Econometric Theory, **27**, 2011, 1279–1319.
doi:10.1017/S0266466611000053

LOCALLY STATIONARY FACTOR MODELS: IDENTIFICATION AND NONPARAMETRIC ESTIMATION

GIOVANNI MOTTA
Maastricht University

CHRISTIAN M. HAFNER AND RAINER VON SACHS
Université Catholique de Louvain

In this paper we propose a new approximate factor model for large cross-section and time dimensions. Factor loadings are assumed to be smooth functions of time, which allows considering the model as *locally stationary* while permitting empirically observed time-varying second moments. Factor loadings are estimated by the eigenvectors of a nonparametrically estimated covariance matrix. As is well known in the stationary case, this principal components estimator is consistent in approximate factor models if the eigenvalues of the noise covariance matrix are bounded. To show that this carries over to our locally stationary factor model is the main objective of our paper. Under simultaneous asymptotics (cross-section and time dimension go to infinity simultaneously), we give conditions for consistency of our estimators. A simulation study illustrates the performance of these estimators.

1. INTRODUCTION

Linear factor models have attracted considerable interest over recent years, especially in the econometrics literature. The intuitively appealing idea of explaining a panel of economic variables by a few common factors is one of the reasons for their popularity. One of the main models in finance, the arbitrage pricing theory (APT) of Ross (1976), is based on a factor model. From a statistical viewpoint, the need to reduce the cross-section dimension to a much smaller factor space dimension is obvious considering the large data sets available in economics and

The authors would like to thank Jushan Bai, Michael Eichler, John Geweke, Domenico Giannone, Marc Hallin, Stefano Herzel, Oliver Linton, Marco Lippi, Alexei Onatski, Franz Palm, Juan M. Rodríguez-Poo, Catalin Stărică, Victor Turchin, and Jean-Pierre Urbain for helpful comments and discussions. We are grateful to two anonymous referees whose comments helped to improve the quality of the paper. We gratefully acknowledge financial support by the Fonds Spéciaux de Recherche (FSR 05) of the Université catholique de Louvain, Louvain-la-Neuve, Belgium. Financial support from the contract "Projet d'Actions de Recherche Concertées" n° 07/12-002 of the "Communauté française de Belgique," granted by the Académie universitaire Louvain, is also gratefully acknowledged. Address correspondence to Giovanni Motta, Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands; e-mail: dr.giovannimotta@gmail.com.

finance. The traditional approach of fixing either the time dimension T or the cross-section dimension N and letting the other dimension go to infinity is likely to be inappropriate in situations where both dimensions are large. Large N, T asymptotics were first rigorously treated in Phillips and Moon (1999), which provides a general treatment of joint asymptotics for nonstationary panel data. The large factor model literature (including Stock and Watson, 2002a, 2002b; Forni, Hallin, Lippi, and Reichlin, 2000, 2005; Forni and Lippi, 2001; Bai, 2003; and Bai and Ng, 2002) use the concept of double asymptotics, where both N and T go to infinity simultaneously without restrictions. This seems to be a promising concept, which we adopt in this paper.

One of the characteristics of the traditional factor model is that the process is stationary in the time dimension. This appears restrictive, given the fact that over long time periods it is unlikely that factor loadings remain constant. In cases of, for example, transitions between recessions and booms, the impact of political crises, or changes in the monetary policy of the central bank, there is a need for a nonstationary approach. For example, in the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), a special case of the APT with just one factor, the market portfolio, typical empirical results show that factor loadings, or betas, are time-varying, which in the CAPM is caused by time-varying second moments.

There is indeed an extensive literature in financial econometrics, both theoretical and empirical, on time-varying variances and correlations. In particular, there are two different specifications in the econometric literature: the prevalent *stationary, conditional* paradigm based on autoregressive conditional heteroskedastic (ARCH-type) processes and the *nonstationary, unconditional* alternative approach.

The *conditional* paradigm is mainly based on the multivariate generalized ARCH (GARCH) framework (see, e.g., Engle, Ng, and Rothschild, 1990; Diebold and Nerlove, 1989; Alexander, 2001). These approaches typically specify univariate GARCH models for the factors and keep the factor loadings constant over time. Rather than imposing GARCH-type models for the factors, other models have been proposed to introduce dynamics into the classical factor model. For example, Forni et al. (2005) suggest a dynamic model with autoregressive factors. A potential drawback of observationally driven dynamic models such as stationary autoregressive moving average (ARMA) or GARCH is that dynamics explained by lagged observations are the same over time. Recently, approaches have been made to relax the constant impact and specify time-varying parameters in GARCH-type models (see, e.g., Härdle, Herwartz, and Spokoiny, 2004). Dahlhaus and Subba Rao (2006) generalized the class of ARCH(∞) models to the nonstationary class of ARCH(∞) models with time-varying coefficients.

A promising alternative is to directly model *unconditional* variances and covariances via nonparametric estimation as in Rodríguez-Poo and Linton (2001). It imposes very little structure on the unconditional covariance matrix, while ensuring positive definiteness and being very easy to estimate. Moreover, it does not impose any restrictions on the number of factors that can be estimated from the time-varying covariance matrices. A combination of parametric GARCH models for the variances and nonparametric estimation of correlations has been proposed by Hafner, van Dijk, and Franses (2006). As shown by Mikosch and Stărică (2004), the empirically often observed long-range dependence in volatility can be explained by a time-varying unconditional variance rather than autocorrelation.

The main task when developing a general nonstationary theory is the problem of asymptotics: A classical asymptotic theory based on the assumption that more and more observations of the future become available does not make sense. Consider the multivariate nonstationary process

$$\boldsymbol{X}_t = \mathbf{S}(t)\,\boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{X}_t$ is the $N \times 1$ vector of observations at time t, $\mathbf{S}$ is a square matrix of dimension N of deterministic functions of t, and $\boldsymbol{\varepsilon}_t$ is the $N \times 1$ vector of errors at time t: $\boldsymbol{\varepsilon}_t \overset{\text{iid}}{\sim} (\mathbf{0}, \mathbf{I}_N)$. Future observations of a general nonstationary process do not necessarily contain any information on the structure of the process at present. To overcome this problem, Dahlhaus (1996, 1997, 2000) has suggested an approach based on the concept of *rescaled time*, i.e., $\frac{t}{T} \in (0, 1)$. Analogously to nonparametric regression, the idea is to set down the asymptotic theory in a way that we "observe" the function $\mathbf{S}(t)$ on a finer grid (but on the same interval),

$$\boldsymbol{X}_{t,T} = \mathbf{S}\left(\tfrac{t}{T}\right)\boldsymbol{\varepsilon}_t, \tag{1}$$

where $\mathbf{S}$ is now defined on the rescaled interval $(0, 1)$ and $\boldsymbol{X}_{t,T}$ becomes a triangular array (the structure of $\boldsymbol{X}$ does not only depend on t, but also on T). Letting T tend to infinity does not mean extending the data to the future anymore. In the rescaled time framework, letting T tend to infinity means that we have in the sample $\boldsymbol{X}_{1,T}, \ldots, \boldsymbol{X}_{T,T}$ more and more 'observations' for each value of $\mathbf{S}$. The model in (1) has been applied by Herzel, Stărică, and Tütüncü (2006) to financial returns. As this model stands, it does not allow for dimension reduction ($\mathbf{S}$ is square), and there is no specification of common and idiosyncratic components.

In this paper we propose a factor model with time-varying factor loadings $\boldsymbol{\Lambda}$,

$$\boldsymbol{X}_{t,T} = \boldsymbol{\Lambda}\left(\tfrac{t}{T}\right)\boldsymbol{F}_t + \boldsymbol{E}_t,$$

with factors $\boldsymbol{F}$ and idiosyncratic components $\boldsymbol{E}$. The basic idea is to consider the loadings as smooth functions of rescaled time, rendering the process nonstationary while the factors remain stationary. However, the assumption that loadings are smooth permits considering the process as locally stationary and enables us to estimate the model using nonparametric methods. The technique employed to *model* the nonstationarity is the rescaling of time to the unit interval, as explained by the example in (1). While the nonparametric estimation of locally stationary time series models is now well established, it is new in the context of locally stationary factor models.

The technique employed to *estimate* the operators of our model is based on the well-known principal components regression (PCR). To adapt this technique to nonstationary factor models, we use time-varying (i.e., a localized version of) PCR.

We discuss identification and estimation of the model and derive a theory for the estimated loadings and factors under simultaneous asymptotics. We show that our estimation procedure achieves convergence rates typical for nonparametric estimation. In a Monte Carlo study, we show that for moderately large N and T, the estimators perform very well.

The paper is organized as follows. In Section 2 we recall the traditional factor model with constant factor loadings and then we introduce the locally stationary factor model and proposes an estimator of the factors and factor loadings. In Section 3 we show the asymptotic properties of our estimates. Section 4 provides a simulation study that confirms the theoretical results of Section 3, and Section 5 concludes. Appendix 5 provides an auxiliary lemma that guarantees the identifiability in estimating the time-varying factor loadings. The proofs of our propositions and theorems are given in Appendix B.

Throughout the paper we use bold unslanted letters for matrices, bold slanted letters for vectors, and unbold (normal) letters for scalars. We denote by $\mathrm{tr}(\cdot)$ the trace operator, by $\mathrm{rk}(\mathbf{A})$ the rank of a matrix $\mathbf{A}$, by $\mathbf{I}_n$ the identity matrix of dimension n, by $\otimes$ the Kronecker product, by $\perp\!\!\!\perp$ the stochastic independence, and by $\|\cdot\|$ the Frobenius (euclidean) norm, i.e., $\|\mathbf{A}\| = \sqrt{\mathrm{tr}(\mathbf{A}'\mathbf{A})}$.

2. LOCALLY STATIONARY FACTOR MODELS

To better understand the properties of our locally stationary factor model, we remind the reader of the definition of a stationary factor model. Consider the following r-factor model for the N-dimensional stochastic process $\{\boldsymbol{X}_{t,N}, t \in \mathbb{N}, N \in \mathbb{N}\}$:

$$\boldsymbol{X}_{t,N} = \boldsymbol{C}_{t,N} + \boldsymbol{E}_{t,N} = \boldsymbol{\Lambda}_N \boldsymbol{F}_t + \boldsymbol{E}_{t,N}, \tag{2}$$

where $\boldsymbol{C}_{t,N} := \boldsymbol{\Lambda}_N \boldsymbol{F}_t$ is the $N \times 1$ vector of common components with covariance matrix $\boldsymbol{\Sigma}_N^C$, $\boldsymbol{\Lambda}_N$ is an $N \times r$ matrix of loadings, the $r \times 1$ vector of factors $\boldsymbol{F}_t$ is zero mean with covariance matrix $\boldsymbol{\Sigma}^F$, and the $N \times 1$ vector of idiosyncratic errors $\boldsymbol{E}_{t,N}$ is zero mean with covariance matrix $\boldsymbol{\Sigma}_N^E$. We consider the case where the dimension N is large, say 100 or larger, but the number of factors r is small, say two or three. The idea of the factor model is to explain a substantial part of the variation of $\boldsymbol{X}_{t,N}$ by some factors common to all components of $\boldsymbol{X}_{t,N}$, and some idiosyncratic error term $\boldsymbol{E}_{t,N}$ that is not correlated with the factors and that captures variable-specific variations. In the classical factor model, $\boldsymbol{\Sigma}_N^E$ is assumed to be diagonal, but as we will see it is useful to relax this assumption when dealing with large dimensions. The model implies first that the mean of $\boldsymbol{X}_{t,N}$ is zero, but this is without loss of generality. Furthermore, the covariance matrix of $\boldsymbol{X}_{t,N}$ is given by

$$\boldsymbol{\Sigma}_N := \mathbb{V}\mathrm{ar}\left[\boldsymbol{X}_{t,N}\right] = \boldsymbol{\Sigma}_N^C + \boldsymbol{\Sigma}_N^E = \boldsymbol{\Lambda}_N \boldsymbol{\Sigma}^F \boldsymbol{\Lambda}_N' + \boldsymbol{\Sigma}_N^E,$$

which does not depend on t. The part of the variance explained by the factors is $\mathbf{\Lambda}_N \mathbf{\Sigma}^F \mathbf{\Lambda}'_N$; the remainder $\mathbf{\Sigma}_N^E$ is the part due to idiosyncratic noise.

We now generalize the definition of a factor model to allow for time-varying factor loadings. We assume that, for all N fixed, there exists a function $\mathbf{\Lambda}_N(u)$ defined on the unit interval $u \in (0,1)$ such that

$$\boldsymbol{X}_{t,NT} = \boldsymbol{C}_{t,NT} + \boldsymbol{E}_{t,N} = \mathbf{\Lambda}_N\left(\tfrac{t}{T}\right)\boldsymbol{F}_t + \boldsymbol{E}_{t,N}, \qquad t = 1,\ldots,T. \tag{3}$$

The process $\{\boldsymbol{X}_{t,NT}\}$ is in fact a sequence (triangular array) of doubly indexed processes because the loadings are defined in rescaled time $\frac{t}{T} \in (0,1)$ as in Dahlhaus (1997). In (3) the loadings $\mathbf{\Lambda}_N(u) := \left\{\lambda_{ij}(u)\right\}$, $i = 1,\ldots,N$, $j = 1,\ldots,r$, are assumed to be smooth functions of rescaled time $u \in (0,1)$. Hence, recalling that $\mathbf{\Sigma}_N^E = \mathbb{V}\mathrm{ar}\left[\boldsymbol{E}_{t,N}\right]$, we can define for all $u \in (0,1)$ a matrix-valued smooth function $\mathbf{\Sigma}_N(u)$ given by

$$\mathbf{\Sigma}_N(u) = \mathbf{\Sigma}_N^C(u) + \mathbf{\Sigma}_N^E, \tag{4}$$

where

$$\mathbf{\Sigma}_N^C(u) := \mathbf{\Lambda}_N(u)\mathbf{\Sigma}^F\mathbf{\Lambda}'_N(u), \tag{5}$$

such that the covariance matrix of $\boldsymbol{X}_{t,NT}$ can be defined as

$$\mathbf{\Sigma}_N\left(\tfrac{t}{T}\right) := \mathbb{V}\mathrm{ar}\left[\boldsymbol{X}_{t,NT}\right] = \mathbf{\Sigma}_N^C\left(\tfrac{t}{T}\right) + \mathbf{\Sigma}_N^E, \qquad t = 1,\ldots,T, \tag{6}$$

where $\mathbf{\Sigma}_N^C\left(\frac{t}{T}\right) := \mathbf{\Lambda}_N\left(\frac{t}{T}\right)\mathbf{\Sigma}^F\mathbf{\Lambda}'_N\left(\frac{t}{T}\right)$ is the time-varying covariance matrix of the common components $\boldsymbol{C}_{t,NT}$, $t = 1,\ldots,T$.

The class of locally stationary factor models, which includes the stationary factor model (2) as a special case, is defined as follows.

DEFINITION 1 (Locally stationary approximate factor model). *The sequence $\left\{\boldsymbol{X}_{t,NT}\right\}$ in (3) is a locally stationary approximate factor model (LSAFM) if*

(i) *there exists for all $N \geq 1$ a function*

$$\mathbf{\Lambda}_N(u) \in \mathcal{C}^1\left\{[0,1], \mathbb{R}^{N\times r}\right\}, \qquad \mathrm{rk}\{\mathbf{\Lambda}_N(u)\} = r \ \forall u \in (0,1),$$

such that for all T

$$\mathbf{\Sigma}_N\left(\tfrac{t}{T}\right) = \mathbf{\Lambda}_N\left(\tfrac{t}{T}\right)\mathbf{\Sigma}^F\mathbf{\Lambda}'_N\left(\tfrac{t}{T}\right) + \mathbf{\Sigma}_N^E;$$

(ii) $\mathbf{\Sigma}_N^E := \mathbb{V}\mathrm{ar}\left[\boldsymbol{E}_{t,N}\right]$ *is a sequence of positive definite matrices with uniformly bounded eigenvalues; that is,*

$$\sup_N \mathrm{v}_{1N}^E < \infty, \qquad N = 1,2,\ldots,$$

where v_{1N}^E denotes the largest eigenvalue of the matrix $\mathbf{\Sigma}_N^E$.

We aim to estimate the covariances $\mathbf{\Sigma}_N(u)$, the loadings $\mathbf{\Lambda}_N(u)$, the factors $\boldsymbol{F}_t$, and the common components $\boldsymbol{C}_{t,NT}$. For this we need double asymptotics: both $T \to \infty$ and $N \to \infty$.

First, as in the stationary case, we need that $T \to \infty$ to estimate $\mathbf{\Sigma}_N(u)$ consistently. In the locally stationary framework, the targets are defined in rescaled time, $u \in (0,1)$. There exists a link between the estimation of smooth functions defined in real time, such as $\mathbf{\Sigma}_N(u)$ and $\mathbf{\Lambda}_N(u)$, and stochastic processes defined in discrete time, such as $\boldsymbol{F}_t$ and $\boldsymbol{C}_{t,NT}$. For a given sample size T and given $u \in (0,1)$ in rescaled time, there exists a sequence t_T such that $t_T/T \to u$ as $T \to \infty$. For example, $t_T = [uT]$, where $[x]$ is the largest integer smaller than or equal to x. Consequently, due to the continuity of $\mathbf{\Lambda}_N(u)$ in u, the covariance matrix $\mathbf{\Sigma}_N\left(\frac{t}{T}\right)$ in (6) converges to the matrix $\mathbf{\Sigma}_N(u)$ in (4) as $T \to \infty$.

Second, we need $N \to \infty$ because the PCR estimator is biased if N is fixed and $T \to \infty$, unless $\mathbf{\Sigma}_N^E = \sigma^2 \mathbf{I}_N$ for some $\sigma^2 > 0$. The model allowing for weak cross-correlation of the errors $\boldsymbol{E}_{t,N}$ in the cross-section dimension, i.e., for a nondiagonal covariance matrix $\mathbf{\Sigma}_N^E$, is usually called the *approximate factor model* (AFM); see, e.g., Chamberlain and Rothschild (1983), Stock and Watson (2002a, 2002b) and Bai (2003). For approximate factor models, Bai (2003) has proved that under $N \to \infty$ and $T \to \infty$ the principal components estimator is consistent and asymptotically normal. The PCR estimator is consistent under the assumption that $\mathbf{\Sigma}_N^E$ is a sequence of matrices with uniformly bounded eigenvalues. More precisely, exactly r of the eigenvalues of the sequence of covariance matrices $\mathbf{\Sigma}_N$ increase without bound, and all the other eigenvalues of $\mathbf{\Sigma}_N$ are bounded. In this work we assume that the factors are orthogonal ($\mathbf{\Sigma}^F$ is diagonal), serially independent, and independent from the errors ($\mathbb{C}\text{ov}[\boldsymbol{F}_t, \boldsymbol{E}_t] = 0$ for all t), but we allow for loadings that change over time. We assume that the errors are correlated in the cross-section dimension but not in the time dimension.

Remark 1. The common components $\boldsymbol{C}_{t,NT} := \mathbf{\Lambda}_N\left(\frac{t}{T}\right)\boldsymbol{F}_t$ in model (3) are *locally stationary* in the sense of Definition 2.1 of Dahlhaus (1997). The rescaled time principle is a tool for local estimation: We get more and more observations for the local structure of the loadings $\mathbf{\Lambda}_N\left(\frac{t}{T}\right)$ at each time point. Our estimation target has to be defined properly, which is not the case if we do not rescale in time. A process with time-varying parameters, like those in Fancourt and Principe (1998), Mikosch and Stărică (2004), and Stărică and Granger (2005), is not necessarily locally stationary in the sense of Definition 2.1 of Dahlhaus (1997, 2000).

2.1. Assumptions of the LSAFM

We list here the assumptions of model (3). In terms of model identification, we need the double asymptotics: $T \to \infty$ (for the local stationarity) and $N \to \infty$ (for the approximate factor modeling). In Assumptions A, B, and C we introduce three constants: M_1 and M_2 are needed to bound the fourth moments of the factors and the idiosyncratic errors, respectively, whereas M_3 is needed to bound the eigenvalues of $\boldsymbol{\Sigma}_N^E$.

Assumption A (Factors).

1. $\boldsymbol{F}_t \overset{\text{iid}}{\sim} \left(\mathbf{0}, \boldsymbol{\Sigma}^F\right)$, where the $r \times r$ matrix $\boldsymbol{\Sigma}^F$ is diagonal and positive definite;
2. $\mathbb{E}\|\boldsymbol{F}_t\|^4 \leq M_1 < \infty$.

Assumption B (Factor loadings).

1. $\lambda_{ij}(u) \in \mathcal{C}^1[0,1]\ \forall\, i = 1,\ldots,N,\ j = 1,\ldots,r$;
2. $\sup_{u\in(0,1)} \|\boldsymbol{\lambda}_i(u)\| \leq \bar{\lambda}_0 < \infty$ for all i, and $\sup_{u\in(0,1)} \left\| \frac{\boldsymbol{\Lambda}_N'(u)\boldsymbol{\Lambda}_N(u)}{N} - \boldsymbol{\Sigma}^{\Lambda}(u) \right\| \to 0$ as $N \to \infty$, where $\boldsymbol{\lambda}_i(u)$ is the ith row of $\boldsymbol{\Lambda}_N(u)$ and $\boldsymbol{\Sigma}^{\Lambda}(u)$ is an $r \times r$ positive definite matrix for all u;
3. $\sup_{u\in(0,1)} \left\|\boldsymbol{\lambda}_i^{(1)}(u)\right\| \leq \bar{\lambda}_1 < \infty$ for all i, and $\sup_{u\in(0,1)} \left\| \frac{\boldsymbol{\Lambda}_N^{(1)\prime}(u)\boldsymbol{\Lambda}_N^{(1)}(u)}{N} - \boldsymbol{\Sigma}^{\Lambda^{(1)}}(u) \right\| \to 0$ as $N \to \infty$, where $\boldsymbol{\Lambda}_N^{(1)}(u)$ is the first derivative of $\boldsymbol{\Lambda}_N(u)$, $\boldsymbol{\lambda}_i^{(1)}(u)$ is the ith row of $\boldsymbol{\Lambda}_N^{(1)}(u)$, and $\boldsymbol{\Sigma}^{\Lambda^{(1)}}(u)$ is an $r \times r$ positive definite matrix for all u.

Assumption C (Idiosyncratic errors).

1. $\boldsymbol{E}_{t,N} \overset{\text{iid}}{\sim} \left(\mathbf{0}, \boldsymbol{\Sigma}_N^E\right)$, where $\boldsymbol{E}_{t,N} = \{E_{1t},\ldots,E_{it},\ldots,E_{Nt}\}$;
2. $\mathbb{E}|E_{it}|^4 \leq M_2$ for all $i = 1,\ldots,N$, all $t = 1,\ldots,T$, all $N \in \mathbb{N}$ and all $T \in \mathbb{N}$;
3. $\mathbb{E}(E_{it}E_{jt}) = \sigma_{ij}^E$, $\max_{1\leq i\leq N} \sum_{j=1}^{N} |\sigma_{ij}^E| \leq M_3$, for all $j = 1,\ldots,N$ and all $N \in \mathbb{N}$;
4. $E_{it} \perp\!\!\!\perp F_{jt-k}$ for any i, j, t, and k.

Assumption A1 says that the factors are zero mean, serially independent, and identically distributed. That is, $\mathbb{E}(\boldsymbol{F}_t) = \mathbf{0}$, and for any $k \neq 0$ $F_{it} \perp\!\!\!\perp F_{jt-k}$, for any i, j and t. The fact that $\boldsymbol{\Sigma}^F$ is diagonal means that $\mathbb{C}\text{ov}\left(F_{it}, F_{jt}\right) = 0$ for $i \neq j$, $i, j = 1,\ldots,r$. Assumption A2 says that the factors have finite fourth moments.

In Assumption B1 the model quantities $\lambda_{ij}(u)$ are supposed to be in $\mathcal{C}^1[0,1]$. This condition is given for ease of presentation of the proofs. On the one hand, we note that we could also suppose the time-varying factor loadings to be more regular, i.e., to be in $\mathcal{C}^k[0,1]$ with $k > 1$, if we were interested in deriving optimal rates of convergence of nonparametric estimators of these model quantities. However, in this work we content ourselves to prove consistency of our estimators. On the other hand, we also suspect that even a weaker condition on the regularity of the factor loading, such as Hölder-continuity, would be sufficient to derive our results. The first condition in Assumption B2 ensures that each factor has a nontrivial contribution to the variance of $\boldsymbol{X}_{t,NT}$, and the second condition implies that $\sup_u \|\boldsymbol{\Lambda}_N(u)\| = O\left(\sqrt{N}\right)$. Assumption B3 implies that

$\sup_u \left\|\boldsymbol{\Lambda}_N^{(1)}(u)\right\| = O\left(\sqrt{N}\right)$. This condition parallels the previous Assumption B2, now expressed for the first derivatives of the loading. Although it is mainly technical and essentially convenient for our proofs, it ensures that in matrix norm the contribution to the not-too-abrupt change of the time-varying loadings is homogeneously distributed over all factor directions.

Chamberlain and Rothschild (1983) defined an AFM as having bounded eigenvalues for the $N \times N$ covariance matrix $\boldsymbol{\Sigma}_N^E = \mathbb{E}\left(\boldsymbol{E}_{t,N}\boldsymbol{E}'_{t,N}\right)$. If $\boldsymbol{E}_{t,N}$ is stationary with $\mathbb{E}\left(E_{it}E_{jt}\right) = \sigma_{ij}^E$ for all t, then the largest eigenvalue of $\boldsymbol{\Sigma}_N^E$ is bounded by $\max_{1\le i\le N}\sum_{j=1}^N |\sigma_{ij}^E|$; see, e.g., Lütkepohl (1996). Thus by Assumption C3 model (3) will be an AFM in the sense of Chamberlain and Rothschild. In this paper we assume independence between the factors and the error terms as in Forni et al. (2000) (Assumption C4). Although this appears to be restrictive, it is actually a necessary condition for identification (see Forni and Lippi, 2001, Thm. 4).

2.2. Time-Varying PCR

The estimation of the loadings and the factors in model (3) can be seen as the solution of a weighted least squares criterion. We introduce the problem in the stationary case of model (2), where $\boldsymbol{\Lambda}_N$ and $\boldsymbol{F}_t$ can be estimated by minimizing the nonlinear least squares objective function

$$L_{NT} = (NT)^{-1}\sum_{t=1}^{T}\left(\boldsymbol{X}_{t,N} - \boldsymbol{\Lambda}_N\boldsymbol{F}_t\right)'\left(\boldsymbol{X}_{t,N} - \boldsymbol{\Lambda}_N\boldsymbol{F}_t\right), \tag{7}$$

subject to

$$\boldsymbol{\Lambda}'_N\boldsymbol{\Lambda}_N/N = \mathbf{I}_r, \tag{8}$$

where $\mathbf{F}_T = \{\boldsymbol{F}_1,\ldots,\boldsymbol{F}_T\}'$ is the $T\times r$ matrix containing the factors. Maximizing (7) is equivalent to maximizing

$$N^{-2}\,\mathrm{tr}\left\{\boldsymbol{\Lambda}'_N\mathbf{S}_{NT}\boldsymbol{\Lambda}_N\right\}$$

subject to (8), where $\mathbf{S}_{NT}$ is the sample covariance matrix estimator:

$$\mathbf{S}_{NT} := \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_{t,N}\boldsymbol{X}'_{t,N}. \tag{9}$$

The well-known solution is to set the estimator $\widehat{\boldsymbol{\Lambda}}_N$ of $\boldsymbol{\Lambda}_N$ to be $\sqrt{N}$ times the matrix whose columns are the r orthogonal unit-length eigenvectors $\overline{\boldsymbol{\Lambda}}_N$ corresponding to the largest r-ordered eigenvalues of $\mathbf{S}_{NT}$, that is,

$$\begin{aligned} &N^{-1}\mathbf{S}_{NT}\overline{\boldsymbol{\Lambda}}_N = \overline{\boldsymbol{\Lambda}}_N\widehat{\mathbf{V}}_{NT}, \qquad \overline{\boldsymbol{\Lambda}}'_N\overline{\boldsymbol{\Lambda}}_N = \mathbf{I}_r,\\ &\widehat{\boldsymbol{\Lambda}}_N = \sqrt{N}\,\overline{\boldsymbol{\Lambda}}_N, \qquad \widehat{\boldsymbol{\Lambda}}'_N\widehat{\boldsymbol{\Lambda}}_N/N = \mathbf{I}_r, \end{aligned} \tag{10}$$

where the matrix

$$\widehat{\mathbf{V}}_{NT} = T^{-1}\sum_{t=1}^{T}\widehat{\boldsymbol{F}}_{t,N}\widehat{\boldsymbol{F}}'_{t,N} \tag{11}$$

is the diagonal matrix containing the largest r eigenvalues of $\mathbf{S}_{NT}$ in decreasing order. The resulting estimator of the factors is $\widehat{\boldsymbol{F}}_{t,N} = \widehat{\boldsymbol{\Lambda}}'_N \boldsymbol{X}_{t,N}/N$, which is the vector consisting of the first r principal components of $\boldsymbol{X}_{t,N}$.

We now show that a weighted version (weighted over time) of the loss in (7) applied to the nonstationary model in (3) leads to the nonparametric version of the solutions given in (10) and (9). Let $K(\cdot)$ be a positive kernel over time of finite second moment and symmetric around zero,

$$\int K(u)\,du = 1, \qquad \int K^2(u)\,du < \infty, \qquad K(u) = K(-u),$$

and let $K_h(\cdot) := \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ be the rescaled version of K for a given bandwidth $h > 0$. Define for a fixed point $u \in (0,1)$ the time-varying weighted least squares objective function

$$\begin{aligned} L_{NT}(u;h) = (NT)^{-1}\sum_{s=1}^{T}&\left(\boldsymbol{X}_{s,NT} - \boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right)\boldsymbol{F}_s\right)' \\ &\times \left(\boldsymbol{X}_{s,NT} - \boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right)\boldsymbol{F}_s\right) K_h\left(\tfrac{s}{T} - u\right), \end{aligned} \tag{12}$$

subject to

$$\frac{\boldsymbol{\Lambda}'_N\left(\frac{s}{T}\right)\boldsymbol{\Lambda}_N\left(\frac{s}{T}\right)}{N} = \mathbf{I}_r, \qquad \forall s = 1,\ldots,T, \qquad \forall T \in \mathbb{N}. \tag{13}$$

Without loss of generality, for reasons of ease of presentation we use the same bandwidth h for all $i,j = 1,\ldots,N$. The first-order conditions for maximizing (12) with respect to $\boldsymbol{F}_s$ are

$$\boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right)\widehat{\boldsymbol{F}}_{s,NT} = \boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT},$$

and thus, with the constraint (13), the estimator of the factor satisfies $\forall s = 1,\ldots,T$,

$$\widehat{\boldsymbol{F}}_{s,NT} = \left\{\boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right)\right\}^{-1}\boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT} = N^{-1}\boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT}. \tag{14}$$

Substituting (14) into the objective function (12) yields the concentrated objective function

$$\begin{aligned} \widetilde{L}_{NT}(u;h) = \frac{1}{NT}\sum_{s=1}^{T}\operatorname{tr}&\left\{\boldsymbol{X}_{s,NT}\boldsymbol{X}'_{s,NT} - \frac{1}{N}\boldsymbol{\Lambda}'_N\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT}\boldsymbol{X}'_{s,NT}\boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right)\right\} \\ &\times K_h\left(\tfrac{s}{T} - u\right). \end{aligned} \tag{15}$$

Minimizing the concentrated objective function (15) with respect to $\mathbf{\Lambda}(\frac{s}{T})$ is equivalent to maximizing

$$(NT)^{-1}\sum_{s=1}^{T}\mathrm{tr}\left\{\frac{1}{N}\mathbf{\Lambda}_N'\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT}\boldsymbol{X}_{s,NT}'\mathbf{\Lambda}_N\left(\tfrac{s}{T}\right)\right\}K_h\left(\tfrac{s}{T}-u\right) \tag{16}$$

subject to (13). If we assume that the sequence $h_T \to 0$ as $T \to \infty$, for large values of T the weight $K_h\left(\frac{s}{T}-u\right)$ only takes into account the loadings $\mathbf{\Lambda}_N\left(\frac{s}{T}\right)$ corresponding to the values $\frac{s}{T}$ that are very close to u. Asymptotically, the loadings that minimize the concentrated objective function in (16) depend only on u. This result is formalized in Proposition 1.

PROPOSITION 1. *Under Assumptions A–C, for all* $u \in (0,1)$*,*

$$(NT)^{-1}\sum_{s=1}^{T}\mathrm{tr}\left\{\frac{1}{N}\mathbf{\Lambda}_N'\left(\tfrac{s}{T}\right)\boldsymbol{X}_{s,NT}\boldsymbol{X}_{s,NT}'\mathbf{\Lambda}_N\left(\tfrac{s}{T}\right)-\frac{1}{N}\mathbf{\Lambda}_N'(u)\boldsymbol{X}_{s,NT}\boldsymbol{X}_{s,NT}'\mathbf{\Lambda}_N(u)\right\}$$
$$\times K_{h_T}\left(\tfrac{s}{T}-u\right)=O_p\left(h_T\right).$$

Proof. See Appendix B. ∎

By Proposition 1, maximizing (16) is asymptotically equivalent to maximizing

$$N^{-2}\,\mathrm{tr}\left\{\mathbf{\Lambda}_N'(u)\widehat{\mathbf{\Sigma}}_N(u)\mathbf{\Lambda}_N(u)\right\}$$

subject to (13), where

$$\widehat{\mathbf{\Sigma}}_N\left(u;h_T\right):=T^{-1}\sum_{s=1}^{T}\boldsymbol{X}_{s,NT}\boldsymbol{X}_{s,NT}'\,K_{h_T}\left(\tfrac{s}{T}-u\right), \tag{17}$$

which is the nonparametric estimator of the covariance matrix proposed by Rodríguez-Poo and Linton (2001). The solution is to set $\widehat{\mathbf{\Lambda}}_N(u)$ to be $\sqrt{N}$ times the matrix whose columns are the r orthogonal unit-length eigenvectors $\overline{\mathbf{\Lambda}}_N(u)$ corresponding to the largest r eigenvalues of $\widehat{\mathbf{\Sigma}}_N(u)$. As in the stationary case, by analogous arguments we obtain

$$N^{-1}\widehat{\mathbf{\Sigma}}_N\left(u;h\right)\widehat{\mathbf{\Lambda}}_N(u;h)=\widehat{\mathbf{\Lambda}}_N(u;h)\widehat{\mathbf{V}}_N(u;h),$$
$$\frac{\widehat{\mathbf{\Lambda}}_N'(u)\widehat{\mathbf{\Lambda}}_N(u)}{N}=\mathbf{I}_r,\qquad \forall u\in(0,1), \tag{18}$$

where $\widehat{\mathbf{V}}_N(u)$ is the diagonal matrix containing the largest r-ordered eigenvalues of $N^{-1}\widehat{\mathbf{\Sigma}}_N\left(u;h\right)$. Note that (17) and (18) generalize (9) and (10), respectively.

It is useful to write the LSAFM in a more compact matrix notation. Notice that $N^{-1}\widehat{\mathbf{\Sigma}}_N\left(u;h\right)=(NT)^{-1}\mathbf{X}_{NT}'\mathbf{W}_T(u;h)\mathbf{X}_{NT}$, where the $T\times T$ matrix of weights $\mathbf{W}_T(u;h)$ is defined as

$$\mathbf{W}_T(u;h)=\mathrm{diag}\left\{K_h\left(\tfrac{1}{T}-u\right),K_h\left(\tfrac{2}{T}-u\right),\ldots,K_h\left(\tfrac{T-1}{T}-u\right),K_h\left(1-u\right)\right\}$$

for all $u \in (0,1)$, and the $T \times N$ matrix $\mathbf{X}_{NT}$ collects the data: $\mathbf{X}_{NT} = \left[X_{1,NT}, \ldots, X_{t,NT}, \ldots, X_{T,NT}\right]'$. This means that, for all $u \in (0,1)$,

$$\frac{1}{NT}\mathbf{X}_{NT}'\mathbf{W}_T(u;h)\mathbf{X}_{NT}\widehat{\mathbf{\Lambda}}_N(u) = \widehat{\mathbf{\Lambda}}_N(u)\widehat{\mathbf{V}}_N(u),$$
$$\widehat{\mathbf{\Lambda}}_N'(u)\widehat{\mathbf{\Lambda}}_N(u)/N = \mathbf{I}_r; \tag{19}$$

that is,

$$\widehat{\mathbf{V}}_N(u;h) = N^{-1}\widehat{\mathbf{\Lambda}}_N'(u;h)\frac{1}{NT}\mathbf{X}_{NT}'\mathbf{W}_T(u;h)\mathbf{X}_{NT}\widehat{\mathbf{\Lambda}}_N(u;h),$$
$$\widehat{\mathbf{\Lambda}}_N'(u;h)\widehat{\mathbf{\Lambda}}_N(u;h)/N = \mathbf{I}_r. \tag{20}$$

If we define $\overline{\mathbf{F}}_{t,NT} = \mathbf{X}_{NT}\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)/N$, the result in (20) can be written as

$$\widehat{\mathbf{V}}_N\left(\tfrac{t}{T}\right) = \overline{\mathbf{F}}_{t,NT}'\mathbf{W}_T\left(\tfrac{t}{T};h\right)\overline{\mathbf{F}}_{t,NT}/T, \qquad t = 1,\ldots,T.$$

The matrix $\widehat{\mathbf{V}}_N\left(\frac{t}{T}\right)$ is a localized version of the matrix $\widehat{\mathbf{V}}_{NT}$ defined in (11). In the stationary case the estimated factors contain the same information for all t, while in the locally stationary framework the weights depend on time: They are contained in the matrix $\mathbf{W}_T\left(\frac{t}{T};h\right)$. Notice that the tth column of the $r \times T$ matrix $\overline{\mathbf{F}}_{t,NT}'$ is $\widehat{F}_{t,NT}$ as defined in (14). Indeed, the estimated factors $\widehat{F}_{t,NT}$ in (14) can be written as

$$\widehat{F}_{t,NT} = \widehat{\mathbf{\Lambda}}_N'\left(\tfrac{t}{T}\right)X_{t,NT}/N \qquad t = 1,\ldots,T, \tag{21}$$

since $\widehat{\mathbf{\Lambda}}_N'\left(\frac{s}{T}\right)\widehat{\mathbf{\Lambda}}_N\left(\frac{s}{T}\right)/N = \mathbf{I}_r$ for all $s = 1,\ldots,T$.

Remark 2. The condition $\left\|\frac{\mathbf{\Lambda}'\mathbf{\Lambda}}{N} - \mathbf{\Sigma}^{\Lambda}\right\| \to 0$ is a condition on the behavior of the loadings, while the condition $\frac{\widehat{\mathbf{\Lambda}}'\widehat{\mathbf{\Lambda}}}{N} = \mathbf{I}_r$ is an ingredient for the minimization of the loss functions defined in (7) and (12) of our paper. This condition is exactly the same as in Bai (2003), where the ingredient above becomes $\frac{\widehat{\mathbf{F}}_T'\widehat{\mathbf{F}}_T}{T} = \mathbf{I}_r$ because the matrix to diagonalize is $\mathbf{X}_{NT}\mathbf{X}_{NT}'$ instead of our $\mathbf{X}_{NT}'\mathbf{W}_T(u;h)\mathbf{X}_{NT}$.

3. ASYMPTOTIC ESTIMATION THEORY

The following asymptotic results hold for $N, T \to \infty$ and for a fixed value of $u \in (0,1)$. The proofs are given in Appendix B. Our results hold under the following assumption.

Assumption D (Rate of the bandwidth). Assume $h_T \to 0$ such that $Th_T \to \infty$ and $Th_T^3 \to 0$ as T tends to infinity.

In Assumption D we control the rates of convergence of the bandwidth h_T. We first have the usual requirement in nonparametric curve estimation that the bandwidth goes to zero but at a slower rate than T^{-1}. Second, to control a “multivariate bias" in smoothing a matrix of factor loadings that can grow with order N, we require $Th^3 \to 0$ (see the proof of Theorem 1), in order not to need to correct for the bias due to nonstationarity.

The first result is about consistency of our nonparametric estimator of the time-varying covariance matrix $\mathbf{\Sigma}_N(u)$.

THEOREM 1. *Under Assumptions A–D,*

$$N^{-1}\left\|\widehat{\mathbf{\Sigma}}_N(u;h_T)-\mathbf{\Sigma}_N(u)\right\| = O_p\left[(Th_T)^{-1/2}\right],$$

where $\widehat{\mathbf{\Sigma}}_N(u;h_T)$ *is defined in (17) and* $\mathbf{\Sigma}_N(u)$ *is defined in (4).*

Proof. See Appendix B. ■

The estimator $\widehat{\mathbf{\Sigma}}_N(u;h_T)$ converges to $\mathbf{\Sigma}_N(u)$ at the rate $\frac{1}{\sqrt{Th_T}}$. This means that each element of the matrix $\left[\widehat{\mathbf{\Sigma}}_N(u;h_T)-\mathbf{\Sigma}_N(u)\right]$ is $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. The fact that the euclidean norm of this $N\times N$ matrix grows at the rate N explains the result. The result in Theorem 1 generalizes the consistency of the sample covariance matrix in (9). Similarly to the stationary case, the estimation of the covariance matrix is crucial for factor analysis. It is the first step when estimating factor models, and it is from this estimator that all the other estimators are derived. The consistency of the estimators of the factors and the loadings depends on the consistency of the eigenvalues (see Proposition 2 and Corollary 1) and the eigenvectors (see Theorem 2) of the estimator $\widehat{\mathbf{\Sigma}}_N(u)$ of $\mathbf{\Sigma}_N(u)$. In what follows we denote by ℓ_N the *loadings scale error*, that is, the rate of convergence of the rescaled loadings to the matrix $\mathbf{\Sigma}^{\Lambda}(u)$. To satisfy Assumption B2, we only need that $\ell_N \to 0$ as $N\to\infty$.

PROPOSITION 2. *Under Assumptions A–D,*

$$\min\left(\sqrt{Th_T},N,\ell_N^{-1}\right)\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}(u)\right\| = O_p(1),$$

where $\widehat{\mathbf{V}}_N(u;h_T)$ *is defined in (20),* $\mathbf{V}(u)$ *is the diagonal matrix containing the eigenvalues of* $\mathbf{\Sigma}^{\Lambda}(u)\mathbf{\Sigma}^{F}$, *and*

$$\ell_N := \sup_{u\in(0,1)}\left\|\frac{\mathbf{\Lambda}_N'(u)\mathbf{\Lambda}_N(u)}{N}-\mathbf{\Sigma}^{\Lambda}(u)\right\|. \tag{22}$$

Proof. We can decompose the overall error $\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}(u)\right\|$ as the following:

(i) *estimation error:* $\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N(u)\right\|$, where $\mathbf{V}_N(u)$ is the diagonal matrix with the largest r eigenvalues of the matrix $\frac{1}{N}\mathbf{\Sigma}_N(u)$, $\mathbf{\Sigma}_N(u)$ being defined in (4);

(ii) *approximation error:* $\left\|\mathbf{V}_N(u)-\mathbf{V}(u)\right\|$.

In Appendix B we show that the estimation error is $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$, whereas the approximation error is $O\left(N^{-1}\right)+O\left(\ell_N\right)$. Then we have

$$\begin{aligned}\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}(u)\right\| &\le \left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N(u)\right\|+\left\|\mathbf{V}_N(u)-\mathbf{V}(u)\right\| \\ &= O_p\left(\frac{1}{\sqrt{Th_T}}\right)+O\left(\frac{1}{N}\right)+O\left(\ell_N\right). \qquad \blacksquare\end{aligned}$$

Remark 3 (**Estimation error**). The estimation error depends on the difference between the largest r eigenvalues of $\frac{1}{N}\widehat{\boldsymbol{\Sigma}}_N(u;h_T)$ and the largest r eigenvalues of $\frac{1}{N}\boldsymbol{\Sigma}_N(u)$. It tends to zero by Theorem 1; that is,

$$\begin{aligned}&\sqrt{Th_T}\left\|\frac{1}{N}\left[\widehat{\boldsymbol{\Sigma}}_N(u;h_T)-\boldsymbol{\Sigma}_N(u)\right]\right\|=O_p(1)\\ &\quad\Longrightarrow \min\left(\sqrt{Th_T},N,\ell_N^{-1}\right)\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N(u)\right\|=O_p(1). \qquad \textbf{(23)}\end{aligned}$$

In the stationary case, Kollo and Neudecker (1993) derive a similar result with a parametric rate of convergence. Given the $p\times p$ covariance matrix $\boldsymbol{\Sigma}$ and its eigenvalues $\mathrm{v}_1,\ldots,\mathrm{v}_p$, we have that

$$\sqrt{T}\left\|\mathbf{S}_T-\boldsymbol{\Sigma}\right\|=O_p(1)\Longrightarrow\sqrt{T}\left\|\widehat{\mathbf{V}}-\mathbf{V}\right\|=O_p(1), \qquad \textbf{(24)}$$

where $\mathbf{V}=\mathrm{diag}\left\{\mathrm{v}_1,\ldots,\mathrm{v}_p\right\}$, $\mathbf{S}_T$ is the sample covariance matrix and $\widehat{\mathbf{V}}=\mathrm{diag}\left\{\widehat{\mathrm{v}}_1,\ldots,\widehat{\mathrm{v}}_p\right\}$ contains the eigenvalues of $\mathbf{S}_T$ (the dimension p is finite). The result in (24) has been generalized by Rodríguez-Poo and Linton (2001, Props. 3.2 and 3.3) to the nonstationary case with a nonparametric rate of convergence; that is,

$$\begin{aligned}&\sqrt{Th_T}\left\|\widehat{\boldsymbol{\Sigma}}(u;h_T)-\boldsymbol{\Sigma}(u)\right\|=O_p(1)\Longrightarrow\\ &\quad\sqrt{Th_T}\left\|\widehat{\mathbf{V}}(u;h_T)-\mathbf{V}(u)\right\|=O_p(1), \qquad \textbf{(25)}\end{aligned}$$

where $\boldsymbol{\Sigma}(u)$ is the covariance matrix of a nonstationary process, $\widehat{\boldsymbol{\Sigma}}(u;h_T)$ is its estimator, and the matrices $\mathbf{V}(u)$ and $\widehat{\mathbf{V}}(u;h_T)$ contain the eigenvalues of $\boldsymbol{\Sigma}(u)$ and $\widehat{\boldsymbol{\Sigma}}(u;h_T)$, respectively (no need to rescale $\widehat{\boldsymbol{\Sigma}}(u;h_T)$ and $\boldsymbol{\Sigma}(u)$, their dimension p being finite). The result in (23), together with the approximation error, extends the result in (25) to the framework of factor models.

Remark 4 (**Approximation error**). In contrast to the stochastic estimation error, the approximation error is purely deterministic (it only depends on N and neither on T nor on h_T) and comes from the approximate factor structure of our model, namely the assumption of uniformly bounded eigenvalues (as in Chamberlain and Rothschild, 1983). The approximation error can be decomposed further as

$$\left\|\mathbf{V}_N(u)-\mathbf{V}(u)\right\|\le\left\|\mathbf{V}_N(u)-\mathbf{V}_N^C(u)\right\|+\left\|\mathbf{V}_N^C(u)-\mathbf{V}(u)\right\|,$$

where $\mathbf{V}_N^C(u)$ is the diagonal matrix containing the largest (nonzero) r eigenvalues of $N^{-1}\mathbf{\Sigma}_N^C(u)$, $\mathbf{\Sigma}_N^C(u)$ being defined in (5). The term $\left\|\mathbf{V}_N(u)-\mathbf{V}_N^C(u)\right\|$ is called *idiosyncratic variance error* because it is proportional $N^{-1}\mathrm{v}_{1N}^E$, where we recall that v_{1N}^E is the largest eigenvalue of $N^{-1}\mathbf{\Sigma}_N^E$ (see Appendix B). Then $\left\|\mathbf{V}_N(u)-\mathbf{V}_N^C(u)\right\| = O\left(N^{-1}\right)$ by Definition 1(ii).

The second term, $\left\|\mathbf{V}_N^C(u)-\mathbf{V}(u)\right\|$, is the *common variance error*. It depends on the difference between the (nonzero) eigenvalues of $N^{-1}\mathbf{\Sigma}_N^C(u)$, the rescaled variance of the common components, and the eigenvalues of $\mathbf{\Sigma}^\Lambda(u)\mathbf{\Sigma}^F$, the product between the asymptotic scale of the loadings and the asymptotic variance of factors. In Appendix B we show that the common variance error is $O(\ell_N)$ and therefore it tends to zero as N tends to infinity by Assumption B2 and (22).

Remark 5. The sequence of parameters ℓ_N has been introduced to study the convergence of $\widehat{\mathbf{V}}_N(u;h_T)$ to a well-defined function $\mathbf{V}(u)$ which does not depend on N. This new parameter does not affect the meaning of Assumption B2; it only enters to formalize the results in Theorems 2 and 3, and Corollary 2. In particular, it makes it possible to combine all the rates of convergence in a unified rate, given by the minimum of $\sqrt{Th_T}$, $\sqrt{N}$, and ℓ_N. Without introducing the parameter ℓ_N, we can only estimate the matrix $\mathbf{V}_N^C(u)$ (which depends on N), as the following corollary shows. We also refer to Figure 5 in Section 4.

COROLLARY 1. *Under Assumptions A–D,*

$$\min\left(\sqrt{Th_T},N\right)\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N^C(u)\right\| = O_p(1).$$

Proof. It follows from Proposition 2 that

$$\begin{aligned}\left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N^C(u)\right\| &\le \left\|\widehat{\mathbf{V}}_N(u;h_T)-\mathbf{V}_N(u)\right\| + \left\|\mathbf{V}_N(u)-\mathbf{V}_N^C(u)\right\| \\ &= O_p\left(\tfrac{1}{\sqrt{Th_T}}\right)+\left(\tfrac{1}{N}\right). \qquad \blacksquare\end{aligned}$$

Assumption E (Multiplicity of eigenvalues). Suppose that for the largest r eigenvalues $\{\mathrm{v}_{1N},\ldots,\mathrm{v}_{rN}\}$ of $\mathbf{\Sigma}_N(u)$,

$$\liminf_{\tau\to 0}\frac{\left|\mathrm{v}_{iN}(u+\tau)-\mathrm{v}_{jN}(u+\tau)\right|}{|\tau|} > 0\,,$$

for all $u\in(0,1)$, all $i\neq j,\ i,j=1,\ldots,r$, and all $N\in\mathbb{N}$. This is equivalent to requiring the same condition on the eigenvalues of the $r\times r$ matrix $\mathbf{\Sigma}^\Lambda(u)\mathbf{\Sigma}^F$.

Assumption E is an assumption on the identifiability in estimating the time-varying factor loadings $\lambda_{ij}(u)$ as well as the common components $C_{it,T} := \boldsymbol{\lambda}_i'\left(\frac{t}{T}\right)\boldsymbol{F}_t$. Note that we do not require distinctness of the eigenvalues of $\mathbf{\Sigma}^\Lambda(u)\mathbf{\Sigma}^F$ (as is done, e.g., in Bai, 2003, Assum. D). We rather control the degree of contact of the model eigenvalues as functions of rescaled time u. Note that Assumption E does not permit identical eigenvalues to accumulate over time. However, eigenvalues can intersect each other, but in order to have continuous corresponding eigenvectors it is not allowed that in the points of intersection the derivatives are equal. By Theorems 2.4 and 2.7 of Chern and Dieci (2000, pp. 774–778), under this assumption the matrix of orthogonal eigenvectors of $\mathbf{\Sigma}_N(u)$ can be taken to be a continuous function in u. For details we refer to Lemma 1 in Appendix A, which we apply in the case where $k = 1$ and $e = 1$.

We are now in a position to develop consistent estimation theory for the eigenvalues and the eigenvectors of the estimator $\widehat{\mathbf{\Sigma}}(u;h)$ with well-defined target functions that are continuous in rescaled time. The appropriateness of this condition for our purposes of consistent estimation theory results from the following observation. Assumption E is in fact a sufficient condition to guarantee that not only our (estimated) eigenvalues but also our estimated eigenvectors (i.e., the loadings) converge to continuous functions in time $u \in (0,1)$. It is given in terms of the eigenvalues of $\mathbf{\Sigma}_N(u)$, which by Assumption B.1 are in $\mathcal{C}^1[0,1]$ (as the eigenvalues depend continuously on the regularity of the elements of the covariance matrix $\mathbf{\Sigma}_N(u)$, which in turn is determined by the regularity of the model loadings $\lambda_{ij}(u)$). In a point of intersection u_0 of eigenvalues $\gamma_i(u_0)$, we define the ordering of corresponding eigenvectors to be the same as the one for eigenvectors corresponding to eigenvalues $\gamma_i(u_0^-)$, where u_0^- means points leading up to the point u_0 from below.

We end this remark by adding that, as already in the stationary case, the loadings can only be estimated up to a transformation, and it is only the product $\mathbf{\Sigma}_N^C(u) = \mathbf{\Lambda}_N(u)\mathbf{\Sigma}^F\mathbf{\Lambda}_N'(u)$ that is identifiable; compare also the formulation of our Theorem 2.

The following two theorems are our main results, which are about weak consistency of estimated loadings and factors.

THEOREM 2. *Under Assumptions A–E,*

$$\text{(i)}\quad \min\left(\sqrt{N}, \sqrt{Th_T}, \ell_N^{-1}\right)\left\{\frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_N(u;h_T) - \mathbf{\Lambda}_N(u)\mathbf{R}(u)\right\|\right\} = O_p(1) \qquad \textbf{(26)}$$

$$\text{(ii)}\quad \min\left(\sqrt{N}, \sqrt{Th_T}, \ell_N^{-1}\right)\left\|\widehat{\boldsymbol{\lambda}}_i(u;h_T) - \mathbf{R}'(u)\boldsymbol{\lambda}_i(u)\right\| = O_p(1), \qquad \textbf{(27)}$$

where $\widehat{\mathbf{\Lambda}}_N(u;h_T)$ *is defined in* (18), $\widehat{\boldsymbol{\lambda}}_i'(u;h_T)$ *is the* i*th row of* $\widehat{\mathbf{\Lambda}}_N(u;h_T)$, $\mathbf{R}(u) := \left(\mathbf{\Sigma}^F\right)^{1/2}\mathbf{\Upsilon}(u)\mathbf{V}^{-1/2}(u)$, *and* $\mathbf{\Upsilon}(u)$ *is the* $r \times r$ *matrix containing the orthonormal eigenvectors of the* $r \times r$ *matrix* $\left(\mathbf{\Sigma}^F\right)^{1/2}\mathbf{\Sigma}^\Lambda(u)\left(\mathbf{\Sigma}^F\right)^{1/2}$.

Proof. See Appendix B. ■

The result in (26) shows that the appropriately scaled norm of the distance between the estimated loading matrix and a linear transformation of the true loading matrix converges to zero in probability, where the rate is given by the minimum of $\sqrt{Th}$, $\sqrt{N}$, and ℓ_N^{-1}. The norm in (27) does not need to be rescaled because its argument is a vector of length r. The transformation matrix $\mathbf{R}(u)$ is invertible, so that we have consistency up to a linear invertible transformation. Note that (27), the formulation in terms of vectors, is the locally stationary analogue to Theorem 1 of Bai (2003). In this sense, we are locally (i.e., in a neighborhood of each time t of effective sample size of order Th) in a similar situation as Bai (2003), with either the cross-sectional dimension dominating the sample size or vice versa. Note that N or T are allowed to grow to infinity without any restriction. We now state the analogous result for the estimates of the factors, which are consistent up to the inverse transformation $\mathbf{R}^{-1}(u)$. For this we need the following assumption.

Assumption F (Linear combination of the errors weighted by the loadings). The loadings and the errors are such that

$$\left\| \frac{\boldsymbol{\Lambda}_N'\left(\frac{t}{T}\right)\boldsymbol{E}_{t,N}}{\sqrt{N}} \right\| = O_p(1) \quad \text{for all } t \text{ and for all } T \text{ as } N \to \infty.$$

Assumption F is needed to prove the consistency of the estimated factors (Theorem 3). This is a weaker version of the following assumption (see Bai, 2003, Assum. F3, p. 144):

$$\frac{\boldsymbol{\Lambda}_N'\boldsymbol{E}_{t,N}}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0,\boldsymbol{\Gamma}_t) \quad \text{for each } t \text{ as } N \to \infty$$

for a given covariance matrix $\boldsymbol{\Gamma}_t$.

THEOREM 3. *Under Assumptions A–F,*

$$\min\left(\sqrt{Th_T},\sqrt{N},\ell_N^{-1}\right)\left\|\widehat{\boldsymbol{F}}_{t,NT} - \mathbf{R}^{-1}\left(\tfrac{t}{T}\right)\boldsymbol{F}_t\right\| = O_p(1),$$

where $\widehat{\boldsymbol{F}}_{t,NT}$ is defined in (21), *$\boldsymbol{F}_t$ obeys Assumption A of the* (LSAFM) *in Definition 1, and $\mathbf{R}^{-1}\left(\frac{t}{T}\right)$ is the inverse of the transformation matrix $\mathbf{R}\left(\frac{t}{T}\right)$ defined in Theorem 2.*

Proof. See Appendix B. ∎

Note that the rate of convergence $\left(\sqrt{Th_T},\sqrt{N},\ell_N^{-1}\right)$ is the same as that of the loadings. Finally, similar to Bai (2003), we can give a consistency result for the estimation of the common components in the next proposition.

COROLLARY 2. *Under Assumptions A–F,*

$$\min\left(\sqrt{Th_T},\sqrt{N},\ell_N^{-1}\right)\left|\widehat{C}_{it,T} - C_{it,T}\right| = O_p(1), \quad i=1,\ldots,N, \quad t=1,\ldots,T,$$

where $C_{it,T} := \boldsymbol{\lambda}_i'\left(\frac{t}{T}\right)\boldsymbol{F}_t$ is the common component of the ith series at time t and $\widehat{C}_{it,T} := \widehat{\boldsymbol{\lambda}}_i'\left(\frac{t}{T}\right)\widehat{\boldsymbol{F}}_{t,NT}$ is its estimator.

Proof. This corollary follows directly from Slutsky's theorem applied to Theorems 2(ii) and 3. ∎

Corollary 2 states that $\widehat{C}_{it,T}$ consistently estimates the common component $C_{it,T}$. Note that, unlike the estimation of $\mathbf{\Lambda}_N\left(\frac{t}{T}\right)$ or $\boldsymbol{F}_t$, $C_{it,T}$ and $\widehat{C}_{it,T}$ are well identified. $C_{it,T}$ is identified because the indeterminacy of $\mathbf{\Lambda}_N\left(\frac{t}{T}\right)$ and $\boldsymbol{F}_t$ by the $r \times r$ transformation matrix $\mathbf{R}\left(\frac{t}{T}\right)$ cancels out. $\widehat{C}_{it,T}$ is identified because, by equation (21), the sign of $\widehat{\boldsymbol{F}}_{t,NT}$ depends on the sign of $\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)$. Due to Assumption E this unicity continues to hold for the limit of the estimators as functions of rescaled time tending to a continuous limit because the limit of the estimated eigenvectors are continuous functions of time. We also refer to the discussion on the interpretation of Assumption E above (below Remark 4).

4. SIMULATION STUDY

In this section we illustrate the performance of the estimators $\widehat{\mathbf{\Sigma}}_N(u)$, $\widehat{\mathbf{\Lambda}}_N(u)$, and $\widehat{\mathbf{V}}_N(u)$ defined in (17) and (18). The estimator $\widehat{\mathbf{\Sigma}}_N(u;h_T)$ in (17) depends on the bandwidth sequence h_T. In our simulations the bandwidth $\widehat{h}$ is selected data-dependent with the local *plug-in algorithm*. This method was introduced in the global case ($h(u) \equiv h$) by Gasser, Kneip, and Köhler (1991) and generalized to the local case by Brockmann, Gasser, and Herrmann (1993). The technical computation of the local bandwidth selection procedure is described in Herrmann (1997). The basic idea of plug-in estimation is to obtain a large-sample approximation to the mean integrated squared error (MISE) of the estimator of the entries $\widehat{\sigma}_{i,j}(u;h)$ of $\widehat{\mathbf{\Sigma}}_N(u;h)$, then to minimize the resulting analytical expression with respect to h to obtain the asymptotic optimal bandwidth $\widehat{h}$, and last to replace the unknown terms in $\widehat{h}$ by estimators. For ease of presentation, in this section we skip the dependence of the estimates on the parameter h.

4.1. Two Examples

In Theorem 2 the matrix $\widehat{\mathbf{\Lambda}}_N(u)$ is only able to identify the matrix $\mathbf{\Lambda}_N(u)$ *up to transformation* (i.e., $\mathbf{\Lambda}_N(u)\mathbf{R}(u)$), and *up to sign* (note, however, that the squared difference between estimator and transformed loadings is uniquely defined in this Theorem 2). Indeed, the matrices $\widehat{\mathbf{\Lambda}}_N(u)$ and $-\widehat{\mathbf{\Lambda}}_N(u)$ are both solutions of (18). This explains why in the simulation we still need to take the absolute values of these vectors (see Section 4.1.1 and Figure 2). To show the performance of the estimator $\widehat{\mathbf{\Lambda}}_N(u)$ we consider a first set of simulations with $\mathbf{\Sigma}^{\Lambda}(u) = \mathbf{I}_r$ for all $u \in (0,1)$ (see Section 4.1.1 and Figures 1–2). In this case the transformation matrix $\mathbf{R}(u)$ is — up to sign — the identity matrix $\mathbf{I}_r$ (i.e., $\mathbf{R}(u) = \pm\mathbf{I}_r$), and thus the matrix $\widehat{\mathbf{\Lambda}}_N(u)$ is able to identify the matrix $\mathbf{\Lambda}_N(u)$ up to sign.

By Proposition 2 the matrix $\widehat{\mathbf{V}}_N(u)$ is only able to identify the eigenvalues of the product $\mathbf{\Sigma}^{\Lambda}(u)\mathbf{\Sigma}_F$. To show the performance of the estimator $\widehat{\mathbf{V}}_N(u)$ we

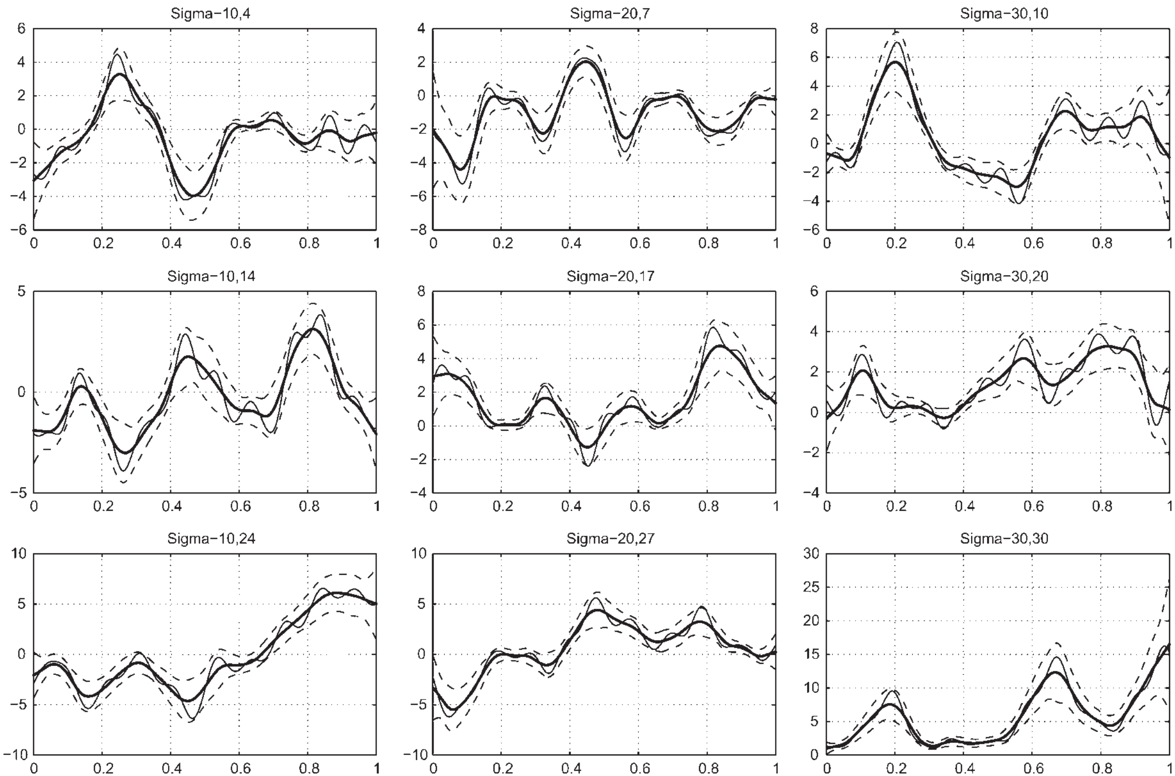

FIGURE 1. First Example, Section 4.1.1. Time-varying entries of the covariance matrix. Solid line: $\mathbf{\Sigma}_N\left(\frac{t}{T}\right)$. Bold line: $\overline{\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}\right)}$. Dashed lines: pointwise 95% confidence intervals.

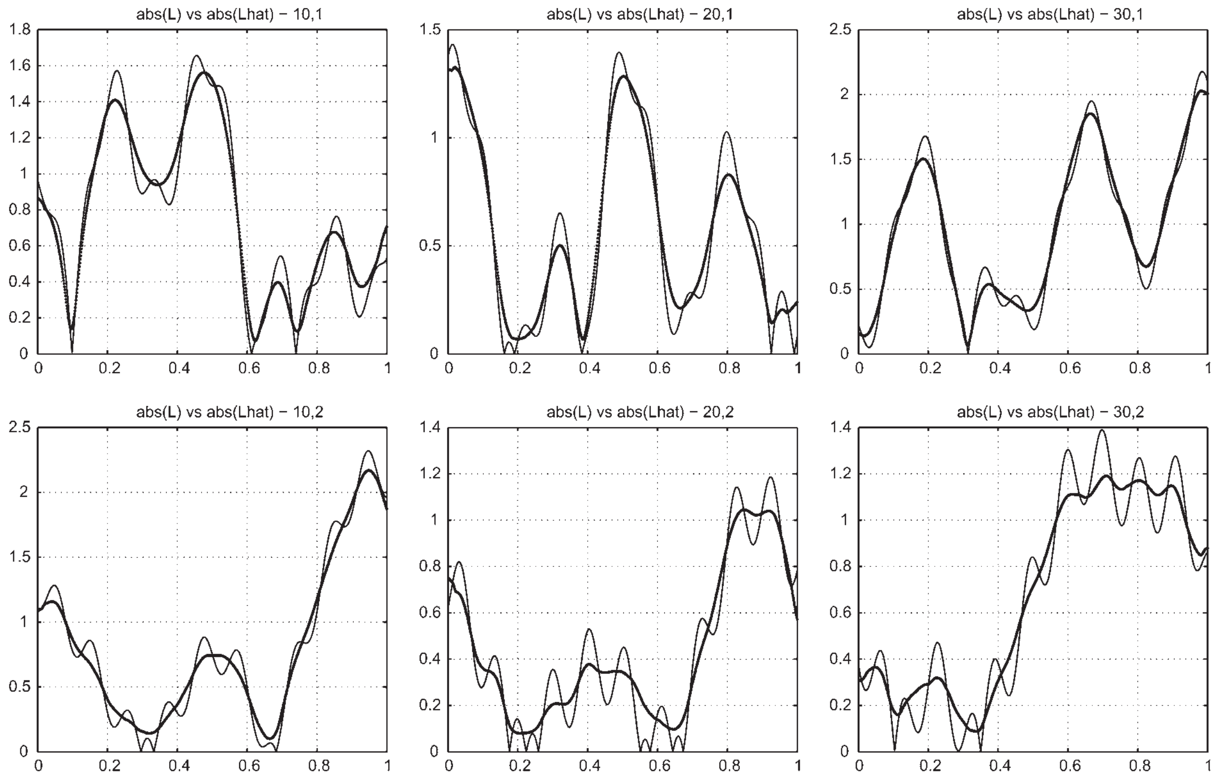

FIGURE 2. First Example, Section 4.1.1. Time-varying entries of the matrices $\left|\mathbf{\Lambda}_N\left(\frac{t}{T}\right)\right|$ (solid line) and $\overline{\left|\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)\right|}$ (bold line).

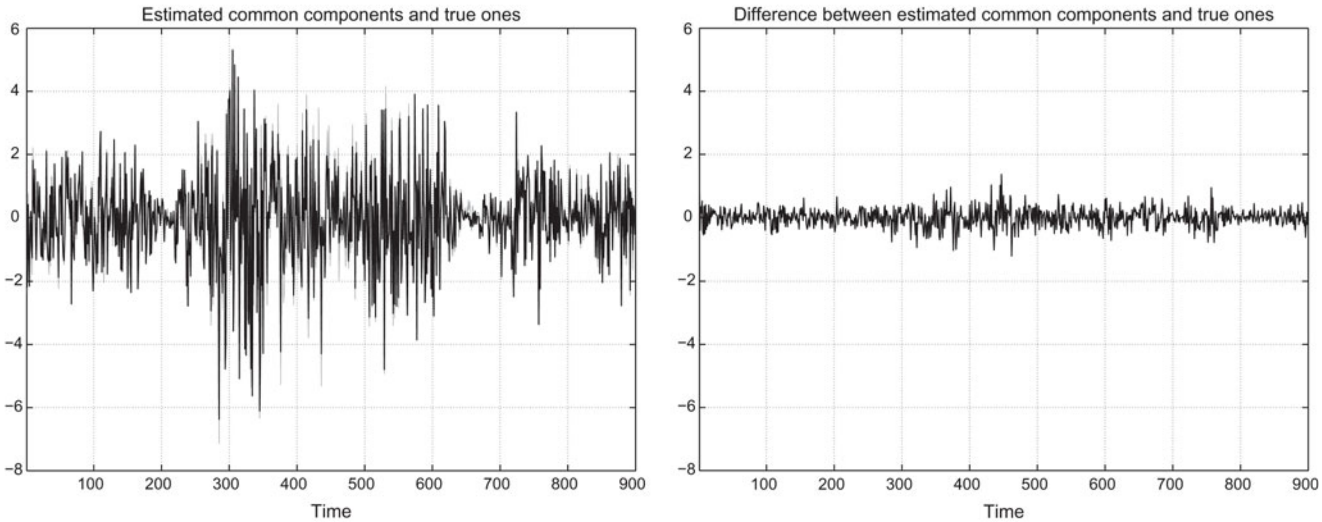

FIGURE 3. First Example, Section 4.1.1. Left: first common component $C_{1t,T}$ (black) and estimated common component $\widehat{C}_{1t,T}$ (grey). Right: $\widehat{C}_{1t,T} - C_{1t,T}$.

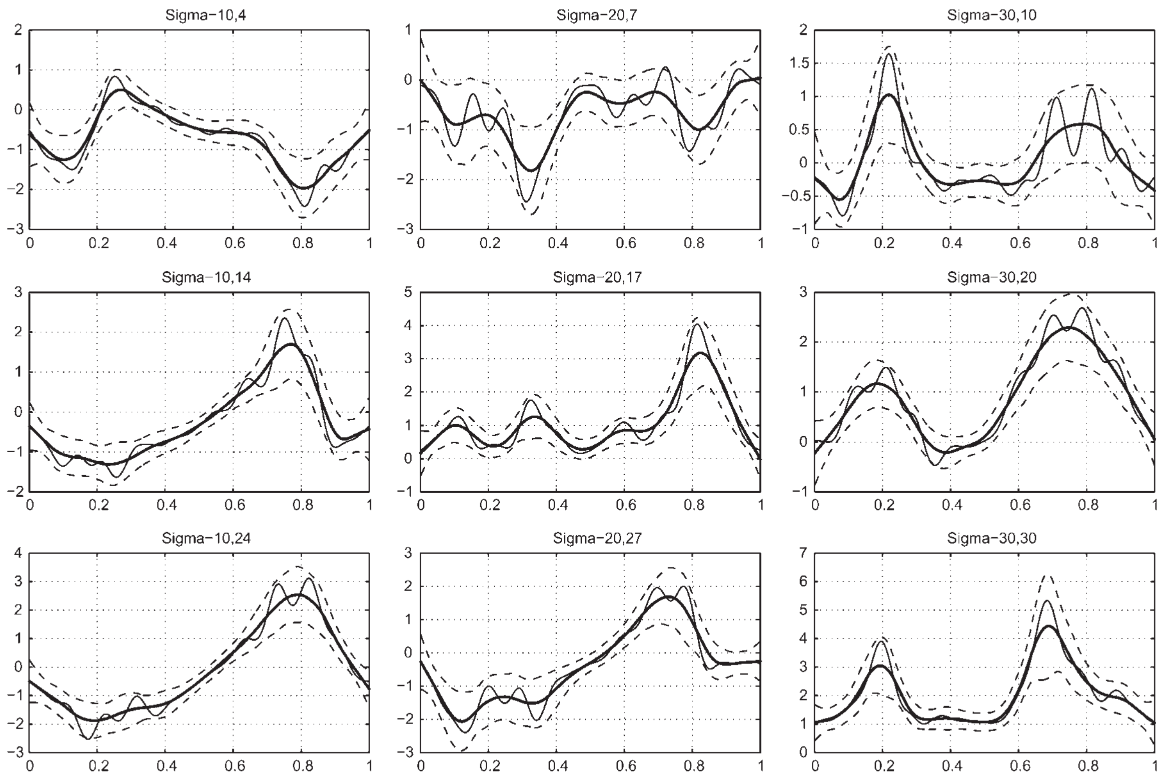

FIGURE 4. Second Example, Section 4.1.2. Time-varying entries of the covariance matrix. Solid line: $\boldsymbol{\Sigma}_N\left(\frac{t}{T}\right)$. Bold line: $\overline{\widehat{\boldsymbol{\Sigma}}_N\left(\frac{t}{T}\right)}$. Dashed lines: pointwise 95% confidence intervals.

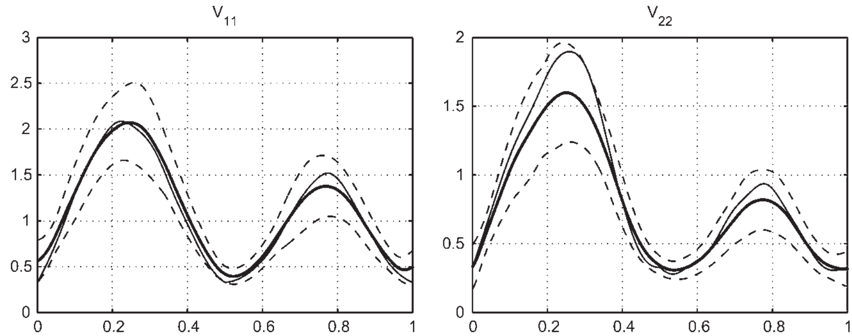

FIGURE 5. Second Example, Section 4.1.2. Time-varying entries of the diagonal matrices $\mathbf{V}_N^C\left(\frac{t}{T}\right)$ (solid line) and $\widehat{\mathbf{V}}_N\left(\frac{t}{T}\right)$ (bold line) corresponding to Figure 4. Dashed lines: pointwise 95% confidence intervals.

consider a second set of simulations with $\boldsymbol{\Sigma}_F = \mathbf{I}_r$ (see Section 4.1.2 and Figures 4–5). In this case the matrix $\widehat{\mathbf{V}}_N(u)$ is an estimate of the eigenvalues of $\boldsymbol{\Sigma}^{\Lambda}(u)$.

We recall that N and T are allowed to grow to infinity without any restriction, and N could even be larger than T. However, in the following examples we use $N = \sqrt{T}$ to avoid a computational burden, as the numerical implementation becomes time demanding for large values of N.

4.1.1. First Example. We generate the data according to model (3) with $N = 30$, $T = 900$, and $r = 2$, and consider the particular case

$$\boldsymbol{\Sigma}^{\Lambda}(u) = \mathbf{I}_r \qquad \forall u \in (0,1). \tag{28}$$

To satisfy (28) we define

$$\boldsymbol{\Lambda}_N\left(\tfrac{t}{T}\right) = \sqrt{N}\exp\left(\pi\,\tfrac{t}{T}\mathbf{Y}_N\right)\mathbf{A}_N, \qquad t = 1,\ldots,T, \tag{29}$$

where $\mathbf{Y}_N$ is an antisymmetric matrix of dimension N, $\mathbf{A}_N$ is an $N \times r$ matrix such that $\mathbf{A}_N'\mathbf{A}_N = \mathbf{I}_r$, and $\exp\left(\pi\,\frac{t}{T}\mathbf{Y}_N\right)$ is the $N \times N$ matrix whose (i,j)th element is $\exp\left(\pi\,\frac{t}{T}y_{ij}\right)$. Then we have $N^{-1}\boldsymbol{\Lambda}_N'\left(\frac{t}{T}\right)\boldsymbol{\Lambda}_N\left(\frac{t}{T}\right) = \mathbf{I}_r$ for all $t = 1\ldots,T$.

An antisymmetric matrix is a square matrix that satisfies $\mathbf{Y} = -\mathbf{Y}'$. In component notation, $y_{ij} = -y_{ji}$. Letting $k = i = j$, the requirement becomes $y_{kk} = -y_{kk}$, so an antisymmetric matrix must have zeros on its diagonal. In our simulations $y_{ij} = 1$ for $i < j$, $y_{ij} = 0$ for $i = j$ and $y_{ij} = -1$ for $i > j$ $(i,j = 1\ldots,N)$. In order to have nontrivial loadings, we multiply the matrix $\exp\left(\pi\,\frac{t}{T}\mathbf{Y}_N\right)$ by the matrix $\mathbf{A}_N$. To obtain the matrix $\mathbf{A}_N$, we simulate $n = 50$ independent and identically distributed (i.i.d.) realizations $\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_N$ of an N-dimensional normal random vector $\boldsymbol{Z}_N \sim \mathcal{N}(\mathbf{0},\mathbf{I}_N)$ and take the r eigenvectors $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_r$ corresponding to the largest eigenvalues of the sample covariance matrix $\mathbf{S}^Z$ of the $\boldsymbol{Z}_i$'s: $\mathbf{S}^Z := n^{-1}\sum_{i=1}^n \boldsymbol{Z}_i\boldsymbol{Z}_i{}'$ and $\mathbf{A}_N := [\boldsymbol{A}_1,\ldots,\boldsymbol{A}_r]$. By construction $\mathbf{A}_N'\mathbf{A}_N = \mathbf{I}_r$.

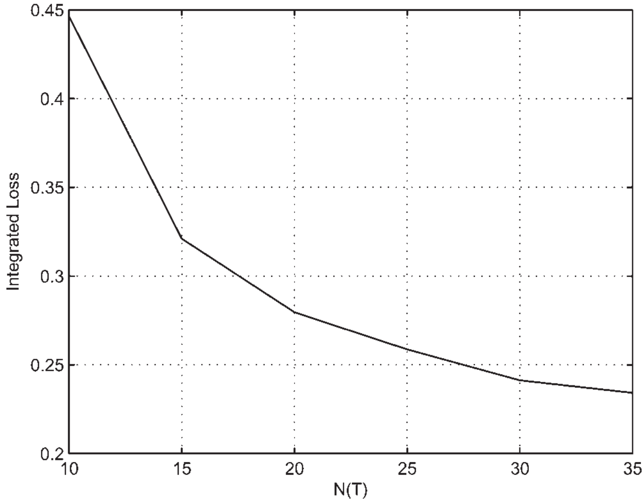

FIGURE 6. Second Example, Section 4.1.2. Integrated Loss defined in (31).

To represent $\mathbf{\Sigma}_N\left(\frac{t}{T}\right)$ and $\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}\right)$, $N = 30$, we only show some typical elements $\sigma_{ij}(\cdot)$ and $\widehat{\sigma}_{ij}(\cdot)$, in particular for $i = 10, 20, 30$, and $j = 4, 7, 10, 14, 17, 20, 24, 27, 30$ (see Figure 1).

We simulate $M = 100$ times the same model, that is, model (3) with the same deterministic loadings defined in (29) but different (realizations of) factors and errors. In particular, $\boldsymbol{F}_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_F)$ with $\mathbf{\Sigma}_F = \text{diag}\{3.5, 1\}$, and $\boldsymbol{E}_{t,N} \overset{\text{iid}}{\sim} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_N^E\right)$ with $\mathbf{\Sigma}_N^E = \mathbf{I}_N$. For each $m = 1, \ldots, M$ and for all t, we compute the estimate $\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; m\right) = \left\{\widehat{\sigma}_{ij}\left(\frac{t}{T}; m\right)\right\}_{i,j=1}^N$ defined in (17). Then for all t we consider the average (bold line)

$$\overline{\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}\right)} := M^{-1} \sum_{m=1}^{M} \widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; m\right) = \left\{\overline{\widehat{\sigma}}_{ij}\left(\frac{t}{T}\right)\right\}_{i,j=1}^N$$

and construct 95% confidence intervals $\left[\overline{\widehat{\sigma}}_{ij}\left(\frac{t}{T}\right) \pm z_{0.975}\,\widehat{s}_{ij}^{\sigma}\left(\frac{t}{T}\right)\right]$ based on asymptotic normality (dashed lines), where $z_\alpha = \Phi^{-1}(\alpha)$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and

$$\widehat{s}_{ij}^{\sigma}\left(\frac{t}{T}\right) := \sqrt{\frac{1}{M-1} \sum_{m=1}^{M} \left[\widehat{\sigma}_{ij}\left(\frac{t}{T}; m\right) - \overline{\widehat{\sigma}}_{ij}\left(\frac{t}{T}\right)\right]}$$

is the estimator of the standard deviation, say $s_{ij}^{\sigma}\left(\frac{t}{T}\right)$, of the estimator $\widehat{\sigma}_{ij}\left(\frac{t}{T}\right)$. The asymptotic normality of the estimator $\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}\right)$ in (17) can be derived analogously to Rodríguez-Poo and Linton (2001, Prop. 3.2). For each $m = 1, \ldots, M$ and for all $t = 1, \ldots, T$ we define $\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}; m\right)$ as $\sqrt{N}$ times the orthonormal eigenvectors of the estimate $N^{-1}\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; m\right)$ (see Figure 2). If $\mathbf{\Sigma}^{\Lambda}(u) = \mathbf{I}_r$ for all $u \in (0,1)$, then $\mathbf{R}(u)$ is $\pm$ the identity matrix $\mathbf{I}_r$ for all $u \in (0,1)$; i.e., $\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)$ converges to $\pm\mathbf{\Lambda}_N(u)$. To remove the up-to-sign indeterminacy we consider, for all t, the average of the absolute values $\overline{\left|\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)\right|} := M^{-1}\sum_{m=1}^{M}\left|\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}; m\right)\right|$.

Finally, in Figure 3 we report an example of estimation of the common components. We just consider the first common component $C_{1t,T}$ and (a realization of) its estimator $\widehat{C}_{1t,T}$, $t = 1, \ldots, T$, where $C_{it,T}$ and $\widehat{C}_{it,T}$ are defined in Corollary 2. As the two series are very close to each other, we also report the difference $\widehat{C}_{1t,T} - C_{1t,T}$ on the original scale of the two (between –8 and 6).

4.1.2. Second Example. We generate the data according to model (3) with loadings

$$\mathbf{\Lambda}_N\left(\tfrac{t}{T}\right) = \sqrt{N}\sin\left(2\pi\tfrac{t}{T}\right)\exp\left(\pi\tfrac{t}{T}\mathbf{Y}_N\right)\mathbf{A}_N + \pi\mathbf{A}_N, \tag{30}$$

where $\mathbf{Y}_N$ and $\mathbf{A}_N$ are the same as in (29), and restrict to the particular case $\mathbf{\Sigma}^F = \mathbf{I}_r$.

The covariance matrix and its estimate are represented in Figure 4 (analogously to the first example). The eigenvalues of the normalized covariance matrix $N^{-1}\mathbf{\Sigma}_N\left(\frac{t}{T}\right)$ are plotted in Figure 5. For each $m = 1, \ldots, M$ and for all $t = 1, \ldots, T$ we compute the eigenvalues $\widehat{\mathbf{V}}_N\left(\frac{t}{T}; m\right)$ of the estimate $N^{-1}\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; m\right)$. Then for all t we consider the average (bold line)

$$\overline{\widehat{\mathbf{V}}_N\left(\tfrac{t}{T}\right)} := M^{-1}\sum_{m=1}^{M}\widehat{\mathbf{V}}_N\left(\tfrac{t}{T}; m\right) = \left\{\overline{\widehat{\nu}}_{ij}\left(\tfrac{t}{T}\right)\right\}_{i,j=1}^{r}$$

and construct 95% confidence intervals $\left[\overline{\widehat{\nu}}_{ij}\left(\frac{t}{T}\right) \pm z_{0.975}\,\widehat{s}_{ij}^{\nu}\left(\frac{t}{T}\right)\right]$ based on asymptotic normality (dashed lines), where $z_{0.975} = 1.96$, and

$$\widehat{s}_{ij}^{\nu}\left(\tfrac{t}{T}\right) := \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}\left[\widehat{\nu}_{ij}\left(\tfrac{t}{T}; m\right) - \overline{\widehat{\nu}}_{ij}\left(\tfrac{t}{T}\right)\right]^2}$$

is the estimator of the standard deviation, say $s_{ij}^{\nu}\left(\frac{t}{T}\right)$, of the estimator $\widehat{\nu}_{ij}\left(\frac{t}{T}\right)$. The asymptotic normality of $\widehat{\mathbf{V}}_N\left(\frac{t}{T}\right)$ comes from the asymptotic normality of $\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; h\right)$ because, by (20), the entries of $\widehat{\mathbf{V}}_N\left(\frac{t}{T}\right)$ are continuous functions of the entries of $\widehat{\mathbf{\Sigma}}_N\left(\frac{t}{T}; h\right)$. Figure 5 shows the performance of the estimator $\widehat{\mathbf{V}}_N(u)$, which approximates $\mathbf{V}_N^C(u)$ by Corollary 1, where we recall that $\mathbf{V}_N^C(u)$ is the diagonal matrix containing the eigenvalues of the matrix $N^{-1}\mathbf{\Lambda}_N'\left(\frac{t}{T}\right)\mathbf{\Lambda}_N\left(\frac{t}{T}\right)$, $\mathbf{\Lambda}_N\left(\frac{t}{T}\right)$ being defined in (30). The solid lines in Figure 5 are the time-varying entries of $\mathbf{V}_N^C(u)$, which approximate the matrix $\mathbf{V}(u)$ by Assumption B2 and (22). We recall that $\mathbf{V}(u)$ is the diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}^{\Lambda}(u)\mathbf{\Sigma}^F$, which is equal to $\mathbf{\Sigma}^{\Lambda}(u)$ in this example.

Figure 2 shows the *local* performance of the estimator $\widehat{\mathbf{\Lambda}}_N\left(\frac{t}{T}\right)$. To have an idea of the *global* performance of this estimator we consider different values of $T = 100, 225, 400, 625, 900, 1225$ and $N(T) = \sqrt{T}$, and for each combination of N and T we compute, in the spirit of Theorem 2, the loss function

$$L(N,T;M) := \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_N\left(\tfrac{t}{T}\right) - \mathbf{\Lambda}_N\left(\tfrac{t}{T}\right)\mathbf{R}_N\left(\tfrac{t}{T}\right)\right\|. \tag{31}$$

By Theorem 2 this loss is decreasing with regard to T and $N(T)$, as also shown in Figure 6. This loss function is the sample counterpart of the integrated loss

$$L(N,T;M) := \frac{1}{M}\sum_{m=1}^{M}\int_0^1 \frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_N(u) - \mathbf{\Lambda}_N(u)\mathbf{R}_N(u)\right\| du.$$

We used the model defined in the second example (i.e., with $\mathbf{\Sigma}^F = \mathbf{I}_r$) because in this case the matrix $\mathbf{R}(u)$ is in general different than $\mathbf{I}_r$ (we recall that $\mathbf{R}(u) = \pm\mathbf{I}_r$ if $\mathbf{\Sigma}^{\Lambda}(u) = \mathbf{I}_r$).

5. CONCLUSIONS

In this paper we have proposed a new locally stationary factor model that allows for smoothly time-varying factor loadings. We showed consistency of the principal components estimator under double asymptotics, up to an invertible transformation and sign. Rates are parallel to the stationary setup with the sample size T replaced by the "effective" sample size Th. No restrictions for T and N growing to infinity simultaneously are necessary. In a simulation study, our estimator was shown to work very well for two alternative scenarios: one where the transformation matrix is identity, another where it is not.

As mentioned in the Introduction, in practice one does not know the number of factors, and a test such as that of Bai and Ng (2002) is required. We are confident that analogous tests can be developed for our model framework, but that is beyond the scope of the paper. Furthermore, it would be interesting to compare our model more explicitly with the dynamic factor model of Forni et al. (2000), especially in empirical applications. If in reality factor loadings are smoothly changing, then quite likely this would show in positive autocorrelations of factors in the dynamic factor model. To distinguish both types of dynamic properties, we can extend our model to allow for autocorrelations of the factors, which is left for future research.

There are many potential applications of our model in macroeconomics and finance. For example, the dynamic factor model of Forni et al. (2000) has been applied to forecasting, monetary policy analysis business cycle analysis, and construction of economic indicators; see Breitung and Eickmeier (2006) for a recent review. In finance, applications to asset pricing and portfolio selection are obvious, and this is also left for future research.

REFERENCES

Alexander, C.O. (2001) Orthogonal GARCH. In C. O. Alexander (ed.), *Mastering Risk*, vol. 2, pp. 21–38. Prentice Hall.

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

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

Breitung, J. & S. Eickmeier (2006) Dynamic factor models. In O. Hübler and J. Frohn (eds.), *Modern Econometric Analysis*. Springer.

Brockmann, M., T. Gasser, & E. Herrmann (1993) Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* 88, 1302–1309.

Chamberlain, G. & M. Rothschild (1983) Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica* 51(5), 1281–1304.

Chern, J.-L. & L. Dieci (2000) Smoothness and periodicity of some matrix decompositions. *SIAM Journal on Matrix Analysis and Applications* 22(3), 772–792.

Dahlhaus, R. (1996) Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In P.M. Robinson & M. Rosenblatt (eds.), *Athens Conference on Applied Probability and Time Series Analysis*, vol. II. Springer-Verlag.

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25(1), 1–37.

Dahlhaus, R. (2000) A likelihood approximation for locally stationary processes. *Annals of Statistics* 28(6), 1762–1794.

Dahlhaus, R. & S.S. Subba Rao (2006) Statistical inference for time-varying arch processes. *Annals of Statistics* 34(3), 1075–1114.

Diebold, F.X. & M. Nerlove (1989) The dynamics of exchange rate volatility: A multivariate latent factor arch model. *Journal of Applied Econometrics* 4, 1–21.

Engle, R.F., V. Ng, & M. Rothschild (1990) Asset pricing with a factor-arch structure: Empirical estimates for Treasury bills. *Journal of Econometrics* 45, 213–237.

Fancourt, C.L. & J.C. Principe (1998) Competitive principal components analysis for locally stationary time series. *IEEE Transactions on Signal Processing* 46, 3068–81.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2000) The generalized dynamic factor model: Identification and estimation. *Review of Economics and Statistics* 82(4), 540–554.

Forni, M., M. Hallin, M. Lippi, & L. Reichlin (2005) The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association* 100(471), 830–840.

Forni, M. & M. Lippi (2001) The generalized dynamic factor model: Representation theory. *Econometric Theory* 17, 1113–1141.

Gasser, T., A. Kneip, & W. Köhler (1991) A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association* 86, 643–652.

Hafner, C.M., D. van Dijk, & P.H. Franses (2006) Semiparametric modeling of correlation dynamics. In T. Fomby & C. Hill (eds.), *Advances in Econometrics*, vol. 20, part A, pp. 59–103. Emerald Group.

Härdle, W., H. Herwartz, & V. Spokoiny (2004) Time inhomogeneous multiple volatility modeling. *Journal of Financial Econometrics* 1, 55–95.

Herrmann, E. (1997) Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics* 6, 35–54.

Herzel, S., C. Stărică, & R. Tütüncü (2006) A non-stationary paradigm for the dynamics of multivariate financial returns. In P. Bertail, P. Doukhan, & P. Soulier (eds.), *Statistics for Dependent Data, Lecture Notes in Statistics*, vol. 187. Springer.

Kollo, T. & H. Neudecker (1993) Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices. *Journal of Multivariate Analysis* 47, 283–300.

Lintner, J. (1965) The valuation of risky assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47, 13–37.

Lütkepohl, H. (1996) *Handbook of Matrices*. Wiley.

Mikosch, T. & C. Stărică (2004) Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics* 86, 378–390.

Phillips, P.C.B. & H.R. Moon (1999) Linear regression limit theory for nonstationary panel data. *Econometrica* 67(5), 1057–1111.

Rodríguez-Poo, J.M. & O. Linton (2001) Nonparametric factor analysis of residual time series. *Test* 10, 161–182.

Ross, S. (1976) The arbitrage theory of capital asset pricing. *Journal of Economic Theory* 13, 341–360.

Sharpe, W. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.

Stărică, C. & Granger, C. (2005) Nonstationarities in stock returns. *Review of Economics and Statistics* 87, 503–522.

Stock, J.H. & M.W. Watson (2002a) Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association* 97(460), 1167–1179.

Stock, J.H. & M.W. Watson (2002b) Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics* 20(2), 147–162.

APPENDIX A: Smooth Spectral Decomposition

LEMMA 1 (Takagi's factorization). *Let* $\mathbf{A} \in \mathcal{C}^k\left(\mathbb{R}, \mathbb{C}^{n\times n}\right)$ *be a complex symmetric matrix valued function of constant rank:* $\mathrm{rk}[\mathbf{A}(x)] \equiv r$ *for all* x *for fixed* $r : 1 \le r \le n$. *Then there exists unitary* $\mathbf{U} \in \mathcal{C}^k\left(\mathbb{R}, \mathbb{C}^{n\times n}\right)$ *such that*

$$\mathbf{A}(x) = \mathbf{U}(x)\begin{bmatrix}\mathbf{S}_+ & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{U}'(x) \qquad \forall x,$$

and $\mathbf{S}_+ \in \mathcal{C}\left(\mathbb{R}, \mathbb{R}^{r\times r}\right)$ *is symmetric positive definite. Moreover, suppose that the continuous eigenvalues* $\{\gamma_1, \ldots, \gamma_r\}$ *of* $\mathbf{S}_+$ *satisfy*

$$\liminf_{\tau\to 0} \frac{\left|\gamma_i(x+\tau) - \gamma_j(x+\tau)\right|}{|\tau^e|} \in (0, +\infty]$$

for some nonnegative integers $e \le k$ *and for all* x *and* $i \ne j$. *Then there exists orthogonal* $\mathbf{\Gamma} \in \mathcal{C}^{k-e}\left(\mathbb{R}, \mathbb{R}^{r\times r}\right)$ *such that* $\mathbf{\Gamma}'\mathbf{S}_+\mathbf{\Gamma} = \mathrm{diag}(\gamma_1, \ldots, \gamma_r)$. *The eigenvalues can be taken to be* $\mathcal{C}^k$ *functions.*

Proof. See Chern and Dieci (2000, Thms 2.4 and 2.7, pp. 774–778). ∎

APPENDIX B: Proofs

The asymptotic theory in Section 3 is given for a fixed value of $u \in (0, 1)$ in rescaled time. For ease of presentation, we give the proofs for the corresponding value of $t = [uT]$ in real time, where $[x]$ is the largest integer smaller than or equal to x. We recall that for a given sample size T and a given value of $u \in (0,1)$, the value of $t = [uT]$ is unique. For the proofs we use the following abbreviations. For $s = 1, \ldots, T$:

$$\boldsymbol{X}_s := \boldsymbol{X}_{s,NT}, \quad \mathbf{X} := \mathbf{X}_{NT}, \quad \mathbf{F} := \mathbf{F}_T, \quad \boldsymbol{E}_s := \boldsymbol{E}_{s,N}, \quad \boldsymbol{\Sigma}^E := \boldsymbol{\Sigma}^E_N,$$
$$\boldsymbol{\Lambda}_s := \boldsymbol{\Lambda}_N\left(\tfrac{s}{T}\right), \quad \boldsymbol{\Lambda}^{(1)}_s := \boldsymbol{\Lambda}^{(1)}_N\left(\tfrac{s}{T}\right), \quad \widehat{\boldsymbol{\Lambda}}_s := \widehat{\boldsymbol{\Lambda}}_N\left(\tfrac{s}{T}\right), \quad \boldsymbol{\Sigma}_s := \boldsymbol{\Sigma}_N\left(\tfrac{s}{T}\right),$$
$$\widetilde{\boldsymbol{\Sigma}}_s := \boldsymbol{X}_{s,NT}\boldsymbol{X}'_{s,NT}, \quad \widehat{\boldsymbol{\Sigma}}_s := \widehat{\boldsymbol{\Sigma}}_N\left(\tfrac{s}{T};h\right), \quad \mathbf{W}_s := \mathbf{W}_T\left(\tfrac{s}{T};h\right), \quad \mathbf{R}_s := \mathbf{R}\left(\tfrac{s}{T}\right),$$
$$\widehat{\mathbf{V}}_s := \widehat{\mathbf{V}}_N\left(\tfrac{s}{T};h\right), \quad \widehat{\boldsymbol{F}}_s := \widehat{\boldsymbol{F}}_{s,NT}, \quad \boldsymbol{\Sigma}^{\Lambda}_s := \boldsymbol{\Sigma}^{\Lambda}\left(\tfrac{s}{T}\right), \quad \mathbf{V}_{s,N} := \mathbf{V}_N\left(\tfrac{s}{T}\right),$$
$$\mathbf{V}^C_{s,N} := \mathbf{V}^C_N\left(\tfrac{s}{T}\right), \quad \mathbf{V}_s := \mathbf{V}\left(\tfrac{s}{T}\right).$$

In what follows a matrix is $O_p(r_T)$ if each element of that matrix goes to zero in probability at the rate r_T. In the sequel the remainder terms are bounded uniformly in t and T since

$$\left\|\frac{\boldsymbol{\Lambda}_t}{\sqrt{N}}\right\| = O(1) \quad \text{uniformly in } t \text{ and } T \text{ by Assumption B2,}$$
$$\left\|\frac{\boldsymbol{\Lambda}^{(1)}_t}{\sqrt{N}}\right\| = O(1) \quad \text{uniformly in } t \text{ and } T \text{ by Assumption B3,}$$
$$\left\|\frac{\widehat{\boldsymbol{\Lambda}}_t}{\sqrt{N}}\right\| = O_p(1) \quad \text{uniformly in } t \text{ and } T \text{ by the constraint in (13).}$$

Proof of Proposition 1. The assertion of this proposition can be written as

$$(NT)^{-1}\sum_{s=1}^{T} \operatorname{tr}\left\{\frac{\boldsymbol{\Lambda}'_s}{\sqrt{N}}\widetilde{\Sigma}_s\frac{\boldsymbol{\Lambda}_s}{\sqrt{N}} - \frac{\boldsymbol{\Lambda}'_t}{\sqrt{N}}\widetilde{\Sigma}_s\frac{\boldsymbol{\Lambda}_t}{\sqrt{N}}\right\}K_h\left(\frac{t-s}{T}\right) = O_p\left(h_T\right).$$

The idea of the proof is to use the inequality $\operatorname{tr}\left\{\frac{1}{N}\boldsymbol{\Lambda}_s\boldsymbol{\Lambda}'_s\widetilde{\boldsymbol{\Sigma}}_s\right\} \le \operatorname{tr}\left\{\frac{1}{N}\boldsymbol{\Lambda}_s\boldsymbol{\Lambda}'_s\right\}\operatorname{tr}\left\{\widetilde{\boldsymbol{\Sigma}}_s\right\}$, which holds since $\frac{1}{N}\boldsymbol{\Lambda}_s\boldsymbol{\Lambda}'_s$ and $\widetilde{\boldsymbol{\Sigma}}_s$ are positive semidefinite matrices. The idea is further to use that $\operatorname{tr}\left\{\frac{1}{N}\boldsymbol{\Lambda}_s\boldsymbol{\Lambda}'_s\right\} = O(1)$ and that $\mathbb{E}\left(\operatorname{tr}\left\{\widetilde{\boldsymbol{\Sigma}}_s\right\}\right) = O(N)$, uniformly in s.

More specifically, due to the existence and uniform boundedness of the first derivative of $\boldsymbol{\Lambda}_t$ we have the Taylor expansion

$$\boldsymbol{\Lambda}\left(\tfrac{s}{T}\right) = \boldsymbol{\Lambda}\left(\tfrac{t}{T}\right) + \boldsymbol{\Lambda}^{(1)}\left(\frac{s^*}{T}\right)\frac{t-s}{T} \qquad \text{for } \left|\frac{t-s}{T}\right| \le h,$$

with a mean value $\frac{s}{T} \le \frac{s^*}{T} \le \frac{t}{T}$ without loss of generality (w.l.o.g.).

We plug in this Taylor expansion for each of the $\boldsymbol{\Lambda}_s$ into $\operatorname{tr}\left\{\frac{\boldsymbol{\Lambda}'_s}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}_s}{\sqrt{N}} - \frac{\boldsymbol{\Lambda}'_t}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}_t}{\sqrt{N}}\right\}$, which gives, for arguments t and s with $|\frac{t-s}{T}| \le h$, that

$$\operatorname{tr}\left\{\frac{\boldsymbol{\Lambda}'_s}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}_s}{\sqrt{N}} - \frac{\boldsymbol{\Lambda}'_t}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}_t}{\sqrt{N}}\right\} = \operatorname{tr}\left\{\frac{\boldsymbol{\Lambda}'_t}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}^{(1)}_{s^*}}{\sqrt{N}}h + \frac{\boldsymbol{\Lambda}^{(1)'}_{s^*}}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}_t}{\sqrt{N}}h + \frac{\boldsymbol{\Lambda}^{(1)'}_{s^*}}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}^{(1)}_{s^*}}{\sqrt{N}}h^2\right\}.$$

Here we give only the treatment of the first term; the second is similar due to symmetry, and the third converges even faster. We need to show that

$$\sup_{t,s^*,s}\mathbb{E}\left|\operatorname{tr}\left\{\frac{\boldsymbol{\Lambda}'_t}{\sqrt{N}}\widetilde{\boldsymbol{\Sigma}}_s\frac{\boldsymbol{\Lambda}^{(1)}_{s^*}}{\sqrt{N}}\right\}\right| = O(N), \tag{B.1}$$

as this will imply, by the Markov inequality the desired stochastic convergence of order $O_p(h_T)$ of the whole difference under consideration, which is again only for the first term,

$$(NT)^{-1}\sum_{s=1}^{T}\operatorname{tr}\left\{\frac{\mathbf{\Lambda}_t'}{\sqrt{N}}\widetilde{\mathbf{\Sigma}}_s\frac{\mathbf{\Lambda}_{s*}^{(1)}}{\sqrt{N}}\right\}K_h\left(\tfrac{t-s}{T}\right)\tfrac{t-s}{T}\ .$$

Note that the sum over kernel weights $T^{-1}\sum_{s=1}^{T}K_h\left(\frac{t-s}{T}\right)$ is of order $O(1)$.

In order to show (B.1) we use that $\operatorname{tr}\left\{\frac{\mathbf{\Lambda}_t'}{\sqrt{N}}\widetilde{\mathbf{\Sigma}}_s\frac{\mathbf{\Lambda}_{s*}^{(1)}}{\sqrt{N}}\right\}=\operatorname{tr}\left\{\frac{1}{N}\mathbf{\Lambda}_{s*}^{(1)}\mathbf{\Lambda}_t'\widetilde{\mathbf{\Sigma}}_s\right\}$ and that, as indicated above,

$$\operatorname{tr}\left\{\tfrac{1}{N}\mathbf{\Lambda}_{s*}^{(1)}\mathbf{\Lambda}_t'\widetilde{\mathbf{\Sigma}}_s\right\}\ \le\ \operatorname{tr}\left\{\tfrac{1}{N}\mathbf{\Lambda}_{s*}^{(1)}\mathbf{\Lambda}_t'\right\}\operatorname{tr}\widetilde{\mathbf{\Sigma}}_s\ .$$

Further, we use the Cauchy-Schwarz inequality to bound

$$\left|\operatorname{tr}\left\{\tfrac{1}{N}\mathbf{\Lambda}_{s*}^{(1)}\mathbf{\Lambda}_t'\right\}\right|\ \le\ \left\|\frac{\mathbf{\Lambda}_{s*}^{(1)}}{\sqrt{N}}\right\|\left\|\frac{\mathbf{\Lambda}_t}{\sqrt{N}}\right\|\ =\ O(1)\,,$$

where we recall the definition of the norm $\|\mathbf{A}\|=\sqrt{\operatorname{tr}\{\mathbf{A}'\mathbf{A}\}}$. But both norms on the right-hand side are bounded from above by condition B3, which implies that

$$\sup_{u\in(0,1)}\ \|\mathbf{\Lambda}_N^{(1)}(u)\|\ =\ O\left(\sqrt{N}\right)\ ,$$

and by the property of $\frac{\mathbf{\Lambda}_t'\mathbf{\Lambda}_t}{N}=\mathbf{I}_r$.

To show (B.1) it remains to show that $\mathbb{E}|\operatorname{tr}\left\{\widetilde{\mathbf{\Sigma}}_s\right\}|=\mathbb{E}\left(\operatorname{tr}\left\{\widetilde{\mathbf{\Sigma}}_s\right\}\right)=O(N)$ (as the eigenvalues, and hence the trace, of $\widetilde{\mathbf{\Sigma}}_s$ are nonnegative by construction).

For this we recall that

$$\mathbb{E}\widetilde{\mathbf{\Sigma}}_t=\mathbf{\Sigma}_t=\mathbf{\Lambda}_t\mathbf{\Sigma}^F\mathbf{\Lambda}_t'+\mathbf{\Sigma}^E\ ,$$

and that with Assumption C3, $\operatorname{tr}\mathbf{\Sigma}^E=O(N)$. Further, by the orthonormality of $\frac{\mathbf{\Lambda}_t}{\sqrt{N}}$ and the invariance of the trace with respect to orthogonal rotations, we easily conclude that

$$\operatorname{tr}\left\{\mathbf{\Lambda}_t\mathbf{\Sigma}^F\mathbf{\Lambda}_t'\right\}=N\ \operatorname{tr}\left\{\frac{\mathbf{\Lambda}_t}{\sqrt{N}}\mathbf{\Sigma}^F\frac{\mathbf{\Lambda}_t'}{\sqrt{N}}\right\}=N\operatorname{tr}\left\{\mathbf{\Sigma}^F\right\}=O(N)\,,$$

as the trace of $\mathbf{\Sigma}^F$ is bounded from above. ■

Proof of Theorem 1. The rescaled covariance matrix is defined as

$$N^{-1}\mathbf{\Sigma}_t=N^{-1}\mathbf{\Sigma}_t^C+N^{-1}\mathbf{\Sigma}^E=N^{-1}\mathbf{\Lambda}_t\mathbf{\Sigma}^F\mathbf{\Lambda}_t'+N^{-1}\mathbf{\Sigma}^E,$$

and the rescaled estimator $N^{-1}\widehat{\mathbf{\Sigma}}_t$ of $N^{-1}\mathbf{\Sigma}_t$ can be written as

$$N^{-1}\widehat{\mathbf{\Sigma}}_t=(NT)^{-1}\mathbf{X}'\mathbf{W}_t\mathbf{X}=(NT)^{-1}\sum_{s=1}^{T}X_sX_s'K_h\left(\tfrac{t-s}{T}\right).$$

Since $\boldsymbol{X}_s = \boldsymbol{\Lambda}_s \boldsymbol{F}_s + \boldsymbol{E}_s$, we have the following decomposition:

$$N^{-1}\widehat{\boldsymbol{\Sigma}}_t = (NT)^{-1}\mathbf{X}'\mathbf{W}_t\mathbf{X} = \mathbf{S}^C_{t,NT} + \mathbf{S}^{CE}_{t,NT} + \left(\mathbf{S}^{CE}_{t,NT}\right)' + \mathbf{S}^E_{t,NT}, \tag{B.2}$$

where

$$\mathbf{S}^C_{t,NT} = (NT)^{-1}\sum_{s=1}^{T} \boldsymbol{\Lambda}_s \boldsymbol{F}_s \boldsymbol{F}_s' \boldsymbol{\Lambda}_s' K_h\left(\tfrac{t-s}{T}\right),$$

$$\mathbf{S}^{CE}_{t,NT} = (NT)^{-1}\sum_{s=1}^{T} \boldsymbol{\Lambda}_s \boldsymbol{F}_s \boldsymbol{E}_s' K_h\left(\tfrac{t-s}{T}\right),$$

$$\mathbf{S}^E_{t,NT} = (NT)^{-1}\sum_{s=1}^{T} \boldsymbol{E}_s \boldsymbol{E}_s' K_h\left(\tfrac{t-s}{T}\right).$$

By the decomposition in (B.2) we have

$$\begin{aligned}
&\left\| N^{-1}\widehat{\boldsymbol{\Sigma}}_t - N^{-1}\boldsymbol{\Sigma}_t \right\| \\
&\quad := \left\| \mathbf{S}^C_{t,NT} + \mathbf{S}^{CE}_{t,NT} + \left(\mathbf{S}^{CE}_{t,NT}\right)' + \mathbf{S}^E_{t,NT} - N^{-1}\left(\boldsymbol{\Lambda}_t \boldsymbol{\Sigma}^F \boldsymbol{\Lambda}_t' + \boldsymbol{\Sigma}^E\right) \right\| \\
&\quad \le \left\| \mathbf{S}^C_{t,NT} - N^{-1}\boldsymbol{\Lambda}_t \boldsymbol{\Sigma}^F \boldsymbol{\Lambda}_t' \right\| + 2\left\| \mathbf{S}^{CE}_{t,NT} \right\| + \left\| \mathbf{S}^E_{t,NT} - N^{-1}\boldsymbol{\Sigma}^E \right\|.
\end{aligned}$$

We now show that each of the terms above is $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. We apply the same Taylor expansion as in the proof of Proposition 1; i.e.,

$$\lambda_{ij}\left(\tfrac{s}{T}\right) = \lambda_{ij}\left(\tfrac{t}{T}\right) + \lambda^{(1)}_{ij}\left(\tfrac{s^*}{T}\right)\ \tfrac{t-s}{T} \quad \text{for } \left|\tfrac{t-s}{T}\right| \le h,$$

with a mean value $\frac{s}{T} \le \frac{s^*}{T} \le \frac{t}{T}$ (w.l.o.g.), which we write in matrix notation and slightly differently as

$$\boldsymbol{\Lambda}_s = \boldsymbol{\Lambda}_t + h_T z_s (\boldsymbol{\Lambda}^{(1)}_t + o(1)) \ = \boldsymbol{\Lambda}_t + h_T z_s \boldsymbol{\Lambda}^{(1)}_t + o(h_T),$$

where $z_s := \frac{t-s}{Th_T}$. Note again that due to the use of the kernel weights in the matrix $\mathbf{W}_t$, which are essentially zero for arguments s with $\frac{t-s}{T} > h$, we can w.l.o.g. argue that $|z_s| \le 1$.

The first term in (B.2) is thus a sum of 4 terms: $\mathbf{S}^C_{t,NT} = \sum_{k=1}^4 {}_k\mathbf{S}^C_{t,NT}$, where

$$\begin{aligned}
{}_1\mathbf{S}^C_{t,NT} &= (NT)^{-1}\sum_{s=1}^{T} \boldsymbol{\Lambda}_t \boldsymbol{F}_s \boldsymbol{F}_s' \boldsymbol{\Lambda}_t' K_h\left(\tfrac{t-s}{T}\right), \\
{}_2\mathbf{S}^C_{t,NT} &= (NT)^{-1}\sum_{s=1}^{T} \boldsymbol{\Lambda}_t \boldsymbol{F}_s \boldsymbol{F}_s' (\boldsymbol{\Lambda}^{(1)'}_t + o(1)) h_T z_s K_h\left(\tfrac{t-s}{T}\right), \\
{}_3\mathbf{S}^C_{t,NT} &= (NT)^{-1}\sum_{s=1}^{T} (\boldsymbol{\Lambda}^{(1)}_t + o(1))\ \boldsymbol{F}_s \boldsymbol{F}_s' \boldsymbol{\Lambda}_t' h_T z_s K_h\left(\tfrac{t-s}{T}\right), \\
{}_4\mathbf{S}^C_{t,NT} &= (NT)^{-1}\sum_{s=1}^{T} (\boldsymbol{\Lambda}^{(1)}_t + o(1)) \boldsymbol{F}_s \boldsymbol{F}_s' (\boldsymbol{\Lambda}^{(1)'}_t + o(1)) h_T^2 z_s^2 K_h\left(\tfrac{t-s}{T}\right).
\end{aligned}$$

We recall that $Th_T^3 \to 0$, so that $O_p\left(\frac{1}{\sqrt{Th_T}}\right) + O\left(h_T\right) = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. In order to show that $\left\|\mathbf{S}_{t,NT}^C - N^{-1}\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_t'\right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right) + O\left(h_T\right)$, we will only treat the first two terms of the above given sum, as the two other terms behave similarly or converge even faster. First we show that $\left\|{}_1\mathbf{S}_{t,NT}^C - N^{-1}\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_t'\right\|$ tends to zero with the appropriate rate. Indeed,

$$\begin{aligned}
&\left\|\frac{1}{NT}\sum_{s=1}^{T}\boldsymbol{\Lambda}_t\boldsymbol{F}_s\boldsymbol{F}_s'\boldsymbol{\Lambda}_t'K_h\left(\tfrac{t-s}{T}\right) - N^{-1}\boldsymbol{\Lambda}_t\boldsymbol{\Sigma}^F\boldsymbol{\Lambda}_t'\right\| \\
&\qquad \le \left\|\frac{1}{N}(\boldsymbol{\Lambda}_t\otimes\boldsymbol{\Lambda}_t)\right\|\left\|\frac{1}{T}\sum_{s=1}^{T}\operatorname{vec}\left(\boldsymbol{F}_s\boldsymbol{F}_s'\right)K_h\left(\tfrac{t-s}{T}\right) - \operatorname{vec}\left(\boldsymbol{\Sigma}^F\right)\right\|
\end{aligned}$$

and $\left\|\frac{1}{N}(\boldsymbol{\Lambda}_t\otimes\boldsymbol{\Lambda}_t)\right\| = \frac{1}{N}\sqrt{\operatorname{tr}\left\{\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t'\otimes\boldsymbol{\Lambda}_t\boldsymbol{\Lambda}_t'\right\}} = \operatorname{tr}\left\{\frac{1}{N}\boldsymbol{\Lambda}_t'\boldsymbol{\Lambda}_t\right\} \to \operatorname{tr}\{\boldsymbol{\Sigma}_\Lambda(u)\} = O(1)$ by Assumption B2. Then it suffices to show that

$$\sqrt{Th_T}\left\|\frac{1}{T}\sum_{s=1}^{T}\operatorname{vec}\left(\boldsymbol{F}_s\boldsymbol{F}_s'\right)K_h\left(\tfrac{t-s}{T}\right) - \operatorname{vec}\left(\boldsymbol{\Sigma}^F\right)\right\| = O_p(1). \tag{B.3}$$

To do that we show that the expectation and the variance of $\sqrt{Th_T}$ times the argument of the norm in (B.3) tend to zero. We recall that this is sufficient as the considered norm is the euclidean norm in $\mathbb{R}^r$ with r fixed. For the expectation we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{T}\sum_{s=1}^{T}\operatorname{vec}\left(\boldsymbol{F}_s\boldsymbol{F}_s'\right)K_h\left(\tfrac{t-s}{T}\right) - \operatorname{vec}\left(\boldsymbol{\Sigma}^F\right)\right] &= \left[\frac{1}{T}\sum_{s=1}^{T}K_h\left(\tfrac{t-s}{T}\right)\mathbf{I}_{r^2} - \mathbf{I}_{r^2}\right]\operatorname{vec}\left(\boldsymbol{\Sigma}^F\right) \\
&= \left\{\left(1 + O\left(T^{-1}h_T^{-1}\right)\right)\mathbf{I}_{r^2} - \mathbf{I}_{r^2}\right\}\operatorname{vec}\left(\boldsymbol{\Sigma}^F\right) \\
&= O\left(T^{-1}h_T^{-1}\right)\operatorname{vec}\left(\boldsymbol{\Sigma}^F\right) = O\left(T^{-1}h_T^{-1}\right).
\end{aligned}$$

The argument of the norm in (B.3) can be written as

$$\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{G}_sK_h\left(\tfrac{t-s}{T}\right) + O\left(T^{-1}h_T^{-1}\right),$$

where $\boldsymbol{G}_t := \operatorname{vec}\left(\boldsymbol{F}_t\boldsymbol{F}_t' - \boldsymbol{\Sigma}^F\right) \overset{\text{iid}}{\sim} (\mathbf{0},\mathbf{P})$ for a given matrix $\mathbf{P}$. Then

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\frac{1}{T}\sum_{s=1}^{T}\boldsymbol{G}_sK_h\left(\tfrac{t-s}{T}\right) + O\left(T^{-1}h_T^{-1}\right)\right] &= \frac{1}{T^2}\mathbb{E}\left[\sum_{s=1}^{T}\boldsymbol{G}_s\boldsymbol{G}_s'K_h^2\left(\tfrac{t-s}{T}\right)\right] \\
&= \frac{1}{T^2}\mathbb{E}\left[\boldsymbol{G}_t\boldsymbol{G}_t'\right]\sum_{s=1}^{T}K_h^2\left(\tfrac{t-s}{T}\right) \\
&= \mathbf{P}\frac{1}{T^2}\sum_{s=1}^{T}K_h^2\left(\tfrac{t-s}{T}\right) = O\left(T^{-1}h_T^{-1}\right).
\end{aligned}$$

The last assertion is due to a classical argument in nonparametric curve estimation with kernels of finite second moment.

Second, in order to prove that $\left\| {}_2\mathbf{S}^C_{t,NT} \right\| = o_p\left(\frac{1}{\sqrt{Th_T}}\right)$, we show quite analogously to the above and recalling the condition $Th_T^3 = o(1)$ that

$$\sqrt{Th_T}\left\| (NT)^{-1}\sum_{s=1}^{T}\mathbf{\Lambda}_t \boldsymbol{F}_s \boldsymbol{F}_s' (\mathbf{\Lambda}_t^{(1)\prime} + o(1))\, h_T z_s K_h\left(\tfrac{t-s}{T}\right) \right\|$$

$$= \sqrt{Th_T}\; O_p(h_T) \;=\; o_p(1).$$

To do so we essentially have to use that

$$\frac{1}{N}\left\| \mathbf{\Lambda}_t' \mathbf{\Lambda}_t^{(1)} \right\| = \frac{1}{N}\sqrt{\operatorname{tr}\left(\mathbf{\Lambda}_t' \mathbf{\Lambda}_t \mathbf{\Lambda}_t^{(1)\prime} \mathbf{\Lambda}_t^{(1)} \right)}$$

$$= \sqrt{N^{-1}\operatorname{tr}\left(\mathbf{\Lambda}_t'\mathbf{\Lambda}_t\right)}\;\sqrt{N^{-1}\operatorname{tr}\left(\mathbf{\Lambda}_t^{(1)\prime}\mathbf{\Lambda}_t^{(1)}\right)} = O(1)$$

by Assumption B3. Treatment of the terms ${}_3\mathbf{S}^C_{t,NT}$ and ${}_4\mathbf{S}^C_{t,NT}$ would be similar.

By similar arguments, it can be shown that $\left\| \mathbf{S}^{CE}_{t,NT} \right\|$ and $\left\| \mathbf{S}^{E}_{t,NT} - N^{-1}\mathbf{\Sigma}^E \right\|$ are both $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. Let $\mathbf{S}^{CE}_{t,NT}$ be given by the sum ${}_1\mathbf{S}^{CE}_{t,NT} + {}_2\mathbf{S}^{CE}_{t,NT}$, again obtained by using a Taylor expansion of $\mathbf{\Lambda}_s$; that is,

$${}_1\mathbf{S}^{CE}_{t,NT} = \frac{1}{NT}\sum_{s=1}^{T}\mathbf{\Lambda}_t \boldsymbol{F}_s \boldsymbol{E}_s' K_h\left(\tfrac{t-s}{T}\right),$$

$${}_2\mathbf{S}^{CE}_{t,NT} = \frac{1}{NT}\sum_{s=1}^{T} h_T z_s \left(\mathbf{\Lambda}_t^{(1)} + o(1)\right)\boldsymbol{F}_s \boldsymbol{E}_s' K_h\left(\tfrac{t-s}{T}\right).$$

For the first term we have $\left\| {}_1\mathbf{S}^{CE}_{t,NT} \right\| \le \left\| \frac{1}{\sqrt{N}}\mathbf{\Lambda}_t \right\| \left\| \frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\boldsymbol{F}_s\boldsymbol{E}_s' K_h\left(\frac{t-s}{T}\right) \right\|$, where $\left\| \frac{1}{\sqrt{N}}\mathbf{\Lambda}_t \right\| = O(1)$ by Assumption B2, and then $\left\| {}_1\mathbf{S}^{CE}_{t,NT} \right\| \propto \left\| \frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\operatorname{vec}\left[\boldsymbol{F}_s\boldsymbol{E}_s'\right] K_h\left(\frac{t-s}{T}\right) \right\|$. Since $\mathbb{C}\mathrm{ov}[\boldsymbol{F}_t, \boldsymbol{E}_t] = 0$ for all t,

$$\mathbb{E}\left\{ \frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\operatorname{vec}\left[\boldsymbol{F}_s\boldsymbol{E}_s'\right] K_h\left(\tfrac{t-s}{T}\right) \right\} = \mathbf{0} \qquad \forall t.$$

In order to treat the variance now we define $\boldsymbol{H}_t := \operatorname{vec}\left[\boldsymbol{F}_t\boldsymbol{E}_t'\right] \overset{\text{iid}}{\sim} (\mathbf{0}, \mathbf{Q})$ for another given matrix $\mathbf{Q}$, which does indeed exist due to Assumptions A2 and C2:

$$\mathbb{V}\mathrm{ar}\left[\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\boldsymbol{H}_s K_h\left(\tfrac{t-s}{T}\right) \right] = \frac{1}{NT^2}\mathbb{E}\left[\sum_{s=1}^{T}\boldsymbol{H}_s\boldsymbol{H}_s' K_h^2\left(\tfrac{t-s}{T}\right) \right]$$

$$= \frac{1}{NT^2}\mathbb{E}\left[\boldsymbol{H}_t\boldsymbol{H}_t'\right]\sum_{s=1}^{T} K_h^2\left(\tfrac{t-s}{T}\right)$$

$$= N^{-1}\mathbf{Q}\,\frac{1}{T^2}\sum_{s=1}^{T} K_h^2\left(\tfrac{t-s}{T}\right) = O\left(N^{-1}T^{-1}h_T^{-1}\right).$$

This means that each term of the $Nr \times 1$ vectors $\frac{1}{T\sqrt{N}} \sum_{s=1}^{T} \text{vec}\left[\boldsymbol{F}_s \boldsymbol{E}_s'\right] K_h\left(\frac{t-s}{T}\right)$ is $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$, and thus

$$\left\| \frac{1}{T\sqrt{N}} \sum_{s=1}^{T} \text{vec}\left[\boldsymbol{F}_s \boldsymbol{E}_s'\right] K_h\left(\tfrac{t-s}{T}\right) \right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$$

because the norm of a vector of dimension $\propto N$ increases at a rate $\propto \sqrt{N}$.

For the term $\left\| {}_2\mathbf{S}_{t,NT}^{CE} \right\|$ we show quite analogously to the term $\left\| {}_2\mathbf{S}_{t,NT}^{C} \right\|$ that

$$\sqrt{Th_T} \left\| (NT)^{-1} \sum_{s=1}^{T} (\boldsymbol{\Lambda}_t^{(1)} + o(1)) \boldsymbol{F}_s \boldsymbol{E}_s' \, h_T z_s K_h\left(\tfrac{t-s}{T}\right) \right\| = \sqrt{Th_T}\; O_p(h_T) = o_p(1).$$

To do so we essentially have to use that

$$\frac{1}{N} \left\| \boldsymbol{E}_s' \boldsymbol{\Lambda}_t^{(1)} \right\| \leq \left\| \frac{\boldsymbol{\Lambda}_t^{(1)}}{\sqrt{N}} \right\| \left\| \frac{\boldsymbol{E}_s}{\sqrt{N}} \right\| = \left\| \frac{\boldsymbol{\Lambda}_t^{(1)}}{\sqrt{N}} \right\| \sqrt{\text{tr}\left\{ \frac{\boldsymbol{E}_s' \boldsymbol{E}_s}{N} \right\}} = \left\| \frac{\boldsymbol{\Lambda}_t^{(1)}}{\sqrt{N}} \right\| \sqrt{\frac{\boldsymbol{E}_s' \boldsymbol{E}_s}{N}} = O_p(1),$$

because $\left\| \boldsymbol{\Lambda}_t^{(1)} / \sqrt{N} \right\| = O(1)$ by Assumption B3, and $\frac{\boldsymbol{E}_s' \boldsymbol{E}_s}{N} = O_p(1)$ using a law of large numbers.

For the last term we now show that $\mathbf{S}_{t,NT}^{E} = N^{-1} \boldsymbol{\Sigma}^{E} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$. The expectation of the term $\left[\mathbf{S}_{t,NT}^{E} - N^{-1} \boldsymbol{\Sigma}^{E}\right]$ is

$$\begin{aligned}
\left[\frac{1}{NT} \sum_{s=1}^{T} \mathbb{E}\left(\boldsymbol{E}_s \boldsymbol{E}_s'\right) K_h\left(\tfrac{t-s}{T}\right) \right] - N^{-1} \boldsymbol{\Sigma}^{E} &= N^{-1} \boldsymbol{\Sigma}^{E} \left[T^{-1} \sum_{s=1}^{T} K_h\left(\tfrac{t-s}{T}\right) \right] - N^{-1} \boldsymbol{\Sigma}^{E} \\
&= N^{-1} \boldsymbol{\Sigma}^{E} \left[1 + O\left(T^{-1} h_T^{-1}\right) \right] - N^{-1} \boldsymbol{\Sigma}^{E} \\
&= O\left(\tfrac{1}{NTh_T}\right).
\end{aligned}$$

To compute the variance, define, as above, $\boldsymbol{J}_t := \text{vec}\left[\boldsymbol{E}_t \boldsymbol{E}_t' - \boldsymbol{\Sigma}^{E}\right] \overset{\text{iid}}{\sim} (\mathbf{0}, \mathbf{R})$ for another given matrix $\mathbf{R}$, which does exist due to Assumption C2, and obtain

$$\begin{aligned}
&\mathbb{V}\text{ar}\left[\frac{1}{TN} \sum_{s=1}^{T} \boldsymbol{J}_s K_h\left(\tfrac{t-s}{T}\right) + O\left(\tfrac{1}{NTh_T}\right) \right] \\
&\quad = \frac{1}{N^2T^2} \mathbb{E}\left[\sum_{s=1}^{T} \boldsymbol{J}_s \boldsymbol{J}_s' K_h^2\left(\tfrac{t-s}{T}\right) \right] = \frac{\mathbb{E}\left[\boldsymbol{J}_t \boldsymbol{J}_t'\right]}{N^2T^2} \sum_{s=1}^{T} K_h^2\left(\tfrac{t-s}{T}\right) \\
&\quad = N^{-2} \mathbf{R} \frac{1}{T^2} \sum_{s=1}^{T} K_h^2\left(\tfrac{t-s}{T}\right) = O\left(N^{-2} T^{-1} h_T^{-1}\right).
\end{aligned}$$

Each term of the $N^2 \times 1$ vectors $\frac{1}{TN} \sum_{s=1}^{T} \text{vec}\left[\boldsymbol{E}_s \boldsymbol{E}_s' - \boldsymbol{\Sigma}^{E}\right] K_h\left(\frac{t-s}{T}\right)$ is $O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$, and thus

$$\left\| \frac{1}{NT} \sum_{s=1}^{T} \text{vec}\left[\boldsymbol{E}_s \boldsymbol{E}_s' - \boldsymbol{\Sigma}^{E}\right] K_h\left(\tfrac{t-s}{T}\right) \right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right),$$

because the norm of a vector of dimension $\propto N^2$ increases at a rate $\propto N$. ■

Proof of Proposition 2. We have to show that

(i) *estimation error:* $\|\widehat{\mathbf{V}}_t - \mathbf{V}_{t,N}\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$;

(ii) *approximation error:* $\|\mathbf{V}_{t,N} - \mathbf{V}_t\| \leq \|\mathbf{V}_{t,N} - \mathbf{V}^C_{t,N}\| + \|\mathbf{V}^C_{t,N} - \mathbf{V}_t\| = O\left(N^{-1}\right) + O\left(\ell_N\right)$.

The matrices $\widehat{\mathbf{V}}_t$ and $\mathbf{V}_{t,N}$ are continuous functions of $\widehat{\boldsymbol{\Sigma}}_t$ and $\boldsymbol{\Sigma}_t$, respectively. Then the estimation error is $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$ by Theorem 1.

Consider part (ii), and let v_{jt}, v^C_{jt}, and v^E_j denote the jth eigenvalue of $\boldsymbol{\Sigma}_t$, $\boldsymbol{\Sigma}^C_t$, and $\boldsymbol{\Sigma}^E$, respectively ($j = 1, \ldots, N$). By Weyl's theorem applied to the decomposition of $\boldsymbol{\Sigma}_t$ in (4) we have

$$\mathrm{v}_{jt} - \mathrm{v}^C_{jt} \leq \mathrm{v}^C_1, \qquad j = 1, \ldots, N. \tag{B.4}$$

Then we have, collecting the largest r eigenvalues on the left-hand side of (B.4)

$$\|\mathbf{V}_{t,N} - \mathbf{V}^C_{t,N}\| \leq \tfrac{1}{N}\|\mathrm{v}^C_1 \mathbf{I}_r\| = \tfrac{1}{N}\mathrm{v}^C_1\sqrt{r},$$

where we recall that $\mathbf{V}_{t,N}$ and $\mathbf{V}^C_{t,N}$ are the $r \times r$ diagonal matrices containing the largest eigenvalues of $\frac{1}{N}\boldsymbol{\Sigma}_t$ and $\frac{1}{N}\boldsymbol{\Sigma}^C_t$, respectively. Hence by Definition 1(ii), the idiosyncratic variance error $\|\mathbf{V}_{t,N} - \mathbf{V}^C_{t,N}\|$ is $O\left(\frac{1}{N}\right)$.

We now show that the common variance error $\|\mathbf{V}^C_{t,N} - \mathbf{V}_t\|$ is of the same order as the loadings scale error ℓ_N defined in (22). Let $\mathbf{S}^{\Lambda F}_t = \frac{\boldsymbol{\Lambda}'_t\boldsymbol{\Lambda}_t}{N}\boldsymbol{\Sigma}^F$ and define $\boldsymbol{\Delta}_{t,N} := \mathbf{S}^{\Lambda F}_t - \boldsymbol{\Sigma}^{\Lambda}_t\boldsymbol{\Sigma}^F$. The matrix $\mathbf{V}^C_{t,N}$ contains the eigenvalues of $\mathbf{S}^{\Lambda F}_t$ (by the definition of $\mathbf{V}^C_{t,N}$ and by (12) in Sect. 5.2 of Lütkepohl, 1996, p. 65); then, applying Weyl's theorem in the same way as above, we have

$$\|\mathbf{V}^C_{t,N} - \mathbf{V}_t\| \leq \|\mathrm{v}^{\Delta}_{1t}\mathbf{I}_r\| = \sqrt{r}\,|\mathrm{v}^{\Delta}_{1t}|,$$

where v^{Δ}_{1t} is the largest eigenvalue of $\boldsymbol{\Delta}_{t,N}$. If we define $\boldsymbol{P}_{1t}$ as the orthonormal eigenvector corresponding to v^{Δ}_{1t}, we have $\mathrm{v}^{\Delta}_{1t} = \boldsymbol{P}'_{1t}\boldsymbol{\Delta}_{t,N}\boldsymbol{P}_{1t}$, and thus

$$\begin{aligned}|\mathrm{v}^{\Delta}_{1t}| &\leq \|\boldsymbol{P}'_{1t}\boldsymbol{P}_{1t}\|\,\|\boldsymbol{\Delta}_{t,N}\| = \|\boldsymbol{\Delta}_{t,N}\| = \|\mathbf{S}^{\Lambda F}_t - \boldsymbol{\Sigma}^{\Lambda}_t\boldsymbol{\Sigma}^F\| \\ &= \left\|\left(\tfrac{\boldsymbol{\Lambda}'_t\boldsymbol{\Lambda}_t}{N} - \boldsymbol{\Sigma}^{\Lambda}\right)\boldsymbol{\Sigma}^F\right\| \leq \left\|\tfrac{\boldsymbol{\Lambda}'_t\boldsymbol{\Lambda}_t}{N} - \boldsymbol{\Sigma}^{\Lambda}_t\right\|\|\boldsymbol{\Sigma}^F\|.\end{aligned}$$

Hence

$$\|\mathbf{V}^C_{t,N} - \mathbf{V}_t\| \leq \sqrt{r}\,|\mathrm{v}^{\Delta}_{1t}| \leq \sqrt{r}\left\|\tfrac{\boldsymbol{\Lambda}'_t\boldsymbol{\Lambda}_t}{N} - \boldsymbol{\Sigma}^{\Lambda}_t\right\|\|\boldsymbol{\Sigma}^F\| = O(1)\,\ell_N\,O(1) = O\left(\ell_N\right)$$

by (22). Finally, for the overall error we have

$$\begin{aligned}\|\widehat{\mathbf{V}}_t - \mathbf{V}_t\| &\leq \|\widehat{\mathbf{V}}_t - \mathbf{V}_{t,N}\| + \|\mathbf{V}_{t,N} - \mathbf{V}^C_{t,N}\| + \|\mathbf{V}^C_{t,N} - \mathbf{V}_t\| \\ &= O_p\left([Th_T]^{-1/2}\right) + O\left(N^{-1}\right) + O\left(\ell_N\right).\end{aligned}$$

■

Auxiliary Results. Proposition 2 gives the asymptotic behavior of the largest r eigenvalues of $\widehat{\boldsymbol{\Sigma}}_t$. The asymptotic behavior of the corresponding r eigenvectors is given in Theorem 2. In order to prove Theorems 2 and 3, we need to study the behavior of the matrix $\frac{\boldsymbol{\Lambda}_t'\widehat{\boldsymbol{\Lambda}}_t}{N}$. This is provided by Proposition 3 below, which is an auxiliary result used in the proofs of our main results, Theorems 2 and 3. For the proof of Proposition 3 we use the following lemma, which is analogous to Bai (2003, Lem. A.3(ii)).

LEMMA 2. *Let*

$$\widehat{\mathbf{Q}}_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2} \frac{\boldsymbol{\Lambda}_t'\widehat{\boldsymbol{\Lambda}}_t}{N};$$

then we have

$$\min\left(\sqrt{Th_T}, \sqrt{N}, \ell_N^{-1}\right) \left\| \widehat{\mathbf{Q}}_{t,NT}' \widehat{\mathbf{Q}}_{t,NT} - \mathbf{V}_t \right\| = O_p(1).$$

Proof. By (19) we have $\widehat{\mathbf{V}}_t = N^{-1}\widehat{\boldsymbol{\Lambda}}_t' \frac{\mathbf{X}'\mathbf{W}_t\mathbf{X}}{NT}\widehat{\boldsymbol{\Lambda}}_t$, and by Theorem 1

$$\frac{\mathbf{X}'\mathbf{W}_t\mathbf{X}}{NT} = N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\Lambda}_t' + N^{-1}\boldsymbol{\Sigma}^E + \mathbf{Z}_{t,NT},$$

where the $N \times N$ matrix $\mathbf{Z}_{t,NT}$ is $O_p\left[\frac{1}{N\sqrt{Th_T}}\right]$. Then

$$\begin{aligned}
\widehat{\mathbf{V}}_t &= \frac{\widehat{\boldsymbol{\Lambda}}_t'\boldsymbol{\Lambda}_t}{N}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\frac{\boldsymbol{\Lambda}_t'\widehat{\boldsymbol{\Lambda}}_t}{N} + \frac{\widehat{\boldsymbol{\Lambda}}_t'}{\sqrt{N}}\frac{\boldsymbol{\Sigma}^E}{N}\frac{\widehat{\boldsymbol{\Lambda}}_t}{\sqrt{N}} + \frac{\widehat{\boldsymbol{\Lambda}}_t'\mathbf{Z}_{t,NT}\widehat{\boldsymbol{\Lambda}}_t}{N} \\
&= \widehat{\mathbf{Q}}_{t,NT}'\widehat{\mathbf{Q}}_{t,NT} + \frac{\widehat{\boldsymbol{\Lambda}}_t'}{\sqrt{N}}\frac{\boldsymbol{\Sigma}^E}{N}\frac{\widehat{\boldsymbol{\Lambda}}_t}{\sqrt{N}} + \frac{\widehat{\boldsymbol{\Lambda}}_t'\mathbf{Z}_{t,NT}\widehat{\boldsymbol{\Lambda}}_t}{N},
\end{aligned}$$

and thus

$$\begin{aligned}
\left\|\widehat{\mathbf{V}}_t - \widehat{\mathbf{Q}}_{t,NT}'\widehat{\mathbf{Q}}_{t,NT}\right\| &\le \left\|\frac{\widehat{\boldsymbol{\Lambda}}_t'\widehat{\boldsymbol{\Lambda}}_t}{N}\right\|\left\|\frac{\boldsymbol{\Sigma}^E}{N}\right\| + \left\|\frac{\widehat{\boldsymbol{\Lambda}}_t'\widehat{\boldsymbol{\Lambda}}_t}{N}\right\|\left\|\mathbf{Z}_{t,NT}\right\| \\
&= \left\|\frac{\widehat{\boldsymbol{\Lambda}}_t'\widehat{\boldsymbol{\Lambda}}_t}{N}\right\|\left(\left\|\frac{\boldsymbol{\Sigma}^E}{N}\right\| + \left\|\mathbf{Z}_{t,NT}\right\|\right) \\
&= O_p(1)\left[O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{1}{\sqrt{Th_T}}\right)\right] \\
&= O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{Th_T}}\right),
\end{aligned}$$

since

$$\begin{aligned}
\left\|\frac{\boldsymbol{\Sigma}^E}{N}\right\| &= N^{-1}\sqrt{\sum_{i=1}^N\sum_{j=1}^N\left(\sigma_{ij}^E\right)^2} \le N^{-1}\sqrt{N\max_{1\le i\le N}\sum_{j=1}^N\left(\sigma_{ij}^E\right)^2} \\
&\le \frac{1}{\sqrt{N}}\max_{1\le i\le N}\sum_{j=1}^N\left|\sigma_{ij}^E\right| = O\left(\frac{1}{\sqrt{N}}\right)
\end{aligned}$$

by Assumption C3. If we use that

$$\left\|\widehat{\mathbf{Q}}'_{t,NT}\widehat{\mathbf{Q}}_{t,NT}-\mathbf{V}_t\right\| \le \left\|\widehat{\mathbf{Q}}'_{t,NT}\widehat{\mathbf{Q}}_{t,NT}-\widehat{\mathbf{V}}_t\right\| + \left\|\widehat{\mathbf{V}}_t-\mathbf{V}_t\right\|,$$

then with Proposition 2 we have the result. ■

PROPOSITION 3. *Under Assumptions A–E,*

$$\min\left(\sqrt{Th},\sqrt{N},\ell_N^{-1}\right)\left\|\frac{\mathbf{\Lambda}'_t\widehat{\mathbf{\Lambda}}_t}{N}-\left(\mathbf{\Sigma}^F\right)^{-\frac{1}{2}}\Upsilon_t\mathbf{V}_t^{\frac{1}{2}}\right\| = O_p(1),$$

where Υ_t is the $r\times r$ matrix containing the orthonormal eigenvectors of the $r\times r$ matrix $\left(\mathbf{\Sigma}^F\right)^{1/2}\mathbf{\Sigma}_t^{\Lambda}\left(\mathbf{\Sigma}^F\right)^{1/2}$.

Proof. Consider equation (19),

$$(NT)^{-1}\mathbf{X}'\mathbf{W}_t\mathbf{X}\widehat{\mathbf{\Lambda}}_t = \widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t,$$

and multiply it on the left by $\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{N}$ to obtain

$$(NT)^{-1}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{N}\mathbf{X}'\mathbf{W}_t\mathbf{X}\widehat{\mathbf{\Lambda}}_t = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{N}\widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t.$$

By (B.2) we get

$$\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{N}\left[\mathbf{S}^C_{t,NT}+\mathbf{S}^{CE}_{t,NT}+\left(\mathbf{S}^{CE}_{t,NT}\right)'+\mathbf{S}^E_{t,NT}\right]\widehat{\mathbf{\Lambda}}_t = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{N}\widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t\,,$$

which can be written as

$$\widehat{\mathbf{S}}^C_{t,NT}+\widehat{\mathbf{S}}^{CE}_{t,NT}+\widehat{\mathbf{S}}^E_{t,NT} = \widehat{\mathbf{Q}}_{t,NT}\widehat{\mathbf{V}}_t$$

where

$$\widehat{\mathbf{S}}^C_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{\sqrt{N}}\mathbf{S}^C_{t,NT}\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}},$$

$$\widehat{\mathbf{S}}^{CE}_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{\sqrt{N}}\left[\mathbf{S}^{CE}_{t,NT}+\left(\mathbf{S}^{CE}_{t,NT}\right)'\right]\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}},$$

$$\widehat{\mathbf{S}}^E_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\mathbf{\Lambda}'_t}{\sqrt{N}}\mathbf{S}^E_{t,NT}\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}}.$$

By Theorem 1, $\mathbf{S}^C_{t,NT} = N^{-1}\mathbf{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\mathbf{\Lambda}'_t+\mathbf{U}_{NT}$, where $\mathbf{U}_{NT} = O_p\left(\frac{1}{N\sqrt{Th_T}}\right)+O\left(\frac{h_T}{N}\right) = O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$ (we recall that $Th_T^3\to 0$ by Assumption D). Then $\mathbf{U}_{NT}$ is an $N\times N$ matrix, each element being $O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$. Since

$\|\frac{\mathbf{F}'\mathbf{F}}{T}\| = O_p(1)$ by Assumption A1 and a law of large numbers,

$\|\frac{\mathbf{\Lambda}_t}{\sqrt{N}}\| = O(1)$ by Assumption B2,

$\|\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}}\| = O_p(1)$ by the constraint in (18), if we plug the definition of $\mathbf{S}^C_{t,NT}$ from Theorem 1 into the expression of $\widehat{\mathbf{S}}^C_{t,NT}$ above, we have

$$\widehat{\mathbf{S}}^C_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2} \frac{\boldsymbol{\Lambda}_t'\boldsymbol{\Lambda}_t}{N} \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right) \frac{\boldsymbol{\Lambda}_t'\widehat{\boldsymbol{\Lambda}}_t}{N} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$$
$$= \mathbf{P}_{t,NT}\widehat{\mathbf{Q}}_{t,NT} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right),$$

where $\mathbf{P}_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}\frac{\boldsymbol{\Lambda}_t'\boldsymbol{\Lambda}_t}{N}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{1/2}$. The term $\|\widehat{\mathbf{S}}^{CE}_{t,NT}\|$ is $O_p\left(\frac{1}{\sqrt{Th}}\right)$ because $\|\mathbf{S}^{CE}_{t,NT}\|$ is $O_p\left(\frac{1}{\sqrt{Th}}\right)$, and the term $\|\widehat{\mathbf{S}}^{E}_{t,NT}\|$ is $O_p\left(\frac{1}{\sqrt{Th}} + \frac{1}{\sqrt{N}}\right)$ because $\|\mathbf{S}^{E}_{t,NT}\|$ is $O_p\left(\frac{1}{\sqrt{Th}} + \frac{1}{\sqrt{N}}\right)$ (see proof of Theorem 1). Then we have

$$\mathbf{P}_{t,NT}\widehat{\mathbf{Q}}_{t,NT} + \widehat{\mathbf{D}}_{t,NT} = \widehat{\mathbf{Q}}_{t,NT}\widehat{\mathbf{V}}_t, \tag{B.5}$$

where $\widehat{\mathbf{D}}_{t,NT} = \widehat{\mathbf{S}}^{CE}_{t,NT} + \widehat{\mathbf{S}}^{E}_{t,NT}$, and thus $\|\widehat{\mathbf{D}}_{t,NT}\| = O_p\left(\frac{1}{\sqrt{Th}} + \frac{1}{\sqrt{N}}\right)$. From (B.5) we have

$$\left[\mathbf{P}_{t,NT} + \widehat{\mathbf{D}}_{t,NT}\widehat{\mathbf{Q}}^{-1}_{t,NT}\right]\widehat{\mathbf{Q}}_{t,NT} = \widehat{\mathbf{Q}}_{t,NT}\widehat{\mathbf{V}}_t.$$

Note that by Lemma 2, for large values of T and N, the matrix $\widehat{\mathbf{Q}}'_{t,NT}\widehat{\mathbf{Q}}_{t,NT}$ has full rank. This implies that, for large T and N, the matrix $\widehat{\mathbf{Q}}_{t,NT}$ is of full rank, too. For the same reason, the matrix $\widehat{\mathbf{V}}^*_t := \operatorname{diag}\{\widehat{\mathbf{Q}}'_{t,NT}\widehat{\mathbf{Q}}_{t,NT}\}$ is also invertible for large T N. Then we can denote $\boldsymbol{\Upsilon}_{t,NT} = \widehat{\mathbf{Q}}_{t,NT}\widehat{\mathbf{V}}^{*-1/2}_t$ so that each column of $\boldsymbol{\Upsilon}_{t,NT}$ has unit length, and we have

$$\left[\mathbf{P}_{t,NT} + \widehat{\mathbf{D}}_{t,NT}\widehat{\mathbf{Q}}^{-1}_{t,NT}\right]\boldsymbol{\Upsilon}_{t,NT} = \boldsymbol{\Upsilon}_{t,NT}\widehat{\mathbf{V}}_t,$$

where each column of $\boldsymbol{\Upsilon}_{t,NT}$ is an eigenvector of the matrix $\left[\mathbf{P}_{t,NT} + \widehat{\mathbf{D}}_{t,NT}\widehat{\mathbf{Q}}^{-1}_{t,NT}\right]$. Note that $\left[\mathbf{P}_{t,NT} + \widehat{\mathbf{D}}_{t,NT}\widehat{\mathbf{Q}}^{-1}_{t,NT}\right]$ converges to $\mathbf{P}_t = \left(\boldsymbol{\Sigma}^F\right)^{1/2}\boldsymbol{\Sigma}^\Lambda_t\left(\boldsymbol{\Sigma}^F\right)^{1/2}$ by Assumptions A1 and B2 and $\|\widehat{\mathbf{D}}_{t,NT}\| = o_p(1)$, implying that for large values of T and N the diagonal matrix $\widehat{\mathbf{V}}_t$ contains the eigenvalues of $\mathbf{P}_{t,NT}$. Because the ordering of eigenvectors in the points of intersections is identified by our convention (see Sect. 3.1), the matrix $\boldsymbol{\Upsilon}_{t,NT}$ converges to the matrix $\boldsymbol{\Upsilon}_t$, similar to the proof of Proposition 1 of Bai (2003, pp. 161–162). By Assumptions A1 and B2 and Slutsky's theorem,

$$\min\left(\sqrt{T}, \ell_N^{-1}\right)\|\mathbf{P}_{t,NT} - \mathbf{P}_t\| = O_p(1) \quad \text{and} \quad \min\left(\sqrt{T}, \ell_N^{-1}\right)\|\boldsymbol{\Upsilon}_{t,NT} - \boldsymbol{\Upsilon}_t\| = O_p(1).$$

From the definitions of $\widehat{\mathbf{Q}}_{t,NT}$ and $\boldsymbol{\Upsilon}_{t,NT}$ it follows that

$$\frac{\boldsymbol{\Lambda}_t'\widehat{\boldsymbol{\Lambda}}_t}{N} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1/2}\widehat{\mathbf{Q}}_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1/2}\boldsymbol{\Upsilon}_{t,NT}\widehat{\mathbf{V}}^{*1/2}_t.$$

The term $\left\|\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right) - \left(\boldsymbol{\Sigma}^F\right)\right\|$ is $O_p\left(T^{-1/2}\right)$ by Assumption A1 and a law of large numbers. By Proposition 2 and Lemma 2,

$$\min\left(\sqrt{Th_T}, \sqrt{N}, \ell_N^{-1}\right)\|\widehat{\mathbf{V}}^*_t - \mathbf{V}_t\|$$
$$= \min\left(\sqrt{Th_T}, \sqrt{N}, \ell_N^{-1}\right)\left\|\operatorname{diag}\{\widehat{\mathbf{Q}}'_{t,NT}\widehat{\mathbf{Q}}_{t,NT}\} - \mathbf{V}_t\right\| = O_p(1),$$

and the proof of Proposition 3 is complete. ■

Note that due to Assumption E and our identifiability assumption in intersection points of eigenvalues of $\mathbf{\Sigma}^{\Lambda}(u)\mathbf{\Sigma}^{F}$, the limiting function is uniquely defined in each $u \in (0,1)$ and a continuous function of u.

Proof of Theorem 2. For this proof we will use the following decompositions.

(i) $\frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t \mathbf{R}_t\right\| \le \frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t \widehat{\mathbf{R}}_{t,NT}\right\| + \frac{1}{\sqrt{N}}\left\|\mathbf{\Lambda}_t \widehat{\mathbf{R}}_{t,NT} - \mathbf{\Lambda}_t \mathbf{R}_t\right\|$,

(ii) $\left\|\widehat{\boldsymbol{\lambda}}_{it} - \mathbf{R}_t'\boldsymbol{\lambda}_{it}\right\| \le \left\|\widehat{\boldsymbol{\lambda}}_{it} - \widehat{\mathbf{R}}_{t,NT}'\boldsymbol{\lambda}_{it}\right\| + \left\|\widehat{\mathbf{R}}_{t,NT}'\boldsymbol{\lambda}_{it} - \mathbf{R}_t'\boldsymbol{\lambda}_{it}\right\|$,

where the estimator $\widehat{\mathbf{R}}_{t,NT}$ of the transformation matrix $\mathbf{R}_t$ is defined as

$$\widehat{\mathbf{R}}_{t,NT} = \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)\widehat{\mathbf{V}}_t^{-1}, \tag{B.6}$$

and we will show that

(ia) $\frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t \widehat{\mathbf{R}}_{t,NT}\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}}\right)$,

(ib) $\frac{1}{\sqrt{N}}\left\|\mathbf{\Lambda}_t \widehat{\mathbf{R}}_{t,NT} - \mathbf{\Lambda}_t \mathbf{R}_t\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}} + \ell_N^{-1}\right)$,

(iia) $\left\|\widehat{\boldsymbol{\lambda}}_{it} - \widehat{\mathbf{R}}_{t,NT}'\boldsymbol{\lambda}_{it}\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}}\right)$,

(iib) $\left\|\widehat{\mathbf{R}}_{t,NT}'\boldsymbol{\lambda}_{it} - \mathbf{R}_t'\boldsymbol{\lambda}_{it}\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}} + \ell_N^{-1}\right)$.

By (19) and (B.2) we have

$$\widehat{\mathbf{\Lambda}}_t = (NT)^{-1}\mathbf{X}'\mathbf{W}_t\mathbf{X}\widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t^{-1} = \left[\mathbf{S}_{t,NT}^{C} + \mathbf{S}_{t,NT}^{CE} + \left(\mathbf{S}_{t,NT}^{CE}\right)' + \mathbf{S}_{t,NT}^{E}\right]\widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t^{-1};$$

then we can write

$$\widehat{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t\widehat{\mathbf{R}}_{t,NT} = \left[\mathbf{S}_{t,NT}^{C} - N^{-1}\mathbf{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\mathbf{\Lambda}_t' + \mathbf{S}_{t,NT}^{CE} + \left(\mathbf{S}_{t,NT}^{CE}\right)' + \mathbf{S}_{t,NT}^{E}\right]\widehat{\mathbf{\Lambda}}_t\widehat{\mathbf{V}}_t^{-1},$$

which implies that

$$\text{(ia)}\quad \frac{1}{\sqrt{N}}\left\|\widehat{\mathbf{\Lambda}}_t - \mathbf{\Lambda}_t\widehat{\mathbf{R}}_{t,NT}\right\| \le \left(\left\|\mathbf{S}_{t,NT}^{C} - N^{-1}\mathbf{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\mathbf{\Lambda}_t'\right\| + 2\left\|\mathbf{S}_{t,NT}^{CE}\right\| + \left\|\mathbf{S}_{t,NT}^{E}\right\|\right)\left\|\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}}\right\|\left\|\widehat{\mathbf{V}}_t^{-1}\right\|,$$

$$\text{(iia)}\quad \left\|\widehat{\boldsymbol{\lambda}}_{it} - \widehat{\mathbf{R}}_{t,NT}'\boldsymbol{\lambda}_{it}\right\| \le \left\|\widehat{\mathbf{V}}_t^{-1}\right\|\left\|\widehat{\mathbf{\Lambda}}_t\right\| \times \left(\left\|S_{it,NT}^{C} - N^{-1}\mathbf{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\lambda}_{it}\right\| + \left\|S_{it,NT}^{CE}\right\| + \left\|S_{it,NT}^{EC}\right\| + \left\|S_{it,NT}^{E}\right\|\right),$$

where $\widehat{\boldsymbol{\lambda}}_{it}$, $\boldsymbol{\lambda}_{it}$, $\boldsymbol{S}^{C}_{it,NT}$, $\boldsymbol{S}^{CE}_{it,NT}$, $\boldsymbol{S}^{E}_{it,NT}$ are the ith columns of $\widehat{\boldsymbol{\Lambda}}'_t$, $\boldsymbol{\Lambda}'_t$, $\mathbf{S}^{C}_{t,NT}$, $\mathbf{S}^{CE}_{t,NT}$, and $\mathbf{S}^{E}_{t,NT}$, respectively:

$$\boldsymbol{S}^{C}_{it,NT} = (NT)^{-1}\sum_{s=1}^{T}\boldsymbol{\Lambda}_s\boldsymbol{F}_s\boldsymbol{F}'_s\boldsymbol{\lambda}_{is}K_h\left(\tfrac{t-s}{T}\right),$$

$$\boldsymbol{S}^{CE}_{it,NT} = (NT)^{-1}\sum_{s=1}^{T}\boldsymbol{\Lambda}_s\boldsymbol{F}_s E_{is}K_h\left(\tfrac{t-s}{T}\right),$$

$$\boldsymbol{S}^{EC}_{it,NT} = (NT)^{-1}\sum_{s=1}^{T}\boldsymbol{E}_s\boldsymbol{F}'_s\boldsymbol{\lambda}_{is}K_h\left(\tfrac{t-s}{T}\right),$$

$$\boldsymbol{S}^{E}_{it,NT} = (NT)^{-1}\sum_{s=1}^{T}\boldsymbol{E}_s E_{is}K_h\left(\tfrac{t-s}{T}\right).$$

Consider part (ia). By Theorem 1, $\left\|\mathbf{S}^{C}_{t,NT} - N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\Lambda}'_t\right\|$ and $\left\|\mathbf{S}^{CE}_{t,NT}\right\|$ are both $O_p\left(\frac{1}{\sqrt{Th_T}}\right) + O\left(h_T\right) = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$ (we recall that $Th_T^3 \to 0$ by Assumption D). For the last term we have $\mathbf{S}^{E}_{t,NT} = N^{-1}\boldsymbol{\Sigma}^{E} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$; then $\left\|\mathbf{S}^{E}_{t,NT}\right\| \le \left\|N^{-1}\boldsymbol{\Sigma}^{E}\right\| + O_p\left(\frac{1}{\sqrt{Th_T}}\right) = O\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. The terms $\left\|\widehat{\mathbf{V}}_t^{-1}\right\|$ and $\left\|\frac{\widehat{\boldsymbol{\Lambda}}_t}{\sqrt{N}}\right\|$ are both $O_p(1)$, and then $\frac{1}{\sqrt{N}}\left\|\widehat{\boldsymbol{\Lambda}}_t - \boldsymbol{\Lambda}_t\widehat{\mathbf{R}}_{t,NT}\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}}\right)$.

Consider part (ib). To prove that $\frac{1}{\sqrt{N}}\left\|\boldsymbol{\Lambda}_t\widehat{\mathbf{R}}_{t,NT} - \boldsymbol{\Lambda}_t\mathbf{R}_t\right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}} + \ell_N\right)$ we use that

$$\frac{1}{\sqrt{N}}\left\|\boldsymbol{\Lambda}_t\widehat{\mathbf{R}}_{t,NT} - \boldsymbol{\Lambda}_t\mathbf{R}_t\right\| \le \frac{1}{\sqrt{N}}\left\|\boldsymbol{\Lambda}_t\right\|\left\|\widehat{\mathbf{R}}_{t,NT} - \mathbf{R}_t\right\|,$$

that $\frac{1}{\sqrt{N}}\left\|\boldsymbol{\Lambda}_t\right\| = O(1)$ by Assumption B2, that $\sqrt{T}\left\|\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right) - \boldsymbol{\Sigma}^{F}\right\| = O_p(1)$ by Assumption A1 and a law of large numbers, and that

$$\begin{aligned}
&\min\left(\sqrt{Th}, \sqrt{N}, \ell_N^{-1}\right) \left\|\frac{\boldsymbol{\Lambda}'_t\widehat{\boldsymbol{\Lambda}}_t}{N} - \left(\boldsymbol{\Sigma}^{F}\right)^{-1/2}\Upsilon_t\left(\mathbf{V}_t\right)^{1/2}\right\| = O_p(1) &&\text{by Proposition 3,}\\
&\min\left(\sqrt{Th}, N, \ell_N^{-1}\right) \left\|\widehat{\mathbf{V}}_t - \mathbf{V}_t\right\| = O_p(1) &&\text{by Proposition 2.}
\end{aligned}$$

Thus, by Slutsky's theorem applied to (B.6), the proof of Theorem 2(i) is complete.

Consider part (iia). The term $\left\|\boldsymbol{S}^{C}_{it,NT} - N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\lambda}_{it}\right\|$ is $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$. To see this, consider the $N\times 1$ vector $\boldsymbol{S}^{C}_{it,NT}$ as the sum of four terms $\boldsymbol{S}^{C}_{it,NT} = \sum_{k=1}^{4} {}_k\boldsymbol{S}^{C}_{it,NT}$, in the same way as in the proof of Theorem 1 for the $N\times N$ matrix $\mathbf{S}^{C}_{t,NT}$. In order to show that $\left\|\boldsymbol{S}^{C}_{it,NT} - N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\lambda}_{it}\right\| = O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$, we will only treat the first two terms ${}_1\boldsymbol{S}^{C}_{it,NT}$ and ${}_2\boldsymbol{S}^{C}_{it,NT}$ of the above given sum, as the two other terms behave similarly or converge even faster.

For the first term we have

$$\begin{aligned}
&\left\|{}_1\boldsymbol{S}^{C}_{it,NT} - N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\lambda}_{it}\right\|\\
&\quad = \left\|\frac{1}{NT}\sum_{s=1}^{T}\boldsymbol{\Lambda}_t\boldsymbol{F}_s\boldsymbol{F}'_s\boldsymbol{\lambda}_{it}K_h\left(\tfrac{t-s}{T}\right) - N^{-1}\boldsymbol{\Lambda}_t\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\boldsymbol{\lambda}_{it}\right\|
\end{aligned}$$

$$\le \left\| \frac{\Lambda_t \boldsymbol{\lambda}_{it}}{N} \right\| \left\| \frac{1}{T} \sum_{s=1}^{T} \operatorname{vec}\left(\boldsymbol{F}_s \boldsymbol{F}_s'\right) K_h\left(\tfrac{t-s}{T}\right) - \operatorname{vec}\left(\boldsymbol{\Sigma}^F\right) \right\|$$
$$\le \frac{1}{\sqrt{N}} \left\| \frac{\Lambda_t}{\sqrt{N}} \right\| \|\boldsymbol{\lambda}_{it}\| \, O_p\left(\frac{1}{\sqrt{Th_T}}\right) = O_p\left(\frac{1}{\sqrt{NTh_T}}\right),$$

by Assumption B2. For the second term ${}_2\boldsymbol{S}^C_{it,NT}$, again as in the proof of Theorem 1, we show quite analogously to the above (recalling the condition $Th_T^3 = o(1)$) that its norm is $\left\| {}_2\boldsymbol{S}^C_{it,NT} \right\| = o_p\left(\frac{1}{\sqrt{NTh_T}}\right)$:

$$\sqrt{NTh_T} \left\| (NT)^{-1} \sum_{s=1}^{T} \Lambda_t \boldsymbol{F}_s \boldsymbol{F}_s' \left(\boldsymbol{\lambda}_{it}^{(1)} + o(1)\right) h_T z_s \omega_t(s; h_T) \right\|$$
$$= \sqrt{NTh_T} \, O_p\left(\frac{h_T}{\sqrt{N}}\right) = o_p(1).$$

To do so we essentially have to use that, by Assumption B2,

$$\left\| \frac{\Lambda_t \boldsymbol{\lambda}_{it}^{(1)}}{N} \right\| \le \frac{1}{\sqrt{N}} \left\| \frac{\Lambda_t}{\sqrt{N}} \right\| \left\| \boldsymbol{\lambda}_{it}^{(1)} \right\| = \frac{1}{\sqrt{N}} O(1) \left\| \boldsymbol{\lambda}_{it}^{(1)} \right\|.$$

Moreover, by Assumption B3, $\left\| \boldsymbol{\lambda}_{it}^{(1)} \right\| = O(1)$ for all t and all i, and thus we have $\left\| \frac{\Lambda_t \boldsymbol{\lambda}_{it}^{(1)}}{N} \right\| = O\left(\frac{1}{\sqrt{N}}\right)$. Treatment of the terms ${}_3\boldsymbol{S}^C_{it,NT}$ and ${}_4\boldsymbol{S}^C_{it,NT}$ would be similar.

The terms $\left\| \boldsymbol{S}^{CE}_{it,NT} \right\|$ and $\left\| \boldsymbol{S}^{EC}_{it,NT} \right\|$ are both $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$ because $\left\| \Lambda_s E_{is} \right\|$ and $\left\| \boldsymbol{E}_s \boldsymbol{\lambda}_{is}' \right\|$ are $O_p(\sqrt{N})$ and $\left\| \boldsymbol{F}_s \right\| = O_p(1)$ for all s and all i. For the last term we have $\left\| \boldsymbol{S}^E_{it,NT} \right\| = O_p\left(\frac{1}{N}\right)$. Indeed $\boldsymbol{S}^E_{it,NT} = N^{-1} \boldsymbol{\sigma}_i^E + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$, and

$$\left\| \boldsymbol{S}^E_{it,NT} \right\| \le N^{-1} \left\| \boldsymbol{\sigma}_i^E \right\| + O_p\left(\frac{1}{\sqrt{NTh_T}}\right) = O\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NTh_T}}\right),$$

because

$$\left\| \boldsymbol{\sigma}_i^E \right\| = \sqrt{\sum_{j=1}^{N} \left(\sigma_{ij}^E\right)^2} \le \sum_{j=1}^{N} |\sigma_{ij}^E| = O(1) \qquad \text{by Assumption C3.}$$

Then

$$\left\| \widehat{\boldsymbol{\lambda}}_{it} - \widehat{\mathbf{R}}_{t,NT}' \boldsymbol{\lambda}_{it} \right\| = O_p(1) O_p\left(\sqrt{N}\right) \left(O_p\left(\frac{1}{\sqrt{NTh_T}}\right) + O_p\left(\frac{1}{\sqrt{NTh_T}}\right) + O_p\left(\frac{1}{N}\right) \right)$$
$$= O_p\left(\frac{1}{\sqrt{Th_T}}\right) + O_p\left(\frac{1}{\sqrt{N}}\right).$$

Consider part (iib). By the same arguments as for part (ib), $\left\| \widehat{\mathbf{R}}_{t,NT}' \boldsymbol{\lambda}_{it} - \mathbf{R}_t' \boldsymbol{\lambda}_{it} \right\| = O_p\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}} + \ell_N\right)$. ■

Proof of Theorem 3. For this proof we will use the decomposition

$$\left\| \widehat{\boldsymbol{F}}_t - \mathbf{R}_t^{-1} \boldsymbol{F}_t \right\| \le \left\| \widehat{\boldsymbol{F}}_t - \widehat{\mathbf{R}}_{t,NT}^{-1} \boldsymbol{F}_t \right\| + \left\| \widehat{\mathbf{R}}_{t,NT}^{-1} \boldsymbol{F}_t - \mathbf{R}_t^{-1} \boldsymbol{F}_t \right\|$$
$$\le \left\| \widehat{\boldsymbol{F}}_t - \widehat{\mathbf{R}}_{t,NT}^{-1} \boldsymbol{F}_t \right\| + \left\| \widehat{\mathbf{R}}_{t,NT}^{-1} - \mathbf{R}_t^{-1} \right\| \left\| \boldsymbol{F}_t \right\|,$$

where the estimator $\widehat{\mathbf{R}}_{t,NT}$ of the transformation matrix $\mathbf{R}_t$ has been defined in (B.6), and its inverse is given by $\widehat{\mathbf{R}}_{t,NT}^{-1} = \widehat{\mathbf{V}}_t \left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)^{-1} \left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}$. We remark that as N and T increase, $\widehat{\mathbf{R}}_{t,NT}$ becomes invertible, as before in the Proof of Proposition 3.

The term $\|\boldsymbol{F}_t\|$ is $O_p(1)$ (by Assumption A), and the term $\|\widehat{\mathbf{R}}_{t,NT}^{-1} - \mathbf{R}_t^{-1}\|$ is $O_p(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}} + \ell_N)$ (by the same arguments as in the proof of Theorem 2).

We now show that $\|\widehat{\boldsymbol{F}}_t - \widehat{\mathbf{R}}_{t,NT}^{-1}\boldsymbol{F}_t\| = O_p(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{Th}})$. By (21) and (3), the estimated factors can be written as

$$\widehat{\boldsymbol{F}}_t = \frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{\Lambda}_t}{N}\boldsymbol{F}_t + \frac{\widehat{\mathbf{\Lambda}}_t'}{N}\boldsymbol{E}_t,$$

and the matrix $\widehat{\mathbf{R}}_{t,NT}^{-1}$ can be written as (see the proof of Lemma 2)

$$\begin{aligned}\widehat{\mathbf{R}}_{t,NT}^{-1} &= \left[\left(\frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{\Lambda}_t}{N}\right)\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right) + \frac{\widehat{\mathbf{\Lambda}}_t'}{\sqrt{N}}\frac{\mathbf{\Sigma}^E}{N}\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}} + \frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{Z}_{t,NT}\widehat{\mathbf{\Lambda}}_t}{N}\right]\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)^{-1}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\\ &= \left(\frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{\Lambda}_t}{N}\right) + \left[\frac{\widehat{\mathbf{\Lambda}}_t'}{\sqrt{N}}\frac{\mathbf{\Sigma}^E}{N}\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}} + \frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{Z}_{t,NT}\widehat{\mathbf{\Lambda}}_t}{N}\right]\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)^{-1}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}.\end{aligned}$$

Then

$$\widehat{\boldsymbol{F}}_t - \widehat{\mathbf{R}}_{t,NT}^{-1}\boldsymbol{F}_t = \frac{\widehat{\mathbf{\Lambda}}_t'}{N}\boldsymbol{E}_t + \left[\frac{\widehat{\mathbf{\Lambda}}_t'}{\sqrt{N}}\frac{\mathbf{\Sigma}^E}{N}\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}} + \frac{\widehat{\mathbf{\Lambda}}_t'\mathbf{Z}_{t,NT}\widehat{\mathbf{\Lambda}}_t}{N}\right]\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)^{-1}\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\boldsymbol{F}_t,$$

and therefore we have

$$\begin{aligned}\left\|\widehat{\boldsymbol{F}}_t - \widehat{\mathbf{R}}_{t,NT}^{-1}\boldsymbol{F}_t\right\| &\le \left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\| + \left(\left\|\frac{\widehat{\mathbf{\Lambda}}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right\|\left\|\frac{\mathbf{\Sigma}^E}{N}\right\| + \left\|\frac{\widehat{\mathbf{\Lambda}}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right\|\left\|\mathbf{Z}_{t,NT}\right\|\right)\\ &\quad\times\left\|\left(\frac{\mathbf{\Lambda}_t'\widehat{\mathbf{\Lambda}}_t}{N}\right)^{-1}\right\|\left\|\left(\frac{\mathbf{F}'\mathbf{F}}{T}\right)^{-1}\right\|\|\boldsymbol{F}_t\|\\ &= \left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\| + \left[O_p(1)O\left(\frac{1}{\sqrt{N}}\right) + O_p(1)O_p\left(\frac{1}{\sqrt{Th_T}}\right)\right]O_p(1)O_p(1)O_p(1)\\ &= \left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\| + O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{Th_T}}\right).\end{aligned}$$

To complete the proof, we now show that $\left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\| = O_p\left(\frac{1}{\sqrt{N}}\right)$. The terms $\left\|\frac{\mathbf{\Lambda}_t}{\sqrt{N}}\right\|$ and $\left\|\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}}\right\|$ are of the same order, because $\left\|\frac{\mathbf{\Lambda}_t}{\sqrt{N}}\right\| = O_p(1)$ for all t by our Assumption B2, and $\left\|\frac{\widehat{\mathbf{\Lambda}}_t}{\sqrt{N}}\right\| = O_p(1)$ for all t by our constraint in (13). Then the norms $\left\|\frac{\mathbf{\Lambda}_t'\boldsymbol{E}_t}{N}\right\|$ and $\left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\|$ are of the same order as well and therefore, by Assumption F, $\left\|\frac{\widehat{\mathbf{\Lambda}}_t'\boldsymbol{E}_t}{N}\right\| = O_p\left(\frac{1}{\sqrt{N}}\right)$. ■