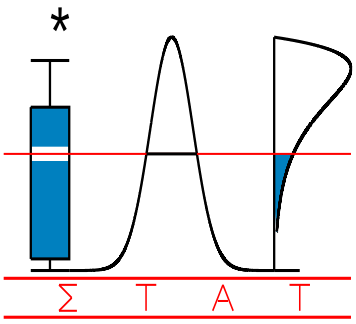
T E C H N I C A L

R E P O R T

0636

LOCALLY STATIONARY FACTOR MODELS : IDENTIFICATION AND NONPARAMETRIC ESTIMATION

MOTTA G., HAFNER C., and R. von SACHS



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Locally Stationary Factor Models: identification and nonparametric estimation

Giovanni Motta* Christian M. Hafner† Rainer von Sachs‡

September 30, 2006

Abstract

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 to consider 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.

Keywords: Approximate Factor Models, Local Stationarity, Principal Components

JEL Classification: C14, C32

*Institut de statistique, Université catholique de Louvain, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium; e-mail motta@stat.ucl.ac.be; corresponding author.

†Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium

‡Institut de statistique, Université catholique de Louvain, Louvain-la-Neuve, Belgium

1 Introduction

Linear factor models have attracted considerable interest over recent years especially in the econometrics literature. The intuitively appealing idea to explain 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 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. Bai (2003) and Bai and Ng (2002) use a new concept of simultaneous asymptotics, where both N and T go to infinity simultaneously, where the rates are rather flexible. 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 e.g. factor loadings remain constant. 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. GARCH, the most popular parametric model in the univariate context, has been generalized to the multivariate case, see e.g. the recent survey by Bauwens et al. (2006). Factor models have also been proposed in the multivariate GARCH framework. For example, Engle, Ng and Rothschild (1990) assume observability of the factors, Diebold and Nerlove (1989) obtain estimates of the factors via the Kalman filter, and in the orthogonal GARCH model of Alexander (2001) factors are the principal components obtained from the unconditional sample covariance matrix. 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, Bai (2003) allows for serial correlation of the error terms and Forni et al. (2000) suggest ARMA models for the factors.

A potential drawback of observationally driven dynamic models such as stationary 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). A promising alternative is to directly model unconditional variances and covariances via nonparametric estimation as in Rodriguez-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. Finally, a combination of parametric GARCH models for the variances and nonparametric estimation of correlations has been proposed by Hafner, van Dijk and Franses (2006).

In this paper, we propose a new approach to allow for time-varying factor loadings. The basic idea is to consider these as smooth functions of time, rendering the process nonstationary while the factors are stationary. However, the assumption that loadings are smooth permits to consider the process as locally stationary and enables us to estimate the model using nonparametric methods. One of the techniques employed is the rescaling of time to the unit interval, as proposed by Dahlhaus (1996), which generates an increasing number of observations in a small neighborhood of the point where the loadings are estimated. The nonparametric estimation of

locally stationary time series models is now well established in the univariate context, see e.g. van Bellegem and von Sachs (2004). It is new in the context of locally stationary factor models. We discuss identification and estimation of the model and derive the asymptotic theory for the estimated loadings and factors. Clearly we have to pay a price for the nonparametric estimation of our model in terms of slower convergence rates. In a Monte Carlo study, we show that for moderately large N and T, the estimators perform very well.

Many issues remain to be resolved such as testing for the number of factors as in Bai and Ng (2002) and Connor and Korajczyk (1993), forecasting as in Stock and Watson (2004), and empirical applications, but we leave this for future research.

The paper is organized as follows. In the following section we first recall the traditional factor model with constant factor loadings. The third section introduces the locally stationary factor model and proposes an estimator of the factors and factor loadings. In the fourth section we provide results on the asymptotic behavior of the estimates. In Section 5 we provide a simulation study and Section 6 concludes. Some auxiliary results needed for the proofs are collected in Appendix A to C, and Appendix D contains the proofs of the propositions.

Throughout the paper we denote by $\mathrm{tr}(\cdot)$ the trace operator, by $\mathrm{rk}(A)$ the rank of a matrix A, by I_n the identity matrix of dimension n and by $||\cdot||$ the L_2 norm, i.e., $||A|| = \sqrt{\mathrm{tr}(A'A)}$.

2 Factor models of large dimensions

Consider the following stationary $r-$factor model for the $N-$dimensional stochastic process $\{X_t\}, t \in \mathbb{Z}$:

$$X_t = \Lambda F_t + e_t \tag{1}$$

where

$$\begin{cases} \Lambda \quad \text{is a } N \times r \text{ matrix of loadings,} \\ \mathbb{E}(F_t) = 0,\ \mathbb{V}\mathrm{ar}(F_t) := \mathbf{\Sigma}_F = \mathrm{diag}\left\{\phi_1, \ldots, \phi_r\right\},\ \phi_j > 0, \quad j = 1, \ldots, r, \\ \mathbb{E}(e_t) = 0,\ \mathbb{V}\mathrm{ar}(U_t) := \Psi,\ \Psi \text{ positive definite} \\ \mathbb{C}\mathrm{ov}(F_t, e_t) = 0, \quad t = 1, \ldots, T. \end{cases}$$

We consider the case where the dimension N is large, say one hundred 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 X_t by some factors common to all components of X_t, and some idiosyncratic error term e_t that is not correlated with the factors and that captures variable-specific variations. In the classical factor model, Ψ 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 X_t is zero, but this is without loss of generality. Furthermore, the covariance matrix of X_t is given by

$$\mathbb{V}\mathrm{ar}\left[X_t\right] := \mathbf{\Sigma} = \Lambda \mathbf{\Sigma}_F \Lambda' + \Psi \quad \forall t = 1, \ldots, T \tag{2}$$

which does not depend on t. The part of the variance explained by the factors is $\Lambda \mathbf{\Sigma}_F \Lambda'$, the remainder Ψ is the part due to idiosyncratic noise.

We can consider estimation of Λ and F_t by minimizing the nonlinear least squares objective function

$$L_{NT}\left(F, \Lambda; r\right) = (NT)^{-1} \sum_{t=1}^{T} \left(X_t - \Lambda F_t\right)' \left(X_t - \Lambda F_t\right), \tag{3}$$

subject to

$$\Lambda'\Lambda / N = I_r. \tag{4}$$

The first order conditions for minimizing the loss function in (3) with respect to F_t are

$$\Lambda'\Lambda\widehat{F}_t = \Lambda' X_t \tag{5}$$

and thus, under (4), the estimator of the factor satisfies

$$\widehat{F}_t = \frac{1}{N}\Lambda' X_t. \tag{6}$$

Substituting (6) into the objective function (3) yields the concentrated objective function

$$\begin{aligned}
\widetilde{L}_{NT}(\Lambda; r) &= (NT)^{-1}\sum_{t=1}^{T}\left(X_t - \frac{1}{N}\Lambda\Lambda' X_t\right)'\left(X_t - \frac{1}{N}\Lambda\Lambda' X_t\right) \\
&= (NT)^{-1}\sum_{t=1}^{T}\left(X_t' X_t - \frac{1}{N}X_t'\Lambda\Lambda' X_t\right) \\
&= (NT)^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{X_t X_t'\right\} - (NT)^{-1}\sum_{t=1}^{T}\operatorname{tr}\left\{\frac{1}{N}\Lambda' X_t X_t'\Lambda\right\}.
\end{aligned} \tag{7}$$

Minimizing $\widetilde{L}_{NT}(\Lambda; r)$ in (7) subject to (4) is equivalent to maximizing $(NT)^{-1}\operatorname{tr}(\sum_{t=1}^{T}\Lambda' X_t X_t'\Lambda)$ subject to (4). By Lemma 4 in Appendix B, the solution is to set the estimator $\widehat{\Lambda}$ of Λ to be $\sqrt{N}$ times the matrix whose columns are the r orthogonal unit-length eigenvectors $\overline{\Lambda}$ corresponding to the largest r ordered eigenvalues of $(NT)^{-1}\sum_{t=1}^{T} X_t X_t'$, that is

$$\left[\frac{1}{NT}\sum_{t=1}^{T} X_t X_t'\right]\overline{\Lambda} = \overline{\Lambda} V_{NT}, \quad \overline{\Lambda}'\overline{\Lambda} = I_r, \quad \widehat{\Lambda} = \sqrt{N}\overline{\Lambda}, \quad \widehat{\Lambda}'\widehat{\Lambda}/N = I_r \tag{8}$$

where V_{NT} is the diagonal matrix containing the largest r eigenvalues of $\frac{1}{NT}\sum_{t=1}^{T} X_t X_t'$ in decreasing order. By equation (6), the resulting estimator of the factors is $\widehat{F}_t = \widehat{\Lambda}' X_t/N$, which is the vector consisting of the first r principal components of X_t. The matrix $V_{NT} = T^{-1}\sum_{t=1}^{T}\widehat{F}_t\widehat{F}_t'$ is the diagonal matrix containing the largest r ordered eigenvalues of $S_{NT} := (NT)^{-1}\sum_{t=1}^{T} X_t X_t' = N^{-1}S_T$, where S_T is the sample covariance matrix estimator:

$$S_T := T^{-1}\sum_{t=1}^{T} X_t X_t'. \tag{9}$$

Unless $\Psi = \sigma^2 I_N$ for some $\sigma^2 > 0$, the principal components estimator is biased if N is fixed and $T \to \infty$. Allowing for weak cross-correlation of the errors e_t, i.e. a non-diagonal covariance matrix Ψ, Bai (2003) has shown that under $N \to \infty$ and $T \to \infty$ the principal components estimator is consistent and asymptotically normal. The model with non-diagonal Ψ is usually called *approximate factor model*, see e.g. Chamberlain and Rothschild (1983), Stock and Watson (2002) and Bai (2003). The assumptions of Bai even allow for weak serial correlation of the errors, as well as weak correlation between the factors and the errors. A further condition is that $\|\Lambda'\Lambda/N - \Sigma_\Lambda\| \to 0$ for some positive definite Σ_Λ, which means that the factor contributions are $O(\sqrt{N})$ for all variables.

In the following section we will generalize this factor model to allow for time-varying factor loadings.

3 Locally Stationary Factor Models

We assume that $X_{t,T}$ is an $N-$dimensional stochastic process explained by r factors, $r < N$:

$$X_{t,T} = \Lambda(u_{t,T})F_t + e_t. \qquad t = 1, \ldots, T \tag{10}$$

This process, in fact a sequence (triangular array) of doubly-indexed processes, is defined in the rescaled time $u_{t,T} := \frac{t}{T} \in (0,1)$ as in Dahlhaus (1996). This class of models allows to model nonstationary multivariate time series because the loadings $\{\lambda_{ij}(u_{t,T})\}$ are assumed to be smooth functions of time ($t = 1, \ldots, T$, $i = 1, \ldots, N$, $j = 1, \ldots, r$). Hence, the covariance of $X_{t,T}$ becomes a matrix-valued smooth function of time

$$\mathbb{V}\text{ar}\left[X_{t,T}\right] := \mathbf{\Sigma}(u_{t,T}) = \Lambda(u_{t,T})\mathbf{\Sigma}_F\Lambda'(u_{t,T}) + \Psi \qquad t = 1, \ldots, T. \tag{11}$$

The class of locally stationary factor models is defined as the following.

Definition 1 (Locally Stationary Factor Model) *The sequence $\{X_{t,T}\}$ in (10) is a Locally Stationary Factor Model* [**LSFM**] *if there exists a function*

$$\Lambda(u) \in \mathcal{C}^k\left\{[0,1], \mathbb{R}^{N\times r}\right\}, k \geq 1, \ \text{rk}\,\Lambda(u) = r \ \forall u \in [0,1], \quad \textit{such that } \forall T$$
$$\Sigma\left(\frac{t}{T}\right) := \mathbb{V}\text{ar}\left(X_{t,T}\right) = \Lambda\left(\frac{t}{T}\right)\Sigma_F\Lambda'\left(\frac{t}{T}\right) + \Psi.$$

We aim to estimate the loadings $\Lambda(\cdot)$ and the Matrix Ψ. Principal Component Regression (PCR) is a well-known estimation approach in the framework of Approximate Factor Models (AFM) defined by Chamberlain & Rothschild (1983). The main difference with respect to the classical factor models is that the AFM allows the errors to be correlated in the cross section dimension, i.e. the matrix Ψ is not diagonal. Recently Bai & Ng (2002) and Bai (2003) generalized the definition of the Approximate Factor Model by also allowing for heteroskedasticity of the errors in the time dimension. In this work we assume that the factors are orthogonal ($\mathbf{\Sigma}_F$ is diagonal) and uncorrelated to the errors ($\mathbb{C}\text{ov}[F_t, e_t] = 0 \ \forall t$), but we allow for loadings that change over time. We assume the errors are correlated in the cross section dimension but not in the time dimension.

The definition of AFM is important because the loadings can be estimated by the eigenvectors of an estimator of the covariance matrix. To do this we need that Ψ_N 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. The following definition combines the definitions of AFM and $LSFM$.

Definition 2 (Locally Stationary Approximate Factor Model) *Let Ψ_N be the sequence of $N \times N$ covariance matrices of the $N \times 1$ vector e_t in (10):*

$$\Psi_N := \mathbb{V}\text{ar}(e_t) \quad \forall t = 1, \ldots, T.$$

The sequence $\{X_{t,T}\}$ in (10) is a Locally Stationary Approximate Factor Model [**LSAFM**] *if*

*(i) $\{X_{t,T}\}$ is a **LSFM** and*

(ii) Ψ_N is a sequence of positive definite matrices with uniformly bounded eigenvalues:

$$\sup_N \text{eval}_1\left(\Psi_N\right) < \infty$$

where $\text{eval}_1(A)$ denotes the largest eigenvalue of the matrix A.

3.1 Assumptions of the LSAFM

ASSUMPTION A - Factors:

1. $\mathbb{E}(F_t) = 0 \ \forall \ t = 1, \ldots, T$;
2. $\mathbb{E}\|F_t\|^4 \leq M < \infty$;
3. $T^{-1}\sum_{t=1}^{T} F_t F_t' \xrightarrow{p} \Sigma_F$, where the $r \times r$ matrix $\mathbf{\Sigma}_F$ is diagonal and positive definite.

ASSUMPTION B - Factor loadings:

1. $\lambda_{ij}(u) \in C^1[0,1] \ \forall \ i, j = 1, \ldots, N$;
2. $\left\|\Lambda'\left(\frac{t}{T}\right)\Lambda\left(\frac{t}{T}\right)/N - \Sigma_\Lambda\left(\frac{t}{T}\right)\right\| \longrightarrow 0$ as $T, N \longrightarrow \infty$, where $\Sigma_\Lambda\left(\frac{t}{T}\right)$ is a positive definite matrix $\forall t$.

ASSUMPTION C - Cross-Section Dependence of the idiosyncratic errors:

1. $\mathbb{E}(e_t) = 0 \ \forall \ t = 1, \ldots, T$;
2. $\mathbb{E}|e_{it}|^4 \leq M \ \forall \ i = 1, \ldots, N, \ \forall t = 1, \ldots, T$;
3. $\mathbb{E}(e_{it}e_{jt}) = \psi_{ij}$, $\sum_{j=1}^{N}|\psi_{ij}| \leq M \ \forall \ i = 1, \ldots, N$ and $\forall N$.

ASSUMPTION D - Rates of convergence:

1. $T \to \infty$, $h_T \to 0$, $N \to \infty$ in such a way that $Th_T \to \infty$, $Th_T^3 \to 0$ and $Nh_T \to 0$.
2. $$\sup_u ||\Lambda^{(1)}(u)|| \ = \ O(\sqrt{N}) \ .$$

ASSUMPTION E - Possible Multiplicity of eigenvalues:

Suppose that for the first ordered r eigenvalues $\{\gamma_1, \ldots, \gamma_r\}$ of $\mathbf{\Sigma}(u)$ there exists an $e \in \{0, 1\}$ such that

$$\lim_{\tau \to 0} \inf \frac{|\gamma_i(u+\tau) - \gamma_j(u+\tau)|}{|\tau^e|} \ > 0 \ , \tag{12}$$

for all $u \in (0,1)$ and $i \neq j, \ i, j = 1, ..., r$. This is equivalent to require the same condition on the eigenvalues of the $r \times r$ matrix $\Sigma_\Lambda(u)\Sigma_F$.

Chamberlain and Rothschild (1983) defined an AFM as having bounded eigenvalues for the $N \times N$ covariance matrix $\Psi = \mathbb{E}(e_t e_t')$. If e_t is stationary with $\mathbb{E}(e_{it}e_{jt}) = \psi_{ij} \ \forall t$, then the largest eigenvalue of Ψ is bounded by $\max_i \sum_{j=1}^{N}|\psi_{ij}|$, see e.g. Lütkepohl (1996). Thus by Assumption C3 model (10) will be an AFM in the sense of Chamberlain and Rothschild.

In Assumption B.1 the model quantities $\lambda_{ij}(u)$ are supposed to be in $C^1[0,1]$. This condition is given for ease of presentation of the proofs, we note that one can more generally suppose the time-varying factor loadings to be in $C^k[0,1]$ with $k \geq 1$. However in this work we are not interested in deriving optimal rates of convergence of non-parametric estimators of these model quantities but content ourselves to prove consistency of our estimators. We suspect that even a weaker condition on the regularity of the factor loading, such as Hölder-continuity, would be sufficient to derive our results.

In Assumption D, 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 which can grow with order N, we require $Nh_T \to 0$ (see Proposition 1) and $Th_T^3 \to 0$ (see proof of Theorem 2). Note that this implies that T grows faster than N but allows either of N and Th to grow faster than the other. 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.

Assumption E is an assumption on the identifiability in estimating the time-varying factor loadings $\lambda_{ij}(u)$ as well as the common components $C_{it,T} := \lambda_i'(u_{t,T})F_t$. Note that we do not require distinctness of the eigenvalues of $\Sigma_\Lambda(u)\Sigma_F$ (as it is done, e.g., in Assumption G of Bai, 2003). 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 from the right" are equal. By Theorems 2.4 and 2.7 of Chern & Dieci (2000, p. 774-778), under this assumption the matrix of orthogonal eigenvectors of $\mathbf{\Sigma}(u)$ can be taken to be a continuous function in u. (For details we refer to Lemma 6 of Appendix C,which we apply in the case $k = 1$ and $e = 1$.)

We are now in a position to develop consistent estimation theory with well defined target functions which 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 $\mathbf{\Sigma}(u)$ which by Assumption B.1 are in $C^1[0, 1]$ (as the eigenvalues depend continuously on the regularity of the elements of the covariance matrix $\mathbf{\Sigma}(u)$ 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 $\gamma_i(u_0^-)$.

We end this remark by adding, that as already in the stationary case of the previous section, the loadings can only be estimated up to a rotation, and it is only the product $\Lambda'\Lambda$ which is identifiable. Hence the above discussion has to be understood in the sense of continuity of the limit $T \to \infty$ of $\widehat{\Lambda}'(t/T)\widehat{\Lambda}(t/T)$; compare also the formulation of our Theorem 5. Note however that it applies as well to the uniquely determined common components $C_{it,T}$ as a function of rescaled time (cf. also our discussion following Corollary 7).

3.2 Time-varying PCA

Consider now the nonstationary model in (10). In order to simplify the notation, we occasionally denote $\Lambda_t = \Lambda(u_{t,T})$, $X_t = X_{t,T}$ and $u_t = u_{t,T}$.

The corresponding time-varying weighted least squares objective function is

$$L_{t,NT}\left(F, \Lambda; r, h\right) = (NT)^{-1} \sum_{s=1}^{T} \left(X_s - \Lambda_s F_s\right)' W_t(u_s; h) \left(X_s - \Lambda_s F_s\right) \tag{13}$$

subject to

$$\Lambda_s'\Lambda_s/N = I_r \quad \forall s = 1, \ldots, T. \tag{14}$$

Without loss of generality, for reasons of ease of presentation we use the same bandwidth h forall $i, j = 1, \ldots, N$. This allows to write the $N \times N$ matrix of weights $W_t(u_s; h)$ as $W_t(u_s; h) = \omega_t(u_s; h)I_N$, where

$$\omega_t(u_s; h) = K_h\left(u_s - u_t\right) > 0 \quad \forall s, t = 1, \ldots, T,$$

$K_h(\cdot) := \frac{1}{h}K\left(\frac{\cdot}{h}\right)$ being the rescaled version of a kernel smoother. The first order conditions for maximizing (13) with respect to F_s are

$$\Lambda_s' W_t(u_s;h)\Lambda_s \widehat{F}_s = \Lambda_s' W_t(u_s;h) X_s \tag{15}$$

and thus the estimator of the factor satisfies

$$\widehat{F}_s = \left\{\Lambda_s' W_t(u_s;h)\Lambda_s\right\}^{-1} \Lambda_s' W_t(u_s;h) X_s. \tag{16}$$

Substituting (16) into the objective function (13) yields the concentrated objective function

$$\begin{aligned} \widetilde{L}_{t,NT}(\Lambda;r) &= (NT)^{-1}\sum_{s=1}^{T} X_s' W_t(u_s;h) X_s \\ &- (NT)^{-1}\sum_{s=1}^{T} X_s' W_t(u_s;h)\Lambda_s \left(\Lambda_s' W_t(u_s;h)\Lambda_s\right)^{-1} \Lambda_s' W_t(u_s;h) X_s. \end{aligned} \tag{17}$$

Since $W_t(u_s;h) = \omega_t(u_s;h) I_N$, the concentrated objective function in (17) reduces to

$$\begin{aligned} \widetilde{L}_{t,NT}(\Lambda;r) &= (NT)^{-1}\sum_{s=1}^{T} \operatorname{tr}\left\{X_s X_s'\right\} \omega_t(u_s;h) \\ &- (NT)^{-1}\sum_{s=1}^{T} \operatorname{tr}\left\{\frac{1}{N}\Lambda_s' X_s X_s' \Lambda_s\right\} \omega_t(u_s;h) \end{aligned} \tag{18}$$

subject to (14). Minimizing the concentrated objective function (18) with respect to $\Lambda(u_{s,T})$ is equivalent to maximizing

$$(NT)^{-1}\sum_{s=1}^{T} \operatorname{tr}\left\{\Lambda_s' X_s X_s' \Lambda_s\right\} \omega_t(u_s;h) \tag{19}$$

subject to (14). If we assume that the sequence $h_T \to 0$ as $T \to \infty$, for large values of T the weight $\omega_t(u_s)$ only takes into account the loadings Λ_s corresponding to the values u_s that are very close to u_t. Asymptotically, the loadings that minimize the concentrated objective function in (19) depend only on t. This result is formalized in the following proposition; the proof is given in Appendix D.

Proposition 1 *Under Assumptions A-D,* $\forall\ t = 1, \ldots, T$,

$$(NT)^{-1}\sum_{s=1}^{T} \operatorname{tr}\left\{\left(\Lambda_s - \Lambda_t\right)' X_s X_s' \left(\Lambda_s - \Lambda_t\right)\right\} \omega_t(u_s;h) = O_p(Nh) = o_p(1)\ .$$

By Proposition 1 maximizing (19) is asymptotically equivalent to maximizing

$$N^{-1} \operatorname{tr}\left\{\Lambda_t' \widehat{\mathbf{\Sigma}}(u_t) \Lambda_t\right\} \qquad t = 1, \ldots, T$$

subject to (14), where

$$\widehat{\mathbf{\Sigma}}(u_t;h) := T^{-1}\sum_{s=1}^{T} X_s X_s' \omega_t(u_s;h) \qquad t = 1, \ldots, T, \tag{20}$$

which is the nonparametric estimator of the covariance matrix proposed by Rodríguez-Poo and Linton (2001). By Lemma 4 in Appendix B the solution is to set $\widehat{\Lambda}_t$ to be $\sqrt{N}$ times the matrix whose columns are the r orthogonal unit-length eigenvectors $\overline{\Lambda}(u_{t,T})$ corresponding to the largest r ordered eigenvalues of $\widehat{\mathbf{\Sigma}}(u_{t,T})$. As for the stationary case, by analogous arguments we obtain

$$N^{-1}\widehat{\mathbf{\Sigma}}(u_t;h)\widehat{\Lambda}_t = \widehat{\Lambda}_t V_{NT}(u_t), \quad \widehat{\Lambda}_t'\widehat{\Lambda}_t/N = I_r, \quad t=1,\ldots,T. \tag{21}$$

where $V_{NT}(u_t)$ is the diagonal matrix containing the largest r ordered eigenvalues of $N^{-1}\widehat{\mathbf{\Sigma}}(u_t;h)$ $(t=1,\ldots,T)$.

It is useful to write the LSFM in a more compact matrix notation. Notice that $N^{-1}\widehat{\mathbf{\Sigma}}(u_t;h) = (NT)^{-1}X'\mathcal{W}_t(h)X$, where the $T\times T$ matrix of weights $\mathcal{W}_t(h)$ is defined as

$$\mathcal{W}_t(h) = \operatorname{diag}\left\{\omega_t\left(u_1;h\right),\ldots,\omega_t\left(u_T;h\right)\right\} \tag{22}$$

where $\omega_t\left(u_s;h\right) := K_h\left(u_s-u_t\right)$, $s,t=1,\ldots,T$, and the $T\times N$ matrix X collects the data:

$$X = \begin{bmatrix} X_{1,T}' \\ \vdots \\ X_{t,T}' \\ \vdots \\ X_{T,T}' \end{bmatrix}. \tag{23}$$

This means that

$$\frac{1}{NT}X'\mathcal{W}_t(h)X\widehat{\Lambda}(u_t) = \widehat{\Lambda}(u_t)V_{NT}(u_t), \quad \widehat{\Lambda}(u_t)\widehat{\Lambda}'(u_t)/N = I_r \quad t=1,\ldots,T \tag{24}$$

that is

$$V_{NT}(u_t) = N^{-1}\widehat{\Lambda}'(u_t)\frac{1}{NT}X'\mathcal{W}_t(h)X\widehat{\Lambda}(u_t), \quad \widehat{\Lambda}'(u_t)\widehat{\Lambda}(u_t)/N = I_r \quad t=1,\ldots,T. \tag{25}$$

If we define $\overline{F}_t = X\widehat{\Lambda}(u_t)/N$, the result in (25) can be written as the following

$$V_{NT}(u_t) = \overline{F}_t'\mathcal{W}_t(h_T)\overline{F}_t/T, \quad t=1,\ldots,T \tag{26}$$

where

$$\overline{F}_t' = \left[\widehat{\Lambda}_t'X_1/N,\ldots,\widehat{\Lambda}_t'X_t/N,\ldots,\widehat{\Lambda}_t'X_T/N\right] \quad t=1,\ldots,T. \tag{27}$$

The matrix $V_{NT}(u_t)$ is a localized version of the matrix V_{NT} (see the definition of V_{NT} in §2). Indeed, 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 $\mathcal{W}_t(h_T)$). Notice that the $t-$th column of the $r\times T$ matrix $\overline{F}_t'$ is $\widehat{F}_t$ as defined in (16). Indeed, the estimated factors $\widehat{F}_t$ in (16) can be written as

$$\widehat{F}_t = \widehat{\Lambda}_t'X_{t,T}/N \quad t=1,\ldots,T \tag{28}$$

since $W_t(u_s;h) = \omega_t(u_s;h)I_N$ $(s,t=1,\ldots,T)$ and $\widehat{\Lambda}_s'\widehat{\Lambda}_s/N = I_r$ $\forall s=1,\ldots,T$.

4 Asymptotic theory

The following asymptotic results hold for $N, T \to \infty$ and for $u_{t,T} = \frac{t}{T} \to u \in (0,1)$. The proofs are given in Appendix D. The first result is about consistency of our nonparametric estimator for $\mathbf{\Sigma}(u_{t,T})$.

Theorem 2 *Under Assumptions A-D,*

$$N^{-1} \left\| T^{-1} X' \mathcal{W}_{t,T}(h) X - \mathbf{\Sigma}(u_{t,T}) \right\| = O_p\left[(Th_T)^{-\frac{1}{2}} \right] \tag{29}$$

where $\mathcal{W}_{\cdot}(h)$ is defined in (22) and $\mathbf{\Sigma}(\cdot)$ is defined in (11).

The estimator $T^{-1} X' \mathcal{W}_{t,T}(h) X$ converges to $\mathbf{\Sigma}(u_{t,T})$ at the rate $\frac{1}{\sqrt{Th_T}}$. This means that each element of the matrix $\left[T^{-1} X' \mathcal{W}_{t,T}(h) X - \mathbf{\Sigma}(u_{t,T}) \right]$ 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 next result is on the consistency of the estimated eigenvalues of $\mathbf{\Sigma}(u_{t,T})$.

Proposition 3 *Under Assumptions A-D,*

$$V_{NT}(u_{t,T}) \xrightarrow{p} V(u) \tag{30}$$

where $V(u)$ is a diagonal matrix containing the eigenvalues of $\mathbf{\Sigma}_\Lambda(u)\mathbf{\Sigma}_F$.

The next proposition is an auxiliary result used in the proofs of our main results.

Proposition 4 *Under Assumptions A-E,*

$$\frac{\Lambda'(u_{t,T}) \widehat{\Lambda}(u_{t,T})}{N} \xrightarrow{p} \mathbf{\Sigma}_F^{-\frac{1}{2}} \Upsilon(u) V^{\frac{1}{2}}(u) \tag{31}$$

where $\Upsilon(u)$ is the $r \times r$ matrix containing the orthonormal eigenvectors of the $r \times r$ matrix $\mathbf{\Sigma}_F^{\frac{1}{2}} \mathbf{\Sigma}_\Lambda(u) \mathbf{\Sigma}_F^{\frac{1}{2}}$.

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

Theorem 5 *Under Assumptions A-E,*

$$(i) \quad \min\left(\sqrt{N}, \sqrt{Th_T}\right) \left\{ \tfrac{1}{\sqrt{N}} \left\| \widehat{\Lambda}(u_{t,T}) - \Lambda(u_{t,T}) R_{NT}(u_{t,T}) \right\| \right\} = O_p(1) \tag{32}$$

$$(ii) \quad \min\left(N, \sqrt{Th_T}\right) \left\{ \tfrac{1}{\sqrt{N}} \left\| \widehat{\lambda}_i(u_{t,T}) - R'_{NT}(u_{t,T}) \lambda_i(u_{t,T}) \right\| \right\} = O_p(1) \tag{33}$$

where $\widehat{\lambda}'_i(u_{t,T})$ is the i-th row of $\widehat{\Lambda}(u_{t,T})$, and where

$$R_{NT}(u_{t,T}) := \left(\frac{F'F}{T} \right) \left(\frac{\Lambda'(u_{t,T}) \widehat{\Lambda}(u_{t,T})}{N} \right) V_{NT}^{-1}(u_{t,T}) \xrightarrow{p} \mathbf{\Sigma}_F^{\frac{1}{2}} \Upsilon(u) V^{-\frac{1}{2}}(u) =: R(u) \ , \tag{34}$$

and where F denotes the $T \times r$ matrix collecting the factors, i.e. $F := \{F_1, \ldots, F_T\}'$.

The result in (32) shows that the appropriately scaled norm of the distance between the estimated loading matrix and a rotated version of the true loading matrix converges to zero in probability. The rate is given by the minimum of $\sqrt{N}$ and $\sqrt{Th}$. Note that (33), the formulation in terms of vectors, is the locally stationary analogue to Theorem 1 of Bai (2003).

Theorem 6 *Under Assumptions A-E,*

$$\min\left(\sqrt{N}, \sqrt{Th_T}\right) \left\| \widehat{F}_{t,T} - R_{NT}^{-1}\left(u_{t,T}\right) F_t \right\| = \; O_p(1), \tag{35}$$

where $\widehat{F}_{t,T}$ is defined in (28), F_t obeys the Assumptions A of the **[LSAFM]** *in (10) and $R_{NT}^{-1}(\cdot)$ is the inverse of the rotation matrix $R_{NT}(\cdot)$ defined in Theorem 5.*

Finally, similar to Bai (2003) we can give a consistency result for the estimation of the common components in the next proposition.

Corollary 7 *Under Assumptions A-E,*

$$\widehat{C}_{it,T} - C_{it,T} \xrightarrow{p} 0 \quad \forall i = 1, \ldots, N \tag{36}$$

where $\widehat{C}_{it,T} := \widehat{\lambda}_i'(u_{t,T})\widehat{F}_{t,T}$ are the estimated common components and $C_{it,T} := \lambda_i'(u_{t,T})F_t$ are the true common components ($i = 1, \ldots, N$).

Corollary 7 states that $\widehat{C}_{it,T}$ consistently estimates the common components. Note that, unlike the estimation of $\Lambda(u)$ or F_t, $C_{it,T}$ and $\widehat{C}_{it,T}$ are well identified. $C_{it,T}$ is identified because the indetermenacy of $\Lambda(u)$ and F_t by the rotation matrix $R(u)$ cancels out. $\widehat{C}_{it,T}$ is identified because $\hat{\Lambda}$ and $\hat{F}_t$ (depending on $\hat{\Lambda}$) have the same sign. 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 in Section 3.1.

The importance of this last result lies in the fact that it is via $\widehat{C}_{it,T}$ that a *predictor* of the originally observed series, i.e. its i-th component $X_{it,T}$, will be given. We will leave an assessment of the qualities of such a predictor for future work, but we like to emphasize that it is the flexibility of a *locally stationary* model with *time-varying* loadings that will prove useful for enhanced prediction qualities (cf also a similar observation in the univariate situation of Van Bellegem and von Sachs, 2004).

5 Simulation study

In this section we illustrate the performance of the estimators $\widehat{\mathbf{\Sigma}}(\cdot; h)$, $\widehat{\Lambda}(\cdot)$ and $V_{NT}(\cdot)$ defined in (20), (21) and (25), respectively. The diagonal $r \times r$ matrix $V_{NT}(\cdot)$ contains the largest r eigenvalues of the $N \times N$ matrix $N^{-1}\widehat{\mathbf{\Sigma}}(\cdot; h)$. The $N \times r$ matrix $\widehat{\Lambda}(\cdot)$ is $\sqrt{N}$ times the orthonormal eigenvectors corresponding to the eigenvalues $V_{NT}(\cdot)$ of the matrix $N^{-1}\widehat{\mathbf{\Sigma}}(\cdot; h)$. To obtain $\widehat{\Lambda}(\cdot)$ and $V_{NT}(\cdot)$ we thus need the estimate $\widehat{\mathbf{\Sigma}}(\cdot; h)$ of the covariance matrix $\mathbf{\Sigma}(\cdot)$.

The estimator $\widehat{\mathbf{\Sigma}}(\cdot; h)$ depends on the bandwidth h. In our simulations, the choice of the bandwidth is based on the local plug-in algorithm described in Brockmann et al. (1993) and in Hermann (1997). We will not enter into the details of how to optimally choose this parameter h because this problem goes beyond the scope of this paper and will be treated elsewhere.

5.1 Two examples

In Theorem 5, the matrix $\widehat{\Lambda}(\frac{t}{T})$ is only able to identify the matrix $\Lambda(u)$ *up to rotation* (i.e. $\Lambda(u)R(u)$), and *up to sign* (note however that the squared difference between estimator and rotated loadings is uniquely defined in this Theorem 5). The problem of an eigenvector to be

unique only up to sign (the linear eigenspace being equivalently represented by two different normalized eigenvectors where one is the negative of the other) is a different problem than the problem of the loadings (and the factors) of a factor model to be identifiable only up to a rotation. In this context the two problems are superimposed: even though we have asymptotic results for the product $\Lambda(u)R(u)$ (which is the identified target), it remains the problem that the estimator $\widehat{\Lambda}(\cdot)$ is defined to be a matrix of eigenvectors which are unique only up to sign. This explains why in the simulation we still would need to take the absolute values of these vectors (see §5.1.1 and Figure 2). To show the performance of the estimator $\widehat{\Lambda}(\cdot)$ we consider a first set of simulations with $\mathbf{\Sigma}_\Lambda(u) = I_r \ \forall u \in (0,1)$ (see §5.1.1, Figures 1-2 and Figure 7). In this case the rotation matrix $R(u)$ is - up to sign - the identity matrix I_r (i.e. $R(u) = \pm I_r$), and thus the matrix $\widehat{\Lambda}(\frac{t}{T})$ is able to identify the matrix $\Lambda(u)$ up to sign.

By Proposition 3, the matrix $V_{NT}(\frac{t}{T})$ is only able to identify the eigenvalues of the product $\mathbf{\Sigma}_\Lambda(u)\mathbf{\Sigma}_F$. To show the performance of the estimator $V_{NT}(\cdot)$ we consider a second set of simulations with $\mathbf{\Sigma}_F = I_r$ (see §5.1.2 and Figures 3-6). In this case the matrix $V_{NT}(\frac{t}{T})$ is an estimate of the eigenvalues of $\mathbf{\Sigma}_\Lambda(u)$.

5.1.1 First example

We generate the data according to model (10) with $N = 30$, $T = 900$ and $r = 2$, and restrict to the particular case

$$\mathbf{\Sigma}_\Lambda(u) = I_r \quad \forall u \in (0,1). \tag{37}$$

To satisfy (37) we define

$$\Lambda(u_t) = \sqrt{N}\exp(\pi u_t Y)A, \tag{38}$$

where

$u_t := \frac{t}{T} \in]0,1]$, is a grid of T equidistant points in the rescaled time ($t = 1\ldots,T$),

Y is an antisymmetric matrix of dimension N,

A is an $N \times r$ matrix such that $A'A = I_r$,

$\exp(\pi u_t Y)$ is the $N \times N$ matrix whose i,j element is $\exp(\pi u_t Y_{ij})$.

Then we have $N^{-1}\Lambda'(u_t)\Lambda(u_t) = I_r \ \forall t = 1\ldots,T$.

An antisymmetric matrix is a square matrix that satisfies $Y = -Y'$. 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. The general 3×3 antisymmetric matrix is of the form

$$Y = \begin{bmatrix} 0 & Y_{12} & Y_{13} \\ -Y_{12} & 0 & Y_{23} \\ -Y_{13} & -Y_{23} & 0 \end{bmatrix}.$$

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 $\exp(\pi u_t Y)$ by the matrix A. To obtain the matrix A, we simulate $n = 50$ iid realizations $Z_1,\ldots Z_N$ of an $N-$dimensional normal random vector $Z \sim \mathcal{N}(0, I_N)$ and take the r eigenvectors $A_1,\ldots,A_r$ corresponding to the largest eigenvalues of the sample covariance matrix S_Z of the Z_i's: $S_Z := n^{-1}\sum_{i=1}^n Z_i Z_i'$ and $A := [A_1,\ldots,A_r]$. By construction $A'A = I_r$.

The covariance matrix is represented in Figure 1. For convenience of presentation (remember that $\mathbf{\Sigma}(u_t)$ is $N \times N$, a very huge matrix) we picked up some typical elements $\sigma_{ij}(u_t)$, $u_t \in]0,1]$, for some combinations of i, j: here $i = 10, 20, 30$, and $j = 4, 7, 10, 14, 17, 20, 24, 27, 30$.

We simulate $M = 100$ times the same model, that is model (10) with the same deterministic loadings defined in (38) but different (realization of) factors and errors. In particular, $F_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_F)$ with $\mathbf{\Sigma}_F = \text{diag}\{3.5, 1\}$, and $e_t \overset{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \Psi)$ with $\Psi = I_N$. For each $m = 1, \ldots, M$ and $\forall u_t \in]0,1]$, we compute the estimate $\widehat{\mathbf{\Sigma}}(u_t; m) = \{\widehat{\sigma}_{ij}(u_t; m)\}_{i,j=1}^{N}$ defined in (20). Then $\forall u_t \in]0,1]$ we consider the average (bold line)

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

and construct 95% confidence intervals based on asymptotic normality (dashed lines)

$$\overline{\widehat{\sigma}}_{ij}(u_t) \pm z_{0.975} \widehat{\nu}_{ij}(u_t) \tag{39}$$

where $z_\alpha = \Phi^{-1}(\alpha)$, $\Phi(\cdot)$ is the standard normal cumulative distribution function, and

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

is the estimator of the standard deviation (say $\nu_{ij}(u_t)$) of the estimator $\widehat{\sigma}_{ij}(u_t)$. The asymptotic normality of the estimator $\widehat{\mathbf{\Sigma}}(\cdot)$ in (20) can be derived analogously to Proposition 3.2 of Rodríquez-Poo & Linton (2001).

For each $m = 1, \ldots, M$ and $\forall u_t \in]0,1]$ we define $\widehat{\Lambda}(u_t; m)$ as $\sqrt{N}$ times the orthonormal eigenvectors of the estimate $N^{-1}\widehat{\mathbf{\Sigma}}(u_t; m)$ (see Figure 2). If $\mathbf{\Sigma}_\Lambda(u) = I_r$ $\forall u \in (0,1)$, then $R(u)$ (the limiting matrix to which $R_{NT}\left(\frac{t}{T}\right)$ converges) is $\pm$ the identity matrix I_r $\forall u \in (0,1)$, i.e. $\widehat{\Lambda}\left(\frac{t}{T}\right)$ converges to $\pm\Lambda(u)$. The definition of $\Lambda(\cdot)$ in (38) removes the indeterminacy due to the rotation; at the same time, since the constraint (14) fixes $\Lambda(\cdot)$ except that any column may have its elements reversed in sign, the loadings are estimated up to sign. We thus consider, $\forall u_t \in]0,1]$, the average $\overline{\left|\widehat{\Lambda}(u_t)\right|} := M^{-1} \sum_{m=1}^{M} \left|\widehat{\Lambda}(u_t; m)\right|$ (the absolute value removes the up-to-sign indeterminacy).

5.1.2 Second example

We generate the data according to model (10) with loadings

$$\Lambda(u_t) = \sqrt{N} \sin(2\pi u_t) \exp(\pi u_t Y) A + \pi A, \tag{40}$$

where Y and A are the same as in (38), and restrict to the particular case $\mathbf{\Sigma}_F = I_r$. Figure 3 shows this example of time-varying loadings in the rescaled time and Figure 4 the corresponding matrix $\mathbf{\Sigma}_\Lambda(u)$.

The covariance matrix and its estimate are represented in Figure 5 (analogously to the first example). The eigenvalues of the normalized covariance matrix $N^{-1}\mathbf{\Sigma}(\cdot)$ are plotted in Figure 6. For each $m = 1, \ldots, M$ and $\forall u_t \in]0,1]$ we compute the eigenvalues $V_{NT}(u_t; m)$ of the estimate $N^{-1}\widehat{\mathbf{\Sigma}}(u_t; m)$. Then $\forall u_t \in]0,1]$ we consider the average $\overline{V_{NT}(u_t)} := M^{-1} \sum_{m=1}^{M} V_{NT}(u_t; m)$ (the bold line) and construct 95% confidence intervals based on asymptotic normality (dashed lines).

Figure 2 shows the *local* performance of the estimator $\widehat{\Lambda}(\cdot)$. 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 5, the loss function

$$L(N,T;M) := \frac{1}{MT}\sum_{m=1}^{M}\sum_{t=1}^{T}\frac{1}{\sqrt{N}}\left\|\widehat{\Lambda}(u_t) - \Lambda(u_t)R_{NT}(u_t)\right\|. \tag{41}$$

By Theorem 5 this loss is decreasing w.r.t. T and $N(T)$, as also shown in Figure 7. This loss function is the sample counterpart of the integrated loss

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

We used the model defined in the second example (i.e. with $\mathbf{\Sigma}_F = I_r$) because in this case the matrix $R(u)$, to which the matrix $R_{NT}(\frac{t}{T})$ converges, is in general different from the identity matrix (we recall that $R(u) = \pm I_r$ if $\mathbf{\Sigma}_\Lambda(u) = I_r$).

Figure 1: Time-varying entries of the covariance matrix. Solid line: $\mathbf{\Sigma}(u_t)$. Bold line: $\overline{\widehat{\mathbf{\Sigma}}(u_t)}$. Dashed lines: confidence intervals.

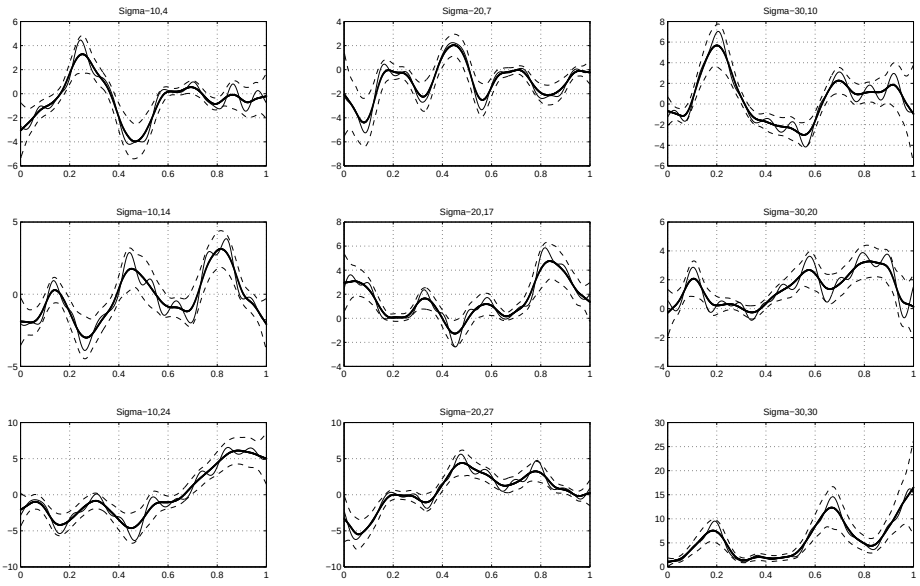


Figure 2: Time-varying entries of the matrices $|\Lambda(u_t)|$ and $\left|\widehat{\Lambda}(u_t)\right|$. Solid line: $|\Lambda(u_t)|$. Bold line: $\overline{\left|\widehat{\Lambda}(u_t)\right|}$.

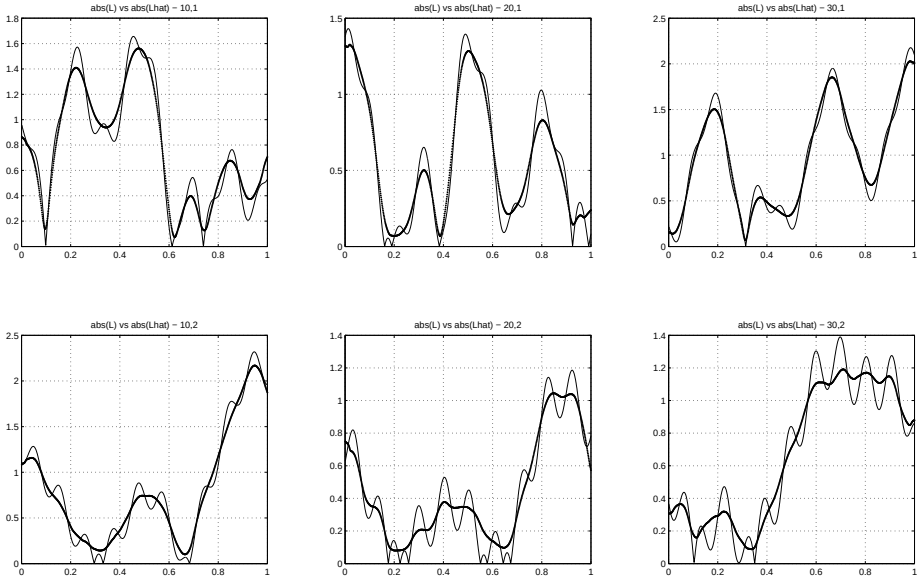


Figure 3: Time-varying entries of the matrix of loadings $\Lambda(u)$.

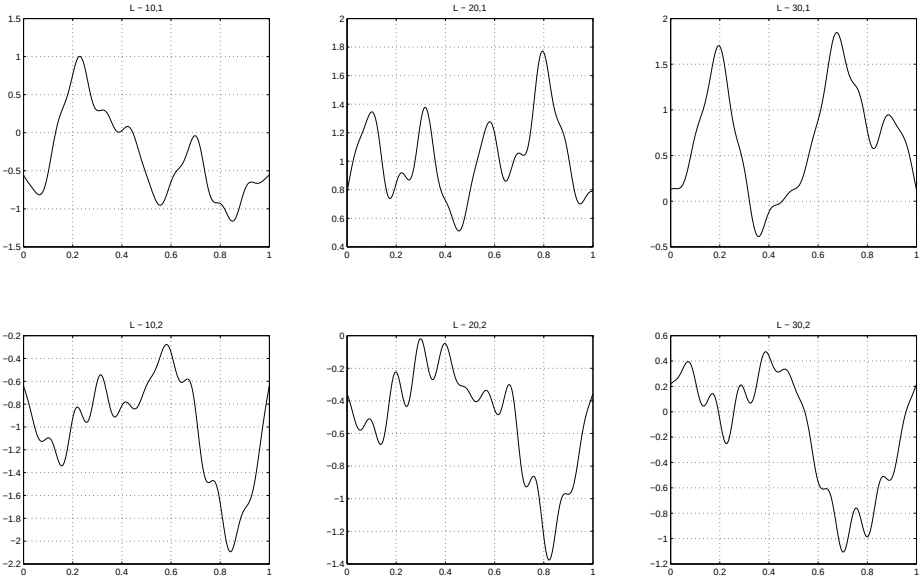


Figure 4: Time-varying entries of the matrix $\mathbf{\Sigma}_\Lambda(u)$.

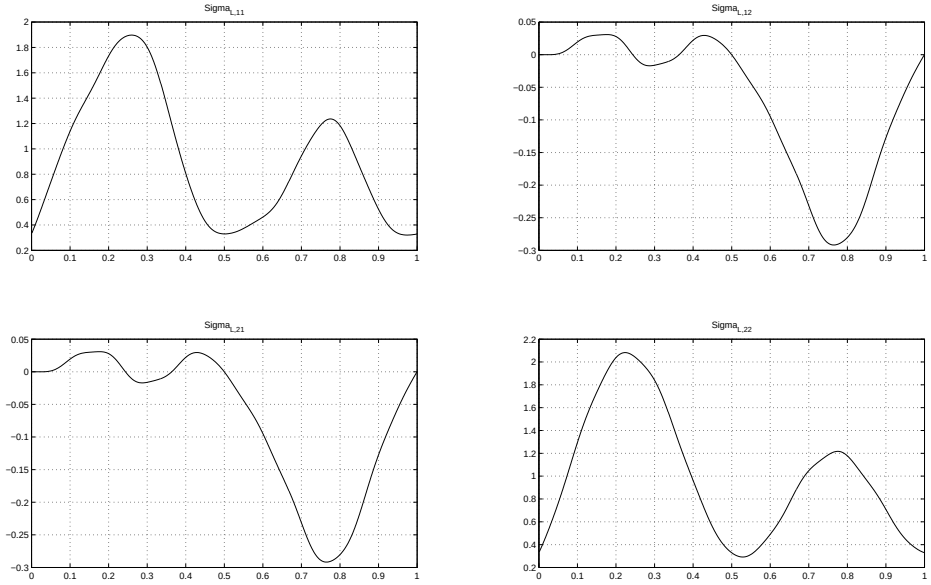


Figure 5: Time-varying entries of the covariance matrix. Solid line: $\mathbf{\Sigma}(u_t)$. Bold line: $\overline{\widehat{\mathbf{\Sigma}}(u_t)}$. Dashed lines: confidence intervals.

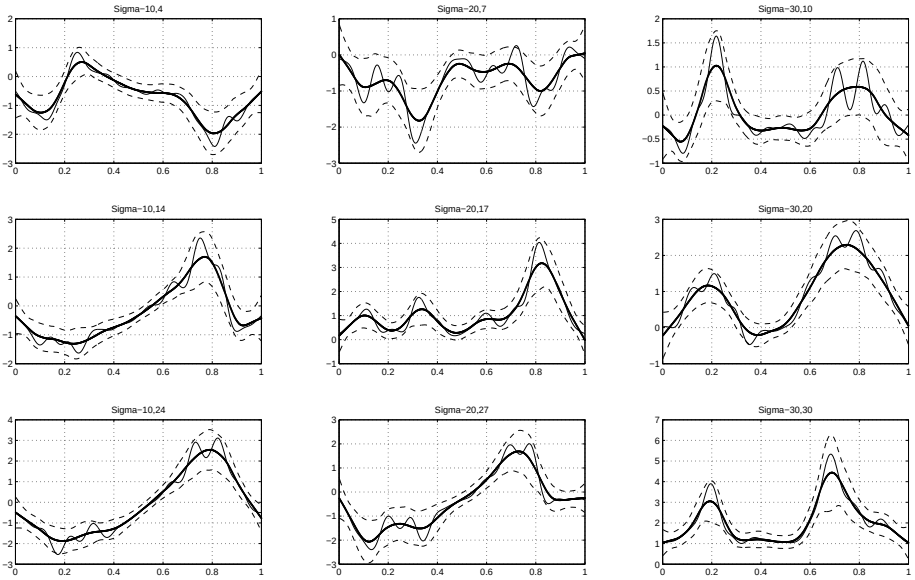


Figure 6: Time-varying entries of the diagonal matrices $V(u_t)$ and $V_{NT}(u_t)$. Solid line: $V(u_t)$. Bold line: $\overline{V_{NT}(u_t)}$. Dashed lines: confidence intervals.

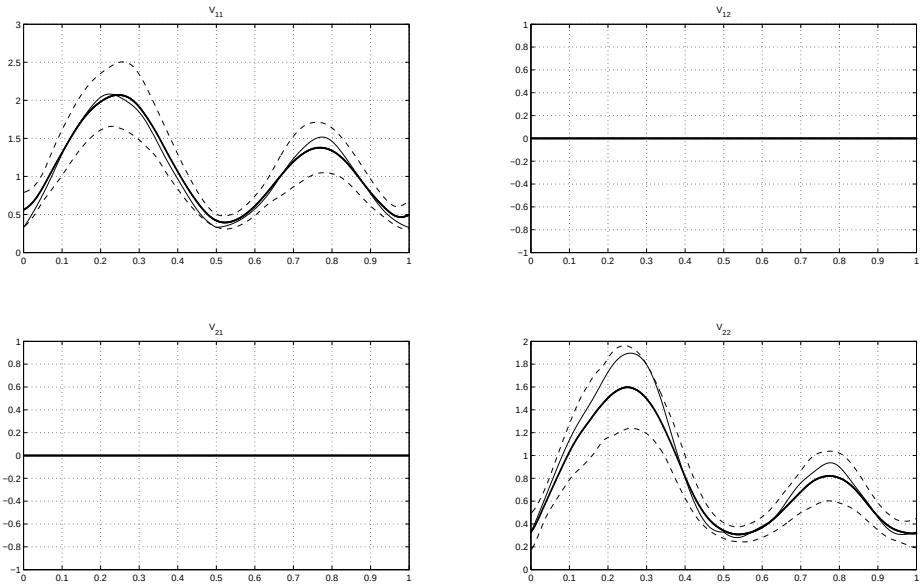
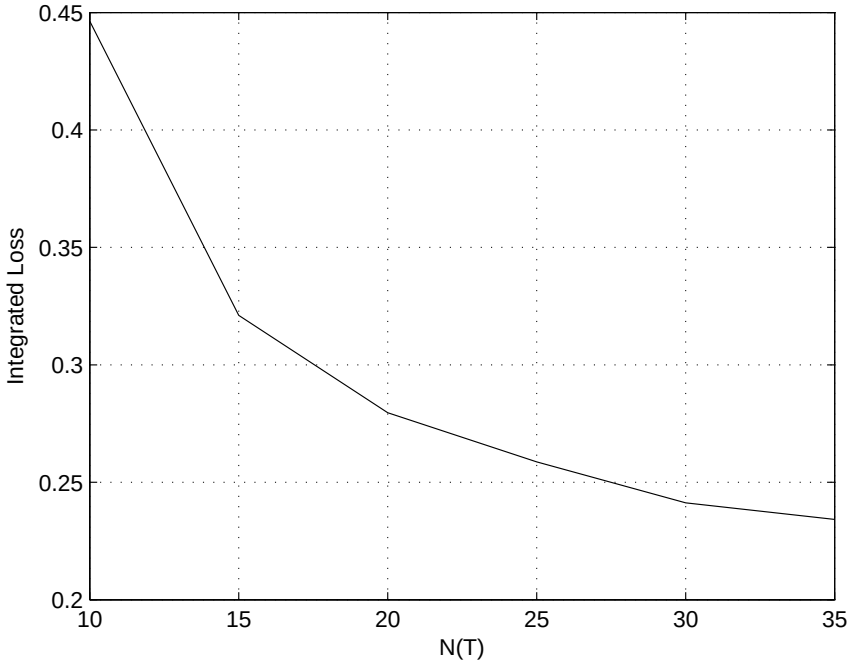


Figure 7: Integrated Loss defined in (41)



6 Conclusions and outlook

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 rotation and sign. Due to the nonparametric estimation of factor loadings, convergence rates are slower than in the stationary factor model with constant loadings. In a simulation study, our estimator was shown to work very well for two alternative scenarios: one where the rotation matrix is identity, another where it is not.

As already 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 analogue tests can be developed for our model framework, but considered it here to be 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.

As implied by Corollary 7, our model is promising for prediction purposes since the directly identifiable common components can be used as predictors. As shown by van Bellegem and von Sachs (2004) in a univariate framework, prediction using locally stationary models performs well when compared with classical stationary models.

There are many potential applications 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. Comparing the empirical results of a factor model using dynamic factors with those of a factor model with time-varying loadings would be an interesting topic for future work. In finance, applications to asset pricing and portfolio selection are obvious and this is also left for future research.

Acknowledgments

The authors would like to thank Michael Eichler, John Geweke, Domenico Giannone, Stefano Herzel, Marco Lippi, Alexei Onatski, and Catalin Stărică for helpful comments and discussions. We gratefully acknowledge financial support by the *Fonds Spéciaux de Recherche* (FSR 05) of the Université catholique de Louvain, Louvain-la-Neuve, Belgium.

A Some elements of matrix algebra

In this section all matrices are real. These results can be found in Lütkepohl (1996): the first two lemmas are given in Section 4.1.2, Lemma 3 is given in Section 5.2.1.

Lemma 1 (Cauchy-Schwarz inequality)

$$\mathrm{tr}^2\left\{A'B\right\} \leq \mathrm{tr}\left\{A'A\right\}\mathrm{tr}\left\{B'B\right\}$$

where A and B are real matrices of dimension $m \times n$.

Lemma 2 *Let A and B be positive semidefinite $(m \times m)$ matrices. Then* $\mathrm{tr}(AB) \leq \mathrm{tr}(A)\,\mathrm{tr}(B)$.

Lemma 3 *Let A and B be $k \times k$ positive definite matrices. Then*

$$\gamma \textit{ is eigenvalue of } BA \iff \gamma \textit{ is eigenvalue of } B^{1/2}AB^{1/2}$$

where $\iff$ means that the two statements are equivalent.

B Optimization of Functions involving the trace

In this section all matrices are real. The following results are given in Section 4.1.3 of Lütkepohl (1996).

Lemma 4 *Let $\mathbf{\Sigma}$ $(N \times N)$ be real symmetric with eigenvalues $\gamma_1 \geq \cdots \geq \gamma_N$ and associated orthonormal $(N \times 1)$ eigenvectors $\Gamma_1, \ldots, \Gamma_N$, $r \in \{1, \ldots, N\}$. Then:*

- $\min_L \{\mathrm{tr}\,(L'\mathbf{\Sigma}L) : L(N \times r)$ *real*$, L'L = I_r\} = \sum_{k=r+1}^{N} \gamma_k$*, the minimizing matrix being* $\widetilde{L} = [\Gamma_{r+1}, \ldots, \Gamma_N]$*;*
- $\max_L \{\mathrm{tr}\,(L'\mathbf{\Sigma}L) : L(N \times r)$ *real*$, L'L = I_r\} = \sum_{k=1}^{r} \gamma_k$*, the maximizing matrix being* $\widehat{L} = [\Gamma_1, \ldots, \Gamma_r]$.

Lemma 5 *Let X be $T \times N$ and $\gamma_1 \geq \cdots \geq \gamma_N$ be the eigenvalues of $X'X$ with associated $(N \times 1)$ orthonormal eigenvectors $\Gamma_1, \ldots, \Gamma_N$, $0 \leq r < N$. Then:*

$$\min_{F,L}\left\{\mathrm{tr}\left(X - FL'\right)'\left(X - FL'\right) : F(T \times r), L(N \times r) \textit{ with } L'L = I_r\right\} = \sum_{k=1}^{r} \gamma_k,$$

the minimizing matrices being

$$\widehat{L} = [\Gamma_1, \ldots, \Gamma_r], \qquad \widehat{F} = X\widehat{L}.$$

Lemma 6 *Let Y be $T \times N$, L be $N \times r$ with* $\mathrm{rk}\,(L) = r$*, and W be $T \times T$ positive definite. Then:*

$$\min_{A}\left\{\mathrm{tr}\left(Y - AL'\right)'W\left(Y - AL'\right) : A(T \times r)\right\} = \mathrm{tr}\left\{WYY' - WYL\left(L'L\right)^{-1}L'Y'\right\}.$$

The minimum is attained for $A = YL\,(L'L)^{-1}$.

C Smoothness of spectral decomposition

Lemma 7 (Takagi's factorization) . *Let $A \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$ be a complex symmetric matrix valued function of constant rank: $\operatorname{rk}[A(x)] \equiv r$ for all x for fixed $r : 1 \leq r \leq n$. Then there exists unitary $U \in \mathcal{C}^k(\mathbb{R}, \mathbb{C}^{n\times n})$ such that*

$$A(x) = U(x) \begin{bmatrix} S_+ & 0 \\ 0 & 0 \end{bmatrix} U'(x) \quad \forall x, \tag{42}$$

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

$$\liminf_{\tau\to 0} \frac{|\gamma_i(x+\tau) - \gamma_j(x+\tau)|}{|\tau^e|} \in (0, +\infty] \tag{43}$$

for some nonnegative integers $e \leq k$ and for all x and $i \neq j$. Then there exists orthogonal $\Gamma \in \mathcal{C}^{k-e}(\mathbb{R}, \mathbb{R}^{r\times r})$ such that $\Gamma' S_+ \Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_r)$. The eigenvalues can be taken to be C^k functions. The proof of this lemma is reported in Theorems 2.4 and 2.7 of Chern & Dieci (2000, p. 774-778).

D Proofs

For the proofs we use the following abbreviations

$$\begin{aligned} X_s &:= X_{s,T}, \qquad \Lambda_s := \Lambda\left(\tfrac{s}{T}\right), \qquad \widehat{\Lambda}_s := \widehat{\Lambda}\left(\frac{s}{T}\right), \quad \Lambda_s^{(1)} := \Lambda^{(1)}\left(\frac{s}{T}\right), \\ \widetilde{\Lambda}_s &:= \Lambda(u_{s,T})/\sqrt{N}, \quad V_{s,NT} := V_{NT}\left(\tfrac{s}{T}\right), \quad R_{s,NT} := R_{NT}\left(\frac{s}{T}\right) \end{aligned}$$

for $s = 1, \ldots, T$ and, following the definition of equation (20),

$$\widehat{\mathbf{\Sigma}}_t := \widehat{\mathbf{\Sigma}}(u_{t,T}; h) = T^{-1} \sum_{s=1}^{T} X_{s,T} X_{s,T}' \omega_{t,T}(u_{s,T}; h) \qquad t = 1, \ldots, T. \tag{44}$$

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.

D.1 Proof of Proposition 1

The assertion of this proposition can be written as

$$\operatorname{tr}\left\{\widetilde{\Lambda}_s' \widehat{\mathbf{\Sigma}}_t \widetilde{\Lambda}_s - \widetilde{\Lambda}_t' \widehat{\mathbf{\Sigma}}_t \widetilde{\Lambda}_t\right\} = O_p(Nh) = o_p(1) \ .$$

The idea of the proof is to apply Lemma 2 to the positive semi-definite matrices $\widetilde{\Lambda}_s \widetilde{\Lambda}_s'$ and $\widehat{\mathbf{\Sigma}}_t$, i.e., $\operatorname{tr}\left\{\widetilde{\Lambda}_s \widetilde{\Lambda}_s' \widehat{\mathbf{\Sigma}}_t\right\} \leq \operatorname{tr}\left\{\widetilde{\Lambda}_s \widetilde{\Lambda}_s'\right\} \operatorname{tr} \widehat{\mathbf{\Sigma}}_t$. The idea is further to use that $\operatorname{tr} \widetilde{\Lambda}_s \widetilde{\Lambda}_s' = O(1)$ and that by Theorem 2 $\mathbb{E} \operatorname{tr} \widehat{\mathbf{\Sigma}}_t = O(N)$.

More specifically due to the existence and uniform boundedness of the first derivative of $\Lambda(\cdot)$ we have the Taylor expansion

$$\lambda_{ij}(s/T) = \lambda_{ij}(t/T) + \lambda_{ij}^{(1)}(s^*/T)\ \frac{t-s}{T} \quad \text{for} \quad \left|\frac{t-s}{T}\right| \leq h\ ,$$

with a mean value $s/T \leq s^*/T \leq t/T$ (w.l.o.g.).

We plug in this Taylor expansion for each of the $\widetilde{\Lambda}_s$ into $\operatorname{tr}\left\{\widetilde{\Lambda}_s'\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_s - \widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_t\right\}$. Note that due to the construction of $\widehat{\mathbf{\Sigma}}_t$ to be a sum over kernel weights which are zero for arguments s with $\frac{s-t}{T} > h$ (essentially), we can directly infer that the above difference is equal to the sum of three traces:

$$\operatorname{tr}\left\{\widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_{s^*}^{(1)}\ h\ + \widetilde{\Lambda}_{s^*}^{(1)'}\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_t\ h\ + \widetilde{\Lambda}_{s^*}^{(1)'}\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_{s^*}^{(1)}\ h^2\right\}\ .$$

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

$$\mathbb{E}\left|\operatorname{tr}\left\{\widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_{s^*}^{(1)}\right\}\right| \ = \ O(N)\ ,$$

as this will imply, by the Markov inequality the desired stochastic convergence of order $O_p(Nh)$ of the whole difference under consideration.

We use that $\operatorname{tr}\left\{\widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\widetilde{\Lambda}_{s^*}^{(1)}\right\} = \operatorname{tr}\left\{\widetilde{\Lambda}_{s^*}^{(1)}\widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\right\}$ and that as indicated above,

$$\operatorname{tr}\left\{\widetilde{\Lambda}_{s^*}^{(1)}\widetilde{\Lambda}_t'\widehat{\mathbf{\Sigma}}_t\right\} \ \leq \ \operatorname{tr}\left\{\widetilde{\Lambda}_{s^*}^{(1)}\widetilde{\Lambda}_t'\right\}\operatorname{tr}\widehat{\mathbf{\Sigma}}_t\ .$$

Further we use the Cauchy-Schwarz inequality Lemma 1 to bound

$$\left|\operatorname{tr}\left\{\widetilde{\Lambda}_{s^*}^{(1)}\widetilde{\Lambda}_t'\right\}\right| \ \leq \ \left\|\widetilde{\Lambda}_{s^*}^{(1)}\right\|\left\|\widetilde{\Lambda}_t\right\| \ = \ O(1)\ ,$$

where we recall the definition of the norm $\|A\| = \sqrt{\operatorname{tr}\{A'A\}}$. But both norms on the right hand side are bounded from above by condition D2, which we recall to be

$$\sup_u \|\Lambda^{(1)}(u)\| \ = \ O(\sqrt{N})\ ,$$

and by the property of $\widetilde{\Lambda}_t'\widetilde{\Lambda}_t = Id_r$.

It remains to show that $\mathbb{E}\operatorname{tr}\widehat{\mathbf{\Sigma}}_t = O(N)$. (Note that in fact the eigenvalues, and hence the trace, of the smoothed covariance matrix estimator $\widehat{\mathbf{\Sigma}}_t$ are non-negative by construction.).

With Theorem 2, $\mathbb{E}\,N^{-1}\operatorname{tr}\widehat{\mathbf{\Sigma}}_t - N^{-1}\mathbf{\Sigma}_t = o(1)$, hence it is sufficient to show that $\operatorname{tr}\mathbf{\Sigma}_t = O(N)$. Recall that

$$\Sigma_t = \Lambda_t\Sigma_F\Lambda_t' + \Psi_N\ ,$$

and that with Assumption C.3, $\operatorname{tr}\Psi_N = O(N)$. Further, by the orthonormality of $\widetilde{\Lambda}_t$ and the invariance of the trace with respect to orthogonal rotations, we easily conclude that

$$\operatorname{tr}\left(\Lambda_t\Sigma_F\Lambda_t'\right) = N\ \operatorname{tr}\left(\widetilde{\Lambda}_t\Sigma_F\widetilde{\Lambda}_t'\right) = N\operatorname{tr}\Sigma_F = O(N)\ ,$$

as the trace of Σ_F is bounded from above. □

D.2 Proof of Theorem 2

Since $X_s = \Lambda_s F_s + e_s$, we have the following decomposition:

$$(NT)^{-1} X' \mathcal{W}_t X = (NT)^{-1} \sum_{s=1}^{T} X_s X_s' \omega_t(s; h_T) = A_{t,NT} + B_{t,NT} + B_{t,NT}' + C_{t,NT} \tag{45}$$

where

$$\begin{aligned} A_{t,NT} &= (NT)^{-1} \sum_{s=1}^{T} \Lambda_s F_s F_s' \Lambda_s' \omega_t(s; h_T), \\ B_{t,NT} &= (NT)^{-1} \sum_{s=1}^{T} \Lambda_s F_s e_s' \omega_t(s; h_T), \\ C_{t,NT} &= (NT)^{-1} \sum_{s=1}^{T} e_s e_s' \omega_t(s; h_T). \end{aligned}$$

This decomposition allows us to write

$$\begin{aligned} \left\| (NT)^{-1} X' \mathcal{W}_t X - N^{-1} \mathbf{\Sigma} \left(\frac{t}{T} \right) \right\| =: \left\| A_{t,NT} + B_{t,NT} + B_{t,NT}' + C_{t,NT} - N^{-1} \left(\Lambda_t \mathbf{\Sigma}_F \Lambda_t' + \Psi \right) \right\| \leq \\ \leq \left\| A_{t,NT} - N^{-1} \Lambda_t \mathbf{\Sigma}_F \Lambda_t' \right\| + \left\| B_{t,NT} \right\| + \left\| C_{t,NT} - N^{-1} \Psi \right\| \end{aligned}$$

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.

$$\lambda_{ij}(s/T) = \lambda_{ij}(t/T) + \lambda_{ij}^{(1)}(s^*/T) \ \frac{t-s}{T} \quad \text{for} \quad \left| \frac{t-s}{T} \right| \leq h \ ,$$

with a mean value $s/T \leq s^*/T \leq t/T$ (w.l.o.g.), which we write in matrix notation and slightly differently here

$$\Lambda_s = \Lambda_t + h_T z_s (\Lambda_t^{(1)} + o(1)) \ = \Lambda_t + h_T z_s \Lambda_t^{(1)} + o(h_T) \ ,$$

with $\Lambda_t := \Lambda(t/T)$, $\Lambda_t^{(1)} := \Lambda^{(1)}(t/T)$ and $z_s := \frac{t-s}{Th_T}$. Note again that due to the use of the kernel weights in the matrix $\mathcal{W}_t$ which are essentially zero for arguments s with $\frac{s-t}{T} > h$, we can w.l.o.g. argue that $|z_s| \leq 1$.

This allows us to consider for the first term in (45) a sum of 4 terms: $A_{t,NT} = \sum_{k=1}^{4} {}_kA_{t,NT}$, where

$$\begin{aligned} {}_1A_{t,NT} &= (NT)^{-1} \textstyle\sum_{s=1}^{T} \Lambda_t F_s F_s' \Lambda_t' \omega_t(s; h_T) \\ {}_2A_{t,NT} &= (NT)^{-1} \textstyle\sum_{s=1}^{T} \Lambda_t F_s F_s' (\Lambda_t'^{(1)} + o(1)) h_T z_s \omega_t(s; h_T) \\ {}_3A_{t,NT} &= (NT)^{-1} \textstyle\sum_{s=1}^{T} (\Lambda_t^{(1)} + o(1)) \ F_s F_s' \Lambda_t' h_T z_s \omega_t(s; h_T) \\ {}_4A_{t,NT} &= (NT)^{-1} \textstyle\sum_{s=1}^{T} (\Lambda_t^{(1)} + o(1)) F_s F_s' (\Lambda_t'^{(1)} + o(1)) h_T^2 z_s^2 \omega_t(s; h_T) \ . \end{aligned}$$

In order to show that $\left\| A_{t,NT} - N^{-1} \Lambda_t \mathbf{\Sigma}_F \Lambda_t' \right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right) + O(h_T)$, 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 $\left\| {}_1A_{t,NT} - N^{-1} \Lambda_t \mathbf{\Sigma}_F \Lambda_t' \right\|$ tends to zero with the appropriate rate. Indeed

$$\left\| \frac{1}{NT} \sum_{s=1}^{T} \Lambda_t F_s F_s' \Lambda_t' \omega_t(s; h_T) - N^{-1} \Lambda_t \mathbf{\Sigma}_F \Lambda_t' \right\| \leq \left\| \frac{1}{N} \Lambda_t' \Lambda_t \right\| \left\| \frac{1}{T} \sum_{s=1}^{T} \operatorname{vec}\left(F_s F_s' \right) \omega_t(s; h_T) - \operatorname{vec}\left(\mathbf{\Sigma}_F \right) \right\|$$

and

$$\left\| \frac{1}{N} \Lambda_t' \Lambda_t \right\| \to \|\mathbf{\Sigma}_\Lambda(u)\| = O(1)$$

by assumption B2. Then it suffices to show that

$$\sqrt{Th_T} \left\| \frac{1}{T} \sum_{s=1}^{T} \operatorname{vec}\left(F_s F_s'\right) \omega_t(s; h_T) - \operatorname{vec}\left(\mathbf{\Sigma}_F\right) \right\| = O_p(1). \tag{46}$$

To do that we show that the expectation and the variance of the argument of the norm in (46) 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

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{T} \sum_{s=1}^{T} \operatorname{vec}\left(F_s F_s'\right) \omega_t(s; h_T) - \operatorname{vec}\left(\mathbf{\Sigma}_F\right) \right] = \left[\frac{1}{T} \sum_{s=1}^{T} \omega_t(s; h_T) I_{r^2} - I_{r^2} \right] \operatorname{vec}\left(\mathbf{\Sigma}_F\right) = \\ &= \left\{ \left(1 + O\left(T^{-1} h_T^{-1}\right)\right) I_{r^2} - I_{r^2} \right\} \operatorname{vec}\left(\mathbf{\Sigma}_F\right) = O\left(T^{-1} h_T^{-1}\right) \operatorname{vec}\left(\mathbf{\Sigma}_F\right) = O\left(T^{-1} h_T^{-1}\right). \end{aligned}$$

The argument of the norm in (46) can be written as

$$\frac{1}{T} \sum_{s=1}^{T} G_s \omega_t(s; h_T) + O\left(T^{-1} h_T^{-1}\right)$$

where $G_t := \operatorname{vec}\left(F_t F_t' - \mathbf{\Sigma}_F\right) \overset{\text{iid}}{\sim} (0, P)$ for a given matrix P. Then

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[\frac{1}{T} \sum_{s=1}^{T} G_s \omega_t(s; h_T) + O\left(T^{-1} h_T^{-1}\right) \right] &= \frac{1}{T^2} \mathbb{E}\left[\sum_{s=1}^{T} G_s G_s' \omega_t^2(s; h_T) \right] = \\ &= \frac{1}{T^2} \mathbb{E}\left[G_t G_t' \right] \sum_{s=1}^{T} \omega_t^2(s; h_T) = \\ &= P \frac{1}{T^2} \sum_{s=1}^{T} \omega_t^2(s; h_T) = O\left(T^{-1} h_T^{-1}\right). \end{aligned}$$

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

Second, in order to treat the term ${}_2A_{t,NT}$ we show quite analogously to the above and recalling the condition $Th_T^3 = o(1)$, that

$$\sqrt{Th_T} \left\| (NT)^{-1} \sum_{s=1}^{T} \Lambda_t F_s F_s' (\Lambda_t'^{(1)} + o(1)) \; h_T z_s \omega_t(s; h_T) \right\| = \sqrt{Th_T}\, O_p(h_T) = o_p(1). \tag{47}$$

To do so we essentially have to use that

$$\frac{1}{N} \left\| \Lambda_t' \Lambda_t^{(1)} \right\| = \frac{1}{N} \sqrt{\operatorname{tr}\left(\Lambda_t' \Lambda_t \Lambda_t'^{(1)} \Lambda_t^{(1)} \right)} = \sqrt{N^{-1} \operatorname{tr}\left(\Lambda_t' \Lambda_t\right)} \sqrt{N^{-1} \operatorname{tr}\left(\Lambda_t'^{(1)} \Lambda_t^{(1)} \right)} = O(1)$$

by assumption D2.

Treatment of the terms ${}_3A_{t,NT}$ and ${}_4A_{t,NT}$ would be similar.

By similar arguments, it can be shown that $\|B_{t,NT}\|$ and $\left\|C_{t,NT} - N^{-1}\Psi\right\|$ are both $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. For the term $\|B_{t,NT}\|$ we have

$$\|{}_1B_{t,NT}\| \leq \left\| \frac{1}{\sqrt{N}} \Lambda_t \right\| \left\| \frac{1}{T\sqrt{N}} \sum_{s=1}^{T} F_s e_s' \omega_t(s; h_T) \right\|$$

where ${}_1B_{t,NT}$ is the first term of the sum $B_{t,NT} = \sum_{k=1}^{4} {}_kB_{t,NT}$, again obtained by using a Taylor expansion of Λ_s, and $\left\|\frac{1}{\sqrt{N}}\Lambda_t\right\| = O(1)$ by assumption B2. Then

$$\|\,{}_1B_{t,NT}\| \propto \left\|\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\text{vec}\left[F_s e_s'\right]\omega_t(s;h_T)\right\|$$

Since $\mathbb{C}\text{ov}[F_t, e_t] = 0\ \forall t$,

$$\mathbb{E}\left\{\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\text{vec}\left[F_s e_s'\right]\omega_t(s;h_T)\right\} = \mathbf{0} \quad \forall t.$$

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

$$\begin{aligned}\mathbb{V}\text{ar}\left[\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}H_s\omega_t(s;h_T)\right] &= \frac{1}{NT^2}\mathbb{E}\left[\sum_{s=1}^{T}H_sH_s'\omega_t^2(s;h_T)\right] = \frac{1}{NT^2}\mathbb{E}\left[H_tH_t'\right]\sum_{s=1}^{T}\omega_t^2(s;h_T) = \\ &= N^{-1}Q\frac{1}{T^2}\sum_{s=1}^{T}\omega_t^2(s;h_T) = O\left(N^{-1}T^{-1}h_T^{-1}\right).\end{aligned}$$

This means that each term of the $Nr \times 1$ vectors $\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\text{vec}\left[F_s e_s'\right]\omega_t(s;h_T)$ is $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$, and thus

$$\left\|\frac{1}{T\sqrt{N}}\sum_{s=1}^{T}\text{vec}\left[F_s e_s'\right]\omega_t(s;h_T)\right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$$

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

For the last term we have

$$\begin{aligned}\mathbb{E}\left[C_{t,NT} - N^{-1}\Psi\right] &= \left[\frac{1}{NT}\sum_{s=1}^{T}\mathbb{E}\left(e_s e_s'\right)\omega_t(s;h_T)\right] - N^{-1}\Psi = \\ &= N^{-1}\Psi\left[T^{-1}\sum_{s=1}^{T}\omega_t(s;h_T)\right] - N^{-1}\Psi = \\ &= N^{-1}\Psi\left[1 + O\left(T^{-1}h_T^{-1}\right)\right] - N^{-1}\Psi = \\ &= O\left(\frac{1}{NTh_T}\right).\end{aligned}$$

As before, define $J_t := \text{vec}\left[e_t e_t' - \Psi\right] \overset{\text{iid}}{\sim} (0, R)$ for another given matrix R which does exist due to Assumption C2 and obtain:

$$\begin{aligned}&\mathbb{V}\text{ar}\left[\frac{1}{TN}\sum_{s=1}^{T}J_s\omega_t(s;h_T) + O\left(\frac{1}{NTh_T}\right)\right] = \frac{1}{N^2T^2}\mathbb{E}\left[\sum_{s=1}^{T}J_sJ_s'\omega_t^2(s;h_T)\right] = \\ &\frac{1}{N^2T^2}\mathbb{E}\left[J_tJ_t'\right]\sum_{s=1}^{T}\omega_t^2(s;h_T) = N^{-2}R\frac{1}{T^2}\sum_{s=1}^{T}\omega_t^2(s;h_T) = O\left(N^{-2}T^{-1}h_T^{-1}\right).\end{aligned}$$

Each term of the $N^2 \times 1$ vectors $\frac{1}{TN}\sum_{s=1}^{T}\text{vec}\left[e_s e_s' - \Psi\right]\omega_t(s;h_T)$ is $O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$, and thus

$$\left\|\frac{1}{NT}\sum_{s=1}^{T}\text{vec}\left[e_s e_s'\right]\omega_t(s;h_T)\right\| = O_p\left(\frac{1}{\sqrt{Th_T}}\right)$$

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

D.3 Proof of Proposition 3

Consider equation (24)

$$(NT)^{-1}X'\mathcal{W}_tX\widehat{\Lambda}_t = \widehat{\Lambda}_tV_{t,NT} \tag{48}$$

where $\mathcal{W}_t := \mathcal{W}_t(h_T)$, and multiply it to the left by

$$\left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'}{N}$$

to obtain

$$(NT)^{-1}\left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'}{N}X'\mathcal{W}_tX\widehat{\Lambda}_t = \left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'}{N}\widehat{\Lambda}_tV_{t,NT}. \tag{49}$$

By (45) we get

$$\left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'}{N}\left[A_{t,NT}+B_{t,NT}+B_{t,NT}'+C_{t,NT}\right]\widehat{\Lambda}_t = \left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'}{N}\widehat{\Lambda}_tV_{t,NT}\ , \tag{50}$$

which can be written as

$$G_{t,NT}+b_{t,NT}+c_{t,NT} = Q_{t,NT}V_{t,NT}$$

where

$$\begin{aligned}
G_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'}{\sqrt{N}}A_{t,NT}\tfrac{\widehat{\Lambda}_t}{\sqrt{N}}\\
b_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'}{\sqrt{N}}\left[B_{t,NT}+B_{t,NT}'\right]\tfrac{\widehat{\Lambda}_t}{\sqrt{N}}\\
c_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'}{\sqrt{N}}C_{t,NT}\tfrac{\widehat{\Lambda}_t}{\sqrt{N}}\\
Q_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'\widehat{\Lambda}_t}{N}.
\end{aligned}$$

By Theorem 2, $A_{t,NT} = N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\Lambda_t' + U_{NT}$, where $U_{NT} = O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$, i.e. U_{NT} is an $N\times N$ matrix each element being $O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$. Since

$$\begin{aligned}
\left\|\frac{F'F}{T}\right\| &= O_p(1) \quad \text{by assumption A3}\\
\left\|\frac{\Lambda_t}{\sqrt{N}}\right\| &= O(1) \quad \text{by assumption B2}\\
\left\|\frac{\widehat{\Lambda}_t}{\sqrt{N}}\right\| &= O_p(1) \quad \text{by the constraint in (21).}
\end{aligned}$$

we have

$$G_{t,NT} = \left(\frac{F'F}{T}\right)^{1/2}\frac{\Lambda_t'\Lambda_t}{N}\left(\frac{F'F}{T}\right)\frac{\Lambda_t'\widehat{\Lambda}_t}{N} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right) = P_{t,NT}Q_{t,NT} + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$$

where

$$\begin{aligned}
P_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'\Lambda_t}{N}\left(\tfrac{F'F}{T}\right)^{1/2}\\
Q_{t,NT} &= \left(\tfrac{F'F}{T}\right)^{1/2}\tfrac{\Lambda_t'\widehat{\Lambda}_t}{N}
\end{aligned}$$

The term $b_{t,NT}$ is $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$ because $B_{t,NT} = O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$, and the term $c_{t,NT}$ is $O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$ because $C_{t,NT} = O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$ (see proof of Theorem 2). Then we have

$$P_{t,NT}Q_{t,NT} + d_{t,NT} = Q_{t,NT}V_{t,NT} \tag{51}$$

where $d_{t,NT} = b_{t,NT} + c_{t,NT} = O_p\left(\frac{1}{\sqrt{NTh_T}}\right) = o_p(1)$.
From (51), using that $Q_{t,NT}$ is of full rank, see e.g. Lütkepohl (1996, 4.3.1. (13) applied to $\frac{\Lambda_t'\widehat{\Lambda}_t}{N}$), we have

$$\left[P_{t,NT} + d_{t,NT}Q_{t,NT}^{-1}\right] Q_{t,NT} = Q_{t,NT}V_{t,NT}$$

Let $V_{t,NT}^{\star}$ be a diagonal matrix consisting of the diagonal elements of $Q_{t,NT}'Q_{t,NT}$. Note that $V_{t,NT}^{\star}$ is invertible because $Q_{t,NT}$ is of full rank. Denote $\Upsilon_{t,NT} = Q_{t,NT}V_{t,NT}^{\star-1/2}$ so that each column of $\Upsilon_{t,NT}$ has unit length, and we have

$$\left[P_{t,NT} + d_{t,NT}Q_{t,NT}^{-1}\right] \Upsilon_{t,NT} = \Upsilon_{t,NT}V_{t,NT}.$$

Thus each column of $\Upsilon_{t,NT}$ is an eigenvector of the matrix $\left[P_{t,NT} + d_{t,NT}Q_{t,NT}^{-1}\right]$. Note that $\left[P_{t,NT} + d_{t,NT}Q_{t,NT}^{-1}\right]$ converges to $P(u_t) = \mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_\Lambda(u_t)\mathbf{\Sigma}_F^{1/2}$ by Assumptions A3 and B2 and $d_{t,NT} = o_p(1)$

For large values of T and N, the diagonal matrix $V_{t,NT}$ contains the eigenvalues of $P_{t,NT}$. By assumptions A3 and B2 and Slutsky's Theorem

$$\text{plim}_{N,T\to\infty} P_{t,NT} = P(u_t) = \mathbf{\Sigma}_F^{1/2}\mathbf{\Sigma}_\Lambda(u_t)\mathbf{\Sigma}_F^{1/2}.$$

Lemma 3 in Appendix A gives the result. □

D.4 Proof of Proposition 4

For this proof we use the following lemma.

Lemma 8

$$Q_{t,NT}'Q_{t,NT} := \frac{\widehat{\Lambda}_t'\Lambda_t}{N}\left(\frac{F'F}{T}\right)\frac{\Lambda_t'\widehat{\Lambda}_t}{N} \xrightarrow{p} V(u).$$

Proof: By (24) we have $V_{t,NT} = \frac{\widehat{\Lambda}_t'}{N}\frac{X'\mathcal{W}_tX}{NT}\widehat{\Lambda}_t$, and by Theorem 2

$$\frac{X'\mathcal{W}_tX}{NT} = N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\Lambda_t' + N^{-1}\Psi + Z_{t,NT}$$

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

$$V_{t,NT} = \frac{\widehat{\Lambda}_t'\Lambda_t}{N}\left(\frac{F'F}{T}\right)\frac{\Lambda_t'\widehat{\Lambda}_t}{N} + \frac{\widehat{\Lambda}_t'}{\sqrt{N}}\frac{\Psi}{N}\frac{\widehat{\Lambda}_t}{\sqrt{N}} + \frac{\widehat{\Lambda}_t'Z_{t,NT}\widehat{\Lambda}_t}{N}$$

and thus

$$\begin{aligned}\left\|V_{t,NT} - Q_{t,NT}'Q_{t,NT}\right\| \leq & \left\|\frac{\widehat{\Lambda}_t'\widehat{\Lambda}_t}{N}\right\| \left\|\frac{\Psi}{N}\right\| + \left\|\frac{\widehat{\Lambda}_t'\widehat{\Lambda}_t}{N}\right\| \|Z_{t,NT}\| = \left\|\frac{\widehat{\Lambda}_t'\widehat{\Lambda}_t}{N}\right\| \left(\left\|\frac{\Psi}{N}\right\| + \|Z_{t,NT}\|\right) = \\ & = O_p(1)\left[O\left(\frac{1}{\sqrt{N}}\right) + O\left(\frac{1}{\sqrt{Th_T}}\right)\right] = o_p(1)\end{aligned}$$

because

$$\left\|\frac{\Psi}{N}\right\| = N^{-1}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}\psi_{ij}^2} \leq N^{-1}\sqrt{N\max_{1\leq i\leq N}\sum_{j=1}^{N}\psi_{ij}^2} \leq \frac{1}{\sqrt{N}}\max_{1\leq i\leq N}\sum_{j=1}^{N}|\psi_{ij}| = O\left(\frac{1}{\sqrt{N}}\right)$$

by assumption C3. With Proposition 3 we have the result. □

Note that our Lemma 8 is analogous to Lemma A.3(ii) of Bai (2003). Because the ordering of eigenvectors in the points of intersections is identified by our convention (see §3.1), the matrix $\Upsilon_{t,NT}$ converges to the matrix $\Upsilon(u)$, similar to the proof of Proposition 1 of Bai (2003, see pages 161-162). We now continue with the proof of Proposition 4.

From the definitions of $Q_{t,NT}$ and $\Upsilon_{t,NT}$ it follows that

$$\frac{\Lambda_t'\widehat{\Lambda}_t}{N} = \left(\frac{F'F}{T}\right)^{-1/2} Q_{t,NT} = \left(\frac{F'F}{T}\right)^{-1/2} \Upsilon_{t,NT} V_{t,NT}^{\star 1/2}.$$

The term $\left(\frac{F'F}{T}\right)^{-1/2}$ converges to $\mathbf{\Sigma}_F^{-1/2}$ by Assumption A3. By Proposition 3 and Lemma 8, $V_{t,NT}^{*} = \text{diag}\left\{Q_{t,NT}'Q_{t,NT}\right\} \xrightarrow{p} V(u)$, and $\Upsilon_{t,NT} \xrightarrow{p} \Upsilon(u)$ by Proposition 3. Note that due to Assumption E and our identifiability assumption in intersection points of eigenvalues of $\Sigma_\Lambda(u)\Sigma_F$, the limiting function is uniquely defined in each $u \in (0,1)$ and a continuous function of u. □

D.5 Proof of Theorem 5

By (24) and (45) we have

$$\widehat{\Lambda}_t = (NT)^{-1}X'\mathcal{W}_t X\widehat{\Lambda}_t V_{t,NT}^{-1} = \left[A_{t,NT} + B_{t,NT} + B_{t,NT}' + C_{t,NT}\right]\widehat{\Lambda}_t V_{t,NT}^{-1}.$$

If we define the rotation matrix

$$R_{t,NT} = \left(\frac{F'F}{T}\right)\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)V_{t,NT}^{-1};$$

we can write

$$\widehat{\Lambda}_t - \Lambda_t R_{t,NT} = \left[A_{t,NT} - N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\Lambda_t' + B_{t,NT} + B_{t,NT}' + C_{t,NT}\right]\widehat{\Lambda}_t V_{t,NT}^{-1}$$

that is

$$(i)\quad \tfrac{1}{\sqrt{N}}\left\|\widehat{\Lambda}_t - \Lambda_t R_{t,NT}\right\| \leq \left(\left\|A_{t,NT} - N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\Lambda_t'\right\| + \|B_{t,NT}\| + \|C_{t,NT}\|\right)\left\|\frac{\widehat{\Lambda}_t}{\sqrt{N}}\right\|\left\|V_{t,NT}^{-1}\right\|$$

$$(ii)\quad \tfrac{1}{\sqrt{N}}\left\|\widehat{\lambda}_{it} - R_{t,NT}'\lambda_{it}\right\| \leq \left\|V_{t,NT}^{-1}\right\|\left\|\frac{\widehat{\Lambda}_t}{\sqrt{N}}\right\| \times \left(\left\|A_{it,NT} - N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\lambda_{it}'\right\| + \|B_{it,NT}\| + \left\|\widetilde{B}_{it,NT}\right\| + \|C_{it,NT}\|\right)$$

where $\widehat{\lambda}_{it}$, λ_{it}, $A_{it,NT}$, $B_{it,NT}$, $C_{it,NT}$ are the $i-$th columns of $\widehat{\Lambda}_t'$, Λ_t', $A_{t,NT}$, $B_{t,NT}$, $C_{t,NT}$, respectively:

$$\begin{aligned}
A_{it,NT} &= (NT)^{-1}\sum_{s=1}^{T}\Lambda_s F_s F_s'\lambda_{is}\omega_t(s;h_T), \\
B_{it,NT} &= (NT)^{-1}\sum_{s=1}^{T}\Lambda_s F_s e_{is}\omega_t(s;h_T), \\
\widetilde{B}_{it,NT} &= (NT)^{-1}\sum_{s=1}^{T} e_s F_s'\lambda_{is}\omega_t(s;h_T), \\
C_{it,NT} &= (NT)^{-1}\sum_{s=1}^{T} e_s e_{is}\omega_t(s;h_T).
\end{aligned}$$

Consider part (i). By Theorem 2, $\left\|A_{t,NT} - N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\Lambda_t'\right\|$ and $\|B_{t,NT}\|$ are both $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$. For the last term we have $C_{t,NT} = N^{-1}\Psi + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$; then

$$\|C_{t,NT}\| \leq \left\|N^{-1}\Psi\right\| + O_p\left(\frac{1}{\sqrt{Th_T}}\right) = O\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{Th_T}}\right) = O\left(\frac{1}{\sqrt{N}}\right).$$

The terms $\left\|V_{t,NT}^{-1}\right\|$ and $\left\|\frac{\widehat{\Lambda}_t}{\sqrt{N}}\right\|$ are both $O_p(1)$, and the proof of Proposition 5 (i) is complete.

Consider part (ii). The term $\left\|A_{it,NT} - N^{-1}\Lambda_t\left(\frac{F'F}{T}\right)\lambda_{it}'\right\|$ is also $O_p\left(\frac{1}{\sqrt{Th_T}}\right)$ because each element of the $r\times 1$ vector λ_{it}, $\forall i$ and $\forall t$, is allowed to grow at the rate $\sqrt{N}$ by the constraint in (21). The terms $\|B_{it,NT}\|$ and $\left\|\widetilde{B}_{it,NT}\right\|$ are both $O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$ because $\|\Lambda_s e_{is}\|$ and $\|e_s\lambda_{is}'\|$ are $O_p(\sqrt{N})$ and $\|F_s\| = O_p(1)$ $\forall s$ and $\forall i$. For the last term we have $\|C_{it,NT}\| = O_p\left(\frac{1}{N}\right)$. Indeed $C_{it,NT} = N^{-1}\psi_i + O_p\left(\frac{1}{N\sqrt{Th_T}}\right)$, and

$$\|C_{it,NT}\| \leq N^{-1}\|\psi_i\| + O_p\left(\frac{1}{\sqrt{NTh_T}}\right) = O\left(\frac{1}{N}\right) + O_p\left(\frac{1}{\sqrt{NTh_T}}\right)$$

because

$$\|\psi_i\| = \sqrt{\sum_{j=1}^{N}\psi_{ij}^2} \leq \sum_{j=1}^{N}|\psi_{ij}| = O(1) \quad \text{by assumption C3.}$$

Then

$$\begin{aligned}
\frac{1}{\sqrt{N}}\left\|\widehat{\lambda}_{it} - R_{t,NT}'\lambda_{it}\right\| &= O_p(1)O_p(1)\left(O_p\left(\tfrac{1}{\sqrt{Th_T}}\right) + O_p\left(\tfrac{1}{\sqrt{NTh_t}}\right) + O_p\left(\tfrac{1}{N}\right)\right) = \\
&= O_p\left(\tfrac{1}{\sqrt{Th_T}}\right) + O_p\left(\tfrac{1}{N}\right).
\end{aligned}$$

□

D.6 Proof of Theorem 6

By (28) and (10), the estimated factors can be written as

$$\widehat{F}_t = \frac{\widehat{\Lambda}_t'\Lambda_t}{N}F_t + \frac{\widehat{\Lambda}_t'}{N}e_t.$$

The rotation matrix is $R_{t,NT} = \left(\frac{F'F}{T}\right)\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)V_{t,NT}^{-1}$, and its inverse is given by $R_{t,NT}^{-1} = V_{t,NT}\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)^{-1}\left(\frac{F'F}{T}\right)^{-1}$. The inverse of the rotation matrix can be written as (see the proof of Lemma 8)

$$\begin{aligned} R_{t,NT}^{-1} = & \left[\left(\frac{\widehat{\Lambda}_t'\Lambda_t}{N}\right)\left(\frac{F'F}{T}\right)\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right) + \frac{\widehat{\Lambda}_t'}{\sqrt{N}}\frac{\Psi}{N}\frac{\widehat{\Lambda}_t}{\sqrt{N}} + \frac{\widehat{\Lambda}_t'Z_{NT}\widehat{\Lambda}_t}{N}\right]\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)^{-1}\left(\frac{F'F}{T}\right)^{-1} = \\ = & \left(\frac{\widehat{\Lambda}_t'\Lambda_t}{N}\right) + \left[\frac{\widehat{\Lambda}_t'}{\sqrt{N}}\frac{\Psi}{N}\frac{\widehat{\Lambda}_t}{\sqrt{N}} + \frac{\widehat{\Lambda}_t'Z_{NT}\widehat{\Lambda}_t}{N}\right]\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)^{-1}\left(\frac{F'F}{T}\right)^{-1}. \end{aligned}$$

Thus

$$\widehat{F}_t - R_{t,NT}^{-1}F_t = \frac{\widehat{\Lambda}_t'}{N}e_t + \left[\frac{\widehat{\Lambda}_t'}{\sqrt{N}}\frac{\Psi}{N}\frac{\widehat{\Lambda}_t}{\sqrt{N}} + \frac{\widehat{\Lambda}_t'Z_{NT}\widehat{\Lambda}_t}{N}\right]\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)^{-1}\left(\frac{F'F}{T}\right)^{-1}F_t$$

and thus

$$\begin{aligned} \left\|\widehat{F}_t - R_{t,NT}^{-1}F_t\right\| \leq & \left\|\frac{\widehat{\Lambda}_t'e_t}{N}\right\| + \left(\left\|\frac{\widehat{\Lambda}_t'\widehat{\Lambda}_t}{N}\right\|\left\|\frac{\Psi}{N}\right\| + \left\|\frac{\widehat{\Lambda}_t'\widehat{\Lambda}_t}{N}\right\|\|Z_{NT}\|\right)\left\|\left(\frac{\Lambda_t'\widehat{\Lambda}_t}{N}\right)^{-1}\right\|\left\|\left(\frac{F'F}{T}\right)^{-1}\right\|\|F_t\| = \\ = & O_p\left(\frac{1}{\sqrt{N}}\right) + \left[O_p(1)O\left(\frac{1}{\sqrt{N}}\right) + O_p(1)O_p\left(\frac{1}{\sqrt{Th_T}}\right)\right]O_p(1)O_p(1)O_p(1) = \\ = & O_p\left(\frac{1}{\sqrt{N}}\right) + O_p\left(\frac{1}{\sqrt{Th_T}}\right). \end{aligned}$$

Thus

$$\min\left(\sqrt{N}, \sqrt{Th_T}\right)\left\|\widehat{F}_t - R_{t,NT}^{-1}F_t\right\| = O_p(1)$$

and the proof of Theorem 6 is complete. □

D.7 Proof of Corollary 7

This corollary follows directly from Slutsky's Theorem applied to Theorems 5 and 6. □

References

Alexander C.O. (2001). Orthogonal GARCH, in *Mastering Risk*, Vol. II, edited by C.O. Alexander, pp. 21-38, Prentice Hall.

Bai, J., and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica*, Vol. 70, 191-221.

Bai, J. (2003). Inferential Theory for Factor Models of large dimension, *Econometrica*, Vol. 71, No. 1, 135-171.

Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics*, 21, 79-109.

Breitung, J. and Eickmeier, S. (2006). Dynamic factor models, in: Hübler and J. Frohn (eds.), *Modern Econometric Analysis*, Chapter 3, Springer 2006.

Brockmann, M., Gasser, T., and Hermann, E. (1993). Locally Adaptive Bandwidth Choice for Kernel Regression Estimators, *JASA*, Vol. 88, 1302-1309.

Chamberlain, G., Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets, *Econometrica*, Vol. 51, no.5, 1281-1304.

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

Connor, G., Korajczyk, R. A. (1993). A test for the number of factors in an approximate factor model, *Journal of Finance*, 48, 1263-1291.

Dahlhaus, R. (1996). Asymptotic statistical inference for nonstationary processes with evolutionary spectra. In: Robinson P. & Rosenblatt M. (ed.), *Athens conference on applied probability and time series analysis*, Vol.2. Springer New York 1996, pp. 145-159.

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

Engle R.F., Ng V. and Rothschild M. (1990). Asset pricing with a factor-ARCH structure: empirical estimates for treasury bills. *Journal of Econometrics*, 45, 213-237.

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

Hafner, C.M., van Dijk, D.J. and Franses, Ph.H. (2006). Semiparametric modelling of correlation dynamics, in T. Fomby and C. Hill (eds.) *Advances in Econometrics* 20, Part A (2006), 59–103.

Härdle, W., Herwartz, H. and Spokoiny, V. (2004). Time Inhomogeneous Multiple Volatility Modeling, *Journal of Financial Econometrics*, 1, 55-95.

Härdle, W. and Simar, L. (2003). *Applied multivariate statistical analysis*, Springer.

Hermann, E. (1997). Local Bandwidth Choice in Kernel Regression Estimation, *Journal of Computational and Graphical Statistics*, Vol. 6, 35-54.

Herzel, S., Stărică C, and Tütüncü, R. (2006). A non-stationary paradigm for the dynamics of multivariate financial returns. Statistics for dependent data, Springer-Verlag, *Lecture Notes in Statistics*, Vol. 187. Bertail, Patrice; Doukhan, Paul; Soulier, Philippe (Eds.).

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*, John Wiley & Sons.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*, Academic Press.

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

Ross, S. (1976). The arbitrage theory of capital asset pricing, *Journal of Finance*, 13, 341-360.

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

Stock, J., Watson, M. (2004). Forecasting with many predictors, *Handbook of Economic Forecasting*, forthcoming.

Van Bellegem, S., and von Sachs, R. (2004). Forecasting economic time series with unconditional time-varying variance. *International Journal of Forecasting*, Vol. 20, 611-627.