



Fitting dynamic factor models to non-stationary time series

Michael Eichler^{a,*}, Giovanni Motta^a, Rainer von Sachs^b

^a Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

^b Institut de statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium

ARTICLE INFO

Article history:

Available online 16 November 2010

JEL classification:

C14

C32

Keywords:

Approximate factor models

Local stationarity

Principal components

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 generalize the 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 a 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. The performance of the estimators is illustrated by a simulation study and an application to a macroeconomic data set.

© 2010 Elsevier B.V. All rights reserved.

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 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 really high-dimensional data sets available nowadays 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 lets the filters become (smooth) functions of time.

With this approach, we provide for a true generalization of two classes of existing factor models in the literature: static factors with time-varying loadings recently developed by Motta et al. (2011), 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

* Corresponding author.

E-mail address: m.eichler@maastrichtuniversity.nl (M. Eichler).

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 optimization algorithm on expanding the “white noise space”: this allows for a generalization 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 formalize 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 behaviour 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 behaviour 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 behaviour 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 behaviour 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 $\hat{\mathbf{X}}_N(t)$ to the common components $\mathbf{X}_N(t)$ of model (1.1), where the estimator $\hat{\mathbf{X}}_N(t)$ is based on the $N \times r$ matrix containing the eigenvectors corresponding to the largest r eigenvalues of the sample covariance matrix of $\mathbf{Y}_N(t)$.

If the true data generating process is dynamic, that is, 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

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

Here, the one-sided filters $\mathbf{\Psi}_{ij}(B)$ – with B being the back-shift operator – have square summable coefficients and the factors $\mathbf{F}(t)$ are orthonormal white noise. Again, no mutual orthogonality of the idiosyncratic components is imposed, 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. This approach requires consistent estimation of the spectral density matrix of the data-generating process. For the statistical properties of dynamic principal components, we also refer to Brillinger (1981).

In this paper now, we generalize the factor model in (1.1) further by allowing the second-order structure of the dynamic factor model to change smoothly over time. More precisely, we consider processes of the form

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

where the loadings $\mathbf{\Psi}_N(t, B)$ – dynamic filters as in the model of Forni et al. (2000) – depend now on time. Additionally, the idiosyncratic components $\mathbf{Z}_N(t)$ are also allowed to be non-stationary with time-varying dynamics. Consequently the spectral density matrix $\mathbf{\Sigma}_N(t, \omega)$ of the process $\mathbf{Y}_N$ becomes dependent on time, and in order to estimate it, we use a time-localized empirical spectrum (a localized periodogram over time). Our contribution is the practical and theoretical treatment of this localized 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 $\mathbf{\Sigma}_N$ 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). We note that our approach is also related to the time-varying static approach of Motta et al. (2011), which generalizes the static factor model of Bai (2003).

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 in Section 5 we compare the application of different factor models to macroeconomic data. Section 6 concludes and mentions the main open problems that are left for future research.

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 , $\lfloor x \rfloor$ denotes the largest integer smaller or equal to x . For a matrix $\mathbf{A}$, the trace and the conjugate transpose are denoted by $\text{tr}(\mathbf{A})$ and $\mathbf{A}^*$, respectively, $\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}\mathbf{A}^*)}$ is the Euclidean norm, and $\mathbf{I}_n$ is the identity matrix of dimension n .

One important tool will be the convolution of time-varying linear filters. For a given point in time t , let $\Phi(t, B)$ and $\Psi(t, B)$

be two time-varying one-sided filters. Then the convolution $(\Phi \star \Psi)(t, B)$ of the two filters is defined by

$$(\Phi \star \Psi)(t, B) = \sum_{k=0}^{\infty} \left[\sum_{j=0}^{\infty} \Phi_j(t) \Psi_{k-j}(t) \right] B^k.$$

The obtained filter $(\Phi \star \Psi)(t, B)$ has the transfer function

$$(\Phi \star \Psi)(t, \omega) = \Phi(t, \omega) \Psi(t, \omega),$$

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

$$\begin{aligned} Y(t) &= X(t) + Z(t) = \Psi(t, B)F(t) + Z(t) \\ &= \sum_{k=0}^{\infty} \Psi_k(t)F(t) + Z(t) \end{aligned} \quad (2.1)$$

for $1 \leq t \leq T$, where $X(t)$ is the common component originating from the common factors $F(t)$ while $Z(t)$ is the idiosyncratic component. We assume that the model satisfies the following conditions:

- (i) $F(t)$ is an r -dimensional stationary white noise process with $\mathbb{E}(F(t)) = 0$ and $\mathbb{E}(F(t)F(t)') = \mathbf{I}_r$;
- (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) $Z(t) = (Z_i(t))_{i \in \mathbb{N}}$ is uncorrelated to the factor process $F(t)$.

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

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

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) $Z(t)$ is non-stationary and only defined over the period $1 \leq t \leq T$. For the above definition of idiosyncratic processes, we therefore set $Z(t) = 0$ for $t < 1$ or $t > T$.

The factor process $F(t)$ is assumed to be a stationary orthonormal white noise process as in Forni et al. (2000). This is, however, not a serious constraint. For example, suppose that $F(t)$ is a non-stationary factor process with time-varying moving average representation $F(t) = \Xi_T(t, B)\eta(t)$, where $\eta(t)$ is a stationary orthonormal white noise process and $\Xi_T(t, B)$ is a square-summable time-varying one-sided linear filter. Then the common component $X(t)$ can be rewritten as

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

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

For the idiosyncratic component $Z(t)$, we assume further that for any $N \in \mathbb{N}$ the subprocess $Z_N(t) = (Z_1(t), \dots, Z_N(t))'$ has a

time-varying moving average representation of the form

$$Z_N(t) = \Upsilon_N(t, B)\varepsilon_N(t),$$

where $\varepsilon_N(t)$ is a stationary white noise process with mean $\mathbb{E}(\varepsilon_N(t)) = 0$ and covariance matrix $\mathbb{E}(\varepsilon_N(t)\varepsilon_N(t)') = \mathbf{I}_N$. Then $Y_N(t)$ is obtained from the $r + N$ -dimensional stationary process $(F(t), \varepsilon_N(t))$ by application of the time-varying linear filter $(\Psi_N(t, B), \Upsilon_N(t, B))$. Under the additional assumption that the time-varying linear filters $\Psi_N(t, B)$ and $\Upsilon_N(t, B)$ are square summable, this leads to the time-varying spectral representation

$$Y_N(t) = \int_{-\pi}^{\pi} e^{i\omega t} \Psi_N(t, \omega) d\xi^F(\omega) + \int_{-\pi}^{\pi} e^{i\omega t} \Upsilon_N(t, \omega) d\xi_N^\varepsilon(\omega),$$

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

$$\Sigma_N(t, \omega) = \Psi_N(t, \omega)\Psi_N(t, \omega)^* + \Upsilon_N(t, \omega)\Upsilon_N(t, \omega)^*.$$

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 $\hat{\Sigma}_N(t, \omega)$ be a consistent estimator of the time-varying spectral matrix $\Sigma_N(t, \omega)$; for details on the estimation of time-varying spectral matrices we refer to Appendix A. Then $\hat{\Sigma}_N(t, \omega)$ has a spectral decomposition

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

where $\hat{\Lambda}_N(t, \omega) = \text{diag}(\hat{\lambda}_{N1}(t, \omega), \dots, \hat{\lambda}_{NN}(t, \omega))$ is the diagonal matrix containing the eigenvalues of $\hat{\Sigma}_N(t, \omega)$ in descending order of magnitude and $\hat{\mathbf{P}}_N(t, \omega)$ is the unitary $N \times N$ matrix whose i -th row is the row eigenvector of $\hat{\Sigma}_N(t, \omega)$ corresponding to the i -th eigenvalue $\hat{\lambda}_{Ni}(t, \omega)$. 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 $\hat{\Sigma}_N(t, \omega)$. In the frequency domain this is accomplished by the projection

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

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,

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

where the filter coefficients $\hat{\Phi}_{N,k}(t)$ are given by

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

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

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

be the truncated filter. Then the common components $X_N(t)$ are estimated by

$$\hat{X}_N(t) = \hat{\Phi}_{NT}(t, B) Y_N(t).$$

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. First, 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. Second, 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.

The above method for recovering the common component works only for the “central part” of the observation period since the two-sided filters use information from the past as well as the future of the process. In contrast, at or close to the boundaries of the observation period, the projection filters must be adapted appropriately to include only observed and hence past (respectively future) values of the series. This would lead to an approach similar to Forni et al. (2005), which will be investigated in a future research project.

Finally, we note that our approach does not work well for time series that exhibit certain types of non-stationary behaviour such as deterministic or stochastic trends even though these may be covered by the theoretical framework we have described. An integrated process, for example, indeed permits a time-varying moving average representation satisfying the above assumptions. The changes in the corresponding linear filters and thus in the dynamics of the process, however, are too fast and make a localized estimation of the time-varying spectrum infeasible. For a successful application of the method, the time series should show locally an approximately stationary behaviour. This is also the key to the consistency result in the next section.

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 $\Psi_N(t, B) = \Psi_N(\frac{t}{T}, B)$ and $\Upsilon_N(t, B) = \Upsilon_N(\frac{t}{T}, B)$. 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 estimators 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 general locally stationary processes.

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 $u \in [0, 1]$

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

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

- (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

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

and

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

where $\xi^{\mathbf{F}}(\omega)$ and $\xi_N^{\boldsymbol{\varepsilon}}(\omega)$ 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

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

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

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

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

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

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 the time-varying spectral matrix

$$\begin{aligned}\Sigma_N(u, \omega) &= \Sigma_N^X(u, \omega) + \Sigma_N^Z(u, \omega) \\ &= \Psi_N(u, \omega) \Psi_N(u, \omega)^* + \Upsilon_N(u, \omega) \Upsilon_N(u, \omega)^*.\end{aligned}\quad (3.2)$$

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 $\tilde{\mathbf{e}}_N(t)$. 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

$$\sup_{u \in [0, 1]} \sup_{\omega \in [-\pi, \pi]} \sigma_{ii}(u, \omega) \leq \bar{\sigma}_i. \quad (3.3)$$

In order to obtain sensible estimates for the 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 $\Sigma_N^X(u, \omega)$ and $\Sigma_N^Z(u, \omega)$.

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

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

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

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

The following proposition shows that, under these 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 1–3**, 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 $\bar{\lambda}_Z$, that is, for $j > r$

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

Proof. See Appendix B.1. $\square$

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 $\mathbf{A}_N(B)$ such that the sequences of integrals $\int_{-\pi}^{\pi} \|\mathbf{A}_N(\omega)\|^2 d\omega$ tend to zero as N tends to infinity, the filtered processes $\mathbf{A}_N(B)\mathbf{Z}_{NT}(t)$ 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. $\square$

3.2. Decomposition of the overall estimation error

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

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

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

$$\hat{\mathbf{X}}_{NT,N}(t) = \hat{\Phi}_{NT} \left(\frac{t}{T}, B \right) \mathbf{Y}_{NT}(t), \quad (3.4)$$

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

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

is the truncated time domain filter obtained from the transfer function

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

The main objective of this section is to prove the consistency of the common component estimator $\hat{\mathbf{X}}_{NT,N}(t)$. As in Forni et al. (2000), the proof of consistency is based on a decomposition of the overall estimation error $\|\hat{\mathbf{X}}_{NT,N}(t) - \mathbf{X}_{NT}(t)\|$. 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 errors.

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

$$\Sigma_N(u, \omega) = \mathbf{P}_N(u, \omega)^* \mathbf{\Lambda}_N(u, \omega) \mathbf{P}_N(u, \omega) \quad (3.5)$$

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

$$\mathbf{P}_{Nj}(u, \omega) = [p_{Nj,1}(u, \omega), p_{Nj,2}(u, \omega), \dots, p_{Nj,N}(u, \omega)] \quad (3.6)$$

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

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

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

$$\begin{aligned} \mathbf{X}_{NT,N}(t) &= \int_{-\pi}^{\pi} e^{i\omega t} \Phi_N\left(\frac{t}{T}, \omega\right) \Psi_N\left(\frac{t}{T}, \omega\right) d\xi^F(\omega) \\ &\quad + \int_{-\pi}^{\pi} e^{i\omega t} \Phi_N\left(\frac{t}{T}, \omega\right) \Upsilon_N\left(\frac{t}{T}, \omega\right) d\xi^g(\omega). \end{aligned}$$

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

$$(\mathbf{P}_N \star \Psi_N)(u, B) = \sum_{k=0}^{\infty} \left[\sum_{j=-\infty}^{\infty} \Phi_{Nj}(u) \Psi_{N,k-j}(u) \right] B^k;$$

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

$$\begin{aligned} \mathbf{X}_{NT,N}(t) &= (\Phi_N \star \Psi_N)\left(\frac{t}{T}, B\right) \mathbf{F}(t) \\ &\quad + (\Phi_N \star \Upsilon_N)\left(\frac{t}{T}, B\right) \varepsilon_N(t). \end{aligned} \quad (3.7)$$

The approximation $\mathbf{X}_{NT,N}(t)$ of the common component $\mathbf{X}_{NT}(t)$ is of purely theoretical interest as it cannot be computed from $\Sigma_N(\frac{t}{T}, \omega)$ for two reasons: first, the filter $\Phi_N(\frac{t}{T}, \omega)$ cannot be applied pointwise as suggested above since this would require knowledge of the filters $\Psi_N(u, B)$ and $\Upsilon_N(u, B)$ and of the latent stationary processes $\mathbf{F}(t)$ and $\varepsilon_N(t)$. 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. Second, the involved filters are of infinite length and thus must be truncated before they can be applied to the data.

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

- (i) *approximation error*: $\|\mathbf{X}_{NT,N}(t) - \mathbf{X}_{NT}(t)\|$, where $\mathbf{X}_{NT,N}(t)$ is given by (3.7);

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

$$\begin{aligned} \bar{\mathbf{X}}_{NT,N}(t) &= (\Phi_{NT} \star \Psi_N)\left(\frac{t}{T}, B\right) \mathbf{F}(t) \\ &\quad + (\Phi_{NT} \star \Upsilon_N)\left(\frac{t}{T}, B\right) \varepsilon_N(t) \end{aligned}$$

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

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

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

- (iv) *estimation error*: $\|\widehat{\mathbf{X}}_{NT,N}(t) - \tilde{\mathbf{X}}_{NT,N}(t)\|$, where $\widehat{\mathbf{X}}_{NT,N}(t)$ is the common component estimator in (3.4).

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 defining time-varying principal components by

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

There are, however, two problems with this approach. First, 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. Second, application of the filter $\mathbf{P}_N(\frac{t}{T}, B)$ 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 $\mathbf{P}_N(\frac{t}{T}, B)$ in the frequency domain. Thus, we define the *time-varying dynamic principal components* process at rescaled time $u \in [0, 1]$ by

$$\begin{aligned} \mathbf{F}_N(u, s) &= (\mathbf{P}_N \star \Psi_N)(u, B) \mathbf{F}(s) \\ &\quad + (\mathbf{P}_N \star \Upsilon_N)(u, B) \varepsilon_N(s), \quad s \in \mathbb{Z}. \end{aligned} \quad (3.8)$$

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

$$\mathcal{F}_N^r(u) = \overline{\text{sp}}\{\mathbf{F}_{Nj}(u, s), j = 1, \dots, r, s \in \mathbb{Z}\}, \quad (3.9)$$

where $\mathbf{F}_{Nj}(u, s)$ 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 $\Sigma_N(u, \omega)$ 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 $\mathcal{F}_N^r(\frac{t}{T})$, that is, we have

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

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

$$\mathbf{Y}_{NT}(t) = \mathbf{X}_{NT,N}(t) + \mathbf{Z}_{NT,N}(t) \quad (3.10)$$

where

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

and

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

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

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

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, $\mathbf{X}_{NT,N}(t) = \text{proj}(\mathbf{Y}_{NT}(t) | \mathcal{F}_N^r(\frac{t}{T}))$ as required.

The decomposition in (3.10) 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 $\Sigma_N(u, \omega)$

$$X_{iT,N}(t) = (\phi_{Ni} \star \Psi_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t) + (\phi_{Ni} \star \Upsilon_N) \left(\frac{t}{T}, B \right) \boldsymbol{\varepsilon}_N(t)$$

where the filter ϕ in time domain is given by

$$\phi_{Ni}(u, B) = (\mathbf{P}_{Ni}^* \star (\mathbf{Q}'_N \mathbf{P}_N))(u, B) \quad (3.11)$$

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

Proposition 3. Suppose that Assumptions 1–3 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:

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

Proof. The proof is given in Appendix B.3. $\square$

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

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

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

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

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

For the discussion of the filtering error, let

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

and

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

and define the filters $\tilde{\phi}_{NTi}^{\Upsilon}(t, B)$ and $\phi_{Ni}^{\Upsilon}(t, B)$ analogously with Ψ_N replaced by Υ_N . Then the filtering error can be written as

$$\begin{aligned} \tilde{X}_{iT,N}(t) - \bar{X}_{iT,N}(t) &= [\tilde{\phi}_{NTi}^{\Psi}(t, B) - \phi_{Ni}^{\Psi}(t, B)] \mathbf{F}(t) \\ &\quad + [\tilde{\phi}_{NTi}^{\Upsilon}(t, B) - \phi_{Ni}^{\Upsilon}(t, B)] \boldsymbol{\varepsilon}_N(t). \end{aligned} \quad (3.12)$$

Notice that the filtering error vanishes if the two filters $\Psi_N(u, B)$ and $\Upsilon_N(u, B)$ 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 1–3 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:

$$\lim_{T \rightarrow \infty} \sup_{M_T \leq t \leq T - M_T} \mathbb{E}|\tilde{X}_{iT,N}(t) - \bar{X}_{iT,N}(t)|^2 = 0, \quad (3.13)$$

where

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

Proof. Since the processes $\mathbf{F}(t)$ and $\boldsymbol{\varepsilon}_N(t)$ are uncorrelated, we have

$$\begin{aligned}
& \mathbb{E}[\tilde{X}_{IT,N}(t) - X_{IT,N}(t)]^2 \\
&= \int_{-\pi}^{\pi} \left\| \tilde{\varphi}_{NTi}^{\Psi}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\|^2 d\omega \\
&+ \int_{-\pi}^{\pi} \left\| \tilde{\varphi}_{NTi}^{\Upsilon}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Upsilon}\left(\frac{t}{T}, \omega\right) \right\|^2 d\omega. \quad (3.14)
\end{aligned}$$

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

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

which implies

$$\left\| \tilde{\varphi}_{NTi}^{\Psi}\left(\frac{t}{T}, \omega\right) - \varphi_{Ni}^{\Psi}\left(\frac{t}{T}, \omega\right) \right\| \leq C \frac{M_T^{3/2}}{T}. \quad (3.15)$$

Thus, we have shown that the first term in (3.14) converges to zero as T tends to infinity. Since the second term can be treated similarly, this proves (3.13). $\square$

The upper bound in (3.15) suffices also to show that the expectation

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

converges to zero as $T \rightarrow \infty$. 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,

$$\lim_{T \rightarrow \infty} \sup_{M_T \leq t \leq T - M_T} \mathbb{E}|\tilde{X}_{IT,N}(t) - X_{IT,N}(t)|^2 = 0, \quad (3.16)$$

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 Appendix A.

Assumption 4 (Spectral Matrix Estimator). There exists a sequence u_T with $u_T \rightarrow 0$ and $Tu_T \rightarrow \infty$ 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) $N \in \mathbb{N}$

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

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

We note that Assumption 4 is fulfilled, for example, by the two spectral estimators $\hat{\Sigma}_N(u, \omega)$ discussed in Appendix A; the required conditions are listed in Assumption 5 in the Appendix A.

The above assumption also implies that the matrix $\hat{\mathbf{P}}_{NT}(u, \omega)$ of the 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 ρ_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 1–4 hold. In addition assume for the truncation parameter M_T of the estimated filters $\hat{\Phi}_{NT}$ that $M_T \rightarrow \infty$ and $M_T/\rho_T \rightarrow 0$ as $T \rightarrow \infty$. Then, for all $\delta > 0$, $1 \leq i \leq N$, and all $N \in \mathbb{N}$, the estimation error satisfies

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

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 $\hat{\Phi}_{NT}$. For more details, we refer to Appendix B.4.

3.6. Consistency of the estimated common components

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

Theorem 6. Suppose that Assumptions 1–4 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 $T \geq T_N^*$

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

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

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

In Proposition 3, we have shown that the approximation error $|\widehat{X}_{iT,N}(t) - X_{iT}(t)|$ 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

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

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 (3.16) implies that for all $N \in \mathbb{N}$, $\epsilon > 0$, and $\eta > 0$ there exists a $T_{1,N}$ such that

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

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 $T > T_{2,N}$

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

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

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

which concludes the proof. $\square$

4. Simulation results

For the investigation of the small finite sample behaviour of the proposed approach, we consider a 2-factor model given by

$$\mathbf{Y}_{NT}(t) = \mathbf{X}_{NT}^{(1)}(t) + \mathbf{X}_{NT}^{(2)}(t) + \mathbf{Z}_T(t), \quad t = 1, \dots, T, \quad (4.1)$$

where the two N -dimensional processes $\mathbf{X}_{NT}^{(1)}(t)$ and $\mathbf{X}_{NT}^{(2)}(t)$ consist of univariate locally stationary AR(p) processes as in (4.1) of Dahlhaus (1997), that is,

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

for $i = 1, \dots, N$, $j = 1, 2$, and $t = 1, \dots, T$. We assume that the factors $F_j(t)$ as well as the idiosyncratic components $\mathbf{Z}_T(t)$ are stationary and normally distributed, $\mathbf{F}(t) \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_r)$ and $\mathbf{Z}_N(t) \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_N)$. For the simulations, we have set $p = 2$ with autoregressive coefficients given by

$$a_{ij,1}(u) = \mp \frac{2}{\alpha} \cos(\theta_{ij}(u)) \quad \text{and} \quad a_{ij,2}(u) = \pm \frac{1}{\alpha^2}, \quad u \in [0, 1],$$

where $|\alpha| > 1$ and $\theta_{ij}(u) = \phi_{ij} - \cos(v_{ij}u)$ for $i = 1, \dots, N$ and $j = 1, 2$. For fixed u , the roots of the characteristic polynomials are $\rho_{ij}(u) = \alpha e^{\pm i\theta_{ij}(u)}$, which implies $|\rho_{ij}(u)| = |\alpha| > 1$ for all $u \in [0, 1]$, $i = 1, \dots, N$ and $j = 1, 2$. Thus, the auto-spectra of the processes $X_{iT}^{(j)}(t)$ have single peaks with time-varying locations $\theta_{ij}(u)$.

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 to a smoothing parameter in time. We also note that for convenience we choose the truncation parameter M of the filters $\widehat{\phi}_{NT,i}(\frac{t}{T}, B)$ 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 in 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 Fig. 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 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 (non-stationarity) 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 $\mathbf{X}_{N,T-2L} = (\mathbf{X}_{NT}(L), \dots, \mathbf{X}_{NT}(T-L))$ be the $N \times (T-2L+1)$ matrix containing the vectors of common components $\mathbf{X}_{NT}(t)$, $t = L, \dots, T-L$, and let $\widehat{\mathbf{X}}_{N,T-2L}$ be the corresponding estimator.

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 (3.4). For $q = 1, \dots, Q$, define the following loss function

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

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

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

and the standard deviation

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

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

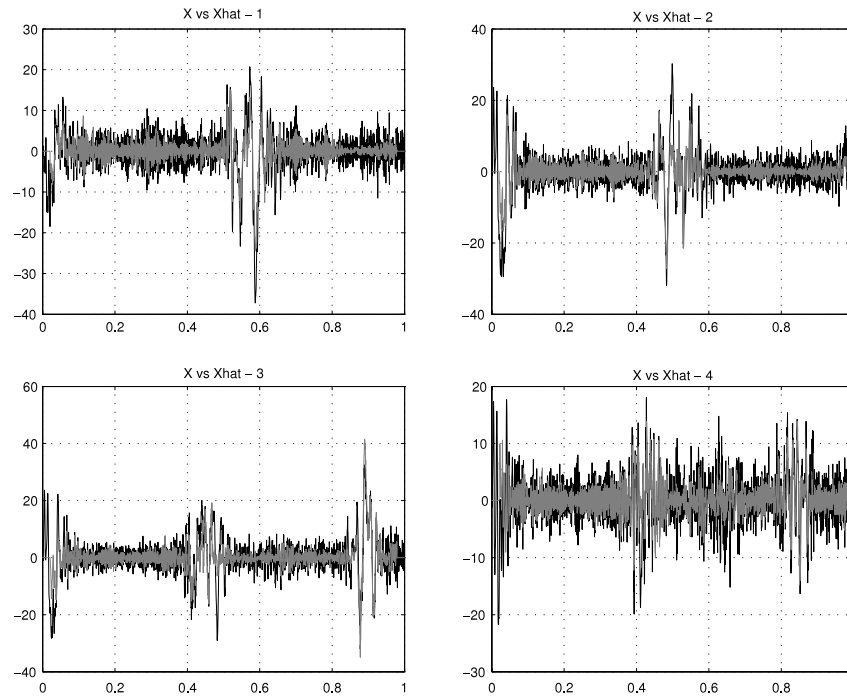


Fig. 4.1. True common components $X_{IT}(t)$ (black) and estimated ones $\hat{X}_{IT}(t)$ (grey), $i = 1, \dots, 4$.

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)

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. Illustrative example

In this section we apply the methodology described in the previous sections to the macroeconomic data¹ analysed in Stock and Watson (2005, Section 4), Jungbacker and Koopman (2008, Section 5), and Hallin and Liška (2007). From this dataset we constructed a balanced panel of 132 monthly US macroeconomic time series from January 1960 to December 2003 (44 years, $T = 528$). The original data have been transformed according to Stock and Watson (2005, Table A.1 in Appendix A).

For the application of our non-stationary dynamic factor modelling approach, the panel size needs to be reduced further. This is due to the fact that the estimated time-varying common components are based on estimation of the whole spectral density matrix (cf Assumption 4) and, more precisely, of its eigenvalues and eigenvectors, whose variance increases with the panel size N . Consequently, our approximation results require that for fixed panel size N the sample length must be larger than some $T = T_N$. Conversely, we find that for fixed T the panel size must not exceed $N = N_T$. We note that the same kind of restriction also applies to the stationary case discussed by Forni et al. (2000) although less severely as in our case the spectral density matrix is estimated locally and hence is based on a much smaller effective sample size. For instance, in case of the segmented periodogram estimator (cf Appendix A.2) the estimation is based on segments of length L_T with $L_T/T \rightarrow 0$ as $T \rightarrow \infty$.

From the original panel of 132 series, we therefore selected a subpanel of $N = 22$ time series, which consists of two groups:

- (1) Industrial production (IP) variables ($i = 6, \dots, 18$);
- (2) Interest Rate (IR) variables ($i = 86, \dots, 94$).

Though homogeneous within each group, these time series are characterized by different dynamics for the two groups as shown by Fig. 5.1. Before studying and comparing these dynamics, we show that the data are non-stationary; we applied (for each group) two tests (Sections 5.1 and 5.2). For more details about the formal definitions and the numerical outcomes of the tests, see Motta (2009, Chapter 4).

5.1. Testing for shifts in the variance

Hallin and Liška (2007, Section 5.2) found three factors $\hat{\tau}_1 = 3$ for 1960–1982 ($T_1 = 276$), and one factor ($\hat{\tau}_2 = 1$) for 1983–2003 ($T_2 = 252$). We believe that the different behaviour of the series in the two subsamples is due to non-stationarity rather than to a different number of underlying factors. To have an idea of the behaviour of the variance in the two subsamples, we plotted the selected series together with their standard deviation, see Fig. 5.1. We found that except for one variable ($i = 17$), the standard deviation in the first subsample is larger than that of the second subsample; this empirical evidence motivated us to test the null hypothesis of equal variance in the two subperiods.

The test proposed by Hoffman and Pagan (1989) is performed by splitting the sample into two parts, and the test-statistic is based on the difference between the two sample variances. We rejected the null hypothesis of constant unconditional variance in the two parts for almost all of the investigated series.

5.2. Testing for time-varying dynamics

The second test that we considered has been proposed by Priestley and Subba Rao (1969) and takes into account the full dynamic structure of the process. The test is based on a factorization of the time-varying spectrum into two components:

a time-dependent factor describing the non-stationarity and a frequency-dependent factor characterizing the dynamics of the process. Given the estimated time-varying spectral density $\hat{\Sigma}_N(u, \omega)$, we can reject the null hypotheses of a constant spectrum over time as well as over frequency.

The results of the tests provide strong evidence against the null hypotheses of stationarity. In Section 5.3 we apply some statistical tools that illustrate how to detect and interpret the non-stationarity.

More generally, the outcomes of the tests indicate that the data are characterized by a time-dependence as well as by a frequency-dependence. In order to understand how these two components interact, we compare the estimation of different factor models (see Section 5.4).

5.3. Statistical tools to detect non-stationarity

In this subsection we apply some non-parametric statistical tools to detect non-stationarity: time-varying covariances, time-varying cross-autocorrelations, time-varying cross-spectra and time-varying cross-coherences. The time-varying covariance matrix is defined as the expectation of the cross-products at rescaled time $u \in (0, 1)$

$$\Sigma_N(u) = \mathbb{E}[\mathbf{Y}_{NT}(\lfloor uT \rfloor) \mathbf{Y}_{NT}(\lfloor uT \rfloor)'], \quad (5.1)$$

and it is estimated non-parametrically according to Motta et al. (2011) (the computation of confidence intervals is based on the asymptotic normality of the estimator). Fig. 5.2 shows that the autocovariances change slowly over time.

The estimated time-varying spectra $\hat{\Sigma}(u, \omega)$, for $u \in (0, 1)$ and $\omega \in (0, \pi)$, are presented in Fig. 5.3 (in absolute value). The estimator of the time-varying spectral density matrix is the smoothed segmented periodogram (see Appendix A.2) with $L_T = 160$ and $b_T = 0.8$. The corresponding estimated coherences $\hat{\mathbf{C}}_N(u, \omega)$, where

$$\hat{c}_{ij}(u, \omega) = \frac{\hat{\sigma}_{ij}(u, \omega)}{\sqrt{\hat{\sigma}_{ii}(u, \omega) \hat{\sigma}_{jj}(u, \omega)}}, \quad (5.2)$$

are given in Fig. 5.3 (in absolute value). The coherences are standardized measures of the dynamics in the frequency domain. Analogously to the matrix $\hat{\mathbf{C}}_N(u, \omega)$, which is defined in the frequency domain, we can define in the time domain the time-varying autocorrelation matrix $\mathbf{R}_N(u, k)$ at lag k and rescaled time u as the $N \times N$ matrix with entries

$$r_{ij}(u, k) = \frac{\mathbb{E}[Y_i(\lfloor uT \rfloor + k) Y_j(\lfloor uT \rfloor)]}{\sqrt{\mathbb{E}[Y_i^2(\lfloor uT \rfloor + k)] \mathbb{E}[Y_j^2(\lfloor uT \rfloor)]}}, \quad (5.3)$$

$$i, j = 1, \dots, N.$$

In Fig. 5.4, we plot the estimated functions $r_{ij}(u, k)$ for lags $0 \leq k \leq 10$. For both IP and IR, the spectra and the coherences show (within each subgroup) a common structure over time. Although the spectra for the IR variables are less homogeneous over time than the spectra of the IP variables, investigation of the coherences suggests that the dynamic interrelationships among the IR variables are smoother than those of the IP variables (both over time and over frequency). The non-stationarity of the IR variables is mainly due to the central part of the series. This is reflected by the behaviour of the (cross-) autocorrelation functions in Fig. 5.4, which vary (though smoothly) more for the IP variables than for the IR variables.

5.4. Comparison of different models

In order to compare various methods for recovering the common components of the time series, we consider the one-step ahead (in-sample) forecast $\hat{\mathbf{Y}}_{NT}(T' + h) = \mathbf{X}_{NT}(T' + h - 1)$, that is,

¹ The data can be downloaded at <http://www.princeton.edu/~mwatson>.

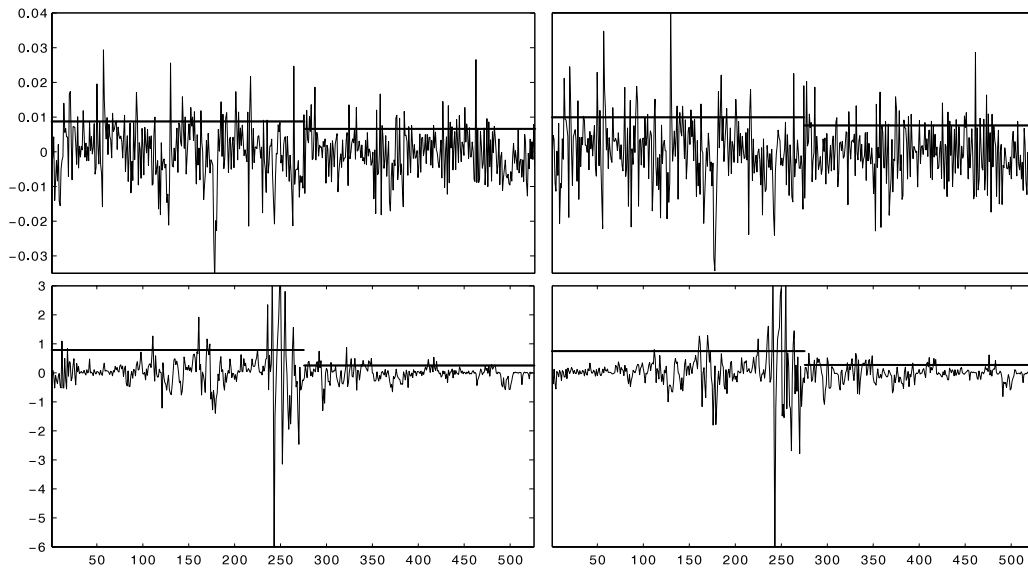


Fig. 5.1. Typical examples of IP (top, $i = 8, 9$) and IR (bottom, $i = 86, 87$) series $Y_{it}(t)$ and their standard deviation.

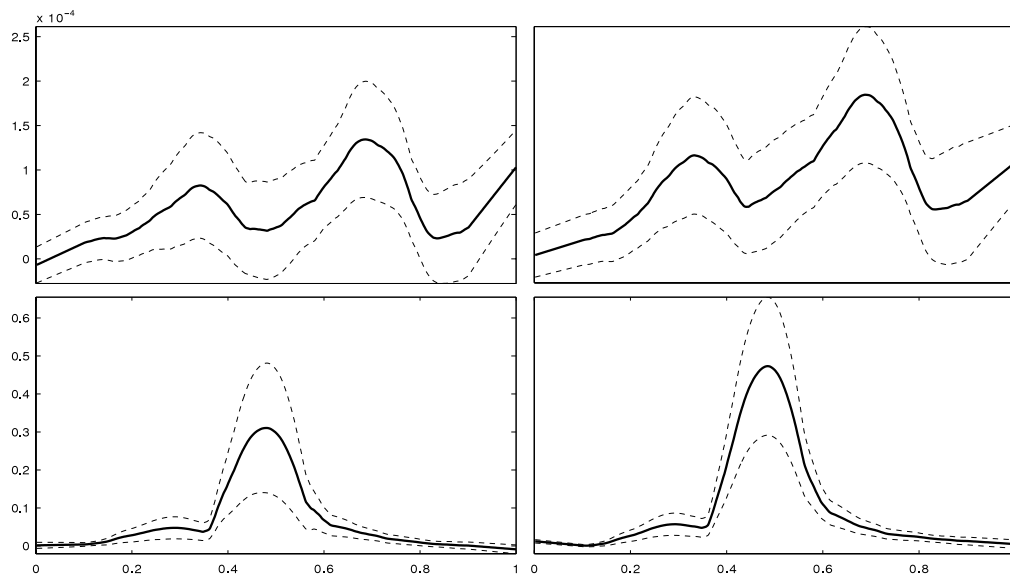


Fig. 5.2. Estimated time-varying covariances. Abscissas: rescaled time $u \in (0, 1)$; ordinates: $\hat{\sigma}_{ij}(u)$. Top: Industrial Production, $i = 8, j = 17$. Bottom: Interest Rate, $i = 86, j = 92$. Bold lines: estimated covariance functions; dashed lines: confidence intervals.

we predict $\hat{Y}_{NT}$ at times $T' + h$ by the common component $\hat{X}_{NT}$ at times $T' + h - 1$. The forecasting performance of each method can be measured by the averaged forecast error

$$FE(T', H) = \frac{1}{H} \sum_{h=1}^H \|\hat{Y}_{NT}(T' + h) - Y_{NT}(T' + h)\|. \quad (5.4)$$

We note that $\hat{Y}_{NT}(T' + h)$ is not a true forecast since computation of the estimate $\hat{X}_{NT}(T' + h - 1)$ also involves the value $Y_{NT}(T' + h)$ due to the two-sided projection filters that we use.

In our application, we used $r = 3$ factors and computed $FE(T', H)$ with $T' = 500$ and $H = 12$. The choice $T' = 500$ is motivated by our interest in measuring the forecasting performance of the four methods for the 12 months ($H = 12$) after September 2001. We remark that [Stock and Watson \(2005\)](#) are mainly interested in forecasting the growth of industrial production, which is one of the series, whereas we look at forecasting a whole group of variables.

Table 5.1

Forecasting error in (5.4) estimated on the basis of different methods and different choices of L_T and b_T .

Method	L_T	b_T	IP	IR
Static stationary ^a			0.4573	5.0946
Static evolutionary ^b			0.4590	5.0526
Dynamic stationary ^c	16		0.4501	5.0423
	32		0.4505	5.0944
	48		0.4497	5.1615
Dynamic evolutionary	16	$T^{-1/2}$	0.4718	5.9310
	16	$T^{-1/4}$	0.4390	5.5544
	16	$T^{-1/6}$	0.4295	5.4929

^a Bai (2003).

^b Motta et al. (2011).

^c Forni et al. (2000).

The results in Table 5.1 show that the dynamic non-stationary approach outperforms the others for the industrial production variables if the bandwidths are chosen properly (for $b_T = T^{-1/4}$, $T^{-1/6}$), whereas for the interest rate variables the non-stationary

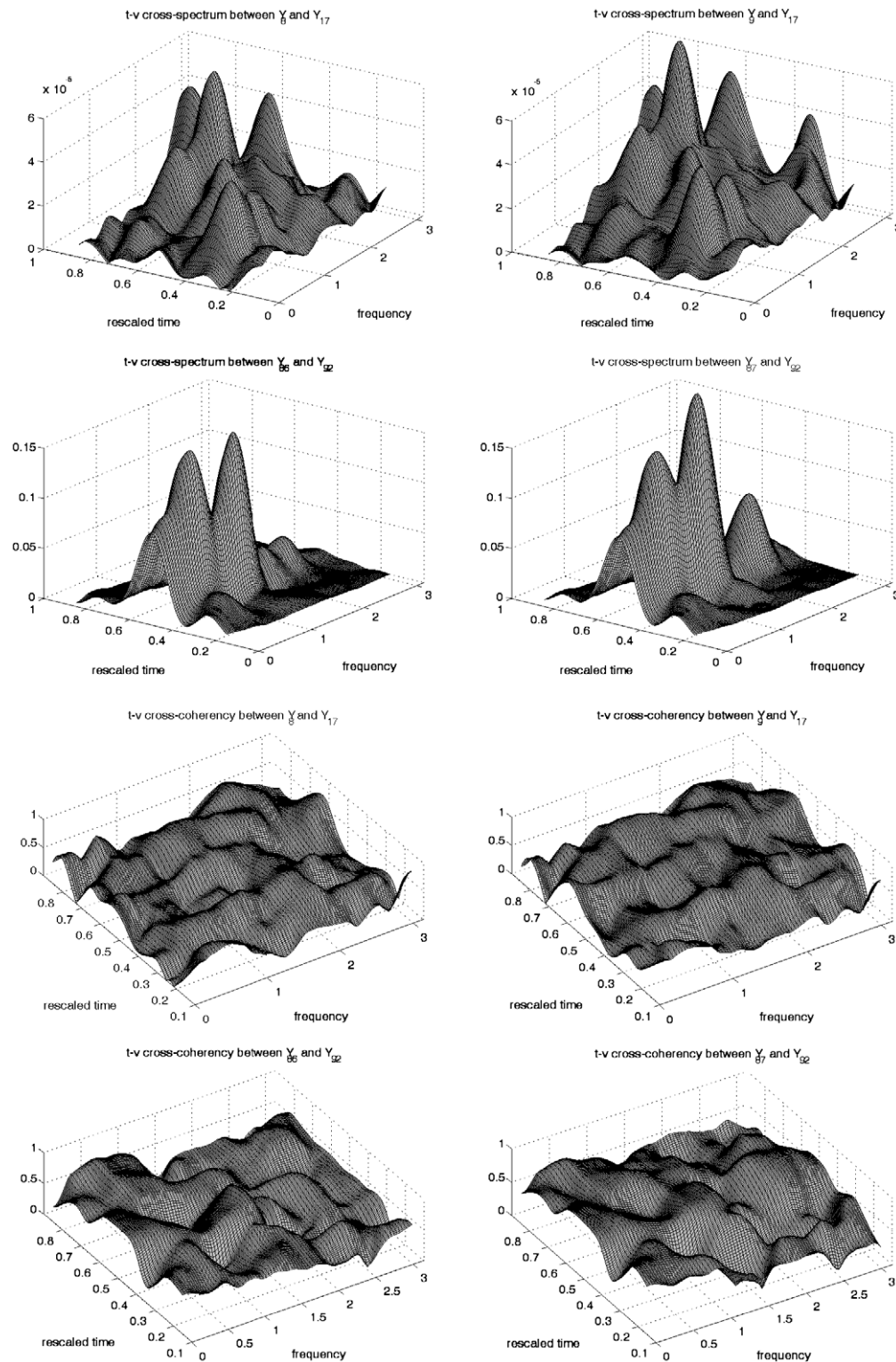


Fig. 5.3. Estimated time-varying spectra (top) and coherences (bottom). Left abscissas: rescaled time $u \in (0, 1)$; right abscissas: frequency $\omega \in (0, \pi)$. Top ordinates: $|\hat{\sigma}_{ij}(u, \omega)|$, where $\hat{\sigma}_{ij}(u, \omega)$ are defined in (A.2); bottom ordinates: $|\hat{c}_{ij}(u, \omega)|$, where $\hat{c}_{ij}(u, \omega)$ are defined in (5.2). Top variables: Industrial Production, $i = 8, 9, j = 17$. Bottom variables: Interest Rate, $i = 86, 87, j = 92$.

static and the stationary dynamic approaches seem to outperform the others. This could be due to that the IR variables exhibit much less variation over time in their interrelationships and thus are well approximated by the more parsimonious stationary dynamic model. In contrast, the time-varying coherences and correlations for the IP variables are clearly time-dependent and require a non-stationary model.

6. Conclusions

With this paper we have delivered an important generalization 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

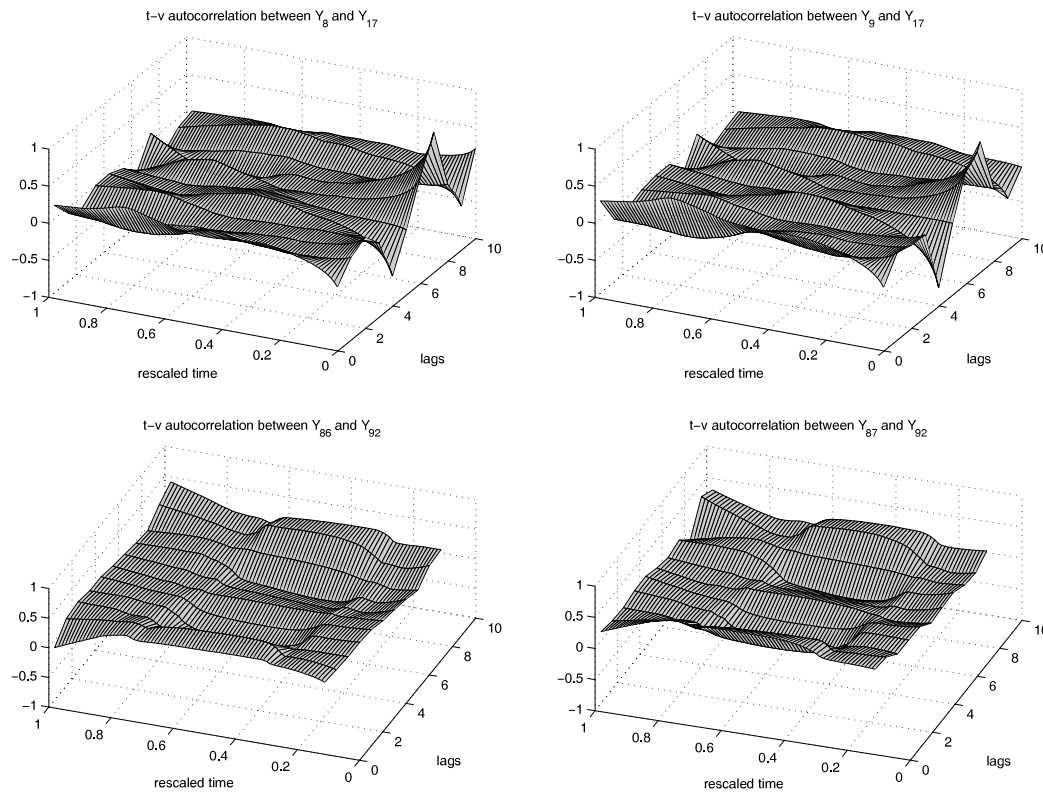


Fig. 5.4. Estimated time-varying autocorrelations defined in (5.3). Left abscissas: rescaled time $u \in (0, 1)$; right abscissas: lags $k = 1, 2, \dots, 10$; ordinates: $\hat{r}_{ij}(u, k)$. Top: Industrial Production, $i = 8, 9, j = 17$. Bottom: Interest Rate, $i = 86, 87, j = 92$.

time resolution) – macroeconomic 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 generalize both the work by Motta et al. (2011) 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 a 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 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 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.2).

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 (3.8), the estimation error is based on the new definition in (3.11) 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 filtering error: this is due to the problem of time-varying filtering of a non-stationary process, which is new compared to the dynamic factor model of Forni et al. (2000).

A few important problems have not been addressed in this paper. First, our approach for recovering the time-varying common components is largely affected by the quality of estimation of the spectral density matrix. Two commonly used estimators are the segmented periodogram and the smoothed pre-periodogram (see Appendix A). Since the performance of these non-parametric estimators depends crucially on the localization in time and frequency direction, it would be desirable to have an optimal procedure for selecting the corresponding smoothing bandwidths. However, in the case of time-varying spectra, this is still an open problem. The optimal values of the corresponding smoothing bandwidths are known to be proportional to $T^{-1/6}$ (see Dahlhaus, 1996, Theorem 2.3) but depend on the derivatives of the unknown true spectrum $\Sigma(u, \omega)$. In our case, the matter is further complicated by the fact that we are interested not directly in the entries of the spectral matrix but in certain joint features such as eigenvalues and eigenvectors, which needs to be taken into account when selecting the bandwidths. Second, we have not discussed the important problem of using dynamic factor models for the purpose of prediction. Since our approach is based on PCR in the frequency domain, the resulting projection filters are two-sided and thus ill-suited for forecasting. In contrast to the stationary dynamic factor model, for which Forni et al. (2005) have developed a forecasting methods based on one-sided projections, we face an additional problem due to the non-stationarity and the lack of information about how the dynamics of the process evolves outside the observed time period. Finally, we have considered the number of factors as fixed. In applications, the exact number of factors is usually unknown and needs to be determined from the data. We believe that recent work on data-based selection

of the number of factors in static and dynamic PCR-based factor models (e.g. Bai and Ng, 2002; Hallin and Liška, 2007) can be generalized, but this is left for future research as well.

Acknowledgements

The authors thank Christian Hafner, Marc Hallin and Marco Lippi for many stimulating discussions and valuable remarks, and two anonymous referees and the editor, Franz Palm, for their helpful comments on an earlier version of this article. Giovanni Motta and Rainer von Sachs 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 non-parametrically. 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 $\hat{\Sigma}_N(u, \omega)$.

A.1. Smoothed pre-periodogram

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

$$\tilde{S}_N\left(\frac{t}{T}, \omega\right) = \frac{1}{2\pi} \sum_{k: 1 \leq \lfloor t + \frac{1}{2} \pm \frac{k}{2} \rfloor \leq T} \mathbf{Y}_{NT}\left(\left\lfloor t + \frac{1+k}{2} \right\rfloor\right) \mathbf{Y}_{NT}'\left(\left\lfloor t + \frac{1-k}{2} \right\rfloor\right) e^{-i\omega k},$$

where $\lfloor x \rfloor$ denotes the largest integer smaller than x . The product $\mathbf{Y}_{NT}(\lfloor t + \frac{1+k}{2} \rfloor) \mathbf{Y}_{NT}'(\lfloor t + \frac{1-k}{2} \rfloor)$ can be regarded as a raw estimator of the local autocovariance $\Gamma_N(\frac{t}{T}, k)$ given by

$$\Gamma_N(u, k) = \int_{-\pi}^{\pi} \Sigma_N(u, \omega) e^{i\omega k} d\omega.$$

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

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

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 $\int_{-1/2}^{1/2} xK(x)dx = 0$ 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 $\Sigma_N(u, \omega)$ can be treated as stationary over short time intervals. This suggests using ordinary kernel estimators applied locally to the time series to obtain a time-varying estimate

for the spectral density. More precisely, let (for L even)

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

be the localized Fourier transform of the process over the segment from $\lfloor uT \rfloor - L/2 + 1$ to $\lfloor uT \rfloor + L/2$, 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 localized or segmented periodogram is defined as

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

Thus the periodogram estimator of the time-varying spectrum $\Sigma_N(u, \omega)$ is obtained from a segment of length L with midpoint $\lfloor uT \rfloor$. Since the periodogram is not a consistent estimator, the segmented periodogram needs to be smoothed (see Dahlhaus, 1996), which leads to the kernel estimator

$$\hat{\Sigma}_{NT}^L(u, \omega) = \frac{1}{b_T} \int_{-\pi}^{\pi} K^f\left(\frac{\omega - \mu}{b_T}\right) \hat{\Sigma}_{NT}^L(u, \mu) d\mu, \quad (\text{A.2})$$

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 $b_T T \rightarrow \infty$ as $T \rightarrow \infty$.

A.3. Uniform consistency

In Assumption 4, we require that the estimator of the spectral matrix is uniformly consistent. For the smoothed pre-periodogram and the smoothed segmented periodogram, uniform consistency has been shown in an unpublished manuscript by Dahlhaus. For convenience, 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 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

$$\mathbf{X}_{NT}(t) = \sum_{k=0}^{\infty} \Phi_{N,k}\left(\frac{t}{T}\right) \mathbf{e}_N(t-k), \quad (\text{A.3})$$

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

(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

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

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 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, unpublished manuscript, 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 $Th_T b_T \gg \log^2 T$

$$\sup_{u \in \left[\frac{h_T}{2}, 1 - \frac{h_T}{2}\right]} \sup_{\omega \in (-\pi, \pi]} \|\widehat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)\| = O_p((h_T b_T T)^{-1/2} + h_T^2 + b_T^2). \quad (\text{A.4})$$

A similar result holds for the kernel estimator based on the segmented periodogram (Dahlhaus, unpublished manuscript, 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.3). 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 these processes, we can also establish uniform consistency for the spectral estimator $\widehat{\Sigma}_N(u, \omega)$.

More precisely, let $\mathbf{V}_N(t)$ be the $2N$ -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.3) and has a time-varying spectral matrix given by

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

where $\Sigma_N^X(\frac{t}{T}, \omega)$ and $\Sigma_N^Z(u, \omega)$ 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 $\widehat{\Sigma}_N^V(u, \omega)$ 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

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

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

$$\Sigma_N(u, \omega) = (\mathbf{I}_N \mathbf{I}_N) \Sigma_N^V(u, \omega) (\mathbf{I}_N \mathbf{I}_N)',$$

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

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 by Dahlhaus.

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)

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

On 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

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

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

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

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 (3.10) we have the following decompositions

$$\mathbf{Y}_{iT}(t) = \mathbf{X}_{iT}(t) + \mathbf{Z}_{iT}(t) = \mathbf{X}_{iT,N}(t) + \mathbf{Z}_{iT,N}(t),$$

which gives for the approximation error

$$|\mathbf{X}_{iT}(t) - \mathbf{X}_{iT,N}(t)|^2 = |\mathbf{Z}_{iT}(t) - \mathbf{Z}_{iT,N}(t)|^2. \quad (\text{B.1})$$

Recalling the definition of the approximate common component

$$\mathbf{X}_{iT,N}(t) = (\phi_{Ni} \star \Psi_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t) + (\phi_{Ni} \star \Upsilon_N) \left(\frac{t}{T}, B \right) \mathbf{e}_N(t),$$

as well as the definition of the filter ϕ in (3.11),

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

we find

$$\begin{aligned} \mathbb{E} \left| (\phi_{Ni} \star \Upsilon_N) \left(\frac{t}{T}, B \right) \mathbf{e}_N(t) \right|^2 \\ = \mathbb{E} |r_{iT}(t)|^2 + \mathbb{E} |\mathbf{Z}_{iT}(t) - \mathbf{Z}_{iT,N}(t)|^2 \\ + \mathbb{E} [r_{iT}(t)(\mathbf{Z}_{iT}(t) - \mathbf{Z}_{iT,N}(t))], \end{aligned} \quad (\text{B.2})$$

where we have set $r_{iT}(t) = \mathbf{X}_{iT}(t) - (\phi_{Ni} \star \Psi_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t)$. In view of (B.1), 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 a zero mean, 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,

$$\lim_{N \rightarrow \infty} \sup_{u \in [0, 1]} \int_{-\pi}^{\pi} \|\phi_{Ni}(u, \omega)\|^2 d\omega = 0. \quad (\text{B.3})$$

We observe that

$$\begin{aligned} \|\phi_{Ni}(u, \omega)\|^2 &= \phi_{Ni}(u, \omega) \phi_{Ni}(u, \omega)^* \\ &= \sum_{j=1}^r |p_{Ni,j}(u, \omega)|^2. \end{aligned} \quad (\text{B.4})$$

By (3.5), (3.6), and Assumption 2, the terms in the sum on the right-hand side satisfy

$$|p_{Ni,j}(u, \omega)|^2 \leq \frac{\bar{\sigma}_i}{\inf_{v \in [0, 1]} \lambda_{Nj}(v, \omega)}, \quad j = 1, \dots, r, \quad (\text{B.5})$$

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.4) converges to zero uniformly in $u \in [0, 1]$. As it is also bounded by 1, condition (B.3) now follows from the application of Lebesgue's dominated convergence theorem.

Next, we have to show that the mixed term on the right-hand side of (B.2) converges to zero. The proof is based on the next Lemma 9, for which we may assume without loss of generality that

$$\lambda_{Nj}(u, \omega) \geq 1 \quad \text{for all } j = 1, \dots, N, \quad N \in \mathbb{N}, \quad u \in [0, 1], \quad \omega \in [-\pi, \pi]. \quad (\text{B.6})$$

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

$$\begin{aligned} \tilde{\mathbf{Z}}_{NT}(t) &= \mathbf{Z}_{NT}(t) + \mathbf{U}_N(t) = \Upsilon_N \left(\frac{t}{T}, B \right) \varepsilon_N(t) + \mathbf{U}_N(t) \\ &= \tilde{\Upsilon}_N \left(\frac{t}{T}, B \right) \tilde{\varepsilon}_N(t) \end{aligned}$$

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

$$\tilde{\mathbf{Y}}_{NT}(t) = \mathbf{X}_{NT}(t) + \tilde{\mathbf{Z}}_{NT}(t), \quad (\text{B.7})$$

we find that the process $\tilde{\mathbf{Y}}_{NT}(t)$ in (B.7) fulfills Assumptions 1–3 with

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

and thus $\lambda_{Nj}^{\tilde{\mathbf{Z}}}(u, \omega) = \lambda_{Nj}^Z(u, \omega) + 1$ and $\lambda_{Nj}^{\tilde{\mathbf{Y}}}(u, \omega) = \lambda_{Nj}(u, \omega) + 1$.

Furthermore, $\mathbf{P}_{Nj}^{\tilde{\mathbf{Y}}} = \mathbf{P}_{Nj}$ and thus also $\phi_{Nj}^{\tilde{\mathbf{Y}}} = \phi_{Nj}$ for all $1 \leq j \leq N$ and $N \in \mathbb{N}$. Consequently, if Proposition 3 holds for the process $\tilde{\mathbf{Y}}_{NT}(t)$, that is,

$$\mathbb{E} \left\| (\phi_{Ni} \star \Psi_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t) + (\phi_{Ni} \star \tilde{\Upsilon}) \left(\frac{t}{T}, B \right) \tilde{\varepsilon}_N(t) \right\|^2 \rightarrow 0$$

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

$$\mathbb{E} \left\| \phi_{Ni} \left(\frac{t}{T}, B \right) \mathbf{U}_N(t) \right\|^2 \rightarrow 0$$

as N tends to infinity.

Under condition (B.6), the function $\mu_{Nj}(u, \omega) = [\lambda_{Nj}(u, \omega)]^{-1/2}$ is well-defined for all $\omega \in [-\pi, \pi]$ and is bounded. Therefore it has a mean square convergent Fourier representation. Denote by $\mu_{Nj}(u, B)$ 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

$$\begin{aligned} \mathbf{W}_N(u, s) &= (W_{N1}(u, s), \dots, W_{Nr}(u, s))' \\ &= (\mathbf{M}_N^r \star \mathbf{P}_N^r \star \Psi_N)(u, B) \mathbf{F}(s) \\ &\quad + (\mathbf{M}_N^r \star \mathbf{P}_N^r \star \Upsilon_N)(u, B) \varepsilon_N(s) \\ &= \mathbf{M}_N^r(u, B) \mathbf{F}_N^r(u, s), \end{aligned}$$

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 (3.8),

$$\begin{aligned} \mathbf{M}_N^r(u, \omega) &= \text{diag}\{\mu_{N1}(u, \omega), \dots, \mu_{Nr}(u, \omega)\} \\ &= [\Lambda_N^r(u, \omega)]^{-\frac{1}{2}} \end{aligned} \quad (\text{B.8})$$

$$\Lambda_N^r(u, \omega) = \text{diag}\{\lambda_{N1}(u, \omega), \dots, \lambda_{Nr}(u, \omega)\}$$

are r -dimensional diagonal matrices and $\mathbf{P}_N^r(u, \omega) = [\mathbf{P}_{N1}(u, \omega)', \dots, \mathbf{P}_{Nr}(u, \omega)']'$ is the $r \times N$ matrix containing the first r rows of the $N \times N$ matrix $\mathbf{P}_N(u, \omega)$. Using (3.2), (3.5), and (B.8), we find

$$\begin{aligned} \mathbf{M}_N^r(u, \omega) [\mathbf{P}_N^r(u, \omega) \Sigma_N^X(u, \omega) \mathbf{P}_N^r(u, \omega)^*] \mathbf{M}_N^r(u, \omega)^* \\ + \mathbf{M}_N^r(u, \omega) [\mathbf{P}_N^r(u, \omega) \Sigma_N^Z(u, \omega) \mathbf{P}_N^r(u, \omega)^*] \mathbf{M}_N^r(u, \omega)^* \\ = \mathbf{M}_N^r(u, \omega) [\mathbf{P}_N^r(u, \omega) \Sigma_N(u, \omega) \mathbf{P}_N^r(u, \omega)^*] \mathbf{M}_N^r(u, \omega)^* \\ = \mathbf{M}_N^r(u, \omega) \Lambda_N^r(u, \omega) \mathbf{M}_N^r(u, \omega)^* \\ = \mathbf{I}_r. \end{aligned}$$

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 (3.8), 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 (3.9), 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 $\mathbf{F}_j(s)$, $j = 1, \dots, r$, $s \in \mathbb{Z}$,

$$\mathbf{W}_N(u, s) = \mathbf{G}_N(u, B)\mathbf{F}(s) + \mathbf{R}_N(u, s) \quad (\text{B.9})$$

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

$$\mathbf{F}(s) = \mathbf{G}_N(u, B^{-1})^* \mathbf{W}_N(u, s) + \mathbf{S}_N(u, s), \quad (\text{B.10})$$

where $\mathbf{G}_N(u, B) = (\mathbf{M}_N \star \mathbf{P}_N^r \star \Psi_N)(u, B)$ is an $r \times r$ two-sided square-summable filter, and where $\mathbf{R}_N(u, s) = (\mathbf{M}_N \star \mathbf{P}_N^r \star \Upsilon_N)(u, B)\mathbf{e}_N(s)$ 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

$$\begin{aligned} \mathbf{W}_{Nj}(u, s) &= (\mu_{Nj} \star \mathbf{P}_{Nj} \star \Psi_N)(u, B)\mathbf{F}(s) \\ &\quad + (\mu_{Nj} \star \mathbf{P}_{Nj} \star \Upsilon_N)(u, B)\mathbf{e}_N(s) \end{aligned} \quad (\text{B.11})$$

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

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

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.6) the supremum is also bounded by $\bar{\lambda}_Z$, we get by Lebesgue's dominated convergence theorem

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

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.9) and (B.10) we have

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

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 $\Sigma_N^R(u, s)$ and $\Sigma_N^S(u, s)$ are the spectral densities of $\mathbf{R}_N(u, s)$ and $\mathbf{S}_N(u, s)$, respectively. This implies $\text{tr}(\Sigma_N^S(u, \omega)) = \text{tr}(\Sigma_N^R(u, \omega))$, 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 $Z_{iT,N}(t)$, $f_{Ni}^{12}(u, \omega)$ 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 $|f_{Ni}^{12}(u, \omega)|$ is bounded by an integrable function, Lebesgue's dominated convergence theorem implies that $\int_{-\pi}^{\pi} f_{Ni}^{12}(u, \omega) d\omega$ converges to zero uniformly in $u \in [0, 1]$, which proves the desired mean square convergence.

Recall that

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

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 $(\phi_{Ni} \star \Psi_N)(\frac{t}{T}, B)\mathbf{F}(t)$.

The cross-spectrum between $Z_{iT,N}(t)$ and $X_{iT}(t)$ is equal to the cross-spectrum between $Z_{iT,N}(t)$ and $\psi_i(\frac{t}{T}, B)\mathbf{S}_N(\frac{t}{T}, t)$ because, by (B.10), $X_{iT}(t)$ can be written as

$$\begin{aligned} X_{iT}(t) &= \psi_i \left(\frac{t}{T}, B \right) \mathbf{F}(t) = \psi_i \left(\frac{t}{T}, B \right) \mathbf{G} \left(\frac{t}{T}, B^{-1} \right)^* \mathbf{W}_N \left(\frac{t}{T}, t \right) \\ &\quad + \psi_i \left(\frac{t}{T}, B \right) \mathbf{S}_N \left(\frac{t}{T}, t \right). \end{aligned}$$

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

$$\begin{aligned} F_{Nj} \left(\frac{t}{T}, t \right) \\ = \left[(\mathbf{P}_{Nj} \star \Psi_N) \left(\frac{t}{T}, B \right) \mathbf{F}(t) + (\mathbf{P}_{Nj} \star \Upsilon_N) \left(\frac{t}{T}, B \right) \mathbf{e}_N(t) \right], \end{aligned}$$

for $j = 1, \dots, r$ at any lead and lag; then by (B.11) it is also orthogonal at any lead and lag to the terms $W_{Nj}(\frac{t}{T}, t)$. Let $g_{Ni}^{12}(\frac{t}{T}, \omega)$ be the time-varying cross-spectrum between $Z_{iT,N}(t)$ and $\psi_i(\frac{t}{T}, B)\mathbf{S}_N(\frac{t}{T}, t)$, and let $\sigma_{i,N}^Z(\frac{t}{T}, \omega)$ be the time-varying spectrum of $Z_{iT,N}(t)$. We have

$$|g_{Ni}^{12}(u, \omega)|^2 \leq \sigma_{i,N}^Z(u, \omega) \psi_i(u, \omega) \Sigma_N^S(u, \omega) \psi_i(u, \omega)^*.$$

By (3.3) $\sigma_{i,N}^Z(u, \omega) \leq \sigma_i(u, \omega) \leq C_{\Sigma,N}$ uniformly in u and ω . Furthermore,

$$\begin{aligned} &\sup_{u \in [0, 1]} \|\psi_i(u, \omega) \Sigma_N^S(u, \omega) \psi_i(u, \omega)^*\| \\ &\leq \sup_{u \in [0, 1]} \|\psi_i(u, \omega)\|^2 \sup_{u \in [0, 1]} \|\Sigma_N^S(u, \omega)\|, \end{aligned}$$

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

By the same arguments, the time-varying cross-spectrum between $Z_{iT,N}(t)$ and $(\phi_{Ni} \star \Psi_N)(\frac{t}{T}, B)\mathbf{F}(t)$ equals that between $Z_{iT,N}(t)$ and $(\phi_{Ni} \star \Psi_N)(\frac{t}{T}, B)\mathbf{S}_N(\frac{t}{T}, t)$. 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 $\Sigma_N(u, \omega) = \Sigma_N^X(u, \omega) + \Sigma_N^Z(u, \omega)$, 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.5), 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

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

satisfies

$$\lim_{T \rightarrow \infty} \sup_{Tu_T < t < T - Tu_T} \mathbb{P}[|\widehat{X}_{iT,N}(t) - \widetilde{X}_{iT,N}(t)| > \epsilon] = 0 \quad (\text{B.12})$$

for all $\epsilon > 0$. 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.12)

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

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

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

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

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

which then can simply be bounded by

$$\begin{aligned} & N^2 \sup_{1 \leq m \leq N} \mathbb{E}\left[\sum_{k=-M_T}^{M_T} \left(\Delta \hat{\phi}_{Nim,k}\left(\frac{t}{T}\right)\right)^2\right. \\ & \quad \times \left.\sum_{k=-M_T}^{M_T} (Y_{mT}(t-k))^2\right]. \end{aligned} \quad (\text{B.13})$$

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

$$\mathbb{E}\left[\left(\sum_{k=-M_T}^{M_T} \left(\Delta \hat{\phi}_{Nim,k}\left(\frac{t}{T}\right)\right)^2\right)\right] = O((\rho_T)^{-2}) \quad (\text{B.14})$$

and that

$$\mathbb{E}\left[\left(\sum_{k=-M_T}^{M_T} (Y_{mT}(t-k))^2\right)\right] = O(M_T^2), \quad (\text{B.15})$$

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

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

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \|\Delta \hat{\phi}_{Ni}(u, \omega)\| = O_p((\rho_T)^{-1}), \quad (\text{B.16})$$

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

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \mathbb{E}[\Delta \hat{\phi}_{Ni,m}(u, \omega)] = O((\rho_T)^{-1}), \quad (\text{B.17})$$

and

$$\sup_{u \in [u_T, 1-u_T]} \sup_{\omega \in [-\pi, \pi]} \text{Var}(\Delta \hat{\phi}_{Ni,m}(u, \omega)) = O((\rho_T)^{-2}). \quad (\text{B.18})$$

Further note that

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

and that this random variable is stochastically bounded from above because $\Delta \hat{\phi}_{Ni,m}(t/T, \omega)$ is a difference of (a function of) eigenvectors of empirical and true spectral matrix of fixed dimension N . Hence we obtain, using (B.17) and (B.18), that

$$\begin{aligned} & \mathbb{E}\left[\left(\sum_{k=-M_T}^{M_T} \left(\Delta \hat{\phi}_{Nim,k}\left(\frac{t}{T}\right)\right)^2\right)\right] \\ & \leq C \mathbb{E}\left[\sum_{k=-M_T}^{M_T} \left(\Delta \hat{\phi}_{Nim,k}\left(\frac{t}{T}\right)\right)^2\right] = O((\rho_T)^{-2}), \end{aligned}$$

and hence (B.14).

Next, the expectation in (B.15) is simply bounded by $CM_T^2 \mathbb{E}(Y_{mT}(t)^4)$, which is of order $O(M_T^2)$ since $\mathbf{Y}_{NT}(t)$ has uniformly bounded fourth moments. Consequently, the upper bound in (B.13) converges to zero as $T \rightarrow \infty$, which completes the proof.

It remains to show that equations (B.16)–(B.18) 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 the rates of convergence of eigenvalues $\hat{\lambda}_j(u, \omega)$ and eigenvectors $\mathbf{P}_{Nj}(u, \omega)$ of $\hat{\Sigma}_{NT}(u, \omega)$ to the limiting ones $\lambda_j(u, \omega)$ and $\mathbf{P}_{Nj}(u, \omega)$ of $\Sigma_N(u, \omega)$ by relating their difference in first order to the difference $\hat{\Sigma}_{NT}(u, \omega) - \Sigma_N(u, \omega)$. (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.)

$$\begin{aligned} \hat{\lambda}_j(u, \omega) - \lambda_j(u, \omega) &= \mathbf{P}_{Nj}(u, \omega)(\hat{\Sigma}_{NT}(u, \omega) \\ & \quad - \Sigma_N(u, \omega))\mathbf{P}_{Nj}(u, \omega)^* + \dots \end{aligned}$$

and

$$\begin{aligned} \hat{\mathbf{P}}_{Nj}(u, \omega) - \mathbf{P}_{Nj}(u, \omega) &= \sum_{i \neq j} [\mathbf{P}_{Ni}(u, \omega)(\hat{\Sigma}_{NT}(u, \omega) \\ & \quad - \Sigma_N(u, \omega))\mathbf{P}_{Nj}(u, \omega)] \frac{\mathbf{P}_{Ni}(u, \omega)}{\lambda_j(u, \omega) - \lambda_i(u, \omega)} + \dots \end{aligned}$$

(In both expansions the $\dots$ denote terms of higher order in the difference between $\hat{\Sigma}_{NT}(u, \omega)$ and $\Sigma_N(u, \omega)$.) 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 $Zf_j = \psi_j f_j$ about $Z = \Sigma_N$ which gives:

$$(\hat{\Sigma}_{NT} - \Sigma_N)\mathbf{P}_{Nj}^* + \Sigma_N(\hat{\mathbf{P}}_{Nj} - \mathbf{P}_{Nj}) = (\hat{\lambda}_j - \lambda_j)\mathbf{P}_{Nj}^* + \lambda_j(\hat{\mathbf{P}}_{Nj} - \mathbf{P}_{Nj})^*,$$

and after multiplication of this equation from left by any other $\mathbf{P}_{Ni}$, with $i \neq j$, a system of equations of the type

$$\mathbf{P}_{Ni}(\hat{\Sigma}_{NT} - \Sigma_N)\mathbf{P}_{Nj}^* = (\lambda_j - \lambda_i)\mathbf{P}_{Ni}(\hat{\mathbf{P}}_{Nj} - \mathbf{P}_{Nj})^*.$$

Note that in the cited sources, no multiplicity of the eigenvalues has been considered. However, there exist 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.16) can be derived, with a rate of consistency which is possibly only a fractional power of the rate $(\rho_T)^{-1}$ 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 pre-periodogram defined in (A.1), fulfilling Assumption 4, with a rate $\rho_T = (h_T b_T T)^{1/2}$ delivered by Theorem 7.

- (i) controlling the “smoothing” bias: Let $\Sigma_{NT}^{cc}(u, \omega) = K_h^\tau K_b^\phi * \Sigma_N(u, \omega)$ denote the convolution of Σ_N with the product kernel in time $K_h^\tau(u)$ and frequency $K_b^\phi(\omega)$ (see, e.g., also (A.1)), and let $\mathbf{P}_{Nj}^{cc}(u, \omega)$ 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 ,

$$\|\mathbf{P}_{Nj}^{cc}(u, \omega) - \mathbf{P}_{Nj}(u, \omega)\| = O(h_T^2) + O(b_T^2),$$

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 ,

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

essentially paralleling Forni et al. (2004), Lemma 4.2(iii) and Eq. (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.4.

- (iii) controlling uniformly the bias

$$\|\mathbb{E}\widehat{\mathbf{P}}_{Nj}(u, \omega) - \mathbf{P}_{Nj}^{cc}(u, \omega)\| = O((h_T b_T T)^{-1}),$$

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 $\widehat{\mathbf{P}}_{Nj} - \mathbf{P}_{Nj}$, 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:

$$\sup_u \sup_\omega \|\mathbf{P}_{Ni}(u, \omega)(d_u^2 \Sigma_N(u, \omega))\mathbf{P}_{Nj}(u, \omega)^* \mathbf{P}_{Ni}(u, \omega)\| \leq C_u(N)$$

and

$$\sup_u \sup_\omega \|\mathbf{P}_{Ni}(u, \omega)(d_\omega^2 \Sigma_N(u, \omega))\mathbf{P}_{Nj}(u, \omega)^* \mathbf{P}_{Ni}(u, \omega)\| \leq C_\omega(N)$$

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., 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., 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: Robinson, P.M., Rosenblatt, M. (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., 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., Principe, J.C., 1998. Competitive principal components analysis for locally stationary time series. *IEEE Transactions on Signal Processing* 46, 3068–3081.
- Forni, M., Hallin, M., Lippi, M., 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., Reichlin, L., 2004. The generalized dynamic factor model: consistency and rates. *Journal of Econometrics* 119, 231–255.
- Forni, M., Hallin, M., Lippi, M., 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., Lippi, M., 2001. The generalized dynamic factor model: representation theory. *Econometric Theory* 17, 1113–1141.
- Hallin, M., Liška, R., 2007. Determining the number of factors in the general dynamic factor model. *Journal of the American Statistical Association* 102 (478), 603–617.
- Hoffman, D., Pagan, A.R., 1989. Post-sample prediction tests for generalized method of moment estimators. *Oxford Bulletin of Economics and Statistics* 51, 333–343.
- Jungbacker, B., Koopman, S.J., 2008. Likelihood-based analysis for dynamic factor models. Technical Report. Department of Econometrics, VU University Amsterdam.
- Lütkepohl, H., 1996. *Handbook of Matrices*. John Wiley & Sons.
- Molenaar, P.C.M., de Gooijer, J.G., Schmitz, B., 1992. Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika* 57, 333–349.
- Motta, G., 2009. Evolutionary factor analysis. Ph.D. Thesis. Institut de statistique, Université catholique de Louvain, Belgium.
- Motta, G., Hafner, C., von Sachs, R., 2011. Locally stationary factor models: identification and nonparametric estimation. *Econometric Theory* 27 (6).
- Neumann, M., 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., Yao, Q., 2008. Modelling multiple time series via common factors. *Biometrika* 95, 365–379.
- Priestley, M.B., Subba Rao, T., 1969. A test for non-stationarity of time series. *Journal of the Royal Statistical Society B* 31, 140–149.
- Stock, J.H., Watson, M.W., 2008. Forecasting in dynamic factor models subject to structural instability. In: Castle, J., Shephard, N. (Eds.), *The Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*. Oxford University Press, Oxford.
- Stock, J.H., Watson, M.W., 2005. Implications of dynamic factor models for VAR analysis. Working Paper Series 11467. National Bureau of Economic Research.
- Wilkinson, J.H., 1965. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford.