can be seen as time-varying principal components approach in the frequency domain. We derive consistency of this estimator in a "double-asymptotic" framework of both cross-section and time dimension tending to infinity. A simulation study illustrates the performance of our estimators.

Keywords: Approximate Factor Models, Local Stationarity, Principal Components

JEL Classification: C14, C32

1. INTRODUCTION

Factor modelling plays an important role in the analysis of high-dimensional multivariate time series. This is based on a clear empirical observation made for many applications, foremost in economics and finance, but also in various other fields such as psychometrics or biomedical signal processing. It can be observed that often only a small number of (latent, i.e. unobservable) factors is sufficient to explain a certain common behaviour of the second-order structure of a large time series panel, up to

Corresponding author: Michael Eichler (m.eichler@ke.unimaas.nl)

*Financial support from the contract "Projet d'Actions de Recherche Concertées" nr. 07/12/002 of the "Communauté française de Belgique", granted by the "Académie universitaire Louvain", is gratefully acknowledged

an idiosyncratic behaviour of each individual series to be cross-sectional uncorrelated with the rest. Moreover, from a statistical, or even a data processing point of view, in view of the nowadays really high-dimensional data sets available in the aforementioned applications, there is an obvious need to reduce the cross-section dimension to a much smaller factor space dimension to overcome the estimation curse of dimensionality.

Furthermore in many of the given applications where data are observed over long time periods (or on a sufficiently fine time resolution), it appears that these data exhibit some time variation in their serial variance-covariance structure. This is plausible if one wants to model, for example, macroeconomic data, in the presence of transitions between economic recessions and booms or external events such as the impact of political crises or changes in the monetary policy of central banks. Consequently, it seems certainly restrictive to assume that the common structure explained by fitting a latent factor model to the data remains constant over time.

These observations have motivated us to develop a fairly general dynamic factor model which takes the time evolution of the underlying serial second-order structure of the data into account by allowing for time-varying factor loadings: the common components are dynamic and non-stationary at the same time through a model that let the filters become (smooth) functions of time.

With this approach, we provide for a true generalisation of two classes of existing factor models in the literature: static factors with time-varying loadings recently developed by Motta et al. (2006), where no serial correlation has been allowed in the common components; and stationary, i.e. time-constant, dynamic factor models with a dynamic structure using (possibly infinite-length) filters of the factors via dynamic loadings, such as in Forni et al. (2000).

In the existing yet rather recent literature on non-stationary factor modelling, there exist a variety of different approaches: random, deterministic or a combination of both. Molenaar et al. (1992) just treat non-stationarity in the mean by a linear time dependent trend function. Del Negro and Otrok (2008) developed a dynamic factor model with time-varying factor loadings following a random walk and with stochastic volatility in both the latent factors and idiosyncratic components. The situation where breaks occur at random time points has already been covered by Fancourt and Principe (1998) with a model based on piecewise constant loadings. A recent approach of Stock and Watson (2008) investigates the effect of structural instabilities on the forecasting ability of a dynamic factor model with a possible abruptly changing parameter in an autoregressive factor specification. Pan and Yao (2008), finally, choose again a different approach on estimating the factor loading space by a stepwise optimisation algorithm on expanding the "white noise space": this allows for a generalisation of the traditional factor approach without specifying any (parametric or non-parametric) structure on the departure from stationarity in the autocorrelation structure between the observations and the unobserved factors.

In contrast, in this paper we consider smooth evolutions of the dynamics of the process with a non-stationarity which is purely deterministic, and which allows to include the stationary model as a true submodel of our more general model: the factor loadings would then simply become functions which are constant over time. Our main motivation comes from the empirical evidence in economic data observed

over long time periods, which show smooth transitions over time in the covariance structure of an observed multivariate time series.

To formalise now a bit more what is behind a factor model let us state the main idea that a very large number N of series can be explained by a small number r of factors. The behavior of the observed N -dimensional stochastic process \mathbf{Y} is driven by two components: the common component \mathbf{X} describes the co-movements of all the series, the idiosyncratic component \mathbf{Z} is specific to each particular series

$$\mathbf{Y}_N(t) = \mathbf{\Lambda}_N \mathbf{F}(t) + \mathbf{Z}_N(t) = \mathbf{X}_N(t) + \mathbf{Z}_N(t) \quad (1.1)$$

where \mathbf{X} and \mathbf{Z} are both unobserved. The N -dimensional common component $\mathbf{X}_N(t)$ is a linear combination of the latent r -dimensional vector of factors $\mathbf{F}(t)$, which drives the joint behavior of all the series and is ‘loaded’ by the $N \times r$ matrix of loadings $\mathbf{\Lambda}_N$. Model (1.1) is *static* in the sense that \mathbf{Y} at time t depends only on the value at time t of \mathbf{F} .

The main task in factor analysis is separating the common components from the specific ones. Chamberlain and Rothschild (1983) defined the class of *approximate* factor models as a sequence of factor models where the covariance matrix $\mathbf{\Sigma}_N^Z$ of the idiosyncratic components is a sequence of non-diagonal matrices with uniformly bounded eigenvalues. Principal components regression (PCR) is a consistent estimation technique for approximate factor models under the assumption that only the largest r of the N eigenvalues of the covariance matrix $\mathbf{\Sigma}_N$ of the observations are unbounded as $N \rightarrow \infty$. A sufficient condition for such a behavior of the eigenvalues of $\mathbf{\Sigma}_N$ is that the eigenvalues of $\mathbf{\Sigma}_N^Z$ are uniformly bounded, i.e. the eigenvalues of $\mathbf{\Sigma}_N^Z$ stay bounded as $N \rightarrow \infty$ for all $N \in \mathbb{Z}$.

Under this assumption on the behavior of the eigenvalues and very mild conditions on the loadings, the factors, and the idiosyncratic components of a static factor model, Bai (2003) derived the convergence of the estimated common components $\widehat{\mathbf{X}}_N(t)$ to the common components $\mathbf{X}_N(t)$ of model (1.1), where the estimator $\widehat{\mathbf{X}}_N(t)$ is based on the $N \times r$ matrix containing the eigenvectors corresponding to the largest r eigenvalues of the sample covariance matrix of $\mathbf{X}_N(t)$.

If the true data generating process is dynamic, i.e. in the presence of serial correlation in the common driving force, a shortcoming of static factor analysis is that one can end-up with a very high number of static principal components. This motivated Forni et al. (2000) to generalize model (1.1) to the dynamic case

$$\mathbf{Y}_N(t) = \mathbf{\Psi}_N(B) \mathbf{F}(t) + \mathbf{Z}_N(t) = \mathbf{X}_N(t) + \mathbf{Z}_N(t), \quad (1.2)$$

where the filters $\mathbf{\Psi}_{ij}(B)$ have square summable coefficients and the factors $\mathbf{F}(t)$ are orthonormal white noise. Again, no mutual orthogonality of the idiosyncratic components is assumed, and the separation from the common components is achieved by assuming that only r of the eigenvalues of the spectral density $\mathbf{\Sigma}_N(\omega)$ of $\mathbf{Y}_N(t)$ diverge as $N \rightarrow \infty$. The authors of Forni et al. (2000) use PCR in the frequency domain to construct weakly consistent estimators of the common components for which the spectral density matrix of the data-generating process needs to be consistently estimated. For the statistical properties of dynamic principal components, see also Brillinger (1981).

In this paper now, we generalise further the approach of Forni et al. (2000) by allowing for a smooth time-variation of the serial correlation in the dynamics of the

factor model:

$$\mathbf{Y}_N(t) = \boldsymbol{\Psi}_N(t, B) \mathbf{F}(t) + \mathbf{Z}_N(t) = \mathbf{X}_N(t) + \mathbf{Z}_N(t). \quad (1.3)$$

This model allows for second order non-stationarity: on the one hand because the loadings, dynamic filters as in Forni et al. (2000), are now assumed to be functions of time $\boldsymbol{\Psi}_N(t, B)$, and on the other hand because the idiosyncratic components $\mathbf{Z}_N(t)$ are also allowed to be non-stationary with time-varying dynamics. Consequently the spectral density matrix $\boldsymbol{\Sigma}_N(t, \omega)$ of $\{\mathbf{Y}_N\}$ becomes dependent on time, and in order to estimate it, we use a time-localised empirical spectrum (a localized periodogram over time). Our contribution is a practical and theoretical treatment of this localised PCR in the frequency domain. We show that, under simultaneous asymptotics (both $N \rightarrow \infty$ and $T \rightarrow \infty$), the PCR based on a consistent estimator of $\boldsymbol{\Sigma}_N$ is consistent for the (dynamic and) non-stationary common components. As such, the development of a rigorous asymptotic theory for consistently estimating the time-varying loadings has been possible by embedding this model into the framework of locally stationary processes as derived by a series of papers by Dahlhaus (1996, 1997, 2000). Compare also with the time-varying static approach of Motta et al. (2006) which generalizes the static factor model of Bai (2003). For further explanations of this framework for theoretical consistency treatment, including a rigorous statement of our model replacing the slightly simplified equation (1.3) above, we refer to section 3.1.

Our paper is organized as follows. In section 2 we introduce a general non-stationary dynamic factor model and explain the proposed methods to estimate the non-stationary common components. In section 3 we show that for an increasing panel size the principal components of the time-varying spectral matrix converge to the true common components (see section 3.3). Furthermore we show the consistency of the estimated common components, that is the convergence of the principal components of the estimated time-varying spectral matrix to the principal components of the true time-varying spectral matrix (see section 3.5). The finite sample behaviour of our method is investigated with a simulation study in section 4, while section 5 concludes.

Throughout the paper we use bold-unslanted letters for matrices, bold-slanted letters for vectors and unbold (normal) letters for scalars. For a real number x , $[x]$ denotes the largest integer smaller or equal to x . For a matrix A , the trace and the conjugate transpose are denoted by $\text{tr}(A)$ and A^* , respectively, $\|A\| = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^*)}$ is the Euclidean norm, and \mathbf{I}_n is the identity matrix of dimension n . Furthermore, we denote by \perp orthogonality in the Hilbert space of real-valued square integrable random variables.

One important tool will be the convolution of time-varying linear filters. For a given value of u in rescaled time, let $\boldsymbol{\Phi}(u, B)$ and $\boldsymbol{\Psi}(u, B)$ be two time-varying filters (B is the backshift operator). Then the convolution $(\boldsymbol{\Phi} \star \boldsymbol{\Psi})(u, B)$ of the two filters is defined by

$$(\boldsymbol{\Phi} \star \boldsymbol{\Psi})(u, B) = \sum_{k=0}^{\infty} \left[\sum_{j=0}^{\infty} \boldsymbol{\Phi}_j(u) \boldsymbol{\Psi}_{k-j}(u) \right] B^k.$$

The obtained filter $(\boldsymbol{\Phi} \star \boldsymbol{\Psi})(u, B)$ has the transfer function

$$(\boldsymbol{\Phi} \star \boldsymbol{\Psi})(u, \omega) = \boldsymbol{\Phi}(u, \omega) \boldsymbol{\Psi}(u, \omega),$$

where, for ease of notation, we use $\Phi(u, B)$ and $\Phi(u, \omega)$ to denote, respectively, the filter in the time-domain and the transfer function in the frequency domain, while $\Phi_k(u)$ denote the filter coefficients.

2. A NON-STATIONARY DYNAMIC FACTOR MODEL

Let $Y_i(t)$, $i \in \mathbb{N}$, be a panel of non-stationary time series at time points $1 \leq t \leq T$ that are driven by a few common factors. Following Forni et al. (2000), we assume that the time series can be described by an approximate dynamic factor model, that is, the idiosyncratic components are allowed to be correlated to some extent. Thus we consider the model

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{Z}(t) = \Psi(t, B) \mathbf{F}(t) + \mathbf{Z}(t), \quad 1 \leq t \leq T, \quad (2.1)$$

where $\mathbf{X}(t)$ is the common component originating from the common factors $\mathbf{F}(t)$ while $\mathbf{Z}(t)$ is the idiosyncratic component. We assume that the model satisfies the following conditions:

- (i) $\mathbf{F}(t)$ is an r -dimensional stationary white noise process with $\mathbb{E}(\mathbf{F}(t)) = 0$ and $\mathbb{E}(\mathbf{F}(t)\mathbf{F}(t)') = \mathbf{I}_r$;
- (ii) $\Psi(t, B) = (\psi_{ij}(t, B))_{i \in \mathbb{N}, j=1, \dots, r}$ is a time-varying one-sided linear filter of dimension $\mathbb{N} \times r$;
- (iii) $\mathbf{Z}(t) = (Z_i(t))_{i \in \mathbb{N}}$ is uncorrelated to the factor process $\mathbf{F}(t)$.

Non-stationarity enters the model in two ways: Firstly, the loadings $\Psi(t, B)$, which determine the influence of the factor process $\mathbf{F}(t)$ on the given process, are time-varying, that is, the influence of the factors may change over time. Secondly, the idiosyncratic component $\mathbf{Z}(t)$ may also be non-stationary with time-varying dynamics. Here, we call a process $\mathbf{Z}(t)$ idiosyncratic if

$$\lim_{N \rightarrow \infty} \mathbb{E} \|\mathbf{A}_N(B)\mathbf{Z}(t)\|^2 = 0$$

for $1 \leq t \leq T$ and all sequences of filters $\mathbf{A}_N(B)$, $N \in \mathbb{N}$, such that

$$\lim_{N \rightarrow \infty} \int_{-\pi}^{\pi} \|\mathbf{A}_N(\omega)\|^2 d\omega = 0.$$

Notice that in contrast to the stationary case in Forni and Lippi (2001) $\mathbf{Z}(t)$ is non-stationary and only defined over the period $1 \leq t \leq T$. For the above definition of idiosyncratic processes, we therefore set $\mathbf{Z}(t) = 0$ for $t < 1$ or $t > T$.

The factor process $\mathbf{F}(t)$ is assumed to be an stationary orthonormal white noise process as in Forni et al. (2000). This is, however, not a serious constraint since, for example, if $\mathbf{F}(t)$ is a non-stationary factor process with time-varying moving average representation $\mathbf{F}(t) = \Xi_T(t, B)\boldsymbol{\eta}(t)$, where $\boldsymbol{\eta}(t)$ is a stationary orthonormal white noise process, the common component $\mathbf{X}(T)$ can be rewritten as

$$\mathbf{X}(t) = \Psi(t, B) \Xi(t, B) \boldsymbol{\eta}(t) = \tilde{\Psi}(t, B) \boldsymbol{\eta}(t).$$

Thus the process can be represented as in (2.1) with factor process $\tilde{\mathbf{F}}(t) = \boldsymbol{\eta}(t)$ satisfying the above assumptions.

For the idiosyncratic component $\mathbf{Z}(t)$, we assume further that for any $N \in \mathbb{N}$ the subprocess $\mathbf{Z}_N(t) = (Z_1(t), \dots, Z_N(t))'$ has a time-varying moving average representation of the form

$$\mathbf{Z}_N(t) = \mathbf{\Upsilon}_N(t, B) \boldsymbol{\varepsilon}_N(t),$$

where $\boldsymbol{\varepsilon}_N(t)$ is a stationary white noise process with mean $\mathbb{E}(\boldsymbol{\varepsilon}_N(t)) = 0$ and covariance matrix $\mathbb{E}(\boldsymbol{\varepsilon}_N(t)\boldsymbol{\varepsilon}_N(t)') = \mathbf{I}_N$. Then $\mathbf{Y}_N(t)$ is obtained from the $r + N$ -dimensional stationary process $(\mathbf{F}(t), \boldsymbol{\varepsilon}_N(t))$ by application of the time-varying linear filter $(\boldsymbol{\Psi}_N(t, B), \mathbf{\Upsilon}_N(t, B))$ and thus has time-varying spectral representation

$$\mathbf{Y}_N(t) = \int_{-\pi}^{\pi} e^{i\omega t} \boldsymbol{\Psi}_N(t, \omega) d\boldsymbol{\xi}^{\mathbf{F}}(\omega) + \int_{-\pi}^{\pi} e^{i\omega t} \mathbf{\Upsilon}_N(t, \omega) d\boldsymbol{\xi}_N^{\boldsymbol{\varepsilon}}(\omega),$$

where $\boldsymbol{\xi}^{\mathbf{F}}(\omega)$ and $\boldsymbol{\xi}_N^{\boldsymbol{\varepsilon}}(\omega)$ are the orthogonal increment processes associated with the stationary processes $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$, respectively. The corresponding time-varying spectral density is given by

$$\boldsymbol{\Sigma}_N(t, \omega) = \boldsymbol{\Psi}_N(t, \omega)\boldsymbol{\Psi}_N(t, \omega)^* + \mathbf{\Upsilon}_N(t, \omega)\mathbf{\Upsilon}_N(t, \omega)^*.$$

Assuming that the factor loadings and thus the dynamics of the process change slowly over time, we can treat the process as if it were stationary over small time intervals and estimate the spectral matrix locally. Then we can proceed for each time-point t as in the stationary case. More precisely, let $\widehat{\boldsymbol{\Sigma}}_N(t, \omega)$ be a consistent estimator of the time-varying spectral matrix $\boldsymbol{\Sigma}_N(t, \omega)$; for details on the estimation of time-varying spectral matrices we refer to the appendix A. Then $\widehat{\boldsymbol{\Sigma}}_N(t, \omega)$ has a spectral decomposition

$$\widehat{\boldsymbol{\Sigma}}_N(t, \omega) = \widehat{\mathbf{P}}_N(t, \omega)^* \widehat{\boldsymbol{\Lambda}}_N(t, \omega) \widehat{\mathbf{P}}_N(t, \omega),$$

where $\widehat{\boldsymbol{\Lambda}}_N(t, \omega) = \text{diag}(\widehat{\lambda}_{N1}(t, \omega), \dots, \widehat{\lambda}_{NN}(t, \omega))$ is the diagonal matrix containing the eigenvalues of $\widehat{\boldsymbol{\Sigma}}_N(t, \omega)$ in descending order of magnitude and $\widehat{\mathbf{P}}_N(t, \omega)$ is the unitary $N \times N$ matrix whose i -th row is the row eigenvector of $\widehat{\boldsymbol{\Sigma}}_N(t, \omega)$ corresponding to the i -th eigenvalue $\widehat{\lambda}_{Ni}(t, \omega)$. As in the stationary case of Forni et al. (2000), we estimate the factor space or, equivalently, the space spanned by the common components by the space spanned by the eigenvectors corresponding to the r largest eigenvalues of $\widehat{\boldsymbol{\Sigma}}_N(t, \omega)$. In the frequency domain this is accomplished by the projection

$$\widehat{\boldsymbol{\Phi}}_N(t, \omega) = \widehat{\mathbf{P}}_N(t, \omega)^* \mathbf{Q}_N^r \widehat{\mathbf{P}}_N(t, \omega),$$

where \mathbf{Q}_N^r is the block matrix given by

$$\mathbf{Q}_N^r = \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

By inverse Fourier transformation, we obtain the corresponding projection filter in the time-domain,

$$\widehat{\boldsymbol{\Phi}}_N^{(\infty)}(t, B) = \sum_{k=-\infty}^{\infty} \widehat{\boldsymbol{\Phi}}_{N,k} B^k,$$

where the filter coefficients $\widehat{\boldsymbol{\Phi}}_{N,k}$ are given by

$$\widehat{\boldsymbol{\Phi}}_{N,k}(t) = \int_{-\pi}^{\pi} \widehat{\boldsymbol{\Phi}}_N(t, \omega) e^{i\omega k} d\omega.$$

Since this is a two-sided filter of infinite length, it needs to be truncated before it can be applied to the data. Let

$$\widehat{\Phi}_{NT}(t, B) = \widehat{\Phi}_N^{(M_T)}(t, B) = \sum_{k=-M_T}^{M_T} \widehat{\Phi}_{N,k} B^k$$

be the truncated filter. Then the common components $\mathbf{X}_N(t)$ are estimated by

$$\widehat{\mathbf{X}}_N(t) = \widehat{\Phi}_{NT}(t, B) \mathbf{Y}_N(t).$$

The estimation method is a localized version of the method proposed by Forni et al. (2000). However, the non-stationarity gives rise to some questions. Firstly, while working in the frequency domain, we could restrict ourselves to quantities at some point in time. However, the estimated spectral matrix is obtained from data that are only close to this time point and thus have different dynamics. Secondly, the filter is applied in the time domain and thus again to data within some time range about t . Thus, although the estimation method for the common component $\mathbf{X}_N(t)$ is a straightforward extension of the estimation method in the stationary case, its properties are unclear and it is not obvious that the theoretical results from Forni et al. (2000, 2004, 2005) carry over easily. In the next section we introduce an appropriate asymptotic framework and show that the estimated common components are indeed consistent. Furthermore, in section 4, we investigate the finite sample behaviour of the common component estimator by simulation.

3. ASYMPTOTIC BEHAVIOUR OF THE ESTIMATED COMMON COMPONENTS

In this section, we prove consistency of the estimated common components. Since we are dealing with non-stationary processes, we have to specify what is meant by asymptotic behaviour. To this end, we first define an evolutionary dynamic factor model, which provides a theoretical framework for the asymptotics.

3.1. An asymptotic model

In contrast to the non-stationary factor model in the previous section, we study now a family of non-stationary panel time series $\mathbf{Y}_{NT}(t) = (Y_{1T}(t), \dots, Y_{NT}(t))'$, $1 \leq t \leq T$, which is indexed by the number of observations T and the number of cross-sectional variables N . Instead of defining a dynamic factor model for each T simply by (2.1), which would correspond to ordinary asymptotics where the same process is observed over longer time periods, we assume that the processes $\mathbf{Y}_{NT}(t)$, $T \in \mathbb{N}$, are derived all from the same filter functions for the factor loadings and the idiosyncratic components. Thus, we assume that for $T, N \in \mathbb{N}$ the process $\mathbf{Y}_{NT}(t)$ is given by

$$\mathbf{Y}_{NT}(t) = \Psi_N\left(\frac{t}{T}, B\right) \mathbf{F}(t) + \mathbf{Z}_{NT}(t), \quad 1 \leq t \leq T, \quad (3.1a)$$

with idiosyncratic component

$$\mathbf{Z}_{NT}(t) = \Upsilon_N\left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t), \quad 1 \leq t \leq T. \quad (3.1b)$$

For fixed T , the model in (2.1) can be embedded into the above evolutionary model by setting $\Psi_N(t, B) = \Psi_N\left(\frac{t}{T}, B\right)$ and $\Upsilon_N(t, B) = \Upsilon_N\left(\frac{t}{T}, B\right)$. As the number of observations T increases, we obtain locally more and more observations with

approximately the same dynamics provided that, e.g., the linear filters $\Psi_N(\frac{t}{T}, \omega)$ and $\Upsilon_N(\frac{t}{T}, \omega)$ are slowly varying.

We note that in practice we observe $\mathbf{Y}_{NT}(t)$ only for one given T . Thus the above evolutionary model represents a purely theoretical device for evaluating large-sample properties of estimator for non-stationary processes. Furthermore, we note that the evolutionary form of the model does not include processes with non-stationary factor processes. Such processes could be covered by considering dynamic factor models based on locally stationary processes (Dahlhaus 1996). For ease of notation, we consider in this paper only evolutionary processes but all results can be similarly derived for the general case.

We now give the main assumptions on our asymptotic model.

Assumption 1 (Evolutionary dynamic factor model). $\mathbf{Y}_{NT}(t)$, $1 \leq t \leq T$, with $T, N \in \mathbb{N}$ is a family of stochastic processes given by (3.1) and satisfying the following conditions:

- (i) the factor process $\mathbf{F}(t)$ is an r -dimensional stationary white noise process with $\mathbb{E}(\mathbf{F}(t)) = 0$ and $\mathbb{E}(\mathbf{F}(t)\mathbf{F}(t)') = \mathbf{I}_r$;
- (ii) the coefficients $\Psi_{N,k}(u)$ and $\Upsilon_{N,k}(u)$ of the time-varying linear filters $\Psi_N(u, B)$ and $\Upsilon_N(u, B)$, respectively, are square summable uniformly in $u \in [0, 1]$

$$\sup_{u \in [0,1]} \sum_{k=-\infty}^{\infty} \|\Psi_{N,k}(u)\|^2 < \infty \quad \text{and} \quad \sup_{u \in [0,1]} \sum_{k=-\infty}^{\infty} \|\Upsilon_{N,k}(u)\|^2 < \infty; \quad (3.2)$$

- (iii) the factor process $\mathbf{F}(t)$ and the idiosyncratic errors $\boldsymbol{\varepsilon}_N(t)$ are orthogonal at all leads and lags, that is, $\mathbb{E}[\boldsymbol{\varepsilon}_N(t)\mathbf{F}(t-k)'] = 0$ for all $N \in \mathbb{N}$ and $t, k \in \mathbb{Z}$.

The common components $\mathbf{X}_{NT}(t)$ and the idiosyncratic components $\mathbf{Z}_{NT}(t)$ are evolutionary processes in the sense that they admit moving average representations with time-varying coefficients $\Psi_{N,k}(\frac{t}{T})$ and $\Upsilon_{N,k}(\frac{t}{T})$, respectively, that are functions of rescaled time $\frac{t}{T} \in [0, 1]$. As in the previous section, this leads to spectral representations

$$\mathbf{X}_{NT}(t) = \int_{-\pi}^{\pi} \Psi_N(\frac{t}{T}, \omega) e^{i\omega t} d\xi^{\mathbf{F}}(\omega)$$

and

$$\mathbf{Z}_{NT}(t) = \int_{-\pi}^{\pi} \Upsilon_N(\frac{t}{T}, \omega) e^{i\omega t} d\xi^{\boldsymbol{\varepsilon}_N}(\omega),$$

where $\xi^{\mathbf{F}}(\omega)$ and $\xi^{\boldsymbol{\varepsilon}_N}(\omega)$ are the orthogonal increment processes associated with the stationary processes $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$, respectively. Then $\mathbf{X}_{NT}(t)$ has the time-varying spectral density matrix

$$\Sigma_N^{\mathbf{X}}(\frac{t}{T}, \omega) = \Psi_N(\frac{t}{T}, \omega) \Psi_N(\frac{t}{T}, \omega)^*, \quad (3.3)$$

where $\Psi_N(u, \omega)$ is the time-varying transfer function

$$\Psi_N(u, \omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \Psi_{N,k}(u) e^{-i\omega k}. \quad (3.4)$$

Similarly, the idiosyncratic component $\mathbf{Z}_N(t)$ has the time-varying spectral density matrix

$$\Sigma_N^{\mathbf{Z}}(\frac{t}{T}, \omega) = \Upsilon_N(\frac{t}{T}, \omega) \Upsilon_N(\frac{t}{T}, \omega)^*, \quad (3.5)$$

where $\Upsilon_N(u, \omega)$ is the time-varying transfer function associated with the filter $\Upsilon_N(u, B)$. It follows that the process $\mathbf{Y}_{NT}(t)$ has time-varying spectral matrix

$$\begin{aligned}\Sigma_N(u, \omega) &= \Sigma_N^{\mathbf{X}}(u, \omega) + \Sigma_N^{\mathbf{Z}}(u, \omega) \\ &= \Psi_N(u, \omega)\Psi_N(u, \omega)^* + \Upsilon_N(u, \omega)\Upsilon_N(u, \omega)^*.\end{aligned}\quad (3.6)$$

We note that unlike for the non-asymptotic model (2.1), the spectral matrix $\Sigma_N(u, \omega)$ for fixed $N \in \mathbb{N}$ is uniquely determined by the family $\mathbf{Y}_{NT}(t)$, $T \in \mathbb{N}$. Finally, we remark that although evolutionary processes are special cases of locally stationary processes (Dahlhaus 2000, Remark 2.2) the process $\mathbf{Y}_{NT}(t)$ as defined above is not locally stationary in the strict sense of Dahlhaus (1996, 2000) as it does not have a time-varying moving average representation in terms of an N -dimensional white noise process $\tilde{\varepsilon}_N(t)$. Nevertheless, most results on locally stationary processes can be simply extended also to the above class of processes (see Appendix A.4).

Assumption 1 (ii) on the filter coefficients implies that the diagonal entries of the spectral matrix $\Sigma_N(u, \omega) = (\sigma_{ij}(u, \omega))_{i,j=1,\dots,N}$ are uniformly bounded for $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$, that is, for all $n \in \mathbb{N}$, there exists a constant $\bar{\sigma}_i > 0$ such that

$$\sup_{u \in [0,1]} \sup_{\omega \in [-\pi,\pi]} \sigma_{ii}(u, \omega) \leq \bar{\sigma}_i. \quad (3.7)$$

In order to obtain sensible estimates for spectral density matrix $\Sigma_N(u, \omega)$, we require that the dynamics of the process change only slowly over time and that accordingly the process can be viewed as approximately stationary over short time intervals. This can be achieved by a suitable degree of smoothness of the time-varying spectral matrix. This is formalized in the next assumption.

Assumption 2 (Smoothness of spectral matrix and transfer functions).

- (i) The time-varying spectral density $\Sigma_N(u, \omega)$ is two times continuously differentiable for all $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$.
- (ii) The time-varying transfer functions $\Psi_N(u, \omega)$ and $\Upsilon_N(u, \omega)$ are Lipschitz continuous in $u \in (0, 1)$.

Next, we note that Assumption 1 (iii) implies that the factors $\mathbf{F}(t)$ and the idiosyncratic components $\mathbf{Z}_{NT}(t)$ are orthogonal. The orthogonality is important to ensure the identifiability of the common and the idiosyncratic components. Note that, since $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$ are latent processes, a representation with orthogonal components can always be achieved by considering appropriate projections.

For the separation of the common component $\mathbf{X}_{NT}(t)$ and the idiosyncratic component $\mathbf{Z}_{NT}(t)$, we also require the following assumption on the behaviour of the eigenvalues of the spectral matrices $\Sigma_N^{\mathbf{X}}(u, \omega)$ and $\Sigma_N^{\mathbf{Z}}(u, \omega)$.

Assumption 3 (Common and idiosyncratic dynamic eigenvalues). Let $\lambda_{Nj}^{\mathbf{X}}(u, \omega)$ and $\lambda_{Nj}^{\mathbf{Z}}(u, \omega)$, $1 \leq j \leq N$, be the time-varying eigenvalues of $\Sigma_N^{\mathbf{X}}(u, \omega)$ and $\Sigma_N^{\mathbf{Z}}(u, \omega)$, respectively, ordered in descending order of magnitude.

- (i) The first r common time-varying dynamic eigenvalues $\lambda_{Nj}^{\mathbf{X}}(u, \omega)$, $j = 1, \dots, r$, diverge uniformly in $u \in [0, 1]$ as N increases: for $j = 1, \dots, r$

$$\lim_{N \rightarrow \infty} \inf_{u \in [0,1]} \lambda_{Nj}^{\mathbf{X}}(u, \omega) = \infty \quad \text{a.e. in } [-\pi, \pi].$$

- (ii) The first idiosyncratic time-varying dynamic eigenvalue $\lambda_{N1}^{\mathbf{Z}}(u, \omega)$ is uniformly bounded, that is, there exists a positive constant $\bar{\lambda}_{\mathbf{Z}}$ such that $\lambda_{N1}^{\mathbf{Z}}(u, \omega) \leq \bar{\lambda}_{\mathbf{Z}}$ for all $u \in [0, 1]$, $\omega \in [-\pi, \pi]$, and $N \in \mathbb{N}$.

The following proposition shows that, under these made assumptions on the process, the time-varying spectrum, and the eigenvalues, only r eigenvalues of the spectral density matrix of the observations diverge as N increases while the remaining $N - r$ stay bounded. Thus the proposition generalizes Proposition 1 of Forni et al. (2000) to the non-stationary case.

Proposition 1. *Under Assumptions 1 to 3, the first r time-varying dynamic eigenvalues of $\Sigma_N(u, \omega)$ diverge, as $N \rightarrow \infty$, uniformly over $u \in [0, 1]$, that is, for $j = 1, \dots, r$*

$$\lim_{N \rightarrow \infty} \inf_{u \in [0, 1]} \lambda_{Nj}(u, \omega) = \infty \quad \text{a.e. in } [-\pi, \pi].$$

The remaining eigenvalues are uniformly bounded by $\bar{\lambda}_{\mathbf{Z}}$, that is, for $j > r$

$$\limsup_{N \rightarrow \infty} \sup_{\omega \in [-\pi, \pi]} \sup_{u \in [0, 1]} \lambda_{Nj}(u, \omega) \leq \bar{\lambda}_{\mathbf{Z}}.$$

Proof. See Appendix B.1. □

The assumption is in line with empirical evidence that for many panel time series only few eigenvalues diverge as the cross-sectional dimension increases while the others seem to be bounded. This fact could be, for example, exploited for selecting the dimension r of the factor process (Hallin and Liška 2007).

An important consequence of the previous proposition is that the processes $\mathbf{Z}_{NT}(t)$ are indeed idiosyncratic, that is, for all sequences of filters $\mathbf{A}_N(B)$ with $\int_{-\pi}^{\pi} \|A_N(\omega)\|^2 d\omega \rightarrow 0$ as $N \rightarrow \infty$ the filtered process $\mathbf{A}_N(B)\mathbf{Z}_{NT}(t)$ converge to zero in mean square.

Corollary 2. *Under the assumptions of the proposition, the processes $\mathbf{Z}_{NT}(t)$, $T \in \mathbb{N}$, are idiosyncratic.*

Proof. See Appendix B.2. □

3.2. Decomposition of the overall estimation error

Let $\widehat{\Sigma}_N(u, \omega)$ be a consistent estimator of the spectral matrix $\Sigma_N(u, \omega)$, and let

$$\widehat{\Sigma}_N(u, \omega) = \widehat{\mathbf{P}}_N(u, \omega)^* \widehat{\Lambda}_N(u, \omega) \widehat{\mathbf{P}}_N(u, \omega)$$

its spectral decomposition. Then, as described in the previous section the common component $\mathbf{X}_{NT}(t)$ can be estimated by

$$\widehat{\mathbf{X}}_{NT,N}(t) = \widehat{\Phi}_{NT}\left(\frac{t}{T}, B\right) \mathbf{Y}_{NT}(t), \quad (3.8)$$

where for $u \in [0, 1]$

$$\widehat{\Phi}_{NT}(u, B) = \sum_{k=-M_T}^{M_T} \widehat{\Phi}_{N,k}(u) B^k$$

is the truncated time domain filter obtained from the transfer function

$$\widehat{\Phi}_N(u, \omega) = \widehat{\mathbf{P}}_N(u, \omega)^* \mathbf{Q}_N^{(r)} \widehat{\mathbf{P}}_N(u, \omega).$$

The main objective of this section is to prove the consistency of the common component estimator $\widehat{\mathbf{X}}_{NT,N}(t)$. As in Forni et al. (2000), the proof of consistency is based

on a decomposition of the overall estimation error $|\widehat{\mathbf{X}}_{NT,N}(t) - \mathbf{X}_{NT}(t)|$. Whereas in the stationary case covered by Forni et al. (2000) it is sufficient to decompose the error in an approximation and an estimation error, the non-stationary case discussed in this paper imposes additional problems, which lead to a decomposition into four separate errors: approximation, truncation, filtering, and estimation error.

In order to motivate the decomposition, we note that the estimation of the common components is based on a principal component analysis in the frequency domain. Since in contrast to the stationary case the spectral decomposition and thus the dynamic principal components vary over time, the resulting projection onto the space spanned by the first r principal components must be applied to the process at time t , which can be accomplished in the frequency domain but not in the time domain.

More precisely, let

$$\boldsymbol{\Sigma}_N(u, \omega) = \mathbf{P}_N(u, \omega)^* \boldsymbol{\Lambda}_N(u, \omega) \mathbf{P}_N(u, \omega) \quad (3.9)$$

be the spectral decomposition of the true spectral matrix $\boldsymbol{\Sigma}_N(u, \omega)$. Here $\boldsymbol{\Lambda}_N(u, \omega) = \text{diag} \{ \lambda_{N1}(u, \omega), \dots, \lambda_{NN}(u, \omega) \}$ is the diagonal matrix containing the eigenvalues of $\boldsymbol{\Sigma}_N(u, \omega)$ and $\mathbf{P}_N(u, \omega)$ is the $N \times N$ matrix whose j -th row

$$\mathbf{P}_{Nj}(u, \omega) = [p_{Nj,1}(u, \omega), p_{Nj,2}(u, \omega), \dots, p_{Nj,N}(u, \omega)] \quad (3.10)$$

equals the row eigenvector of $\boldsymbol{\Sigma}_N(u, \omega)$ corresponding to $\lambda_{Nj}(u, \omega)$. of $\mathbf{P}_N(u, \omega)$. The vectors $\mathbf{P}_{Nj}(u, \omega)$ are called the time-varying dynamic eigenvectors of $\boldsymbol{\Sigma}_N(u, \omega)$. Then the transfer function

$$\boldsymbol{\Phi}_N(u, \omega) = \mathbf{P}_N(u, \omega)^* \mathbf{Q}_N^{(r)} \mathbf{P}_N(u, \omega)$$

can be viewed as a projection operator projecting the frequency components onto the space spanned by the largest r eigenvectors of $\boldsymbol{\Sigma}_N(u, \omega)$. Applying this operator in the frequency domain, we obtain the following approximation $\mathbf{X}_{NT,N}(t)$ of the true common component $\mathbf{X}_{NT}(t)$,

$$\begin{aligned} \mathbf{X}_{NT,N}(t) &= \int_{-\pi}^{\pi} e^{i\omega t} \boldsymbol{\Phi}_N\left(\frac{t}{T}, \omega\right) \boldsymbol{\Psi}_N\left(\frac{t}{T}, \omega\right) d\xi^F(\omega) \\ &\quad + \int_{-\pi}^{\pi} e^{i\omega t} \boldsymbol{\Phi}_N\left(\frac{t}{T}, \omega\right) \boldsymbol{\Upsilon}_N\left(\frac{t}{T}, \omega\right) d\xi_N^\varepsilon(\omega). \end{aligned}$$

To find an expression in terms of linear filters, let $(\mathbf{P}_N \star \boldsymbol{\Psi}_N)(u, B)$ be the convolution of the two filters $\mathbf{P}_N(u, B)$ and $\boldsymbol{\Psi}_N(u, B)$,

$$(\mathbf{P}_N \star \boldsymbol{\Psi}_N)(u, B) = \sum_{k=0}^{\infty} \left[\sum_{j=-\infty}^{\infty} \boldsymbol{\Phi}_{N,j}(u) \boldsymbol{\Psi}_{N,k-j}(u) \right] B^k;$$

$(\boldsymbol{\Phi}_N \star \boldsymbol{\Upsilon}_N)(u, B)$ is similarly defined. Then $\mathbf{X}_{NT,N}(t)$ can be also written as

$$\mathbf{X}_{NT,N}(t) = (\boldsymbol{\Phi}_N \star \boldsymbol{\Psi}_N)\left(\frac{t}{T}, \omega\right) \mathbf{F}(t) + (\boldsymbol{\Phi}_N \star \boldsymbol{\Upsilon}_N)\left(\frac{t}{T}, \omega\right) \boldsymbol{\varepsilon}_N(t). \quad (3.11)$$

The approximation $\mathbf{X}_{NT,N}(t)$ of the common component $\mathbf{X}_{NT,N}(t)$ is of purely theoretical interest as it cannot be computed from $\boldsymbol{\Sigma}_N\left(\frac{t}{T}, \omega\right)$ for two reasons: firstly, the filter $\boldsymbol{\Phi}_N\left(\frac{t}{T}, \omega\right)$ cannot be applied pointwise as suggested above since this would require knowledge of the filters $\boldsymbol{\Psi}_N(u, B)$ and $\boldsymbol{\Upsilon}_N(u, B)$ and of the latent stationary processes $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$. Instead the filter needs to be applied to the observed data

$\mathbf{Y}_{NT}(t)$ in the time domain. Unlike for stationary processes, application of filters in the time and in the frequency domain differs for non-stationary processes as will be shown in section 3.4, Secondly, the involved filters are of infinite length and thus must be truncated before they can be applied to the the data.

This suggests to decompose the overall estimation error $|\widehat{\mathbf{X}}_{NT,N}(t) - X_{NT}(t)|$ as follows:

(i) *approximation error*: $|\mathbf{X}_{NT,N}(t) - \mathbf{X}_{NT}(t)|$, where $\mathbf{X}_{NT,N}(t)$ is given by (3.11);

(ii) *truncation error*: $|\bar{\mathbf{X}}_{NT,N}(t) - \mathbf{X}_{NT,N}(t)|$, where

$$\bar{\mathbf{X}}_{NT,N}(t) = (\Phi_{NT} \star \Psi_N)\left(\frac{t}{T}, \omega\right) \mathbf{F}(t) + (\Phi_{NT} \star \Upsilon_N)\left(\frac{t}{T}, \omega\right) \varepsilon_N(t) \quad (3.12)$$

is obtained from the truncated filter

$$\Phi_{NT}(u, B) = \sum_{k=-M_T}^{M_T} \Phi_{N,k}(u) B^k;$$

(iii) *filtering error*: $|\widetilde{\mathbf{X}}_{NT,N}(t) - \bar{\mathbf{X}}_{NT,N}(t)|$, where

$$\begin{aligned} \widetilde{\mathbf{X}}_{NT,N}(t) &= \Phi_{NT}\left(\frac{t}{T}, \omega\right) \Psi_N\left(\frac{t}{T}, \omega\right) \mathbf{F}(t) + \Phi_{NT}\left(\frac{t}{T}, \omega\right) \Upsilon_N\left(\frac{t}{T}, \omega\right) \varepsilon_N(t) \\ &= \Phi_{NT}\left(\frac{t}{T}, \omega\right) \mathbf{Y}_{NT}(t); \end{aligned} \quad (3.13)$$

is obtained by application of the filter $\Phi_{NT}(\cdot, B)$ in the time domain;

(iv) *estimation error*: $|\widehat{\mathbf{X}}_{NT,N}(t) - \widetilde{\mathbf{X}}_{NT,N}(t)|$, where $\widehat{\mathbf{X}}_{NT,N}(t)$ is the common component estimator in (3.8).

3.3. Approximation error

We start our discussion of the approximation error $|\mathbf{X}_{NT,N}(t) - \mathbf{X}_{NT}(t)|$ defined in the previous section by giving a more detailed account on how $\mathbf{X}_{NT,N}(t)$ approximates the true common component $\mathbf{X}_{NT}(t)$. This will also show how the non-stationary case differs from the stationary case. As noted before, a key role in the recovering of the common components is played by the dynamic principal components. In analogy to Forni et al. (2000), one might consider to define time-varying principal components by

$$\mathbf{F}_{NT}(t) = \mathbf{P}_N\left(\frac{t}{T}, B\right) \mathbf{Y}_{NT}(t), \quad t = 1, \dots, T. \quad (3.14)$$

There are, however, two problems with this approach. Firstly, since the eigenvectors are time-varying, it is not guaranteed that the time-varying dynamic principal components thus defined are orthogonal at all leads and lags. Secondly, application of the filter $\mathbf{P}_N\left(\frac{t}{T}, B\right)$ will mix the time-varying dynamics of $\mathbf{Y}_{NT}(t)$ at different points in time. The problems can be avoided by applying the filter $\mathbf{P}_N\left(\frac{t}{T}, B\right)$ in the frequency domain. Thus, we define the *time-varying dynamic principal components* process at rescaled time $u \in [0, 1]$ by

$$\mathbf{F}_N(u, s) = (\mathbf{P}_N \star \Psi_N)(u, B) \mathbf{F}(s) + (\mathbf{P}_N \star \Upsilon_N)(u, B) \varepsilon_N(s), \quad s \in \mathbb{Z}. \quad (3.15)$$

Notice that the whole process depends on some fixed $u \in [0, 1]$. From the process $\mathbf{F}_N(u, s)$, we obtain the closed linear subspace of $\mathcal{L}_2(\Omega, \mathcal{F}, P)$ spanned by the first r components of the process $\mathbf{F}_N(u, s)$ by

$$\mathcal{F}_N^r(u) = \overline{\text{sp}} \{F_{Nj}(u, s), j = 1, \dots, r, s \in \mathbb{Z}\}, \quad (3.16)$$

where $F_{Nj}(u, s)$ is the j -th element of the N -dimensional vector $\mathbf{F}_N(u, s)$. Since the dynamic principal components process $\mathbf{F}_N(u, s)$ is generated from the spectral matrix $\boldsymbol{\Sigma}_N(u, \omega)$ at a single point u in (rescaled) time, the dynamic principal components are orthogonal at all leads and lags. Then the true common component $\mathbf{X}_{NT}(t)$ can be approximated by the orthogonal projection of $\mathbf{Y}_{NT}(t)$ onto $\mathcal{F}_{NT}^r(\frac{t}{T})$, that is, we have

$$\mathbf{X}_{NT,N}(t) = \text{proj} \left(\mathbf{Y}_{NT}(t) \middle| \mathcal{F}_{NT}^r \left(\frac{t}{T} \right) \right). \quad (3.17)$$

To see that $\mathbf{X}_{NT,N}(t)$ equals the approximated common component defined in the previous section, we note that the process $\mathbf{Y}_{NT}(t)$ can be decomposed as

$$\mathbf{Y}_{NT}(t) = \mathbf{X}_{NT,N}(t) + \mathbf{Z}_{NT,N}(t). \quad (3.18)$$

where

$$\begin{aligned} \mathbf{X}_{NT,N}(t) &= \left(\mathbf{P}'_N \star (\mathbf{Q}_N^r \mathbf{P}_N) \star \boldsymbol{\Psi}_N \right) \left(\frac{t}{T}, B \right) \mathbf{F}(t) \\ &\quad + \left(\mathbf{P}'_N \star (\mathbf{Q}_N^r \mathbf{P}_N) \star \boldsymbol{\Upsilon}_N \right) \left(\frac{t}{T}, B \right) \boldsymbol{\varepsilon}_N(t) \end{aligned}$$

and

$$\begin{aligned} \mathbf{Z}_{NT,N}(t) &= \left(\mathbf{P}'_N \star [(\mathbf{I}_N - \mathbf{Q}_N^r) \mathbf{P}_N] \star \boldsymbol{\Psi}_N \right) \left(\frac{t}{T}, B \right) \mathbf{F}(t) \\ &\quad + \left(\mathbf{P}'_N \star [(\mathbf{I}_N - \mathbf{Q}_N^r) \mathbf{P}_N] \star \boldsymbol{\Upsilon}_N \right) \left(\frac{t}{T}, B \right) \boldsymbol{\varepsilon}_N(t). \end{aligned}$$

From the definitions of $\mathbf{X}_{NT,N}(t)$ and $\mathbf{Z}_{NT,N}(t)$, it immediately follows that

$$\mathbf{X}_{NT,N}(t) = \left[\mathbf{P}'_N(u, B) \mathbf{Q}_N^r \mathbf{F}_N(u, t) \right]_{u=\frac{t}{T}}$$

is in the subspace $\mathcal{F}_N^r(\frac{t}{T})$ whereas $\mathbf{Z}_{NT,N}(t)$ is orthogonal to $\mathcal{F}_N^r(\frac{t}{T})$. Therefore, $\mathbf{X}_{NT,N}(t) = \text{proj} \left(\mathbf{Y}_{NT}(t) \middle| \mathcal{F}_{NT}^r \left(\frac{t}{T} \right) \right)$ as required.

The decomposition in (3.18) separates the dominant part $\mathbf{X}_{NT,N}(t)$ from the residual one $\mathbf{Z}_{NT,N}(t)$. The approximate common component $X_{iT,N}(t)$ reproduces the r -factor structure of the process, and it is given by a (filtered) projection of the data on the eigenvectors corresponding to the largest r eigenvalues of $\boldsymbol{\Sigma}_N(u, \omega)$

$$X_{iT,N}(t) = (\boldsymbol{\phi}_{Ni} \star \boldsymbol{\Psi}_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t) + (\boldsymbol{\phi}_{Ni} \star \boldsymbol{\Upsilon}_N) \left(\frac{t}{T}, B \right) \boldsymbol{\varepsilon}_N(t) \quad (3.19)$$

where the filter $\boldsymbol{\phi}$ in time domain is given by

$$\boldsymbol{\phi}_{Ni}(u, B) = \left(\mathbf{P}_{Ni}^* \star (\mathbf{Q}_N^r \mathbf{P}_N) \right) (u, B) \quad (3.20)$$

and where the notation emphasizes that the population common components $X_{iT,N}(t)$ depends on N . The following theorem shows that such a projection is able to recover the common components of the process when $N \rightarrow \infty$.

Proposition 3. *Suppose that Assumptions 1 to 3 hold. For all $i \in \mathbb{N}$, the population approximate common component $X_{iT,N}(t)$ converges in mean square to the true common component $X_{iT}(t)$ uniformly over $1 \leq t \leq T$ as N tends to infinity:*

$$\lim_{N \rightarrow \infty} \sup_{T \in \mathbb{N}} \sup_{1 \leq t \leq T} \mathbb{E} \left[X_{iT,N}(t) - X_{iT}(t) \right]^2 = 0.$$

Proof. The proof is given in Appendix B.3. □

3.4. Truncation and filtering error

Next, we show the mean square convergence to zero of the filtering error and the truncation error defined in section 3.2. We start with the truncation error. Let $\phi_{NTi}(u, \omega)$ be the transfer function of the truncated filter

$$\phi_{NTi}(u, B) = \sum_{k=-M_T}^{M_T} \phi_{Ni,k}(u) B^k.$$

Noting that $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$ are orthonormal processes, we obtain for the truncation error by Parseval's identity

$$\begin{aligned} & \mathbb{E} |\bar{X}_{iT,N}(t) - X_{iT,N}(t)|^2 \\ & \leq \int_{-\pi}^{\pi} \|\phi_{NTi}(\frac{t}{T}, \omega) - \phi_{Ni}(\frac{t}{T}, \omega)\|^2 \left[\|\boldsymbol{\Psi}_N(\frac{t}{T}, \omega)\|^2 + \|\boldsymbol{\Upsilon}_N(\frac{t}{T}, \omega)\|^2 \right] d\omega. \end{aligned} \quad (3.21)$$

By Assumption 1 (ii) the factors $\|\boldsymbol{\Psi}_N(\frac{t}{T}, \omega)\|^2$ and $\|\boldsymbol{\Upsilon}_N(\frac{t}{T}, \omega)\|^2$ are uniformly bounded in ω and u . Furthermore, since $\|\phi_{NTi}(u, \omega)\|^2 \leq 1$ as noted before, we have

$$\sup_{u \in [u_t, 1-u_t]} \int_{-\pi}^{\pi} \|\phi_{NTi}(u, \omega) - \phi_{Ni}(u, \omega)\|^2 d\omega \rightarrow 0$$

as $N \rightarrow \infty$, which shows that the truncation error converges to zero in mean square.

For the discussion of the filtering error, let

$$\begin{aligned} \varphi_{Ni}^{\Psi}(\frac{t}{T}, B) &= (\phi_{NTi} \star \boldsymbol{\Psi}_N)(\frac{t}{T}, B) \\ &= \frac{1}{2\pi} \sum_{k=0}^{\infty} \left[\sum_{j=-M_T}^{M_T} \phi_{Ni,j}(\frac{t}{T}) \boldsymbol{\Psi}_{N,k-j}(\frac{t}{T}) \right] B^k, \end{aligned}$$

and

$$\begin{aligned} \tilde{\varphi}_{NTi}^{\Psi}(\frac{t}{T}, B) &= \phi_{NTi}(\frac{t}{T}, B) \boldsymbol{\Psi}_N(\frac{t}{T}, B) \\ &= \frac{1}{2\pi} \sum_{k=-0}^{\infty} \left[\sum_{j=-M_T}^{M_T} \phi_{Ni,j}(\frac{t}{T}) \boldsymbol{\Psi}_{N,k-j}(\frac{t-j}{T}) \right] B^k, \end{aligned}$$

and define the filters $\tilde{\varphi}_{NTi}^{\Upsilon}(t, B)$ and $\varphi_{Ni}^{\Upsilon}(t, B)$ analogously with $\boldsymbol{\Psi}_N$ replaced by $\boldsymbol{\Upsilon}_N$. Then the filtering error can be written as

$$\begin{aligned} \tilde{X}_{iT,N}(t) - X_{iT,N}(t) &= [\tilde{\varphi}_{NTi}^{\Psi}(t, B) - \varphi_{Ni}^{\Psi}(t, B)] \mathbf{F}(t) \\ &\quad + [\tilde{\varphi}_{NTi}^{\Upsilon}(t, B) - \varphi_{Ni}^{\Upsilon}(t, B)] \boldsymbol{\varepsilon}_N(t). \end{aligned} \quad (3.22)$$

Notice that the filtering error vanishes if the two filters $\boldsymbol{\Psi}_N(u, B)$ and $\boldsymbol{\Upsilon}_N(u, B)$ do not depend on u , in which case the common and the idiosyncratic components are both stationary. Otherwise, the expression for filter coefficients of the second filter shows that the sequential application of time-varying linear filters in the time domain mixes the dynamics of different time points whereas application of filters in the frequency domain retains the pure dynamics at any specific point in time.

Since the filter $\phi_{NTi}(u, B)$ has finite width M_T , it will shrink asymptotically to the point u as $T \rightarrow \infty$ provided that M_T is of order $o(T)$. Therefore, the filtering

error will vanish for large T given that the filters $\Psi_N(u, B)$ and $\Upsilon_N(u, B)$ are smooth enough. This is formalized in the following proposition.

Proposition 4. *Suppose that Assumptions 1 to 3 hold. For fixed N and for all $i = 1, \dots, N$, the filtering error tends to zero in mean square uniformly over $M_T \leq t \leq T - M_T$ as T tends to infinity:*

$$\lim_{T \rightarrow \infty} \sup_{M_T \leq t \leq T - M_T} \mathbb{E} |\tilde{X}_{iT,N}(t) - \bar{X}_{iT,N}(t)|^2 = 0,$$

where

$$M_T \rightarrow \infty \quad \text{and} \quad \frac{M_T^{3/2}}{T} \rightarrow 0 \quad \text{as} \quad T \rightarrow \infty.$$

Proof. It remains to show mean square convergence to zero of the filtering error. Since the processes $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$ are uncorrelated, we have

$$\begin{aligned} \mathbb{E} \left[\tilde{X}_{iT,N}(t) - X_{iT,N}(t) \right]^2 &= \int_{-\pi}^{\pi} \left\| \tilde{\varphi}_{NTi}^{\Psi}(t, \omega) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\|^2 d\omega \\ &\quad + \int_{-\pi}^{\pi} \left\| \tilde{\varphi}_{NTi}^{\Upsilon}(t, \omega) - \varphi_{Ni}^{\Upsilon}\left(\frac{t}{T}, \omega\right) \right\|^2 d\omega. \end{aligned} \quad (3.23)$$

We show that the first term converges to zero uniformly in $M_T \leq t \leq T - M_T$ as $T \rightarrow \infty$; for the second term, convergence to zero follows by the same arguments. We have for $M_T \leq t \leq T - M_T$

$$\begin{aligned} &\left\| \tilde{\varphi}_{NTi}^{\Psi}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\| \\ &= \frac{1}{2\pi} \left\| \sum_{j=-M_T}^{M_T} \phi_{Ni,j}\left(\frac{t}{T}\right) e^{-i\omega j} \left[\Psi_N\left(\frac{t-j}{T}, \omega\right) - \Psi_N\left(\frac{t}{T}, \omega\right) \right] \right\| \\ &\leq \frac{1}{2\pi} \sum_{j=-M_T}^{M_T} \left\| \phi_{Ni,j}\left(\frac{t}{T}\right) \right\| \left\| \Psi_N\left(\frac{t-j}{T}, \omega\right) - \Psi_N\left(\frac{t}{T}, \omega\right) \right\|. \end{aligned}$$

Here, the second factor can be bounded by

$$\left\| \Psi_N\left(\frac{t-j}{T}, \omega\right) - \Psi_N\left(\frac{t}{T}, \omega\right) \right\| \leq C \frac{|j|}{T}$$

due to Lipschitz continuity of Ψ_N in its first component on $[0, 1]$. With this we obtain the upper bound

$$\left\| \tilde{\varphi}_{NTi}^{\Psi}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\| \leq \frac{C}{2\pi} \frac{M_T}{T} \sum_{j=-M_T}^{M_T} \left\| \phi_{Ni,j}\left(\frac{t}{T}\right) \right\|.$$

By Cauchy Schwarz inequality, Parseval's identity, and $\left\| \phi_{NTi}(u, \omega) \right\|^2 \leq 1$, we have

$$\sum_{j=-M_T}^{M_T} \left\| \phi_{Ni,j}\left(\frac{t}{T}\right) \right\| \leq \sqrt{M_T} \left[\int_{-\pi}^{\pi} \left\| \phi_{Ni}\left(\frac{t}{T}, \omega\right) \right\|^2 d\omega \right]^{\frac{1}{2}} \leq C \sqrt{M_T},$$

which implies

$$\left\| \tilde{\varphi}_{NTi}^{\Psi}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\| \leq C \frac{M_T^{3/2}}{T}. \quad (3.24)$$

Thus, we have shown that the first term in (3.23) converges to zero as T tends to infinity. Since the second term can be treated similarly, this proves (3.25). \square

The upper bound in (3.24) suffices also to show that the expectation

$$\mathbb{E} \left[\left(\tilde{X}_{iT,N}(t) - \bar{X}_{iT,N}(t) \right) \left(\bar{X}_{iT,N}(t) - X_{iT,N}(t) \right) \right]$$

converges to zero as $T \rightarrow \text{infy}$. Thus, under the conditions of the proposition, we also obtain the mean square convergence of the combined error due to filtering and truncation, that is,

$$\lim_{T \rightarrow \infty} \sup_{M_T \leq t \leq T - M_T} \mathbb{E} \left| \tilde{X}_{iT,N}(t) - X_{iT,N}(t) \right|^2 = 0, \quad (3.25)$$

where M_T satisfies the condition in the above proposition.

3.5. Estimation error

In this section we establish the mean square convergence to zero of the estimation error (see Proposition 5). The first step in the estimation of the common components is the estimation of the time-varying spectral density $\Sigma_N(u, \omega)$. Since our method for estimating the common components does not depend on the particular form of the spectral matrix estimator, we do not discuss the estimation in detail but simply impose the following condition on the spectral estimator $\hat{\Sigma}_N(u, \omega)$. For further details on the estimation of $\Sigma_N(u, \omega)$ we refer to the appendix A.

Assumption 4 (Spectral matrix estimator). There exists a sequence u_T with $u_T \rightarrow 0$ and $T u_T \rightarrow \infty$ as $T \rightarrow \infty$ such that the estimator $\hat{\Sigma}_N(u, \omega)$ of the spectral matrix $\Sigma_N(u, \omega)$ is uniformly consistent in $u \in [u_T, 1 - u_T]$ and $\omega \in [-\pi, \pi]$, that is, for all (fixed) $N \in \mathbb{N}$

$$\sup_{u \in [u_T, 1 - u_T]} \sup_{\omega \in [-\pi, \pi]} \left\| \hat{\Sigma}_N(u, \omega) - \Sigma_N(u, \omega) \right\| = O_p(r_T^{-1})$$

for some sequence $r_T \rightarrow \infty$ as $T \rightarrow \infty$.

Assumption 4, which is fulfilled, for example, for the two spectral estimators $\hat{\Sigma}_N(u, \omega)$ discussed in Appendix A, implies that also the matrix $\hat{\mathbf{P}}_{NT}(u, \omega)$ of estimated eigenvectors and thus the estimated transfer functions $\hat{\phi}_N(u, v)$ of the projection filters are uniformly consistent. Based on this uniform convergence, we show in the next proposition that the estimation error converges to zero in probability. Note that for this result to hold, an arbitrary slow rate r_T is sufficient in Assumption 4 as we do not aim at giving a rate of convergence for the estimation error. We also emphasize that for this result we keep N fixed while T tends to infinity.

Proposition 5. *Suppose that Assumptions 1 to 4 hold. In addition assume for the truncation parameter M_T of the estimated filters $\hat{\Phi}_{NT}$ that $M_T \rightarrow \infty$ and $M_T/r_T \rightarrow 0$ as $T \rightarrow \infty$. Then, for all $\delta > 0$, $1 \leq i \leq N$, and all $N \in \mathbb{N}$, the estimation error satisfies*

$$\lim_{T \rightarrow \infty} \sup_{M_T^* \leq t \leq T - M_T^*} \mathbb{P} \left[\left| \hat{X}_{iT,N}(t) - \tilde{X}_{iT,N}(t) \right| > \delta \right] = 0.$$

where $M_T^* = \max\{u_T T, M_T\}$.

Proof. See Appendix B.4. \square

Remark 1. The proof of Proposition 5 does actually include how Assumption 4 implies derivation of a rate of uniform convergence to zero of the estimated eigenvectors and estimated transfer functions of the projection filters $\widehat{\Phi}_{NT}$. For more details, we refer to Appendix B.4.

3.6. Consistency of the estimated common components

With the results of the previous three sections 3.3, 3.4, and 3.5, we can now show the main result of this paper. The following theorem establishes the convergence in probability of the common components estimator $\widehat{\mathbf{X}}_{NT,N}(t)$ to the true common components $\mathbf{X}_{NT}(t)$.

Theorem 6. *Suppose that Assumptions 1 to 4 hold. Then, for all $\epsilon > 0$ and $\eta > 0$, there exists $N^* \in \mathbb{N}$ and $T_N^* \in \mathbb{N}$, $N \geq N^*$, such that for all $N \geq N^*$ and $T \geq T_N^*$*

$$\sup_{M_T^* \leq t \leq T - M_T^*} \mathbb{P} \left[\left| \widehat{\Phi}_{Ni}^{M_T} \left(\frac{t}{T}, B \right) \mathbf{Y}_{NT}(t) - X_{iT}(t) \right| > \epsilon \right] \leq \eta,$$

where $M_T^* = \max\{u_T T, M_T\}$ with u_T and M_T obeying the same conditions as in Assumption 4 and Proposition 5, respectively.

Proof. In order to prove consistency of the estimated common components we decompose the overall estimation error as in section 3.2, and for all $t \in (M_T^*, T - M_T^*)$ we consider the following time-varying probability

$$\begin{aligned} & \mathbb{P} \left[\left| \widehat{X}_{iT,N}(t) - X_{iT}(t) \right| > \epsilon \right] \\ & \leq \mathbb{P} \left[\left| \widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t) \right| > \frac{\epsilon}{3} \right] + \mathbb{P} \left[\left| \widetilde{X}_{iT,N}(t) - X_{iT,N}(t) \right| > \frac{\epsilon}{3} \right] \\ & \quad + \mathbb{P} \left[\left| X_{iT,N}(t) - X_{iT}(t) \right| > \frac{\epsilon}{3} \right]. \end{aligned}$$

In Proposition 3, we have shown that the approximation error $|X_{iT,N}(t) - X_{iT}(t)|$ vanishes as $N \rightarrow \infty$ uniformly over $T \in \mathbb{N}$ and $1 \leq t \leq T$. This implies that for all $\epsilon > 0$ and $\eta > 0$ there exists an N^* such that

$$\mathbb{P} \left[\left| \widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t) \right| > \frac{\epsilon}{3} \right] \leq \frac{\eta}{3}$$

for all $N \geq N^*$ uniformly over $1 \leq t \leq T$ and $T \in \mathbb{N}$. Next, we have shown in section 3.4 that the combined error due to filtering and truncation, $|\widetilde{X}_{iT,N}(t) - X_{iT,N}(t)|$ vanishes as $T \rightarrow \infty$ uniformly over $1 \leq t \leq T$ for all $N \in \mathbb{N}$. Then the uniform convergence in (3.25) implies that for all $N \in \mathbb{N}$, $\epsilon > 0$, and $\eta > 0$ there exists a $T_{1,N}$ such that

$$\mathbb{P} \left[\left| \widetilde{X}_{iT,N}(t) - X_{iT,N}(t) \right| > \frac{\epsilon}{3} \right] \leq \frac{\eta}{3}$$

for all $T > T_{1,N}$. Finally, we have shown in section 3.5 that for fixed N the estimation error tends to zero in probability as T tends to infinity. More precisely, it follows from Proposition 5 that for all $\epsilon > 0$, $\eta > 0$, and $N > N^*$ there exists $T_{2,N} \in \mathbb{N}$ such that for all $T > T_{2,N}$

$$\mathbb{P} \left[\left| X_{iT,N}(t) - X_{iT}(t) \right| > \frac{\epsilon}{3} \right] \leq \frac{\eta}{3}.$$

Altogether, we find that for all $N > N^*$ and all $T > \max\{T_{1,N}, T_{2,N}\}$

$$\mathbb{P} \left[\left| \widehat{X}_{iT,N}(t) - X_{iT}(t) \right| > \epsilon \right] < \eta,$$

which concludes the proof. \square

4. SIMULATION RESULTS

We simulate an r -factor model

$$Y_{iT}(t) = \sum_{j=1}^r X_{iT,j}(t) + Z_i(t), \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (4.1)$$

with *locally stationary AR(p) common components*, as in (4.1) of Dahlhaus (1997)

$$X_{iT,j}(t) = \sum_{k=1}^p a_{ij,k} \left(\frac{t}{T} \right) X_{iT,j}(t-k) + F_j(t) \quad i = 1, \dots, N, \quad j = 1, \dots, r, \quad t = 1, \dots, T, \quad (4.2)$$

where $a_{ij,0} \left(\frac{t}{T} \right) \equiv 1$ for all i, j and t , with stationary factors as well as idiosyncratic errors: $\mathbf{F}(t) \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$ and $\mathbf{Z}_N(t) \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$. For simulation purposes we restrict to the class of AR(2) locally stationary two-factor models (i.e. $r = p = 2$) with the following specifications for the parameters

$$\begin{aligned} a_{ij1}(u) &= \mp \frac{2}{\alpha} \cos(\theta_{ij}(u)) \quad u \in [0, 1] \\ a_{ij2}(u) &= \pm \frac{1}{\alpha^2} \quad \forall u \in [0, 1] \quad \text{and} \quad \forall i, j = 1, \dots, N \end{aligned}$$

with $\theta_{ij}(u) = \phi_{ij} - \cos(\nu_{ij}u)$, $i, j = 1, \dots, N$ and $|\alpha| > 1$. For u fixed, the roots of the characteristics polynomial are

$$\rho_{ij}(u) = \alpha e^{\pm i\theta_{ij}(u)}$$

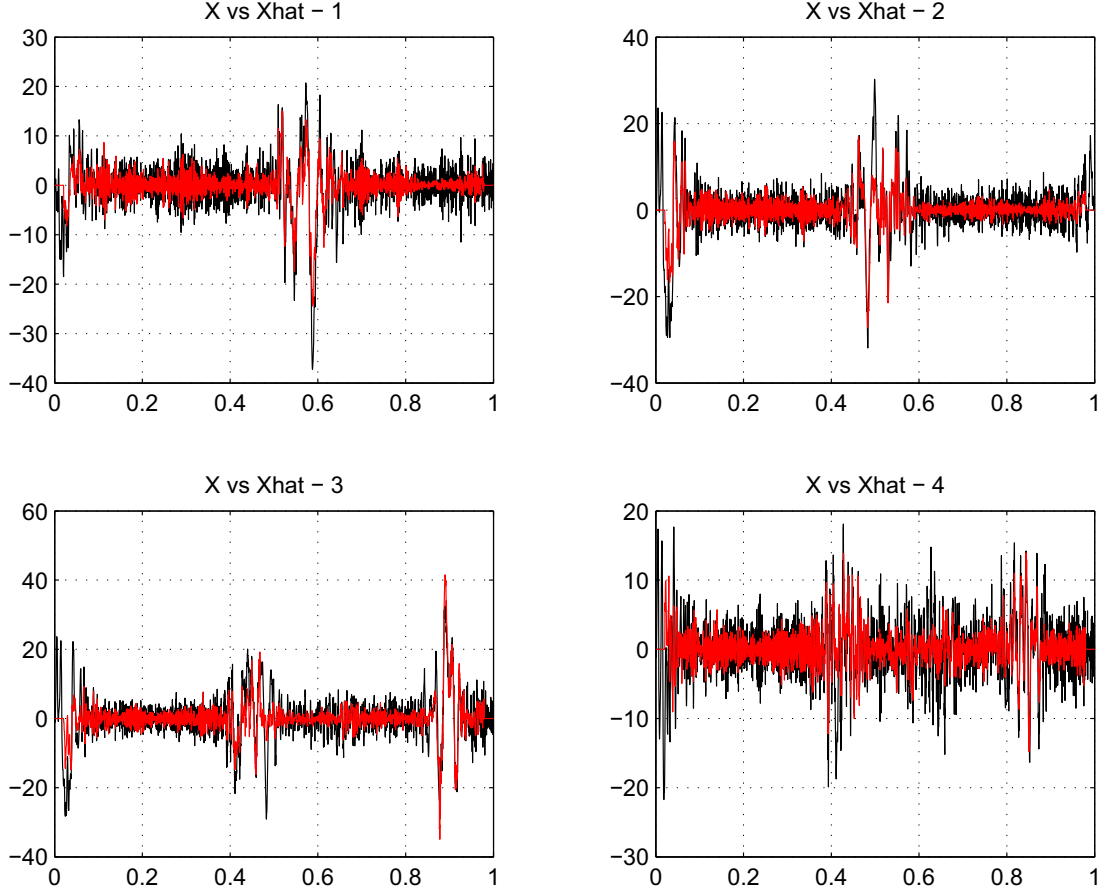
and thus $|\rho_{ij}(u)| = |\alpha| > 1$ for all $u \in [0, 1]$ and for all $i, j = 1, \dots, N$. The wave $\theta(u)$ depends on three parameters: α controls the amplitude, ϕ is the phase and ν is the velocity.

The spectral estimator implemented in the simulations is the smoothed segmented periodogram (see Appendix A.2). This estimator depends on two "smoothing" parameters: first, the parameter b_T , the bandwidth in the frequency direction. Dahlhaus (1996, Theorem 2.3) showed that the optimal value b_T^{opt} for this parameter is of the order $T^{-\frac{1}{6}}$, where b_T^{opt} is optimal in the sense that it minimizes the relative mean squared error of the estimator of the time-varying spectrum. The bandwidth b_T used in our simulations is $b_T = \frac{1}{3}T^{-\frac{1}{6}}$. Second, this estimator depends on the length L of each segment, which is akin a smoothing parameter in time. We also note that for convenience we choose the truncation parameter M of the filters $\widehat{\phi}_{NT,i} \left(\frac{t}{T}, B \right)$ to be equal to $L/2$. Finally, as in Dahlhaus (1996), we allow our estimator to be based on shifted segments with a shift S from segment to segment, resulting into P segments (such that $T = S(P-1) + L$), and with midpoints $t_j = S(j-1) + L/2$, $j = 1, \dots, P$.

We consider two numerical scenarios, both based on $N = 10, 20, 30, 40, 50$ cross-section dimensions. In the first one $T = L^2$ with $L = 16, 32, 48, 64, 80$, $P = L + 1$ and $S = L - 1$. In the second one $T = 8L$ with $L = 32, 64, 96, 128, 160$, $P = L + 1$ and $S = 7$.

The consistency result in Theorem 6 ensures the consistency of the estimated common components $\widehat{X}_{iT}(t)$ *elementwise*, that is for all i and all t . This is confirmed by the simulated example in Figure 4.1, where we report the estimation of the common components for $N = 10$ and $T = 2304$ (this is the parametrization corresponding to our first scenario with $L = 48$ and $N = 10$). The quality of the

FIGURE 4.1. True common components $X_{iT}(t)$ (black) and estimated ones $\widehat{X}_{iT}(t)$ (grey), $i = 1, \dots, 4$.



fit is remarkable, even for a relatively small cross-section size. In this figure we just consider the first four (among 10) common components $X_{iT}(t)$ and (a realization of) its estimator $\widehat{X}_{iT}(t)$, $t = 1, \dots, T$. Our estimator is able to capture the time-varying variance (nonstationarity) as well as the time-varying dynamics (changes in the autocorrelation).

To evaluate the global performance of the estimator we consider the rescaled norm of the difference between the matrix of estimated common components and that containing the true ones. Let $\mathbf{X}_{N,T-2L}$ be the $N \times (T-2L+1)$ matrix containing the $N \times 1$ vectors of common components $\mathbf{X}_{NT}(t)$, $t = L, \dots, T-L$

$$\mathbf{X}_{N,T-2L} := [\mathbf{X}_{NT}(L), \dots, \mathbf{X}_{NT}(T-L)]$$

and $\widehat{\mathbf{X}}_{N,T-2L}$ the corresponding estimator

$$\widehat{\mathbf{X}}_{N,T-2L} := [\widehat{\mathbf{X}}_{NT}(L), \dots, \widehat{\mathbf{X}}_{NT}(T-L)].$$

We now report simulations of $\mathbf{X}_{N,T-2L}$ based on Q replications. Let $X_{iT}^{(q)}(t)$ be the (i, t) entry of $\mathbf{X}_{N,T-2L}$ ($i = 1, \dots, N$, $t = L, \dots, T-L$) obtained from the q -th realization of the common component $X_{iT}(t)$ simulated from the locally stationary

AR(p) model in (4.1)-(4.2), with $r = p = 2$, and define $\widehat{X}_{iT}^{(q)}(t)$ as the estimator of $X_{iT}^{(q)}(t)$ based on (3.8). For $q = 1, \dots, Q$, define the following loss function

$$\begin{aligned} \Delta_{NT}^{(q)}(\beta) &= [(T - 2L + 1)]^{-\beta} \left\| \widehat{\mathbf{X}}_{N,T-2L}^{(q)} - \mathbf{X}_{N,T-2L}^{(q)} \right\| \\ &= [(T - 2L + 1)]^{-\beta} \sqrt{\sum_{i=1}^N \sum_{t=L}^{T-L} \left(\widehat{X}_{iT}^{(q)}(t) - X_{iT}^{(q)}(t) \right)^2} \end{aligned} \quad (4.3)$$

depending on N , T and β . In Tables 4.1 and 4.2 we consider, for both scenarios and for different values of β , the average

$$\bar{\Delta}_{NTQ}(\beta) = \frac{1}{Q} \sum_{q=1}^Q \Delta_{NT}^{(q)}(\beta)$$

and the standard deviation

$$S_{NTQ}^{\Delta}(\beta) = \sqrt{\frac{1}{Q-1} \sum_{q=1}^Q \left[\Delta_{NT}^{(q)}(\beta) - \bar{\Delta}_{NTQ}(\beta) \right]^2}$$

of the loss in (4.3) over Q experiments.

In the case of *mean-square convergence* (of the estimated common components to the true ones), the loss function in (4.3) with $\beta = \frac{1}{2}$ can be interpreted as the sample counterpart of the (square root of the) mean squared error of the estimator $\widehat{\mathbf{X}}$ with respect to \mathbf{X} . However, Theorem 6 guarantees only *consistency in probability* (of the estimated common components to the true ones). This explains why we cannot expect $\bar{\Delta}_{NTQ}(0.5)$ in generality to be decreasing with increasing N and/or T . We expect however, on the other hand, that $\bar{\Delta}_{NTQ}(0.5)$ should not increase with these parameters, or to phrase it differently that once β becomes larger than 0.5 the loss $\bar{\Delta}_{NTQ}(\beta)$ should begin to decrease. Hence we add a small investigation to confirm this conjecture. Whereas for $\beta = 0.6$, the values of $\bar{\Delta}_{NTQ}(0.6)$ begin to decrease w.r.t. T (starting from $N = 20$), the values $\bar{\Delta}_{NTQ}(0.7)$ can be observed to monotonically decrease in both N and T .

5. CONCLUSIONS

With this paper we have delivered an important generalisation of factor modelling as it is currently available in the literature. We allow for a second-order non-stationarity of our data-generating process in order to cope with the empirical observation that - being observed over long time periods (or on a sufficiently fine time resolution) - macro economical or financial data can exhibit some time variation in their serial variance-covariance structure. Furthermore we do not content ourselves to work with a static factor model in order to avoid taking a potentially large number of factors into account. In this respect we generalise both the work by Motta et al. (2006) on time-varying static factors and by Forni et al. (2000) on stationary dynamic factor modelling. With our approach we contribute to the yet recent literature on dimension-reduction of multivariate time series with possibly time-varying correlation, where the latter one will allow to work with a considerably smaller number of factors (common components) to explain the co-movements in a

TABLE 4.1. First scenario. Average and standard deviation (in brackets) of the loss defined in (4.3) over $Q = 100$ experiments for different values of β

$\beta = 0.5$	$T = 256$	$T = 1024$	$T = 2304$	$T = 4096$	$T = 6400$
$N = 10$	3.3208 (0.3810)	3.3127 (0.2899)	3.3696 (0.2841)	3.2455 (0.2117)	3.3220 (0.1845)
$N = 20$	3.6029 (0.5169)	3.7409 (0.3238)	3.6789 (0.2352)	3.7103 (0.2177)	3.6548 (0.2414)
$N = 30$	3.6423 (0.2589)	3.8204 (0.2272)	3.8148 (0.2180)	3.7878 (0.1896)	3.7324 (0.1803)
$N = 40$	3.6112 (0.2706)	3.8746 (0.2338)	3.8280 (0.2159)	3.9015 (0.1892)	3.8897 (0.1875)
$N = 50$	3.5562 (0.2586)	3.8887 (0.1827)	3.8896 (0.1814)	3.9669 (0.1847)	3.9551 (0.1655)
$\beta = 0.6$	$T = 256$	$T = 1024$	$T = 2304$	$T = 4096$	$T = 6400$
$N = 10$	1.5347 (0.1761)	1.3241 (0.1159)	1.2393 (0.1045)	1.1257 (0.0734)	1.1012 (0.0612)
$N = 20$	1.5536 (0.2229)	1.3951 (0.1207)	1.2624 (0.0807)	1.2007 (0.0705)	1.1304 (0.0747)
$N = 30$	1.5081 (0.1072)	1.3681 (0.0814)	1.2570 (0.0718)	1.1771 (0.0589)	1.1086 (0.0536)
$N = 40$	1.4529 (0.1089)	1.3482 (0.0813)	1.2256 (0.0691)	1.1780 (0.0571)	1.1225 (0.0541)
$N = 50$	1.3992 (0.1017)	1.3232 (0.0622)	1.2178 (0.0568)	1.1714 (0.0545)	1.1162 (0.0467)
$\beta = 0.7$	$T = 256$	$T = 1024$	$T = 2304$	$T = 4096$	$T = 6400$
$N = 10$	0.7093 (0.0814)	0.5292 (0.0463)	0.4558 (0.0384)	0.3904 (0.0255)	0.3650 (0.0203)
$N = 20$	0.6699 (0.0961)	0.5203 (0.0450)	0.4332 (0.0277)	0.3886 (0.0228)	0.3496 (0.0231)
$N = 30$	0.6245 (0.0444)	0.4899 (0.0291)	0.4142 (0.0237)	0.3658 (0.0183)	0.3292 (0.0159)
$N = 40$	0.5845 (0.0438)	0.4691 (0.0283)	0.3924 (0.0221)	0.3557 (0.0173)	0.3239 (0.0156)
$N = 50$	0.5505 (0.0400)	0.4503 (0.0212)	0.3813 (0.0178)	0.3459 (0.0161)	0.3150 (0.0132)

TABLE 4.2. Second scenario. Average and standard deviation (in brackets) of the loss defined in (4.3) over $Q = 100$ experiments for different values of β

$\beta = 0.5$	$T = 256$	$T = 512$	$T = 768$	$T = 1024$	$T = 1280$
$N = 10$	2.9580 (0.2251)	3.3043 (0.3142)	3.2508 (0.2313)	3.4686 (0.2303)	3.5640 (0.2790)
$N = 20$	3.3768 (0.2428)	3.5524 (0.1892)	3.6772 (0.2897)	3.7479 (0.2485)	3.7753 (0.2561)
$N = 30$	3.5409 (0.2428)	3.7406 (0.2390)	3.8534 (0.2426)	3.7542 (0.1886)	3.8762 (0.2924)
$N = 40$	3.5707 (0.1669)	3.8851 (0.1989)	3.9483 (0.2926)	3.9142 (0.2432)	3.8172 (0.2462)
$N = 50$	3.7174 (0.2334)	3.9705 (0.2692)	3.9183 (0.2127)	3.8137 (0.2236)	3.8464 (0.2599)
$\beta = 0.6$	$T = 256$	$T = 512$	$T = 768$	$T = 1024$	$T = 1280$
$N = 10$	1.3882 (0.1056)	1.4472 (0.1376)	1.3673 (0.0973)	1.4176 (0.0941)	1.4245 (0.1115)
$N = 20$	1.4786 (0.1063)	1.4517 (0.0773)	1.4431 (0.1137)	1.4292 (0.0947)	1.4079 (0.0955)
$N = 30$	1.4888 (0.1021)	1.4678 (0.0938)	1.4521 (0.0914)	1.3747 (0.0691)	1.3881 (0.1047)
$N = 40$	1.4588 (0.0682)	1.4813 (0.0758)	1.4457 (0.1071)	1.3927 (0.0865)	1.3282 (0.0857)
$N = 50$	1.4852 (0.0933)	1.4805 (0.1004)	1.4031 (0.0762)	1.3270 (0.0778)	1.3088 (0.0884)
$\beta = 0.7$	$T = 256$	$T = 512$	$T = 768$	$T = 1024$	$T = 1280$
$N = 10$	0.6515 (0.0496)	0.6339 (0.0603)	0.5751 (0.0409)	0.5794 (0.0385)	0.5694 (0.0446)
$N = 20$	0.6474 (0.0465)	0.5932 (0.0316)	0.5663 (0.0446)	0.5450 (0.0361)	0.5250 (0.0356)
$N = 30$	0.6260 (0.0429)	0.5760 (0.0368)	0.5472 (0.0345)	0.5034 (0.0253)	0.4971 (0.0375)
$N = 40$	0.5960 (0.0279)	0.5648 (0.0289)	0.5294 (0.0392)	0.4955 (0.0308)	0.4622 (0.0298)
$N = 50$	0.5934 (0.0373)	0.5520 (0.0374)	0.5024 (0.0273)	0.4617 (0.0271)	0.4454 (0.0301)

large panel of observed time series: it is obvious that allowing the dynamics of a few common components to slowly change over time will allow for very sparse modelling.

From a statistical point of view we enlarge the domain of applicability of Principal Components Regression (PCR) in the frequency domain, one of the main tools in estimating dynamic factor models. The PCR technique is based on the spectral decomposition of the matrix describing the dynamics of the underlying process. We allow for non-stationarity of these dynamics by working with time-varying filters, and using the by now standard approach of letting both sample size and cross-sectional dimension tend to infinity, we derive weakly consistent estimators of the common components via the spectral decomposition of a localized smoothed periodogram, as, for example, defined in (A.3).

The main theoretical contribution of the paper is provided by the treatment of non-stationarity, both in the time-varying dynamical factor loadings and in the idiosyncratic components. To develop our asymptotic theory of consistency, the approximation error is based on the new definition of time-varying dynamic principal components given in (3.15), the estimation error is based on the new definition in (3.20) of the filter applied to data. In order to localize (at rescaled time u) the corresponding definitions in the stationary case properly, these definitions are based on the convolution of the eigenvectors. As a result, we have an additional source of error, the representation error: this is due to the problem of time-varying filtering of a non-stationary process, which is novel with respect to the dynamic factor model of Forni et al. (2000).

With our paper we do not touch upon the important problem of using dynamic factor models for the purpose of prediction. Having used two-sided filters in our approach, merely for the comfort of reducing technical complexity when deriving our consistency results, would need to be given up, and some generalisation is called for to be able to adapt our methodology to one-sided filters or alike. Another open problem is the question of how to determine the factor space dimensionality automatically from the data - we believe that recent work in the static and dynamic PCR-based factor modelling, e.g. by Bai and Ng (2002) and by Hallin and Liška (2007) can be generalised, but this is left for future research as well.

ACKNOWLEDGEMENTS

The authors thank Marco Lippi, Marc Hallin, and Christian Hafner for many stimulating discussions and valuable remarks. Rainer von Sachs and Giovanni Motta acknowledge financial support from the contract “Projet d’Actions de Recherche Concertées” nr. 07/12/002 of the “Communauté française de Belgique”, granted by the “Académie universitaire Louvain”.

APPENDIX A. ESTIMATION OF TIME-VARYING SPECTRAL DENSITY MATRICES

The methods presented in the main part of this chapter are based on estimators of the time-varying spectral density matrix. As the approximate factor model leaves the dependence structure of the idiosyncratic components unspecified, the spectral matrix is estimated nonparametrically. In this section, we review two common approaches for estimating time-varying spectral densities and cite relevant results to achieve uniform consistency of the spectral matrix estimator $\widehat{\Sigma}_N(u, \omega)$.

A.1. Smoothed pre-periodogram

Neumann and von Sachs (1997) proposed estimation of time-varying spectral densities based on the pre-periodogram

$$\tilde{\mathbf{S}}_N\left(\frac{t}{T}, \omega\right) = \frac{1}{2\pi} \sum_{k: 1 \leq [t + \frac{1}{2} \pm \frac{k}{2}] \leq T} \mathbf{Y}_{NT}\left([t + \frac{1+k}{2}]\right) \mathbf{Y}_{NT}\left([t + \frac{1-k}{2}]\right) e^{-i\omega k},$$

where $[x]$ denotes the largest integer smaller than x . The product $\mathbf{Y}_{NT}\left([t + \frac{1+k}{2}]\right) \mathbf{Y}_{NT}\left([t + \frac{1-k}{2}]\right)$ can be regarded as a raw estimator of the local autocovariance $\mathbf{\Gamma}_N\left(\frac{t}{T}, k\right)$ given by

$$\mathbf{\Gamma}_N(u, k) = \int_{-\pi}^{\pi} \mathbf{\Sigma}_N(u, \omega) e^{i\omega k} d\omega.$$

Similarly, $\tilde{\mathbf{S}}_N\left(\frac{t}{T}, \omega\right)$ is a raw estimate of the time-varying spectral matrix $\mathbf{\Sigma}_N\left(\frac{t}{T}, \omega\right)$. Like the periodogram in the stationary case, the pre-periodogram $\tilde{\mathbf{S}}_N\left(\frac{t}{T}, \omega\right)$ is not consistent and thus needs to be smoothed in time and frequency direction. This leads to kernel estimators of the form

$$\hat{\mathbf{\Sigma}}_{NT}(u, \omega) = \frac{2\pi}{T^2} \sum_{s, j=1}^T \frac{1}{h_T b_T} K^\tau\left(\frac{u - \frac{s}{T}}{h_T}\right) K^\phi\left(\frac{\omega - \omega_j}{b_T}\right) \tilde{\mathbf{S}}_N\left(\frac{t}{T}, \omega_j\right), \quad (\text{A.1})$$

where $\omega_j = 2\pi j/T$ are the Fourier frequencies, K^τ and K^ϕ are two kernels and h_T and b_T are the corresponding bandwidths in time and frequency direction, respectively. We assume that the kernels K^τ and K^ϕ have compact support on $[-\frac{1}{2}, \frac{1}{2}]$ and are of bounded variation with $\int_{-1/2}^{1/2} x K(x) dx = 0$ and $\int_{-1/2}^{1/2} K(x) dx = 1$.

A.2. Smoothed segmented periodogram

An alternative common approach for estimating time-varying spectral densities is based on the idea that a process with slowly varying spectral matrix $\mathbf{\Sigma}_N(u, \omega)$ can be treated as stationary over short time intervals. This suggests to use of ordinary kernel estimators applied locally to the time series to obtain a time-varying estimate for the spectral density. More precisely, let

$$\mathbf{D}_{NT}^L(u, \omega) = \sum_{s=0}^{L-1} h\left(\frac{s}{L}\right) \mathbf{Y}_{NT}([uT] - L/2 + s + 1) e^{-i\omega s}$$

be the localized Fourier transform of the process over the segment $\{[uT] - L/2 + 1, [uT] + L/2\}$, where $h: \mathbb{R} \rightarrow \mathbb{R}$ is a data taper with $h(x) = 0$ for $x \notin [0, 1]$ and Fourier transform

$$H_k^L(\omega) = \sum_{s=0}^{L-1} h\left(\frac{s}{L}\right)^k e^{-i\omega s}.$$

The the localized or segmented periodogram is defined as

$$\hat{\mathbf{S}}_{NT}^L(u, \omega) = \frac{1}{2\pi H_2^L(0)} \mathbf{D}_{NT}^L(u, \omega) \mathbf{D}_{NT}^L(u, \omega)^*. \quad (\text{A.2})$$

Thus the periodogram estimator of the time-varying spectrum $\mathbf{\Sigma}_N(u, \omega)$ is obtained from a segment of length L with midpoint $[uT]$. Since the periodogram is not a consistent estimator, the segmented periodogram needs to be smoothed (see Dahlhaus

1996), which leads to the kernel estimator

$$\widehat{\Sigma}_{NT}^L(u, \omega) = \frac{1}{b_T} \int_{-\pi}^{\pi} K^f\left(\frac{\omega - \mu}{b_T}\right) \widehat{\mathbf{S}}_{NT}^L(u, \omega) d\mu, \quad (\text{A.3})$$

where K^f is a kernel with compact support $[-\frac{1}{2}, \frac{1}{2}]$ satisfying $K^f(x) = K^f(-x)$ and $\int K^f(x) dx = 1$, and b_T is the bandwidth in frequency direction. Usual assumptions on the parameters $L = L_T$ and b_T to achieve consistency are $L_T \rightarrow \infty$, $L_T/T \rightarrow 0$, $b_T \rightarrow 0$ and $b_T T \rightarrow \infty$.

A.3. Uniform consistency

In Assumption 4, we require that the used estimator of the spectral matrix is uniformly consistent. For the smoothed pre-periodogram and the smoothed segmented periodogram, the uniform consistency has been showed by Dahlhaus (2008). Since this paper is a yet unpublished manuscript, we repeat here the assumptions and the statement. For simplicity, we treat only the case of evolutionary processes, which is sufficient for the purpose of this paper.

The proof by Dahlhaus (2008) requires a number of technical assumptions. Let

$$V(g) = \sup_{m \in \mathbb{N}} \sup_{0 \leq x_0 < \dots < x_m \leq 1} \sum_{k=1}^m |g(x_k) - g(x_{k-1})|$$

be the total variation of a function $g : [0, 1] \rightarrow \mathbb{R}$. Furthermore, we set $l_q(k) = \max\{|k| \log^{1+q} |k|, 1\}$ for $q > 0$.

Assumption 5. $\mathbf{X}_{NT}(t)$, $1 \leq t \leq T$, is an N -dimensional stochastic process satisfying the following conditions:

(i) $\mathbf{X}_{NT}(t)$ has a representation

$$\mathbf{X}_{NT}(t) = \sum_{k=0}^{\infty} \Phi_{N,k}\left(\frac{t}{T}\right) \boldsymbol{\varepsilon}_N(t-k), \quad (\text{A.4})$$

where the errors $\boldsymbol{\varepsilon}_N(t)$, $t \in \mathbb{Z}$, are independent and identically distributed with mean $\mathbb{E}(\boldsymbol{\varepsilon}_N(t)) = \mathbf{0}$ and covariance matrix $\mathbb{E}(\boldsymbol{\varepsilon}_N(t) \boldsymbol{\varepsilon}_N(t)') = \mathbf{I}_N$;

(ii) there exists a positive constant $C_{\varepsilon, N}$ such that $\mathbb{E}|\varepsilon_i(t)|^k \leq C_{\varepsilon, N}^k$ for all $i = 1, \dots, N$, $t \in \mathbb{Z}$, and $k \in \mathbb{N}$;

(iii) the coefficients $\Phi_{N,k}(u) = (\phi_{ij,k}(u))_{i,j=1,\dots,N}$ satisfy

$$\sup_{u \in [0,1]} |\phi_{ij,k}(u)| \leq \frac{C_\phi}{l_q(k)} \text{ and } V(\phi_{ij,k}) \leq \frac{C_\phi}{l_q(k)} \quad (\text{A.5})$$

for all $i = 1, \dots, N$, $j = 1, \dots, r$, and $k \in \mathbb{Z}$, where C_ϕ is a positive finite constant not depending on T ;

(iv) the time-varying spectral density $\Sigma_N(u, \omega)$ is twice differentiable with respect to $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$.

Under these conditions, Dahlhaus (2008) showed that the above kernel estimators of the spectral density matrix are uniformly consistent in $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$. More precisely, we have the following result.

Theorem 7 (Dahlhaus 2008, Thm 3.1). *Suppose that $\mathbf{X}_{NT}(t)$, $1 \leq t \leq T$, is an N -dimensional evolutionary process satisfying Assumption 5. Furthermore, let $\widehat{\Sigma}_{NT}(u, \omega)$ be the smoothed pre-periodogram estimator in (A.1) with bandwidth h_T and b_T such that $h_T, b_T \rightarrow 0$ and $Th_T b_T \rightarrow \infty$ as $T \rightarrow \infty$. Then for $Th_T b_T \gg \log^2 T$*

$$\sup_{u \in \left[\frac{h_T}{2}, 1 - \frac{h_T}{2}\right]} \sup_{\omega \in (-\pi, \pi)} \left\| \widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega) \right\| = O_p\left((h_T b_T T)^{-1/2} + h_T^2 + b_T^2\right). \quad (\text{A.6})$$

A similar result holds for the kernel estimator based on the segmented periodogram (Dahlhaus 2008, Remark 3.2) where then $h_T = L_T/T$.

A.4. Application to evolutionary dynamic factor processes

As remarked in section 3, the evolutionary dynamic factor model in (3.1) is not locally stationary according to the definition of Dahlhaus nor has it an evolutionary representation as in (A.4). Thus Theorem 7 is not directly applicable. However, $\mathbf{Y}_{NT}(t)$ can be written as a linear combination of two evolutionary and thus locally stationary processes; applying Theorem 7 to this process, we can establish uniform consistency also for the spectral estimator $\widehat{\Sigma}_N(u, \omega)$.

More precisely, let $\mathbf{V}_N(t)$ be the 2 N -dimensional processes consisting of the common and the idiosyncratic components of $\mathbf{Y}_N(t)$, that is,

$$\mathbf{V}_{NT}(t) = \begin{pmatrix} \mathbf{X}_{NT}(t) \\ \mathbf{Z}_{NT}(t) \end{pmatrix}.$$

Then $\mathbf{V}_{NT}(t)$ is an evolutionary process in the sense of (A.4) and has a time-varying spectral matrix given by

$$\Sigma_N^{\mathbf{V}}(u, \omega) = \begin{pmatrix} \Sigma_N^{\mathbf{X}}(u, \omega) & \mathbf{0} \\ \mathbf{0} & \Sigma_N^{\mathbf{Z}}(u, \omega) \end{pmatrix},$$

where $\Sigma_N^{\mathbf{X}}(u, \omega)$ and $\Sigma_N^{\mathbf{Z}}(u, \omega)$ are the time-varying spectral density matrices of $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$, respectively, and where the off-diagonal zero matrices are due to the orthogonality between $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$. Now suppose that Assumption 5 holds for $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$. Then, by Theorem 7, an estimator $\widehat{\Sigma}_N^{\mathbf{V}}(u, \omega)$ based on $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$ would be uniformly consistent; note that such an estimator is hypothetical since $\mathbf{X}_{NT}(t)$ and $\mathbf{Z}_{NT}(t)$ are unobserved. Nevertheless, since the observed process $\mathbf{Y}_{NT}(T)$ is given by

$$\mathbf{Y}_{NT}(t) = \begin{pmatrix} \mathbf{I}_N & \mathbf{I}_N \end{pmatrix} \mathbf{V}_{NT}(t),$$

its spectral matrix $\Sigma_N(u, \omega)$ can be written as

$$\Sigma_N(u, \omega) = \begin{pmatrix} \mathbf{I}_N & \mathbf{I}_N \end{pmatrix} \Sigma_N^{\mathbf{U}}(u, \omega) \begin{pmatrix} \mathbf{I}_N & \mathbf{I}_N \end{pmatrix}',$$

and $\widehat{\Sigma}_N(u, \omega)$ is related to $\widehat{\Sigma}_N^{\mathbf{V}}(u, \omega)$ analogously. Consequently, the uniform consistency of $\widehat{\Sigma}_N(u, \omega)$ follows immediately from that of $\widehat{\Sigma}_N^{\mathbf{V}}(u, \omega)$.

We note that Assumption 5 applied to the common component $\mathbf{X}_{NT}(t)$ and the idiosyncratic component $\mathbf{Z}_{NT}(t)$ provides a sufficient set of conditions for Assumption 4 to hold. We think that these technical conditions are not necessary to establish uniform consistency but are due to the method of proof in Dahlhaus (2008).

APPENDIX B. PROOFS

B.1. Proof of Proposition 1

The matrices $\Sigma_N^X(u, \omega)$ and $\Sigma_N^Z(u, \omega)$ are spectral matrices and thus Hermitian. By Weyl's Theorem (Lütkepohl 1996, p. 75)

$$\lambda_{Nj}^X(u, \omega) + \lambda_{NN}^Z(u, \omega) \leq \lambda_{Nj}(u, \omega) \leq \lambda_{Nj}^X(u, \omega) + \lambda_{N1}^Z(u, \omega), \quad j = 1, \dots, N.$$

On the one hand, Assumption 3 (i) implies for $j = 1, \dots, r$ that $\lambda_{Nj}(u, \omega) \geq \lambda_{Nj}^X(u, \omega)$ tends to infinity a.e. in $[-\pi, \pi]$ and uniformly in $u \in [0, 1]$ as $N \rightarrow \infty$. On the other hand, if $j > r$, we have $\lambda_{Nj}^X(u, \omega) = 0$ and thus by Assumption 3 (ii) $\lambda_{Nj}(u, \omega) \leq \lambda_{N1}^Z(u, \omega) \leq \bar{\lambda}_Z$ uniformly in $\omega \in [-\pi, \pi]$ and $u \in [0, 1]$ for all $N > r$. \square

B.2. Proof of Corollary 2

Since $\mathbf{Z}_{NT}(t)$ is set to zero for $t < 1$ and $t > T$, we have for linear filters $\mathbf{A}_N(B)$

$$\mathbf{A}_N(B) \mathbf{Z}_{NT}(t) = \mathbf{A}_{N,t-1} \mathbf{Z}_{NT}(1) + \dots + \mathbf{A}_{N,T-t} \mathbf{Z}_{NT}(T).$$

Since T is fixed, it therefore suffices to show that the cross-sectional averages $\mathbf{A}_N \mathbf{Z}_{NT}(t)$ converge to zero in mean square for all sequences of vectors \mathbf{A}_N with $\|\mathbf{A}_N\| \rightarrow 0$ as $N \rightarrow \infty$.

The assumptions on the spectral matrix $\Sigma_N^Z(u, \omega)$ imply that for fixed $u \in [0, 1]$ the stationary process $\tilde{\mathbf{Z}}_{NT}(u, t)$, $t \in \mathbb{Z}$, given by

$$\tilde{\mathbf{Z}}_{NT}(u, t) = \Upsilon_N(u, B) \boldsymbol{\varepsilon}_N(t)$$

is idiosyncratic according to the original definition of Forni and Lippi (2001). In particular, it follows that

$$\mathbf{A}_N \mathbf{Z}_{NT}(t) = \mathbf{A}_N \Upsilon_N\left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t)$$

converges to zero in mean square for all \mathbf{A}_N with $\|\mathbf{A}_N\| \rightarrow 0$ as $N \rightarrow \infty$. \square

B.3. Proof of Proposition 3

From (3.1) and (3.18) we have the following decompositions

$$Y_{iT}(t) = X_{iT}(t) + Z_{iT}(t) = X_{iT,N}(t) + Z_{iT,N}(t), \quad (\text{B.1})$$

which gives for the approximation error

$$\|X_{iT}(t) - X_{iT,N}(t)\|^2 = \|Z_{iT}(t) - Z_{iT,N}(t)\|^2. \quad (\text{B.2})$$

Recalling the definition of the approximate common component

$$X_{iT,N}(t) = (\boldsymbol{\phi}_{Ni} \star \boldsymbol{\Psi}_N)\left(\frac{t}{T}, B\right) \mathbf{F}(t) + (\boldsymbol{\phi}_{Ni} \star \boldsymbol{\Upsilon}_N)\left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t),$$

as well as the definition of the filter $\boldsymbol{\phi}$ in (3.20),

$$\boldsymbol{\phi}_{Ni}\left(\frac{t}{T}, B\right) = [\mathbf{P}'_{Ni} \star (\mathbf{Q}_N^r \mathbf{P}_N)]\left(\frac{t}{T}, B\right),$$

we find

$$\begin{aligned} \mathbb{E} \|(\boldsymbol{\phi}_{Ni} \star \boldsymbol{\Upsilon}_N)\left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t)\|^2 &= \mathbb{E} \|r_{iT}(t)\|^2 + \mathbb{E} \|Z_{iT}(t) - Z_{iT,N}(t)\|^2 \\ &\quad + \mathbb{E} [r_{iT}(t)(Z_{iT}(t) - Z_{iT,N}(t))], \end{aligned} \quad (\text{B.3})$$

where we have set $r_{iT}(t) = X_{iT}(t) - (\phi_{Ni} \star \Psi_{Ni})(\frac{t}{T}, B) \mathbf{F}(t)$. In view of (B.2), the stated mean square convergence follows if we show that the term on the left hand side and the third term on the right hand side both converge to zero as N tends to infinity.

We start with the term on the left side.

Lemma 8. *For $N \in \mathbb{N}$, let $\varphi_N(u, B)$ be an N -dimensional two-sided square-summable filter, and assume that*

$$\lim_{N \rightarrow \infty} \sup_{u \in [0,1]} \int_{-\pi}^{\pi} \|\varphi_N(u, \omega)\|^2 d\omega = 0.$$

Then

$$\lim_{N \rightarrow \infty} \sup_{T \in \mathbb{N}} \sup_{1 \leq t \leq T} \mathbb{E} \left\| (\varphi_N \star \Upsilon_N) \left(\frac{t}{T}, B \right) \varepsilon_N(t) \right\|^2 = 0.$$

Proof. Since $\mathbf{Z}_N(t)$ has mean zero, it follows that

$$\begin{aligned} \sup_{T \in \mathbb{N}} \sup_{1 \leq t \leq T} \mathbb{E} \left\| (\varphi_N \star \Upsilon_N) \left(\frac{t}{T}, B \right) \varepsilon_N(t) \right\|^2 &\leq \sup_{u \in [0,1]} \int_{-\pi}^{\pi} \varphi_N(u, \omega) \Sigma_N^Z(u, \omega) \varphi_N(u, \omega)^* d\omega \\ &\leq \sup_{u \in [0,1]} \int_{-\pi}^{\pi} \lambda_{N1}^Z(u, \omega) \|\varphi_N(u, \omega)\|^2 d\omega. \end{aligned}$$

Assumption 3 (ii) gives the result. \square

Thus it suffices to show that the conditions of the lemma hold for $\varphi_N(u, \omega) = \phi_{Ni}(u, \omega)$, that is,

$$\lim_{N \rightarrow \infty} \sup_{u \in [0,1]} \int_{-\pi}^{\pi} \|\phi_{Ni}(u, \omega)\|^2 d\omega = 0. \quad (\text{B.4})$$

We observe that

$$\|\phi_{Ni}(u, \omega)\|^2 = \phi_{Ni}(u, \omega) \phi_{Ni}(u, \omega)^* = \sum_{j=1}^r |p_{Ni,j}(u, \omega)|^2. \quad (\text{B.5})$$

By (3.9), (3.10), and Assumption 2, the terms in the sum on the right hand side satisfy

$$|p_{Ni,j}(u, \omega)|^2 \leq \frac{\bar{\sigma}_i}{\inf_{v \in [0,1]} \lambda_{Nj}(v, \omega)}, \quad j = 1, \dots, r, \quad (\text{B.6})$$

uniformly in $u \in [0, 1]$ for almost all $\omega \in [-\pi, \pi]$. Since by Proposition 1 the denominator on the right hand side diverges as $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} \sup_{u \in [0,1]} |p_{Ni,j}(u, \omega)| = 0$$

for almost all $\omega \in [-\pi, \pi]$. Consequently, (B.5) converges to zero uniformly in $u \in [0, 1]$. As it is also bounded by 1, condition (B.4) now follows from application of Lebesgue's dominated convergence theorem.

Next, we have to show that the mixed term on the right hand side of (B.3) converges to zero. The proof is based on the next Lemma 9, for which we may assume without loss of generality that

$$\lambda_{Nj}(u, \omega) \geq 1 \quad \text{for all } j = 1, \dots, N, N \in \mathbb{N}, u \in [0, 1], \omega \in [-\pi, \pi]. \quad (\text{B.7})$$

Indeed, let $\mathbf{U}_N(t)$ be an N -dimensional orthonormal white noise process that is orthogonal to the factors $\mathbf{F}(t)$ and the idiosyncratic component $\mathbf{Z}_{NT}(t)$ at all leads and lags, and define

$$\tilde{\mathbf{Z}}_{NT}(t) = \mathbf{Z}_{NT}(t) + \mathbf{U}_N(t) = \mathbf{\Upsilon}_N\left(\frac{t}{T}, B\right)\boldsymbol{\varepsilon}_N(t) + \mathbf{U}_N(t) = \tilde{\mathbf{\Upsilon}}_N\left(\frac{t}{T}, B\right)\tilde{\boldsymbol{\varepsilon}}_N(t)$$

with $\tilde{\mathbf{\Upsilon}}_N(u, B) = (\mathbf{\Upsilon}_N(\frac{t}{T}, B), \mathbf{I}_N)$ and $\tilde{\boldsymbol{\varepsilon}}_N(t) = (\boldsymbol{\varepsilon}_N(t)', \mathbf{U}_N(t)')$. Then, setting

$$\tilde{\mathbf{Y}}_{NT}(t) = \mathbf{X}_{NT}(t) + \tilde{\mathbf{Z}}_{NT}(t), \quad (\text{B.8})$$

we find that the process $\tilde{\mathbf{Y}}_{NT}(t)$ in (B.8) fulfills Assumptions 1, 2, and 3 with

$$\boldsymbol{\Sigma}_{\tilde{\mathbf{Z}}_N}(u, \omega) = \boldsymbol{\Sigma}_{\mathbf{Z}_N}(u, \omega) + \mathbf{I}_N \quad \text{and} \quad \boldsymbol{\Sigma}_{\tilde{\mathbf{Y}}_N}(u, \omega) = \boldsymbol{\Sigma}_N(u, \omega) + \mathbf{I}_N$$

and thus $\lambda_{\tilde{\mathbf{Z}}_N}^j(u, \omega) = \lambda_{\mathbf{Z}_N}^j(u, \omega) + 1$ and $\lambda_{\tilde{\mathbf{Y}}_N}^j(u, \omega) = \lambda_{N_j}(u, \omega) + 1$. Furthermore, $\mathbf{P}_{N_j}^{\tilde{\mathbf{Y}}} = \mathbf{P}_{N_j}$ and thus also $\boldsymbol{\phi}_{N_j}^{\tilde{\mathbf{Y}}} = \boldsymbol{\phi}_{N_j}$ for all $1 \leq j \leq N$ and $N \in \mathbb{N}$. Consequently, if Theorem 3 holds for the process $\tilde{\mathbf{Y}}_{NT}(t)$, that is,

$$\mathbb{E} \left\| (\boldsymbol{\phi}_{N_i} \star \boldsymbol{\Psi}_N)\left(\frac{t}{T}, B\right) \mathbf{F}(t) + (\boldsymbol{\phi}_{N_i} \star \tilde{\mathbf{\Upsilon}})\left(\frac{t}{T}, B\right) \tilde{\boldsymbol{\varepsilon}}_N(t) \right\|^2 \rightarrow 0 \quad \text{as } N \rightarrow \infty,$$

then the same result also holds for the original process $\mathbf{Y}_{NT}(t)$ since the process $\mathbf{U}_N(t)$ is orthogonal to the factors and idiosyncratic components and satisfies by Lemma 8

$$\mathbb{E} \left\| \boldsymbol{\phi}_{N_i}\left(\frac{t}{T}, B\right) \mathbf{U}_N(t) \right\|^2 \rightarrow 0$$

as N tends to infinity.

Under condition (B.7), the function $\mu_{N_j}(u, \omega) = [\lambda_{N_j}(u, \omega)]^{-1/2}$ is well defined for all $\omega \in [-\pi, \pi]$ and is bounded. Therefore it has a mean-square convergent Fourier representation. Denote by $\mu_{N_j}(u, B)$ the corresponding square-summable filter and for a fixed value of $u \in (0, 1)$ consider the vector process $\{\mathbf{W}_N(u, s), s \in \mathbb{Z}\}$ of the first r normalized principal components

$$\begin{aligned} \mathbf{W}_N(u, s) &= (W_{N_1}(u, s), \dots, W_{N_r}(u, s))' \\ &= (\mathbf{M}_N^r \star \mathbf{P}_N^r \star \boldsymbol{\Psi}_N)(u, B) \mathbf{F}(s) + (\mathbf{M}_N^r \star \mathbf{P}_N^r \star \mathbf{\Upsilon}_N)(u, B) \boldsymbol{\varepsilon}_N(s) \\ &= \mathbf{M}_N^r(u, B) \mathbf{F}_N^r(u, s), \end{aligned}$$

where $\mathbf{F}_N^r(u, s)$ is the r -dimensional vector containing the first r elements of the vector $\mathbf{F}_N(u, s)$ defined in (3.15),

$$\begin{aligned} \mathbf{M}_N^r(u, \omega) &= \text{diag} \{ \mu_{N_1}(u, \omega), \dots, \mu_{N_r}(u, \omega) \} = \left[\boldsymbol{\Lambda}_N^r(u, \omega) \right]^{-\frac{1}{2}} \\ \boldsymbol{\Lambda}_N^r(u, \omega) &= \text{diag} \{ \lambda_{N_1}(u, \omega), \dots, \lambda_{N_r}(u, \omega) \} \end{aligned} \quad (\text{B.9})$$

are r -dimensional diagonal matrices and $\mathbf{P}_N^r(u, \omega) = \left[\mathbf{P}_{N_1}(u, \omega)', \dots, \mathbf{P}_{N_r}(u, \omega)' \right]'$ is the $r \times N$ matrix containing the first r rows of the $N \times N$ matrix $\mathbf{P}_N(u, \omega)$. Using

(3.6), (3.9), and (B.9), we find

$$\begin{aligned}
& \mathbf{M}_N^r(u, \omega) \left[\mathbf{P}_N^r(u, \omega) \boldsymbol{\Sigma}_N^X(u, \omega) \mathbf{P}_N^r(u, \omega)^* \right] \mathbf{M}_N^r(u, \omega)^* \\
& + \mathbf{M}_N^r(u, \omega) \left[\mathbf{P}_N^r(u, \omega) \boldsymbol{\Sigma}_N^Z(u, \omega) \mathbf{P}_N^r(u, \omega)^* \right] \mathbf{M}_N^r(u, \omega)^* \\
& = \mathbf{M}_N^r(u, \omega) \left[\mathbf{P}_N^r(u, \omega) \boldsymbol{\Sigma}_N(u, \omega) \mathbf{P}_N^r(u, \omega)^* \right] \mathbf{M}_N^r(u, \omega)^* \\
& = \mathbf{M}_N^r(u, \omega) \boldsymbol{\Lambda}_N^r(u, \omega) \mathbf{M}_N^r(u, \omega)^* \\
& = \mathbf{I}_r,
\end{aligned}$$

Thus the process $\mathbf{W}_{NT}(u, s)$ is an orthonormal r -dimensional white noise since its spectrum is the identity matrix for all u and for all ω .

Remark 2. Analogously to the (rescaled) time-varying principal components process $\mathbf{F}_N(u, s)$ defined in (3.15), the process $\mathbf{W}_{NT}(u, s)$ fixes the dynamics at rescaled time point u and generates the whole process for all $s \in \mathbb{Z}$.

The following Lemma shows that the space spanned by the normalized principal components $\mathbf{W}_N(u, s)$, which is identical to the space spanned by the principal components $\mathbf{F}_N(u, s)$ themselves, i.e. $\mathcal{F}_N^r(u)$ defined in (3.16), converges to the space spanned by the true factors, \mathcal{F} .

Lemma 9. Consider the orthogonal projection of $\mathbf{W}_N(u, s)$ on the factor space \mathcal{F} spanned by the factors $F_j(s)$, $j = 1, \dots, r$, $s \in \mathbb{Z}$,

$$\mathbf{W}_N(u, s) = \mathbf{G}_N(u, B) \mathbf{F}(s) + \mathbf{R}_N(u, s) \quad (\text{B.10})$$

and the orthogonal projection of $\mathbf{F}(s)$ on the space spanned by the normalized principal components

$$\mathbf{F}(s) = \mathbf{G}_N(u, B^{-1})^* \mathbf{W}_N(u, s) + \mathbf{S}_N(u, s), \quad (\text{B.11})$$

where $\mathbf{G}_N(u, B) = (\mathbf{M}_N \star \mathbf{P}_N^r \star \boldsymbol{\Psi}_N)(u, B)$ is an $r \times r$ two-sided square-summable filter, and where $\mathbf{R}_N(u, s) = (\mathbf{M}_N \star \mathbf{P}_N^r \star \boldsymbol{\Upsilon}_N)(u, B) \boldsymbol{\varepsilon}_N(s)$ is orthogonal to \mathcal{F} . Then, as $N \rightarrow \infty$,

$$(i) \sup_{u \in [0,1]} \mathbb{E} \|\mathbf{R}_N(u, s)\|^2 = o(1);$$

$$(ii) \sup_{u \in [0,1]} \mathbb{E} \|\mathbf{S}_N(u, s)\|^2 = o(1).$$

Proof. Noting that

$$\mathbf{W}_{Nj}(u, s) = \left(\mu_{Nj} \star \mathbf{P}_{Nj} \star \boldsymbol{\Psi}_N \right)(u, B) \mathbf{F}(s) + \left(\mu_{Nj} \star \mathbf{P}_{Nj} \star \boldsymbol{\Upsilon}_N \right)(u, B) \boldsymbol{\varepsilon}_N(s) \quad (\text{B.12})$$

we find for the time-varying spectral density of the j -th element of $\mathbf{R}_N(u, s)$, $j = 1, \dots, r$,

$$\begin{aligned}
\sigma_{Nj}^R(u, \omega) &= \mu_{Nj}^2(u, \omega) \mathbf{P}_{Nj}(u, \omega) \boldsymbol{\Sigma}_N^Z(\omega) \mathbf{P}_{Nj}(u, \omega)^* \\
&\leq \bar{\lambda}_Z \mu_{Nj}^2(u, \omega) \left| \mathbf{P}_{Nj}(u, \omega) \right|^2 \leq \bar{\lambda}_Z \lambda_{Nj}^{-1}(u, \omega).
\end{aligned}$$

It follows from Proposition 1 that $\sup_{u \in [0,1]} \sigma_{Nj}^R(u, \omega) \rightarrow 0$ for all ω as $N \rightarrow \infty$. Since by (B.7) the supremum is also bounded by $\bar{\lambda}_Z$, we get by Lebesgue's dominated

convergence theorem

$$\sup_{u \in [0,1]} \int_{-\pi}^{\pi} \sigma_{N_j}^R(u, \omega) d\omega \rightarrow 0$$

and the proof of (i) is completed. For the second part, we note that the spectral density of $\mathbf{W}_N(u, s)$ as well as that of $\mathbf{F}(s)$ is the identity matrix for all u and for all s and thus, by (B.10) and (B.11) we have

$$\begin{aligned} \mathbf{I}_r &= \mathbf{G}_N(u, \omega) \mathbf{G}_N(u, \omega)^* + \boldsymbol{\Sigma}_N^R(u, \omega) \\ &= \mathbf{G}_N(u, \omega)^* \mathbf{G}_N(u, \omega) + \boldsymbol{\Sigma}_N^S(u, \omega) \end{aligned}$$

for all $u \in [0, 1]$ and $\omega \in [-\pi, \pi]$, where $\mathbf{G}_N(u, \omega)$ is the transfer function of the filter $\mathbf{G}_N(u, B)$, and $\boldsymbol{\Sigma}_N^R(u, s)$ and $\boldsymbol{\Sigma}_N^S(u, s)$ are the spectral densities of $\mathbf{R}_N(u, s)$ and $\mathbf{S}_N(u, s)$, respectively. This implies $\text{tr}(\boldsymbol{\Sigma}_N^S(u, \omega)) = \text{tr}(\boldsymbol{\Sigma}_N^R(u, \omega))$, and (ii) is proved. \square

In order to show that $\mathbb{E}[r_{iT}(t)(Z_{iT}(t) - Z_{iT,N}(t))]$ tends to zero, it suffices to show that the cross-spectrum between $r_{iT}(t)$ and $Z_{iT,N}(t)$, $f_{N_i}^{12}(u, \omega)$ say, tends to zero uniformly in $u \in [0, 1]$ and almost everywhere in $[-\pi, \pi]$ since $Z_{iT}(t)$ is orthogonal to $\mathbf{F}(t)$ and thus to $r_{iT}(t)$ at all leads and lags. Since $|f_{N_i}^{12}(u, \omega)|$ is bounded by an integrable function, Lebesgue's dominated convergence theorem implies that $\int_{-\pi}^{\pi} f_{N_i}^{12}(u, \omega) d\omega$ converges to zero uniformly in $u \in [0, 1]$, which proves the desired mean square convergence.

Recall that

$$\begin{aligned} Z_{iT,N}(t) &= (\mathbf{P}_{N_i}^* \star [(\mathbf{I}_N - \mathbf{Q}_{N_i}^r) \mathbf{P}_N] \star \boldsymbol{\Psi}_N) \left(\frac{t}{T}, B\right) \mathbf{F}(t) \\ &\quad + (\mathbf{P}_{N_i}^* \star [(\mathbf{I}_N - \mathbf{Q}_{N_i}^r) \mathbf{P}_N] \star \boldsymbol{\Upsilon}_{N_i}) \left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t). \end{aligned}$$

First we study the cross spectrum between $Z_{iT,N}(t)$ and $X_{iT}(t)$, and then we study the cross spectrum between $Z_{iT,N}(t)$ and $(\boldsymbol{\phi}_{N_i} \star \boldsymbol{\Psi}_N) \left(\frac{t}{T}, B\right) \mathbf{F}(t)$.

The cross spectrum between $Z_{iT,N}(t)$ and $X_{iT}(t)$ is equal to the cross spectrum between $Z_{iT,N}(t)$ and $\boldsymbol{\psi}_i \left(\frac{t}{T}, B\right) \mathbf{S}_N \left(\frac{t}{T}, t\right)$ because, by (B.11), $X_{iT}(t)$ can be written as

$$X_{iT}(t) = \boldsymbol{\psi}_i \left(\frac{t}{T}, B\right) \mathbf{F}(t) = \boldsymbol{\psi}_i \left(\frac{t}{T}, B\right) \mathbf{G} \left(\frac{t}{T}, B^{-1}\right)^* \mathbf{W}_N \left(\frac{t}{T}, t\right) + \boldsymbol{\psi}_i \left(\frac{t}{T}, B\right) \mathbf{S}_N \left(\frac{t}{T}, t\right).$$

As explained in section 3.3, $Z_{iT,N}(t)$ is orthogonal to the terms

$$F_{N_j} \left(\frac{t}{T}, t\right) = \left[(\mathbf{P}_{N_j} \star \boldsymbol{\Psi}_N) \left(\frac{t}{T}, B\right) \mathbf{F}(t) + (\mathbf{P}_{N_j} \star \boldsymbol{\Upsilon}_N) \left(\frac{t}{T}, B\right) \boldsymbol{\varepsilon}_N(t) \right],$$

for $j = 1, \dots, r$ at any lead and lag; then by (B.12) it is also orthogonal at any lead and lag to the terms $W_{N_j} \left(\frac{t}{T}, t\right)$. Let $g_{N_i}^{12} \left(\frac{t}{T}, \omega\right)$ be the time-varying cross-spectrum between $Z_{iT,N}(t)$ and $\boldsymbol{\psi}_i \left(\frac{t}{T}, B\right) \mathbf{S}_N \left(\frac{t}{T}, t\right)$, and let $\sigma_{i,N}^Z \left(\frac{t}{T}, \omega\right)$ be the time-varying spectrum of $Z_{iT,N}(t)$. We have

$$\left| g_{N_i}^{12}(u, \omega) \right|^2 \leq \sigma_{i,N}^Z(u, \omega) \boldsymbol{\psi}_i(u, \omega) \boldsymbol{\Sigma}_N^S(u, \omega) \boldsymbol{\psi}_i(u, \omega)^*.$$

By (3.7) $\sigma_{i,N}^Z(u, \omega) \leq \sigma_i(u, \omega) \leq C_{\boldsymbol{\Sigma}, N}$ uniformly in u and ω . Furthermore,

$$\sup_{u \in [0,1]} \left\| \boldsymbol{\psi}_i(u, \omega) \boldsymbol{\Sigma}_N^S(u, \omega) \boldsymbol{\psi}_i(u, \omega)^* \right\| \leq \sup_{u \in [0,1]} \left\| \boldsymbol{\psi}_i(u, \omega) \right\|^2 \sup_{u \in [0,1]} \left\| \boldsymbol{\Sigma}_N^S(u, \omega) \right\|,$$

which converges to 0 a.e. in $[-\pi, \pi]$ by the assumption on the coefficients $\boldsymbol{\psi}_i$ in Assumption 1 (ii) and by Lemma 9.

By the same arguments, the time-varying cross-spectrum between $Z_{iT,N}(t)$ and $(\phi_{Ni} \star \Psi_N)\left(\frac{t}{T}, B\right) \mathbf{F}(t)$ equals that between $Z_{iT,N}(t)$ and $(\phi_{Ni} \star \Psi_N)\left(\frac{t}{T}, B\right) \mathbf{S}_N\left(\frac{t}{T}, t\right)$. Denoting it by $h_{Ni}^{12}(u, \omega)$, we have

$$\begin{aligned} |h_{Ni}^{12}(u, \omega)|^2 &= \sigma_{i,N}^Z \phi_{Ni}(u, \omega) \Psi_N(u, \omega) \Sigma_N^S(u, \omega) \Psi_N(u, \omega)^* \phi_{Ni}(u, \omega)^* \\ &\leq \sigma_{i,N}^Z \lambda_{N1}^S(u, \omega) \phi_{Ni}(u, \omega) \Psi_N(u, \omega) \Psi_N(u, \omega)^* \phi_{Ni}(u, \omega)^*. \end{aligned}$$

Since $\Sigma_N^X(u, \omega) = \Psi_N(u, \omega) \Psi_N(u, \omega)^*$ and $\Sigma_N(u, \omega) = \Sigma_N^X(u, \omega) + \Sigma_N^Z(\omega)$, we obtain

$$\begin{aligned} |h_{Ni}^{12}(u, \omega)|^2 &\leq \sigma_{i,N}^Z \lambda_{N1}^S(u, \omega) \phi_{Ni}(u, \omega) \Sigma_N(u, \omega) \phi_{Ni}(u, \omega)^* \\ &= \sigma_{i,N}^Z \lambda_{N1}^S(u, \omega) \sum_{j=1}^r |p_{Ni,j}(u, \omega)|^2 \lambda_{Nj}(u, \omega). \end{aligned}$$

By Lemma 9, $\lambda_{N1}^S(u, \omega)$ tends to zero for almost all $\omega \in [-\pi, \pi]$ uniformly in $u \in [0, 1]$. Furthermore, by (B.6), the sum is bounded by the constant $\bar{\sigma}_i$. Thus $\sup_{u \in [0, 1]} |h_{Ni}^{12}(u, \omega)|$ tends to zero almost everywhere, and the proof of Proposition 3 is complete. \square

B.4. Proof of Proposition 5

We need to show that for all $N \in \mathbb{N}$ the estimation error

$$\widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t) = [\widehat{\phi}_{NT,i}\left(\frac{t}{T}, B\right) - \phi_{Ni}\left(\frac{t}{T}, B\right)] \mathbf{Y}_{NT}(t)$$

satisfies

$$\lim_{T \rightarrow \infty} \sup_{Tu_T < t < T - Tu_T} \mathbb{P}[|\widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t)| > \varepsilon] = 0 \quad (\text{B.13})$$

for all $\varepsilon > 0$. Recall that

$$\phi_{Ni}\left(\frac{t}{T}, B\right) = \sum_{h=-M_T}^{M_T} \phi_{Ni,h}\left(\frac{t}{T}\right) B^h$$

and

$$\widehat{\phi}_{NT,i}\left(\frac{t}{T}, B\right) = \sum_{h=-M_T}^{M_T} \widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) B^h.$$

By Chebyshev's inequality we obtain for the probability in (B.13)

$$\begin{aligned} \mathbb{P}[|\widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t)| > \varepsilon] &= \mathbb{P}\left[\left|\sum_{h=-M_T}^{M_T} \Delta \widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) B^h \mathbf{Y}_{NT}(t)\right| > \varepsilon\right] \\ &\leq \frac{1}{\varepsilon^2} \mathbb{E}\left[\left|\sum_{h=-M_T}^{M_T} \Delta \widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) B^h \mathbf{Y}_{NT}(t)\right|^2\right] \end{aligned}$$

where we have set $\Delta \widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) = \left(\widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) - \phi_{Ni,h}\left(\frac{t}{T}\right)\right)$. Analogously, let in the sequel, $\Delta \widehat{\phi}_{Ni}(u, \omega) = \left(\widehat{\phi}_{Ni}(u, \omega) - \phi_{Ni}(u, \omega)\right)$. In order to bound the expectation on the right hand side,

$$\mathbb{E}\left[\left|\sum_{h=-M_T}^{M_T} \Delta \widehat{\phi}_{Ni,h}\left(\frac{t}{T}\right) B^h \mathbf{Y}_{NT}(t)\right|^2\right] = \mathbb{E}\left[\left|\sum_{m=1}^N \sum_{k=-M_T}^{M_T} \Delta \widehat{\phi}_{Ni,m,k}\left(\frac{t}{T}\right) Y_{mT}(t-k)\right|^2\right],$$

we apply Cauchy's inequality on $\sum_k \Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}) Y_{mT}(t-k)$ to get as upper bound

$$\mathbb{E} \left[\left| \sum_{m=1}^N \left(\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \sum_{k=-M_T}^{M_T} (Y_{mT}(t-k))^2 \right)^{1/2} \right|^2 \right],$$

which then can simply be bounded by

$$N^2 \sup_{1 \leq m \leq N} \mathbb{E} \left[\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \sum_{k=-M_T}^{M_T} (Y_{mT}(t-k))^2 \right]. \quad (\text{B.14})$$

To bound, for fixed N , the expectation we apply again Cauchy's inequality and will show that

$$\mathbb{E} \left[\left(\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \right)^2 \right] = O((r_T)^{-2}) \quad (\text{B.15})$$

and that

$$\mathbb{E} \left[\left(\sum_{k=-M_T}^{M_T} (Y_{mT}(t-k))^2 \right)^2 \right] = O(M_T^2), \quad (\text{B.16})$$

which is sufficient to finish the proof because by assumption, $M_T/r_T \rightarrow 0$ as $T \rightarrow \infty$.

We will later show that Assumption 4 implies the following uniform rate of convergence of the estimated filters $\widehat{\phi}_{Ni}(u, \omega)$ to the true filters $\phi_{Ni}(u, \omega)$: for fixed $N \in \mathbb{N}$

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \|\Delta \widehat{\phi}_{Ni}(u, \omega)\| = O_p((r_T)^{-1}), \quad (\text{B.17})$$

and, in particular, for each $1 \leq m \leq N$,

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \mathbb{E} \left[\Delta \widehat{\phi}_{Ni,m}(u, \omega) \right] = O((r_T)^{-1}), \quad (\text{B.18})$$

and

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \text{Var} (\Delta \widehat{\phi}_{Ni,m}(u, \omega)) = O((r_T)^{-2}). \quad (\text{B.19})$$

Further note that

$$\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \leq \int |\Delta \widehat{\phi}_{Ni,m}(t/T, \omega)|^2 d\omega,$$

and that this random variable is stochastically bounded from above because $\Delta \widehat{\phi}_{Nim,k}$ is a difference of (a function of) eigenvectors of empirical and true spectral matrix of fixed dimension N . Hence we obtain, using (B.18) and (B.19), that

$$\mathbb{E} \left[\left(\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \right)^2 \right] \leq C \mathbb{E} \left[\sum_{k=-M_T}^{M_T} (\Delta \widehat{\phi}_{Nim,k}(\frac{t}{T}))^2 \right] = O((r_T)^{-2}),$$

and hence (B.15).

Next, the expectation in (B.16) is simply bounded by $C M_T^2 \mathbb{E}(Y_{mT}(t)^4)$, which is of order $O(M_T^2)$ since $\mathbf{Y}_{NT}(t)$ has uniformly bounded fourth moments. Consequently, the upper bound in (B.14) converges to zero as $T \rightarrow \infty$, which completes the proof.

It remains to show that equations (B.17)-(B.19) are a consequence of Assumption 4. The proof parallels parts of Lemma 4.2 of Forni et al. (2004) which is essentially built on the following Taylor expansions to be found in the proof of

Brillinger (1981), Theorem 9.2.4. The idea is to derive rates of convergence of eigenvalues $\widehat{\lambda}_j(u, \omega)$ and eigenvectors $\widehat{P}_j(u, \omega)$ of $\widehat{\Sigma}_{NT}(u, \omega)$ to the limiting ones $\lambda_j(u, \omega)$ and $P_j(u, \omega)$ of $\Sigma_N(u, \omega)$ by relating their difference in first order to the difference $\widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)$. (Note that this is done for fixed N , so most of the double asymptotic considerations of Lemma 4.2 of Forni et al. (2004) can be ignored.)

$$\widehat{\lambda}_j(u, \omega) - \lambda_j(u, \omega) = P_j(u, \omega)^* (\widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)) P_j(u, \omega) + \dots$$

and

$$\widehat{P}_j(u, \omega) - P_j(u, \omega) = \sum_{i \neq j} [P_i(u, \omega)^* (\widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)) P_j(u, \omega)] P_i(u, \omega) / (\lambda_j(u, \omega) - \lambda_i(u, \omega)) + \dots$$

(In both expansions the \dots denote terms of higher order in the difference of $\widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)$.) For details on the derivation of these Taylor expansions see Wilkinson (1965), Ch. 2, but note that essentially they can be derived from a first order Taylor expansion of the "symbolic" eigenvalue-eigenvector equation $Z f_j = \Psi_j f_j$ about $Z = \Sigma_N$ which gives:

$$(\widehat{\Sigma}_{NT} - \Sigma_N) P_j + \Sigma_N (\widehat{P}_j - P_j) = (\widehat{\lambda}_j - \lambda_j) P_j + \lambda_j (\widehat{P}_j - P_j),$$

and after multiplication of this equation from left by any other P_i^* , with $i \neq j$, a system of equations of the type

$$P_i^* (\widehat{\Sigma}_{NT} - \Sigma_N) P_j = (\lambda_j - \lambda_i) P_i^* (\widehat{P}_j - P_j).$$

Note that in the cited sources, no multiplicity of the eigenvalues has been considered. However, there exists some techniques to group those multiple eigenvalue-eigenvectors pairs into blocks and repeat the same technique over blocks. To avoid lengthy technicalities we just refer to, e.g., Anderson (1963), and, more generally, to Wilkinson (1965, Chap. 2), which essentially suggests that a result of the same type as in (B.17) can be derived, with a rate of consistency which is possibly only a fractional power of the rate $(r_T)^{-1}$ derived for the case of simple eigenvalues. However, note that any arbitrarily slow rate will be sufficient for the purpose of our proof.

To finally give some insights into the principles of this proof, details of which we omit here, we note that as in Forni et al. (2004), Lemma 4.2, we proceed via three steps, which in our situation, again, need to be studied only for a fixed N . W.l.o.g., we formulate a sketch of these constructive steps for the case of using one of our proposed spectral estimator, the smoothed preperiodogram defined in (A.1), fulfilling Assumption 4, with a rate $r_T = (h_T b_T T)^{1/2}$ delivered by Theorem 7.

- (i) controlling the "smoothing" bias: Let $\Sigma_{NT}^{cc}(u, \omega) := K_h^T K_b^\Phi * \Sigma_N(u, \omega)$ denote the convolution of Σ_N with the product kernel in time $K_h^T(u)$ and frequency $K_b^\Phi(\omega)$ (see, e.g., also (A.1)), and let $P_{j,T}^{cc}(u, \omega)$ denote its j -th eigenvector. Then we have (paralleling Forni et al. (2004), Lemma 4.2(ii)) that uniformly in u and ω , for all j ,

$$\|P_{j,T}^{cc}(u, \omega) - P_j(u, \omega)\| = O(h_T^2) + O(b_T^2),$$

a result that is based on the classical Taylor expansion of order 2 of the kernel convolution (in time and in frequency).

(ii) controlling the variance, uniformly in u and ω , for all j ,

$$\text{Var}(\widehat{P}_j(u, \omega)) = O((h_T b_T T)^{-1}),$$

essentially paralleling Forni et al. (2004), Lemma 4.2(iii) and equation (9.4.17) in Brillinger (1981). Obviously, here we use that the asymptotic variance of $\widehat{\Sigma}_{NT} - \Sigma_{NT}^{cc}$ is of the same given order, see (A.6).

(iii) controlling uniformly the bias

$$\|\mathbb{E}\widehat{P}_j(u, \omega) - P_{j,T}^{cc}(u, \omega)\| = O((h_T b_T T)^{-1}),$$

because it is essentially bounded by the bias of $\mathbb{E}\widehat{\Sigma}_{NT} - \Sigma_{NT}^{cc}$ which is also of order $O((h_T b_T T)^{-1})$ (paralleling Brillinger (1981), Theorem 7.4.1 and the proof of Forni et al. (2004), Lemma 4.2(iv)).

Note that, in order to have these results on controlling separately bias and variance of $\widehat{P}_j - P_j$, uniformly over time u and frequency ω , we observe that the following conditions on the second (partial) derivatives of $\Sigma_N(u, \omega)$ with respect to ω and u are indeed fulfilled in our situation:

$$\sup_u \sup_\omega \|P_i(u, \omega)^*(d_u^2 \Sigma_N(u, \omega))P_j(u, \omega) P_i(u, \omega)\| \leq C_u(N)$$

and

$$\sup_u \sup_\omega \|P_i(u, \omega)^*(d_\omega^2 \Sigma_N(u, \omega))P_j(u, \omega) P_i(u, \omega)\| \leq C_\omega(N)$$

with constants $C_u(N)$ and $C_\omega(N)$ which remain bounded over T , for each fixed N .

For further details on this proof we refer to the given references to avoid lengthy technicalities. \square

REFERENCES

- Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics* **34**, 122–148.
- Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica* **71**, 135–171.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica* **70**(1), 191–221.
- Brillinger, D. R. (1981). *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, New York.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* **51**, 1281–1304.
- Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In P. M. Robinson and M. Rosenblatt (eds), *Athens Conference on Applied Probability and Time Series Analysis*, Vol. II, Springer-Verlag, New York.
- Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics* **25**, 1–37.
- Dahlhaus, R. (2000). A likelihood approximation for locally stationary processes. *The Annals of Statistics* **28**(6), 1762–1794.
- Dahlhaus, R. (2008). Uniform convergence of a spectral density estimate for locally stationary processes. Unpublished Manuscript.

- Del Negro, M. and Otrok, C. (2008). Dynamic factor models with time-varying parameters: Measuring changes in international business cycles. *Staff Report 326*, Federal Reserve Bank of New York.
- Fancourt, C. L. and Principe, J. C. (1998). Competitive principal components analysis for locally stationary time series. *IEEE Transactions on Signal Processing* **46**, 3068–81.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics* **82**, 540–554.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2004). The generalized dynamic factor model: Consistency and rates. *Journal of Econometrics* **119**, 231–255.
- Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* **100**(471), 830–840.
- Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* **17**, 1113–1141.
- Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* **102**(478), 603–617.
- Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons.
- Molenaar, P. C. M., de Gooijer, J. G. and Schmitz, B. (1992). Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika* **57**, 333–349.
- Motta, G., Hafner, C. and von Sachs, R. (2006). Locally stationary factor models: Identification and nonparametric estimation. *Discussion Paper 0624*, Institut de Statistique, UCL.
- Neumann, M. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *The Annals of Statistics* **25**(1), 38–76.
- Pan, J. and Yao, Q. (2008). Modelling multiple time series via common factors. *Biometrika* **95**, 365–379.
- Stock, J. H. and Watson, M. W. (2008). Forecasting in dynamic factor models subject to structural instability. In J. Castle and N. Shephard (eds), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, Oxford University Press, Oxford.
- Wilkinson, J. H. (1965). *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.

DEPARTMENT OF QUANTITATIVE ECONOMICS, UNIVERSITY OF MAASTRICHT, P.O.BOX 616, 6200 MD MAASTRICHT, THE NETHERLANDS

INSTITUT DE STATISTIQUE, UNIVERSITÉ CATHOLIQUE DE LOUVAIN, VOIE DU ROMAN PAYS 20, B-1348 LOUVAIN-LA-NEUVE, BELGIUM