Economics School of Louvain - ESL

Economics School of Namur - ESN

Locally Stationary Wavelet Factor Model

Author : LAFONTAINE Guillaume

Thesis Director : VAN BELLEGEM Sébastien

Thesis Reader : HAFNER Christian

Academic Year 2021-2022

Master in Economics, Research focus– 120 credits

Université catholique de Louvain - Economics School of Louvain - Place Montesquieu 3, 1348 Louvain-La-Neuve
Université de Namur – Economics School of Namur – Rempart de la Vierge 8, 5000 Namur

Université Catholique de Louvain

Economics School of Louvain - ESL

Locally Stationary Wavelet Factor Model

Author (NOMA):
LAFONTAINE Guillaume (1197-17-00)

Thesis Director:
VAN BELLEGEM Sébastien

Thesis Reader:
HAFNER Christian

Master in Economics - Research focus - 120 credits
2021 - 2022

Acknowledgment

This second master thesis concludes 9 months of research and study of factor models and wavelet theory. Following my first thesis "*Multiscale Factor Modeling of High Dimensional Locally Stationary Timeseries*" which laid the intuitive foundations of the topic, this sequel delves deeper into the theoretical framework required for a rigorous statistical and asymptotic reasoning.

First of all, I would like to thank my thesis director, Professor Van Bellegem, for pursuing our collaboration during this second year. His helpful guidance and comments were of precious importance in the realization of this work.
I would also like to thank Professor Hafner for his proofreading.

In addition, I would like to thank my parents and my brother for their unconditional love and support during my university studies. They have been and will remain a great source of motivation.

Contents

Abstract

In this paper we propose a new factor model based on Locally Stationary Wavelet processes of Nason et al. (2000). This peculiar decomposition is well-suited for analyzing nonlinear and non-stationary time series. The time-varying factor loadings are estimated by the eigenvectors of the Evolutionary Wavelet Spectrum, the wavelet analogue of the covariance matrix. We prove convergence under general identification conditions without directly solving the rotational indeterminacy inherent to factor models. We allow both time and cross-section dimensions go to infinity and pursue a simulation study to illustrate the convergences.

1 Introduction

In current times significant quantities of data are being accumulated over a longer span and from an ever growing number of sources. Classical quantitative procedures cannot cope with this increase in dimensionality and innovative methods are called to alleviate this curse. The main drawback of this great gathering resides in the sparcification of the common information. In finite sample scenario, when the number of features about a sample becomes large, some of the information provided by an additional feature is redundant. Quantities estimations, which rely —directly or indirectly —on finding the average behavior of some class, stumble on this highly noisy data. Eventhough statistical asymptotic reasoning frames the model consistency, when the number of samples is limited, parameter sparsity reigns. One of the first to exemplify this phenomenon was (Hughes, 1968) who showed the existence of an optimal complexity —number of features —when measuring the performance of classifiers even with theoretical asymptotic support. This *curse of dimensionality* applies not only to classifiers but also to high-dimensional estimation in general. In the context of timeseries analysis, factor models have been renowned for their dimensionality reduction capabilities, thus mitigating the curse (Chamberlain & Rothschild, 1983; Forni, Halli, Lippi, & Reichlin, 2000; Stock & Watson, 2002; Bai, 2003). The essence of this method lies in distilling the relevant information contained in a plethora of timeseries into some representative indices called common factors. Traditionally, processes following such dynamics are assumed stationary but such a restriction doesn't seem appropiate for a lot of real-word applications (e.g. moneraty policy changes, political crisis effects, macroeconomic variables during recessions and expansions, neural activity monitoring . . .). Instabilities in the parameters could be modelled in several ways. For instance, parameters could exhibit structural breaks at certain points in time (Bai, Han, & Shi, 2020; Breitung & Eickmeier, 2009; Chen, Dolado, & Gonzalo, 2014), they could also evolve stochastically by following a simple random walk (Del Negro & Otrok, 2008), a VAR process (Mikkelsen, Hillebrand, & Urga, 2019) or some functional of observable variables (Connor & Linton, 2007; Fan, Liao, & Wang, 2016). Non-parametric methods, which could be seen as a limiting case of the structural break approach when a timeseries experiences changes continuously, have also been successfully applied. By assuming smooth changes in factor loadings, (Motta, Hafner, & von Sachs, 2011) propose a locally stationary factor model based on kernel estimation of the second-order dependence structure of timeseries. (Su & Wang, 2017) further extended the latter analysis by enhancing kernel estimation near the boundaries, providing parameters' asymptotic distributions and related test for structural breaks.

In this paper, analoguously to this non-parametric factor model litterature, we assume the second-order dependence of some random variables to change smoothly through time —more specifically as a Lipschitz continuous function. The latter restricts the parameters' instabilities to be locally small but their cumulative effects could be non negligible, which would discard classical PCA methods (Stock & Watson, 2002). This is realistic given that it seems probable that variables undergo fundamental changes in their time dynamics due to institutional changes, natural disasters, political turmoil ...

In line with (Forni et al., 2000) who conceived a factor model in Fourier domain, ours leverages wavelet functions to carry out non-parameteric estimations. Similarily to sinusoïds, wavelets promote sparsity of representation, hence providing a first aid to the aforementionned curse. Nonetheless, they provide more efficient time-frequency tradeoff than Fourier analysis, which naturally helps handling non-stationarities —i.e. parameters can depend on both time and frequency/scale. The time-localized autocovariance of timeseries process is modelled as a *locally stationary wavelet process* (Nason, von Sachs, & Kroisandt, 2000), which utilizes discrete non-decimated wavelets to create an evolutionary wavelet spectrum, —the wavelet analogue to the time-varying Fourier spectrum. To ease technical proofs, we focus on wavelets that are compactly supported but results could be extended to broader classes of wavelets as long as consistency with (Nason et al., 2000)'s theory holds. Other types of regularity conditions on the time-varying covariance structure could be considered as well. For instance, (Van Bellegem & von Sachs, 2008) explored discountinuities in the wavelet second-order dependence, that is, they obtain consistency of the autocovariance function when wavelet amplitudes have their total variation restricted —*versus* Lipschitz. This extension better suits the analysis of abruptly changing timeseries (e.g. stoch market prices, contemporaneous reaction to monetary policy shocks, neural activity...). Nonetheless, in this paper we stick with Lipschitz regularization, being the most commonly used —hence most easily comparable —restriction for non-parametric covariance structure estimation. Our factor model relies on the multivariate extension of the locally stationary wavelet process developped by (Park, Eckley, & Ombao, 2014).

In addition to providing an efficient way of handling non-stationarities, wavelets also allow one to obtain a scale decomposition of processes. This is appropriate when multiple driving forces are influencing a given random process at different frequency bands. For instance, the effect of macroeconomic expansions and contractions would be prominently active at lower scales while higher scales would presumably harbor contemporaneous shocks (e.g. monetary policy, stock market, supply, demand, etc). It's however more an art than a science to know which precise scales will be influenced by which shocks, since phenomenon could exhibit time-shifting frequency behaviors, such as time-accelerating business cycles. By harnesting this peculiar properties of wavelets, we impose common factors to influence only finitely many —possibly different —scales, so as to ease factors' interpretation.

The rest of this paper is organized as follows. In section 2, we recall the multivariate extension by (Park et al., 2014) of the *locally stationary wavelet process* (Nason et al., 2000). In section 3, we introduce our factor model with time-varying wavelet loadings. In section 4, we propose our estimators and develop the consistency for the estimated factors, loadings and common components. In section 5, we investigate the finite sample properties of our estimation via simulation. Finally, section 6 concludes.

Notations. For an $m \times n$ real matrix $\boldsymbol{A}$, we denote its transpose $\boldsymbol{A}'$, its Frobenius norm $\|\boldsymbol{A}\|$, its i-th largest eigenvalue $\mu_i(\boldsymbol{A})$ and an element as $\boldsymbol{A}^{(\alpha,\beta)}$. The identity and null matrix of rank N are defined as $\boldsymbol{I}_N$ and $\boldsymbol{0}_N$, respectively. The largest and smallest eigenvalues are represented by $\mu_{max}(\boldsymbol{A})$ and $\mu_{min}(\boldsymbol{A})$. Every matrices and vectors are bold.

2 Multivariate Locally Stationary Wavelet Process

To model the second-order dependence structure of multivariate high-dimensional timeseries, an efficient and flexible way of characterizing the covariance structure must be considered. This section introduces such a method based on discrete non-decimated wavelets (Nason et al., 2000;

Park et al., 2014).
Let $\boldsymbol{X}_{t;T}$ be a $(N \times 1)$ vector of stochastic processes, for all $t = 0, \dots, T-1$ and $T = 2^J$, having the representation in the mean-square sense

$$\boldsymbol{X}_{t;T} = \sum_{j=-J}^{-1} \sum_{k=0}^{T-1} \boldsymbol{W}_j\left(k/T\right) \boldsymbol{\xi}_{j,k} \psi_{j,k}(t) \tag{2.1}$$

where $\boldsymbol{\xi}_{j,k}$ is a random $(N \times 1)$ vector of orthonormal increments, $\boldsymbol{W}_j\left(k/T\right)$ is a lower triangular $(N \times N)$ matrix and $\psi_{j,k}(t)$ is a discrete non-decimated wavelet. Together, the increments and wavelet coefficients form the stochastic amplitudes —i.e. $\boldsymbol{W}_j\left(k/T\right) \boldsymbol{\xi}_{j,k}$. Note that since $T = 2^J$, due to wavelet theory, we usually restrict the maximum number of scale decomposition as $J \approx \log_2(T)$. Analoguously to (Nason et al., 2000), the object in the representation (2.1) have the following (distributional) properties.

1. $\mathrm{E}\left[\xi_{j,k}^{(u)}\right] = 0$
2. $\mathrm{Cov}\left[\xi_{j,k}^{(u)}, \xi_{j',k'}^{(u')}\right] = \delta_{j,j'}\delta_{k,k'}\delta_{u,u'}$
3. $W_j^{(u,u')}(z)$ is a lipschitz continuous function for $z \in (0,1), j < 0$ and $u, u' \in \mathbb{N}$ with the additional properties :
   (a) $\displaystyle\sum_{j=-\infty}^{-1} \sum_{u=1}^{\infty} \left|W_j^{(u,u')}(z)\right|^2 < \infty$
   (b) $\displaystyle\sum_{j=-\infty}^{-1} 2^{-j} L_j^{(u,u')} < \infty$
   (c) $\displaystyle\sum_{u=1}^{N} W_j^{(u',u)}(z) W_{j'}^{(u,u'')}(z) = \begin{cases} S_j^{(u',u'')}(z) & \text{if } j = j' \\ 0 & \text{otherwise} \end{cases}$

The wavelet functions applicable to our factor model must be in L^1 such that all compactly supported wavelets satisfy this condition. Note that, this assumption excludes the *Shannon* wavelet from the realm of the possibilities. Yet, the latter function still allows a decomposition of the form (2.1) (see Nason et al., 2000).

Assumption 1 forces the MLSW to have zero-mean which is without loss of generality. One of the most important assumption concerning the random increments lies in their orthogonality (2) —or in the presence case, their orthonormality —such that the dependence structure of the processes is solely encapsulated in the wavelet coefficients matrix, $\boldsymbol{W}_j(z)$. The latter condition is mostly for technical convenience since first, (Sanderson, Fryzlewicz, & Jones, 2010) considered the case where increments are correlated and, second, we will be imposing a factor struture on wavelet coefficient matrix and increments jointly, such that the random amplitudes will be represented as $\boldsymbol{W}_j\left(k/T\right) \boldsymbol{\xi}_{j,k} = \boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k + \boldsymbol{\varepsilon}_{j,k}$.

To quantify the local power at a given scale and location, (Nason et al., 2000; Park et al., 2014) defined the —*Cross —Evolutionary Wavelet Spectrum* (CEWS) as $S_j^{(u,u)}(z) = \left|W_j^{(u,u)}(z)\right|^2$ — and $\boldsymbol{S}_j(z) = \boldsymbol{W}_j(z)\boldsymbol{W}_j(z)'$. This alternative measure parallels a localized cross-covariance in wavelet domain since $c^{(u,u')}(z,\tau) := \sum_{j=-\infty}^{-1} S_j^{(u,u')}(z)\Psi_j(\tau)$, where $\Psi_j(\tau) := \sum_k \psi_{j,k}(0)\psi_{j,k}(\tau)$ defines the autocorrelation wavelet. The mean-square convergence criterion assumed for the process $\boldsymbol{X}_{t;T}$ calls for the restriction that $c^{(u,u')}(z,0) := \sum_{j=-\infty}^{-1} S_j^{(u,u')}(z) < \infty$, $\forall z \in (0,1)$ and $u, u' \in \mathbb{N}$, which is ensured given the finite energy decomposition assumption (3a).

This measure also directly inherits the continuity regularization from the wavelet coefficients, such that smoothly varying second-order processes are naturally characterize by this spectral representation. In practical application, eventhough Lipschitz continuous non-stationarities suits well in normal times, when processes experience shocks, more erractic behaviors of dependences may arise. (Van Bellegem & von Sachs, 2008) permit an easier treatment of jump processes by constraining their total variations.

The lower triangular form of $\boldsymbol{W}_j(z)$ restricts the directionality of the dependencies between the considered processes which eases the interpretation of the structure. One question naturally arises about the potential variability of the spectral representation under permutations. Surprisingly enough, the spectral properties are order-invariant (see Park et al., 2014, proposition 2.2.5). In addition, this matrix restriction favors the use of the Cholesky decomposition to recover the wavelet coefficients matrix from the CEWS matrix. Assumption (3c) discards the possibility to have spectral power between processes at different scales of decomposition, which complements the assumption about the orthonormality of increments. This is restrictive and it should be addressed to pursue any empirical analysis.

Remark. In the non-stationary framework, classical asymptotic reasoning looses its meaning since it becomes impossible to gather increasingly more information of the same kind as the sample size increases. As noted by (Dahlhaus, 1997), "extending the process into the future, will not give any information on the behavior of the process at the beginning of the time interval". He formally develops the concept of local stationarity to alleviate this issue, such that time is rescaled : $z = t/T \in (0,1)$ and, thus information is accumulated on a finer mesh as $T \to \infty$. Consequently he asymptotically deciphers the "stationary" local behavior of the non-stationary object perfectly. (Nason et al., 2000) adopt this abstract framework for constructing their wavelet process representation on which our factor model relies.

Remark. The *non-decimated* wavelet family doesn't form a basis for the $\mathcal{L}^2$ space since those types of wavelets are not orthogonal to each other for differing scales and locations. Consequently, raw wavelet coefficients $\boldsymbol{d}_{j,k}$ are not uniquely identified by the projection of the signals $\boldsymbol{X}_{t;T}$ onto the wavelet $\psi_{j,k}(t)$. Neighbouring coefficients could thus —possibly —contain the same information about the original processes. Nonetheless, (Nason et al., 2000), and (Park et al., 2014), managed to provide a unique representation of the second-order dependence structure for univariate —respectively multivariate —timeseries. This lack of identifiability will have implications for our common factors but not on loadings which will be uniquely defined.

3 Wavelet Factor Model

Let $\{\boldsymbol{X}_{t;T}\}_{t=0,\dots,T-1}$ be T realizations of the N-dimensional timeseries process defined in (2.1). We assume that $\boldsymbol{X}_{t;T}$ admits the following wavelet factor model with K common factors :

$$\boldsymbol{X}_{t;T} = \sum_{j=-J}^{-1} \sum_{k=0}^{T-1} \left(\boldsymbol{\Lambda}_j(\frac{k}{T}) \boldsymbol{F}_k + \boldsymbol{\varepsilon}_{j,k} \right) \psi_{j,k}(t)$$

which is equivalent to imposing a factor structure on the random wavelet amplitudes

$$\boldsymbol{W}_j\left(k/T\right) \boldsymbol{\xi}_{j,k} = \boldsymbol{\Lambda}_j(\frac{k}{T}) \boldsymbol{F}_k + \boldsymbol{\varepsilon}_{j,k} \tag{3.1}$$

where $\mathbf{\Lambda}_j(\frac{k}{T})$ is a $(N \times K)$ matrix of loadings, $\boldsymbol{F}_k$ is a $(K \times 1)$ vector of common factors and $\boldsymbol{\varepsilon}_{j,k}$ is a $(N \times N)$ matrix of idiosyncratic errors, which are assumed independent of the common factors.

The major benefit of assuming such a decomposition on random wavelet amplitudes lies in a sparser and more interpretable charaterization of the processes variability, such that the wavelet spectrum —i.e. CEWS defined in section 2 —can be represented as

$$\boldsymbol{S}_j(z) = \mathbf{\Lambda}_j(z)\mathbf{\Sigma}_F\mathbf{\Lambda}_j(z) + \mathbf{\Sigma}_{j;\varepsilon}(z) \tag{3.2}$$

which depends on rescaled time $z \in (0, 1)$ and where $\mathbf{\Lambda}_j(z)\mathbf{\Sigma}_F\mathbf{\Lambda}_j(z)$ is the second-order dependence structure explained by the factors and $\mathbf{\Sigma}_{j;\varepsilon}(k/T)$ is due to idiosyncratic noise. In the framework of high-dimensional factor modelling, when the number of cross-sections increases, the first K eigenvalues of the CEWS matrix grow without bounds and are associated with the first term of the LHS. However, all eigenvalues of the idiosyncratic matrix are always bounded since, it's believed that information common to all timeseries doesn't accumulate in the errors. Note that we implicitly use the increments' orthonormality —assumption 2 of section 2 —to derive the wavelet spectral representation (3.2)[1].

As in conventional factor model, the loadings and factors are not separately identifiable. This inherent indeterminacy can be exposed by choosing a $K \times K$ invertible matrix $\boldsymbol{R}_{j,k}$ such that $\mathbf{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k = (\mathbf{\Lambda}_j(\frac{k}{T})\boldsymbol{R}_{j,k}^{-1})(\boldsymbol{R}_{j,k}\boldsymbol{F}_k)$ and thus $\mathbf{\Lambda}_j(\frac{k}{T})^* = \mathbf{\Lambda}_j(\frac{k}{T})\boldsymbol{R}_{j,k}^{-1}$ and $\boldsymbol{F}_k^* = \boldsymbol{R}_{j,k}\boldsymbol{F}_k$ span the same factor space. To pinpoint unique —upto a sign change —parameters, we need K^2 restrictions on the rotation matrix $\boldsymbol{R}_{j,k}$. (Bai & Ng, 2008) impose $\frac{1}{T}\sum_k \boldsymbol{F}_k\boldsymbol{F}_k' = I_K$ and $\mathbf{\Lambda}'\mathbf{\Lambda}$ to be a diagonal matrix. They show that they could have equivalently chosen $\mathbf{\Lambda}'\mathbf{\Lambda}/N = I_K$ and $\frac{1}{T}\sum_k \boldsymbol{F}_k\boldsymbol{F}_k'$ diagonal as valid restrictions. The relevant choice of identification conditions hangs on the particular empirical situation, such that, analoguous to (Motta et al., 2011), we build our factor model on a general invertible rotation matrix which combines the latter pairs of restrictions.

3.1 Assumptions of the WFM

In this subsection, we list and discuss the assumptions on which our factor model relies. The asymptotic reasoning is underlined by T and N tending to infinity, if not jointly, at least over some of their paths. More specifically, the number of timeseries observations will first grow without bounds, followed by the number of cross-sections. This constrainted asymptotic framework, as exposed in subsequent theorems, is mainly inherited by the behavior of the CEWS in high-dimensional settings, which if enhanced, would make our factor model more asymptotically robust.

3.1.1 Assumptions F - Factors

F1. $\boldsymbol{F}_k \sim (\mathbf{0}, \mathbf{\Sigma}_F), \forall k$, where $\mathbf{\Sigma}_F$ is a diagonal positive definite $(K \times K)$ matrix.

[1] By taking the variance-covariance matrix on both sides of (3.1) and noting the independence between the factors and idiosyncratic errors and assumption 2 of section 2, we get

$$\mathrm{Var}\left[\boldsymbol{W}_j\,(k/T)\,\boldsymbol{\xi}_{j,k}\right] = \mathrm{Var}\left[\mathbf{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k\right] + \mathrm{Var}\left[\boldsymbol{\varepsilon}_{j,k}\right]$$

$$\boldsymbol{W}_j\,(k/T)\,\mathrm{Var}\left[\boldsymbol{\xi}_{j,k}\right]\boldsymbol{W}_j\,(k/T)' = \mathbf{\Lambda}_j(\frac{k}{T})\mathbf{\Sigma}_F\mathbf{\Lambda}_j(\frac{k}{T})' + \mathbf{\Sigma}_{j;\varepsilon}(k/T)$$

F2. $\sup_u \mathrm{E}\left[{F_k^{(u)}}^4\right] \leq M_1 < \infty$

F3. $\frac{1}{T}\sum_{k=0}^{T-1} \boldsymbol{F}_k \boldsymbol{F}_k' \xrightarrow{p} \boldsymbol{\Sigma}_F$

F4. $\lim_{i\to\infty} \mathrm{Cov}\left[\boldsymbol{F}_{k_0}\boldsymbol{F}_{k_1}', \boldsymbol{F}_{k_2+i}\boldsymbol{F}_{k_3+i}'\right] = \boldsymbol{0}_K \quad \forall k_0, k_1, k_2, k_3$

3.1.2 Assumptions L - Factor Loadings

L1. $\left\|\frac{\boldsymbol{\Lambda}_j(z)'\boldsymbol{\Lambda}_j(z)}{N} - \boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)\right\| \to 0$

L2. $\sup_k \sum_{j=-\infty}^{-1} \left\|\boldsymbol{\lambda}_j^{(\alpha)}(\frac{k}{T})\right\| \leq \bar{\lambda} < \infty$, where $\boldsymbol{\lambda}_j^{(\alpha)}(\frac{k}{T})$ is the α-th row of the non-random matrix $\boldsymbol{\Lambda}_j(\frac{k}{T})$.

3.1.3 Assumptions E - Idiosyncratic errors

E1. $\boldsymbol{\varepsilon}_{j,k} \sim (\boldsymbol{0}, \boldsymbol{\Sigma}_{j,k;\varepsilon})$, where $\boldsymbol{\Sigma}_{j,k;\varepsilon}$ is a positive definite $(N\times N)$ matrix for every j, k.

E2. $\sup_u \mathrm{E}\left[{\varepsilon_{j,k}^{(u)}}^4\right] < \infty \quad \forall j, k$

E3. $\mathrm{E}\left[\varepsilon_{j,k}^{(u)}\varepsilon_{j,k}^{(v)}\right] = \sigma_{j,k}^{(u,v)}$ such that $\sup_u \sum_{v=1}^{N} \left|\sigma_{j,k}^{(u,v)}\right| \leq \bar{\sigma} < \infty$ and $\forall N \in \mathbb{N}, \forall j, k$

E4. $\sup_u \sum_{j=-\infty}^{-1} \left|\sigma_{j,k}^{(u,v)}\right| \leq M_2 < \infty$

E5. $F_k^{(u)} \perp \varepsilon_{j,k'}^{(v)}, \forall j, k, k', u, v$

E6. $\lim_{i\to\infty} \mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}_{l,m+i}'\right] = \boldsymbol{0}_N \quad \forall j, l, k, m$

E7. $\lim_{i\to\infty} \mathrm{Cov}\left[\boldsymbol{F}_{k_0}\boldsymbol{\varepsilon}_{j,k_1}', \boldsymbol{F}_{m_0+i}\boldsymbol{\varepsilon}_{l,m_1+i}'\right] = \boldsymbol{0}_{K\times N} \quad \forall k_0, k_1, m_0, m_1, j, l$

Assumption F1. states that factors are zero mean and cross-sectionnaly uncorrelated —i.e. $\mathrm{Cov}\left[F_k^{(u)}, F_k^{(u')}\right] = 0, \forall k, u \neq u'$. Note that we don't require factors to be *iid* as in (Motta et al., 2011). Assumption F2. is widely used in the literature and restricts their fourth moments. The next general assumption underpins the existence of a consistent estimator for the factors' covariance matrix, such that, in practical application further conditions must be provided for convergence. Assumption F4. is more general than (Su & Wang, 2017) and further restricts the fourth moments but only asymptotically.

Assumption L1., as in the standard litterature, highlights the intrinsic nature of the common factors' impacts on each timeseries. Namely, their existence is supported if they exhibit a non-negligible contribution to the variances of $\boldsymbol{X}_{t;T}$. Moreover, notice that we don't restict the inter-scale contributions, which allows "spectral" dependences between high and low frequencies of processes, such that potential exogeneous shocks to high-frequency phenomenon could influence slow-moving variables, and conversely. This restriction could be further strenghened to obtain an accurate factor structure representation of MLSW processes (see Appendix).

Assumption L2. is analoguous to B2 of (Motta et al., 2011) with the further restriction on wavelet scales. This is not restrictive as it follows from assumption 3a (section 2) and wavelet theory underlines the exponential decay of wavelet coefficients with respect to scale anyway (Daubechies, 1992). This last assumption on the loadings staightforwardly entails $\sup_z \sum_{j=-\infty}^{-1} \|\boldsymbol{\Lambda}_j(z)\| = O(\sqrt{N})$, which is the wavelet equivalent loadings growth condition as in (Motta et al., 2011), and also that $\|\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)\| = o\left(\frac{1}{|j|^2}\right)$ since $\left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\right\| = o\left(\frac{1}{|j|}\right)$, such that each factor only contributes to finitely many scales, this is again not restrictive.

Assumptions E governs the behavior of the idiosyncratic error terms and their interactions with the common factors. Those errors are zero-mean and have a scale- and time-dependent covariance matrix. In the high-dimensional setting, constraining the latter covariance matrix to be diagonal is too stringent, as it's increasingly probable that exogeneous shocks influence multiple timeseries as the number of cross-section elements grows. Therefore, we ground our factor model in the *approximate* framework of (Chamberlain & Rothschild, 1983) by constraining the eigenvalues of the latter error covariance matrix for every j, k (assumption E3.) —i.e. only *weak* cross-section dependence is allowed. Assumption E2. restricts their fourth moments. Analoguously to assumption L2. on the loadings, assumption E4. is inherited from the (Nason et al., 2000)'s processes characterization in L^2 space —i.e. assumption 3a.

As in conventional —and dynamic —factor model, the common factors and idiosyncratic errors are orthogonal to each other (assumption E5.). This condition, as stringent as it is, becomes essential for identification of the structure (Forni & Lippi, 2001). Assumption E6. limits the degree of serial dependencies by requiring the cross-covariance matrix to vanish, which is more general than the usual absolute summability in —time-varying —models (Motta et al., 2011) (Su & Wang, 2017) (Bai et al., 2020). Once one remembers the compactness of the wavelet's support and their decorrelation abilities in analyzing long-term processes (Dijkerman & Mazumdar, 1994), this latter weak restriction doesn't seem too quixotic. Finally, assumption E7., which is mainly technical, asymptotically conditions the fourth order dependence between factors and errors.

4 Estimation

In this section, we introduce the estimation procedures for the wavelet spectral representation of processes as well as the estimators of the time-varying loadings and common factors. As it will be shown, the most important aspect of obtaining a precise estimation of the factor structure lies in efficient estimation of the second-order structure.

4.1 Cross-Evolutionary Wavelet Spectrum

To estimate the CEWS exposed in section 2, define the raw wavelet coefficients as well as the raw wavelet periodogram,

$$\boldsymbol{d}_{j,k} = \sum_{t=0}^{T-1} \boldsymbol{X}_{t;T} \psi_{j,k}(t) \tag{4.1}$$

$$\boldsymbol{I}_{j,k} = \boldsymbol{d}_{j,k} \boldsymbol{d}'_{j,k} \tag{4.2}$$

As straightforward as this spectral estimator is, we cannot directly rely on the latter since there is non-negligible bias coupled with a non-vanishing variance (see Park et al., 2014). The consistent bias-corrected spectrum estimator is thus given by

$$\widehat{\boldsymbol{S}}_j(k/T) = \frac{1}{2M+1}\sum_{s=-M}^{M}\sum_{l=-J}^{-1} A_{j,l}^{-1}\boldsymbol{I}_{j,k+s} \tag{4.3}$$

where $A_{j,l} := \sum_\tau \Psi_j(\tau)\Psi_l(\tau)$ is the inner product matrix of discrete autocorrelation wavelets which is invertible for every Daubechies compactly supported wavelets (Nason et al., 2000). This estimator is asymptotically unbiaised and consistent under the factor structure decomposition (see Theorem 1 later). This estimator is further analysed by (Park et al., 2014), as well as its univariate version by (Nason et al., 2000).
As mentionned in section 2, under the invertibility assumption of the inner product matrix $A_{j,l}$, there is an equivalence relation between the cross-evolutionary wavelet spectrum and the cross-covariance between two processes,

$$c^{(u,u')}(z,\tau) = \sum_{j=-\infty}^{-1} S_j^{(u,u')}(z)\Psi_j(\tau) \qquad \Longleftrightarrow \qquad S_j^{(u,u')}(z) = \sum_{l=-\infty}^{-1} A_{j,l}^{-1}\sum_\tau c^{(u,u')}(z,\tau)\Psi_j(\tau) \tag{4.4}$$

such that, we can supply an estimator of the cross-covariance, $\widehat{c}^{(u,u')}(z,\tau)$, by replacing $S_j^{(u,u')}(z)$ with its estimate $\widehat{S}_j^{(u,u')}(k)$, and *vice versa*. This representation characterizes much more than just the dependences between contemporaneous processes. Due to (4.4), the factor structure also encompasses dependences at different times, since $\Psi_j(\tau)$ is solely responsible to account for the time lag.

4.2 Wavelet Factor Model

To estimate the loadings $\boldsymbol{\Lambda}_j(\frac{k}{T})$ and factors $\boldsymbol{F}_k$, we consider the following non-linear least squares optimization problem,

$$\begin{aligned}\min_{\{\bar{\boldsymbol{\Lambda}}_{j,k}\}_{\forall j,k},\{\bar{\boldsymbol{F}}_k\}_{\forall j,k}} \quad & (NT)^{-1}\sum_t\left[\boldsymbol{X}_{t;T} - \sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left(\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k\right)\psi_{j,k}(t)\right]'\left[\boldsymbol{X}_{t;T} - \sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left(\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k\right)\psi_{j,k}(t)\right]\\ \text{s.t.} \quad & \frac{2^{|j|}\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N} = \boldsymbol{I}_K\end{aligned} \tag{4.5}$$

The solution to this minimization problem is only approximate[2] and can be obtained via PCA applied to the CEWS matrix for every k. Taking the First Order Condition with respect to $\bar{\boldsymbol{F}}_k$ for every k and using the constraint, we get

$$\bar{\boldsymbol{F}}_k = \frac{1}{N}\frac{2^J}{2^J-1}\sum_{J=-J}^{-1}\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k} \tag{4.6}$$

Substituting this solution in the objective function yields,

$$\min_{\{\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})\}_{\forall j,k}} (NT)^{-1}tr\left\{\left[\sum_t \boldsymbol{X}_{t;T}'\boldsymbol{X}_{t;T} - \frac{1}{N}\frac{2^J}{2^J-1}\sum_{j=-J}^{-1}\sum_{l=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}_{j,k}'\bar{\boldsymbol{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\right]\right\}$$

[2] see Appendix for the derivation.

which is equivalent to maximizing the second term only and by properties of the trace operator,

$$\max_{\{\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\}\forall j,k} \frac{1}{N^2T}\frac{2^J}{2^J-1}\sum_{j=-J}^{-1}\sum_{l=-J}^{-1}\sum_{k=0}^{T-1} tr\left\{\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\boldsymbol{d}'_{j,k}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\right\} \tag{4.7}$$

Finally, replace the raw wavelet *cross-periodogram* —i.e. $\boldsymbol{I}_{j,l,k} = \boldsymbol{d}_{j,k}\boldsymbol{d}_{l,k}$ —by its consistent estimator (Koch, 2015) and due to assumption (3c), maximizing (4.7) is asymptotically equivalent to maximizing

$$\frac{1}{N^2T}\frac{2^J}{2^J-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1} tr\left\{\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\widehat{\boldsymbol{S}}_j(k/T)\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\right\}$$

which attains its optimal value when all terms are maximized,

$$\max_{\{\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\}\forall j,k} N^{-2}tr\left\{\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\widehat{\boldsymbol{S}}_j(k/T)\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\right\}$$

The conventional PCA problem yields the solution $\widehat{\mathbf{\Lambda}}_j(\frac{k}{T})$ as $\sqrt{2^{-|j|}N}$ times the K largest eigenvectors of the CEWS estimator $\widehat{\boldsymbol{S}}_j(k/T)$ stacked columnwise for each j,k. Define $\widehat{\boldsymbol{V}}_j(z)$ as the matrix with the K largest eigenvalues as its diagonal elements, then the estimation procedure can be summarized in the following pair of equations,

$$N^{-1}\widehat{\boldsymbol{S}}_j(z)\widehat{\mathbf{\Lambda}}_j(z) = \widehat{\mathbf{\Lambda}}_j(z)\widehat{\boldsymbol{V}}_j(z), \qquad \frac{2^{|j|}\widehat{\mathbf{\Lambda}}_j(z)'\widehat{\mathbf{\Lambda}}_j(z)}{N} = I_K, \qquad \forall j,z \tag{4.8}$$

Our estimator of the common factor substitues the estimated loadings in (4.8) into (4.6) to obtain,

$$\widehat{\boldsymbol{F}_k} = \frac{1}{N}\frac{2^J}{2^J-1}\sum_{J=-J}^{-1}\widehat{\mathbf{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k} \tag{4.9}$$

Remark. As already mentionned in section 2, the raw wavelet coefficients $\boldsymbol{d}_{j,k}$ suffer from a lack of identifiability which impedes their interpretability. Consequently, our estimated factors and common components inherit this curse, such that, as a practical tool, they may produce misleading and unreliable conclusions. Notice, however, that our time-varying factor loadings are well-defined and allow a rich and flexible multi-frequency analysis of the dependence structure between the studied processes. To solve this indeterminacy, as any practitionner would feel the necessity, requires working with identified objects. In our case, the time-varying autocovariance —as exposed in (4.4) —presents itself as a natural candidate. One possible solution, that we won't explore in this paper, consists in applying the analoguous principal component decomposition to this measure. However, since it's not possible to improve on every fronts, we will end up trading the clear identifiability for less expressive frequency representation of the dependence structure. Namely, making loadings scale-dependent will become impossible and as a consequence, the factor structure will lose its ability to express the entire time-varying cross-covariance function. But more importantly, loadings and errors' covariances will have to see their time-dependence behavior restricted to be lipschitz, which is currently not the case.

Remark. Embedded in a semiparametric framework, such that mean objects are prevailently estimated through local smoothing, our factor model estimates suffer from the classical boundary bias problem. Said differently, each raw wavelet coefficient $\boldsymbol{d}_{j,k}$ constitutes an average of past and future observations weighted by the chosen wavelet, which by being compactly supported, entails the *local* smoothing property. The first and last estimates of the sampling range cannot, however, be defined properly since no data is available before and after the observation period. Aware and following the nonparametric litterature on this issue, (Su & Wang, 2017) adopted a *boundary kernel* that conserves general kernels' properties where classical methods fail by rescaling the kernel operator near the boundaries. Nonetheless, as attractive and simple as this correction seems, (Nason et al., 2000)'s theory —and also wavelet's theory —doesn't permit us the use of such a rescaling scheme. Ignoring the initial starting position problem, one solution makes use of smoothing filters that depend only on past information. From those *causal* filters, *causal* wavelets forming a complete basis were developed by (Szu, 1992) who, for instance, found an analytical representation for the causal Morlet wavelet. The extension of LSW representation of stochastic processes is yet to be proven for such basis functions —i.e. by proving the invertibility of inner product matrix of discrete autocorrelation wavelets $A_{j,l}$. Enhanced (right-) boundary bias and general forecasting exercices would then be made easier.

5 Asymptotic theory

This section estabilishes the convergence in probability of the various estimators defined in the preceding section. Overall the estimation precision is constrainted by our ability to efficiently estimate the time-varying covariance structure between LSW processes. The latter, called *Cross-Evolutionary Wavelet Spectrum* (CEWS), doesn't behave ideally under a high-dimensional framework, such that the growth of the cross-section dimension will be restricted by $O(T^\nu)$ for some $0 \leq \nu < 1$. Consequently, our asymptotic results only hold for some particular trajectories of (N, T). This contrasts with the classical factor models where the cross-section and time dimensions are unrestricted in their asymptotic behavior, and thus allows situations where N is much larger than T —i.e. $N \gg T$.

Theorem 1. *Under the MLSW framework and assumptions LFE,*

$$N^{-1} \left\| \widehat{\boldsymbol{S}}_j(k/T) - \boldsymbol{S}_j\left(k/T\right) \right\| = O_p\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}} \right)$$

where $\widehat{\boldsymbol{S}}_j(k/T)$ *is defined in* (4.3) *and* $\boldsymbol{S}_j(z) = \boldsymbol{\Lambda}_j(z)\boldsymbol{\Sigma}_F\boldsymbol{\Lambda}_j(z) + \boldsymbol{\Sigma}_{j;\varepsilon}(k/T)$*.*
If one further assumes $N = O(T^\nu)$ *and* $M = O(T^\eta)$ *where* $0 \leq \nu < \eta < 1$*, then the optimal growth of the smoothing parameter is given by* $\eta^* = \frac{1}{3}(\nu - \frac{|j|+1}{\log_2 T} + 2)$ *and thus,*

$$N^{-1} \left\| \widehat{\boldsymbol{S}}_j(k/T) - \boldsymbol{S}_j\left(k/T\right) \right\| = O_p\left(\frac{1}{2^{\frac{|j|}{3}} T^{\frac{1}{3}(1-\nu)}} \right)$$

Proof. See Appendix. □

Given the factor structure, and more predominantly its implied limitations, the CEWS estimator converges to the true time-varying wavelet spectrum —this was already known independently from the factor structure thanks to (Nason et al., 2000) and (Park et al., 2014). As seen from the first part of the result, there exists a biais-variance trade-off which could be resolved by choosing the MSE-minimizing smoothing parameter M, which is computed in the second part. The latter emphasizes the cross-section growth restriction on the average estimation error. As it will be exposed in the simulation results, efficiently smoothing the estimator lies at the heart of a precise estimation of the factor structure. The convergence is also scale-dependent such that, lower wavelet scales of decomposition have faster rates. The worst convergence is achieved when $|j| = 1$. In that case, consistency is preserved but errors are of the order of $O_p\left(T^{\frac{1}{3}(\nu-1)}\right)$. Our estimator isn't guaranteed to be positive definite and issues in practical applications may arise. We may use regularization techniques to ensure this behavior by following (Rothman, 2012), yet we only consider this problem as a side note, since its relevance lies in the realm of efficient CEWS estimation and not directly on factor modelling.

Assumption S - Eigenvalues of $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)\boldsymbol{\Sigma}_F$ are distinct for every $-j \in \mathbb{N}$ and $z \in (0,1)$.

Assumption S, which ensures the uniqueness of the limiting rotated factors and loadings, is standard in factor analysis. (Motta et al., 2011) use a similar assumption but, instead of constraining the eigenvalues to be distinct, they leverage their time-varying aspects by forbidding their derivatives to be equal at any time $z \in (0,1)$. In our case, the lipschitz continuity assumption concerning the CEWS doesn't permit us to assume such a smoothness restriction. However, we believe it may be possible to generalize (Motta et al., 2011)'s condition to this case since lipschitz functions are almost everywhere differentiable. We nonetheless use the stricter assumption S. Note that this assumption is only relevant for identification of the loadings, and factors, since the common components —i.e. $\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k$ —are identified.

The next theorem shows the weak consistency of the rotation matrix that underlies the inherent indeterminacy of factor models.

Theorem 2. *Under the MLSW framework and assumptions LFE-S, define*

$$\begin{aligned}\widehat{\boldsymbol{R}}_{j,k} &= \widehat{\boldsymbol{\Sigma}}_F \frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{2^{\frac{|j|}{2}}N}\widehat{\boldsymbol{V}}_j^{-1}(k/T)\\ \boldsymbol{R}_{j,k} &= 2^{-\frac{|j|}{2}}\boldsymbol{\Sigma}_F^{\frac{1}{2}}\Omega_{j,k}\boldsymbol{V}_j(k/T)^{-\frac{1}{2}}\end{aligned}$$

where $\Omega_{j,k}$ contains the orthonormal eigenvectors of $\boldsymbol{\Sigma}_F\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)$.
If $N = O(T^\nu)$ and $M = O(T^\eta)$ where $0 \leq \nu < \eta < 1$ and $\eta = \frac{1}{3}(\nu - \frac{|j|+1}{\log_2 T} + 2)$ then,

$$\left\|\widehat{\boldsymbol{R}}_{j,k} - \boldsymbol{R}_{j,k}\right\| = O_p\left(\frac{1}{2^{\frac{|j|}{3}}T^{\frac{1}{3}(1-\nu)}}\right)$$

Proof. See Appendix. □

The relevance of assumption S can clearly be seen in the limiting rotation matrix through the matrix of eigenvectors, which are unique if their corresponding eigenvalues are distinct. From its definition, $\boldsymbol{R}_{j,k}$ is invertible such as to correctly characterize the consistency of estimated loadings and factors. In order to stick to the Hilbert space, assumption 3a constrains the series of matrices such that $\sum_{j=-\infty}^{-1} 2^{-\frac{|j|}{2}}\left\|\boldsymbol{R}_{j,k}^{-1}\right\| < \infty$ since $\|\boldsymbol{R}_{j,k}\| = o\left(\frac{|j|}{2^{\frac{|j|}{2}}}\right)$. On the one hand, the dependence on scale of this rotation matrix hinders the estimation of the common factor which are defined on the usual time-domain. However due to the non-identifiability of the estimated factors, stemming from the raw wavelet coefficients, we made the conscious choice to impose further restrictions on the factors' estimates rather than on the well-defined loadings'. On the other hand, the time-varying orthogonal matrix renders the regression forecasting exercice with factor models more challenging. Yet, (Su & Wang, 2017) showed that, with time-varying parameters, this issue is alleviated.

The next two theorems constitute our main results concerning the factor structure defined in wavelet domain.

Theorem 3. *Under the MLSW framework, assumptions LFE-S and if* $N = O(T^{\nu})$*,* $M = O(T^{\eta})$ *where* $0 \leq \nu < \eta < 1$ *and* $\eta = \frac{1}{3}(\nu - \frac{|j|+1}{\log_2 T} + 2)$ *then,*

$$\frac{2^{\frac{|j|}{2}}}{\sqrt{N}} \left\| \widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T}) - \boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{R}_{j,k} \right\| = o\left(|j|^2\right) O_p\left(\frac{1}{2^{\frac{4|j|}{3}} T^{\frac{1}{3}(1-\nu)}}\right)$$

Proof. See Appendix. □

The estimated loadings are asymptotically unbiaised and consistent for the true loadings, up-to an orthogonal transformation. The norm is scaled by a factor $\frac{1}{\sqrt{N}}$ to match the growth of the matrix norm, hence this convergence holds only on average. Note that uniform consistency results could be straightforwardly obtained by following (Su & Wang, 2017). The factor $2^{\frac{|j|}{2}}$ rescales the norm which organically vanishes due to the constraint (4.5) and the definiton of $\boldsymbol{R}_{j,k}$.

Common factor consistency is achieved through additional restrictions.

Assumption D1 - Linear combinaison of loadings and errors. The dependence between loadings and errors is restricted such that

$$\left\| \frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\varepsilon}_{j',k'}}{\sqrt{N}} \right\| = o\left(\frac{1}{|j|\sqrt{|j'|}}\right) \qquad \forall j, k$$

Assumption D2 - Linear combinaison of asynchronous loadings. The dependence between loadings existing at different scale and time is restricted such that

$$\left\| \frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_{j'}(\frac{k'}{T})}{\sqrt{N}} \right\| = o\left(\frac{1}{|j|\,|j'|}\right) \qquad \forall j \neq j', k \neq k'$$

The first assumption, a lighter version of a similar version of (Bai, 2003) and (Su & Wang, 2017), is classical in factor analysis. The less-trivial second assumption is specific to our time-varying wavelet model and, as will be exposed later, goes in the direction of *exact* estimators for the factor structure (see Appendix A.1). As usual, the loadings' absolute summability with respect to scale is inherited from the Hilbert space.

Theorem 4. *Under the MLSW framework, assumptions LFE-S-D1-D2 and if* $N = O(T^{\nu})$*,* $M = O(T^{\eta})$ *where* $0 \leq \nu < \eta < 1$ *and* $\eta = \frac{1}{3}(\nu - \frac{|j|+1}{\log_2 T} + 2)$ *then,*

$$\left\| \widehat{\boldsymbol{F}}_k - 2^{-\frac{|j|}{2}} \boldsymbol{R}_{j,k}^{-1} \boldsymbol{F}_k \right\| = o\left(|j|\right) O_p\left(\frac{1}{2^{\frac{4|j|}{3}} T^{\frac{1}{3}(1-\nu)}}\right)$$

Proof. See Appendix. □

Our estimator is asymptotically consistent but unbiaised up-to a rescaling —and rotation — of the true factors. Yet, as noted earlier, the rotation matrix $\boldsymbol{R}_{j,k}$, whose norm decreases with scale, makes the direct correspondance between estimated and true factors arduous.

Theorem 5. *Under the MLSW framework, assumptions LFE-S-D1-D2 and if* $N = O(T^{\nu})$*,* $M = O(T^{\eta})$ *where* $0 \leq \nu < \eta < 1$ *and* $\eta = \frac{1}{3}(\nu - \frac{|j|+1}{\log_2 T} + 2)$ *then,*

$$\frac{2^{\frac{|j|}{2}}}{\sqrt{N}} \left\| \widehat{\boldsymbol{C}}_{j,k} - 2^{-\frac{|j|}{2}} \boldsymbol{C}_{j,k} \right\| = o\left(|j|\right) O_p\left(\frac{1}{2^{\frac{4|j|}{3}} T^{\frac{1}{3}(1-\nu)}} \right)$$

Proof. See Appendix. □

As anticipated by the two previous theorems, the estimated common components converge in probability to the true ones. In contrast to the factors and their loadings, the common components don't suffer from a lack of rotational identifiability and are thus well defined.

6 Simulation Study

In this section we consider simulations to illustrate the convergence of our estimators and their finite sample properties. The simulations are based on the assumptions that $K = 2$, $N = 15$, $T = 1024$ and $M = 256$ along with the Haar wavelet function.

This example replicates the first simulation of (Motta et al., 2011), which assumed the loadings to satisfy, $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z) = \boldsymbol{I}_K$, such that the rotational indeterminacy becomes $\boldsymbol{R}_{j,k} = \pm \boldsymbol{I}_K$. We extend the latter assumption to the wavelet domain by assuming $\left\| \frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}} \right\| = O\left(\frac{1}{|j|^2}\right)$, such as to ensure the summability $\forall z \in (0,1) : \sum_{j=-\infty}^{-1} \left\| \frac{\boldsymbol{\Lambda}_j(z)}{\sqrt{N}} \right\| = O(1)$. Due to Proposition 3 of (Nason et al., 2000), the latter restriction doesn't imply absolute summability of the underlying autocovariance function of the process. Each factor contributions is thus $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z) = |j|^{-4} \boldsymbol{I}_K$, hence $\boldsymbol{R}_{j,k} = \pm |j|^2 2^{\frac{|j|}{2}} \boldsymbol{I}_K$. To satisfy this condition, we take

$$\boldsymbol{\Lambda}_j(\frac{k}{T}) = |j|^{-2} \sqrt{N} e^{\frac{k}{T}\boldsymbol{A}_N} \boldsymbol{O}_{j;N}$$

where $\boldsymbol{A}_N$ is an $N \times N$ antisymmetric matrix[3] and the matrix exponentiation is defined as $e^{\boldsymbol{X}} = \sum_{n=0}^{\infty} \frac{\boldsymbol{X}^n}{n!}$, and not simply by taking the exponential of each (i,j)-element of the matrix, as in (Motta et al., 2011). Each entry of $\boldsymbol{A}_N$ fulfills

$$a_{ij} = \begin{cases} 1 & \text{if } i < j \\ 0 & \text{if } i = j \\ -1 & \text{if } i > j \end{cases}$$

The orthonormal matrix $\boldsymbol{O}_j$ is random, such that some variation in the loadings construction process exists. This formulation implies the desired property —i.e. $\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T}) = |j|^{-2} N\boldsymbol{I}_K$.

In order to satisfy assumption 3c, which restricts the dependence between different wavelet scales, we define the idiosyncratic errors' cross-covariances as,

$$\boldsymbol{\Sigma}_{j,j';\varepsilon}(k/T) = \begin{cases} \boldsymbol{\Sigma}_{j;\varepsilon}(k/T) & \text{if } j = j' \\ \boldsymbol{S}_{j,j'}\left(k/T\right) - \boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{\Sigma}_F \boldsymbol{\Lambda}_{j'}(\frac{k}{T})' & \text{otherwise} \end{cases}$$

such that, $\boldsymbol{S}_{j,j'}\left(k/T\right) = \delta_{jj'}[\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{\Sigma}_F \boldsymbol{\Lambda}_j(\frac{k}{T})' + \boldsymbol{\Sigma}_{j;\varepsilon}(k/T)]$.

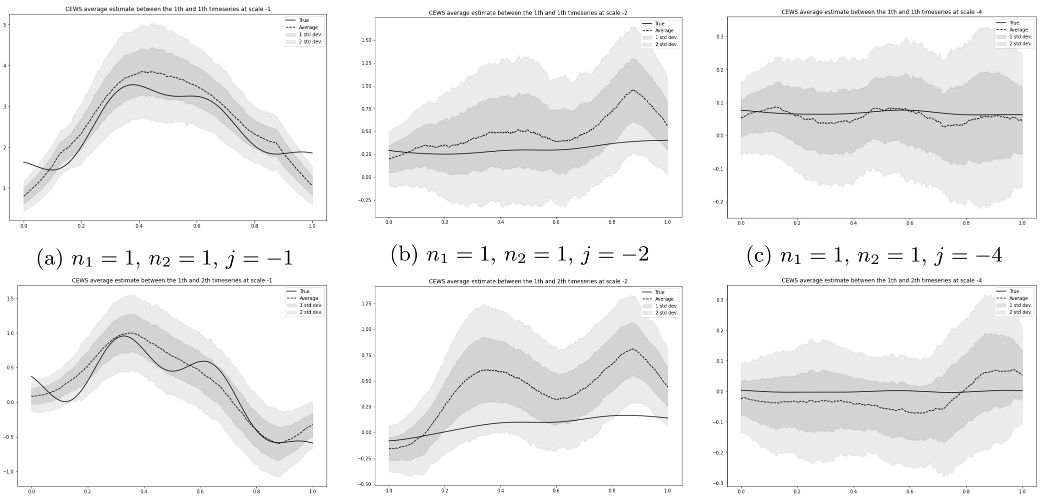
[3]An antisymmetric, or skew-symmetric, matrix fulfills $\boldsymbol{A}' = -\boldsymbol{A}$ and one of its property is $e^{\boldsymbol{A}} = \boldsymbol{Q}$ such that $\boldsymbol{Q}$ is an orthonormal matrix : $\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$.

With the same deterministic loadings, we simulate the estimation process $R = 100$ times by drawing new realizations of the factors and idiosyncratic terms. We assume the following distributional properties of the latter objects for $k = 1, \dots, T$. First, $\boldsymbol{F}_k \overset{iid}{\sim} N(\mathbf{0}_K, \boldsymbol{\Sigma}_F)$, where $\boldsymbol{\Sigma}_F = diag\{1, 0.4\}$. Second, $\boldsymbol{vec}\left(\{\boldsymbol{\varepsilon}_{j,j',k}\}_{j,j'}\right) \overset{iid}{\sim} N(\mathbf{0}_{NJ}, \Omega_\varepsilon(\frac{k}{T}))$, where $\Omega_\varepsilon(\frac{k}{T})$ is a block matrix defined by $\{\boldsymbol{\Sigma}_{j,j';\varepsilon}(k/T)\}_{j,j'} \in \mathbb{R}^{NJ \times NJ}$.

We simulate multivariate processes according to (2.1) and (3.1) with Haar wavelet —i.e. Daubechie's with 1 vanishing moment. For each replication, we compute the estimated CEWS with (4.3). We average the estimates over the 100 simulations to get $\overline{\widehat{\boldsymbol{S}}_j(k/T)}$ and compute estimated standard errors between the αth and βth processes at scale j and time k with,

$$\widehat{\sigma}_j^{(\alpha,\beta)}(k/T) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} \left(\widehat{S}_{j;r}^{(\alpha,\beta)}(k) - \overline{\widehat{S}_{j;r}^{(\alpha,\beta)}(k)} \right)^2}$$

The graphs pictured in Figure 1 depicts the convergence of the smoothed-corrected wavelet periodogram for different scales where we have used a rectangular kernel with a window length of 256. The standard deviation bands makes clearer the estimation uncertainty, yet the asymptotic normality of this estimator is only conjectured and must be theoretically justified. Although lower scales appear to be less precisely estimated, their contributions to the overall spectral power is limited, such that, the most important scale —i.e. $j = -1$ in the present case —have far better finite sample behavior.



(a) $n_1 = 1$, $n_2 = 1$, $j = -1$

(b) $n_1 = 1$, $n_2 = 1$, $j = -2$

(c) $n_1 = 1$, $n_2 = 1$, $j = -4$

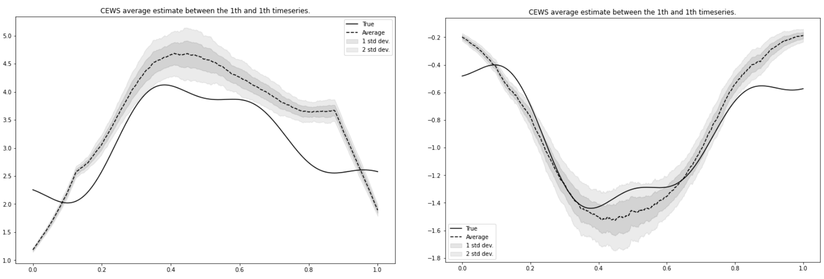
(d) $n_1 = 1$, $n_2 = 2$, $j = -1$

(e) $n_1 = 1$, $n_2 = 2$, $j = -2$

(f) $n_1 = 1$, $n_2 = 2$, $j = -4$

Figure 1: Comparison of the true CEWS (solid) and its replication estimator (dashed) after averaging $R = 100$ realisations between two timeseries (n_1, n_2) for scale j. Darker (lighter) grey region indicates the bootstrapped 1 (2) standard deviation(s) from the estimate.

Giving a straightforward interpretation to the dependencies between processes in the wavelet domain isn't an easy task. Consequently, we get back to the time-domain with the localized cross-covariance formula (4.4) and provide its estimator in Figure 2. As convenient as this representation seems, it suffers from the accumulated non-negligible bias from the aformentionned poor lower-scale estimates. We believe this bias could be reduced by entirely discarding those non-relevant estimates. For instance, the upward bias of the time-varying variance of the first process (Figure 2a) comes partially from the over-estimation of the second scale of decomposition (Figure 1b). The bias-boundary[4] issue is also clearly visible at both end of the timeseries.

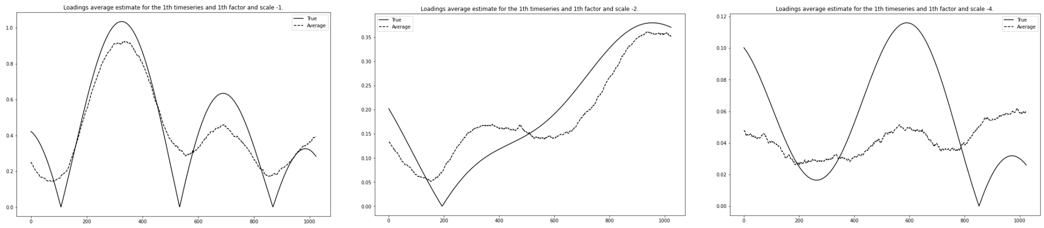


(a) $n_1 = 1$, $n_2 = 1$, $lag = 0$ (b) $n_1 = 1$, $n_2 = 1$, $lag = 1$

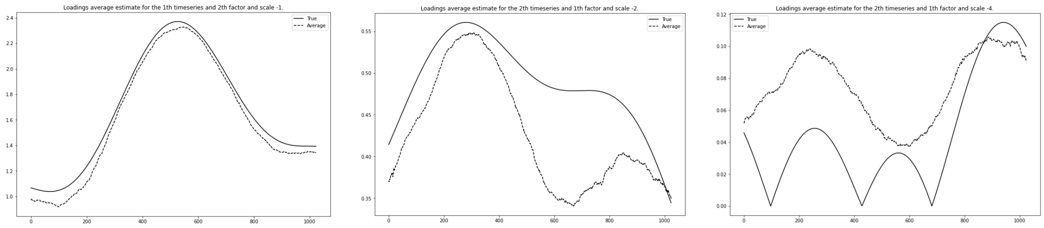
Figure 2: Comparison of the true cross-covariance (solid) and its replication estimator (dashed) after averaging $R = 100$ realisations between two timeseries (n_1, n_2) for a given lag. Darker (lighter) grey region indicates the bootstrapped 1 (2) standard deviation(s) from the estimate.

We estimate factor loadings following the procedures of section 4 and aggregate their *absolute* estimates, since the latter are only identified up-to a sign change, $\overline{\Lambda_j^{(\alpha,\beta)}(\frac{k}{T})} := \frac{1}{R}\sum_{r=1}^{R} \left|\mathbf{\Lambda}_{j;r}^{(\alpha,\beta)}(\frac{k}{T})\right|$. Several replications loadings are pictured in Figure 3. Derived from the CEWS, the latter also inherit its accumulated bias issue at lower-scales of decomposition. Nonetheless, we can observe the consistency of our estimates for the relevant scales. The effect of the smoothing procedure is also noticeable, such that, when enhanced, better estimates should possible.

[4]Before estimation, the timeseries were padded with zeros which, thus attracts the estimates.



(a) $n_1 = 1$, $j = -1$, $K = 1$ (b) $n_1 = 1$, $j = -2$, $K = 1$ (c) $n_1 = 1$, $j = -4$, $K = 1$



(d) $n_1 = 1$, $j = -1$, $K = 2$ (e) $n_1 = 2$, $j = -2$, $K = 1$ (f) $n_1 = 2$, $j = -4$, $K = 1$

Figure 3: Comparison between true and replicated loadings for several (cross)-processes (n_1, n_2), scales j and factors K. Note the different ranges for the y-axis which suggests that more important loadings are more precisely estimated.

7 Conclusion

In this paper, we propose a factor model based on the locally stationary wavelet processes of (Nason et al., 2000) and more specifically on its multivariate extension (Park et al., 2014). This peculiar decomposition gives us the ability to model time-varying factor loadings along with idiosyncratic errors' covariances rather conveniently through wavelets. The latter functions, known for their superior localization, renders possible the treatement of long-memory processes, thus avoiding the need for an absolutely summable errors' covariance function as in previous non-stationary estimation procedures (Su & Wang, 2017). This flexible dependence structure comes at the cost of restricted asymptotic behaviors, namely the number of cross-sections N must not grow faster than the number of observations T. The wavelet framework allows a rich characterisation of the dependencies between processes, so that not only time-varying covariances but also cross-covariances are considered by the factor structure. We believe convergence should be possible over a wider range of (N, T)-paths along with improved rates if one is willing to restrict these scale-dependencies.

We showed the weak convergence of the non-linear least-squares estimators for the loadings, factors and common components under general rotational indeterminacy matrix. At the exception of forbidding the inter-scales spectral power, none of the factor structure restrictions can be seen as restrictive under the LSW framework, hence various distributional properties can be considered for the factors and errors (e.g. heteroskedasticity, serial correlation, ...). Eventhough factors are consistently estimated, their interpretation is hindered by the non-identifiability of raw wavelet coefficients and further research should be pursued on that front. We finally provide some simulations to judge the finite sample properties of our estimators and noted that, eventhough simple averages offer theoretically convenient developments, more efficient and robust smoothing methods must be used in empirical applications. The non-negligible bias observed in the replications plots highlights the most important drawback of our methodology, namely converges hold only for datasets with a large number of cross-sections and time observations. Nonetheless wavelets still provide a unified framework to handle non-stationary, non-linear and multi-frequencies signals and we believe it will play an important role in the future of high-dimensional timeseries analysis.

References

Bai, J. (2003). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, *71*, 135–171.

Bai, J., Han, X., & Shi, Y. (2020). Estimation and inference of change points in high-dimensional factor models. *Journal of Econometrics*, *219*(1), 66–100. doi: 10.1016/j.jeconom.2019.08.013

Bai, J., & Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends® in Econometrics*, *3*(2), 89–163. doi: 10.1561/0800000002

Breitung, J., & Eickmeier, S. (2009). Testing for Structural Breaks in Dynamic Factor Models. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2785334

Chamberlain, G., & Rothschild, M. (1983). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, *51*(5), 1281. doi: 10.2307/1912275

Chen, L., Dolado, J. J., & Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics*, *180*(1), 30–48. doi: 10.1016/j.jeconom.2014.01.006

Connor, G., & Linton, O. (2007). Semiparametric estimation of a characteristic-based factor model of common stock returns. *Journal of Empirical Finance*, *14*(5), 694–717. doi: 10.1016/j.jempfin.2006.10.001

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *The Annals of Statistics*, *25*(1), 1–37. doi: 10.1214/aos/1034276620

Daubechies, I. (1992). *Ten lectures on wavelets* (No. 61). Philadelphia, Pa: Society for Industrial and Applied Mathematics.

Del Negro, M., & Otrok, C. M. (2008). Dynamic Factor Models with Time-Varying Parameters: Measuring Changes in International Business Cycles. *SSRN Electronic Journal*. doi: 10.2139/ssrn.1136163

Dijkerman, R., & Mazumdar, R. (1994). On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory*, *40*(5), 1609–1612. (Conference Name: IEEE Transactions on Information Theory) doi: 10.1109/18.333875

Fan, J., Liao, Y., & Wang, W. (2016). Projected principal component analysis in factor models. *The Annals of Statistics*, *44*(1). doi: 10.1214/15-AOS1364

Forni, M., Halli, M., Lippi, M., & Reichlin, L. (2000). The Generalized Dynamic-Factor Model : Identification and Estimation. *The Review of Economics and Statistics*, *82*(4), 540–554.

Forni, M., & Lippi, M. (2001). The Generalized Dynamic Factor Model: Representation Theory. *Econometric Theory*, *17*(6), 1113–1141. (Publisher: Cambridge University Press)

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions* *on Information Theory*, *14*(1), 55–63. doi: 10.1109/TIT.1968.1054102

Koch, D. (2015). *Multiscale methods for the analysis of high-dimensional locally stationary time series* (Doctoral dissertation, Universite Catholique de Louvain, Louvain-la-Neuve). (Supervisor : Van Bellegem, Sébastien)

Mikkelsen, J. G., Hillebrand, E., & Urga, G. (2019). Consistent estimation of time-varying loadings in high-dimensional factor models. *Journal of Econometrics*, *208*(2), 535–562. doi: 10.1016/j.jeconom.2018.09.020

Motta, G., Hafner, C. M., & von Sachs, R. (2011). Locally Stationary Factor Models : Identification and Nonparametric Estimation. *Econometric Theory*, *27*(6), 1279–1319. doi: 10.1017/S0266466611000053

Nason, G. P., von Sachs, R., & Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 271–292. doi: 10.1111/1467-9868.00231

Park, T., Eckley, I. A., & Ombao, H. C. (2014). Estimating Time-Evolving Partial Coherence Between Signals via Multivariate Locally Stationary Wavelet Processes. *IEEE Transactions on Signal Processing*, *62*(20), 5240–5250. doi: 10.1109/TSP.2014.2343937

Rothman, A. J. (2012). Positive definite estimators of large covariance matrices. *Biometrika*, *99*(3), 733–740. doi: 10.1093/biomet/ass025

Sanderson, J., Fryzlewicz, P., & Jones, M. W. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika*, *97*(2), 435–446. doi: 10.1093/biomet/asq007

Stock, J. H., & Watson, M. W. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. doi: 10.1198/016214502388618960

Su, L., & Wang, X. (2017). On time-varying factor models: Estimation and testing. *Journal of Econometrics*, *198*(1), 84–101. doi: 10.1016/j.jeconom.2016.12.004

Szu, H. H. (1992). Causal analytical wavelet transform. *Optical Engineering*, *31*(9), 1825. doi: 10.1117/12.59911

Van Bellegem, S., & von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *The Annals of Statistics*, *36*(4), 1879–1924. doi: 10.1214/07-AOS524

Appendix A Derivation of estimators

In this section we thoroughly derive our estimators of the loadings and common factors, which is essentially carried out by a non-linear least-squares minimization procedure. The next usefull lemma approximate the optimization problem by a new —simpler —one. We also point out possible restrictions which makes the latter exactly equivalent to the former.

Lemma 1. *Under the MLSW framework and assumptions LFE, the following random variable is stochastically bounded,*

$$(NT)^{-1}\left|\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k-\sum_t\sum_{j,l=-J}^{-1}\sum_{k,m=0}^{T-1}\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{m}{T})\boldsymbol{F}_m\psi_{j,k}(t)\psi_{l,m}(t)\right|=O_p(1)$$

Proof. Due to Markov, the proof summarizes as showing the existence of its expectation —i.e. $\mathrm{E}\left[|\cdot|\right]<\infty$.

The expectation of the random variable can be bounded by,

$$(NT)^{-1}\left(\mathrm{E}\left[\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left|\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k\right|\right]+\mathrm{E}\left[\sum_t\sum_{j,l=-J}^{-1}\sum_{k,m=0}^{T-1}\left|\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{m}{T})\boldsymbol{F}_m\right|\left|\psi_{j,k}(t)\right|\left|\psi_{l,m}(t)\right|\right]\right)$$

Those two terms are $O(1)$, we focus on the second since the first can be treated analoguously and with less technical details.

Note the scalar nature of the random variable, such that, one can freely apply the trace matrix operator and benefit from its cyclic property,

$$\begin{aligned}(NT)^{-1}\mathrm{E}&\left[\sum_t\sum_{j,l=-J}^{-1}\sum_{k,m=0}^{T-1}\left|tr\left\{\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{m}{T})\boldsymbol{F}_m\right\}\right|\left|\psi_{j,k}(t)\right|\left|\psi_{l,m}(t)\right|\right]\\&=(NT)^{-1}\sum_t\sum_{j,l=-J}^{-1}\sum_{k,m=0}^{T-1}\mathrm{E}\left[\left|tr\left\{\boldsymbol{\Lambda}_l(\frac{m}{T})\boldsymbol{F}_m\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\right\}\right|\right]\left|\psi_{j,k}(t)\right|\left|\psi_{l,m}(t)\right|\end{aligned}$$

The expectation term can be further expanded and bounded by repetitively applying Cauchy-Schwarz for summations and expectations,

$$\begin{aligned}\mathrm{E}\left[\left|tr\left\{\boldsymbol{\Lambda}_l(\frac{m}{T})\boldsymbol{F}_m\boldsymbol{F}_k'\boldsymbol{\Lambda}_j(\frac{k}{T})'\right\}\right|\right]&\le\sum_{\alpha=1}^{N}\sum_{i=1}^{K}\sum_{i'=1}^{K}\left|\Lambda_j^{(\alpha,i)}(\frac{k}{T})\right|\left|\Lambda_l^{(\alpha,i')}(\frac{m}{T})\right|\mathrm{E}\left[\left|F_k^{(i)}F_m^{(i')}\right|\right]\\&\le\sum_{\alpha=1}^{N}\left(\sum_{i=1}^{K}\left|\Lambda_j^{(\alpha,i)}(\frac{k}{T})\right|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{K}\mathrm{E}\left[\left|F_k^{(i)}\right|^2\right]\right)^{\frac{1}{2}}\\&\qquad\left(\sum_{i=1}^{K}\left|\Lambda_l^{(\alpha,i)}(\frac{m}{T})\right|^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{K}\mathrm{E}\left[\left|F_m^{(i)}\right|^2\right]\right)^{\frac{1}{2}}\\&\le K\left(\sup_{i,k}\mathrm{E}\left[\left|F_k^{(i)}\right|^2\right]\right)\left\|\boldsymbol{\Lambda}_j(\frac{k}{T})\right\|\left\|\boldsymbol{\Lambda}_j(\frac{m}{T})\right\|\end{aligned}$$

which then implies the bound for the initial sums,

$$K\left(\sup_{i,k}\mathrm{E}\left[\left|F_k^{(i)}\right|^2\right]\right)T^{-1}\sum_t\left(\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\right\|\left|\psi_{j,k}(t)\right|\right)^2$$

and due to the compactly supported wavelet,

$$\leq K\left(\sup_{i,k}\mathrm{E}\left[\left|F_k^{(i)}\right|^2\right]\right)\left(\sup_x|\psi(x)|\right)^2\left(\sum_{k=0}^{\bar{N}}\sum_{j=-J}^{-1}2^{\frac{j}{2}}\left\|\frac{\mathbf{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\right\|\right)^2$$

$$\leq K\bar{N}\left(\sup_{i,k}\mathrm{E}\left[\left|F_k^{(i)}\right|^2\right]\right)\left(\sup_x|\psi(x)|\right)^2\left(\sup_k\sum_{j=-J}^{-1}2^{\frac{j}{2}}\left\|\frac{\mathbf{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\right\|\right)^2$$

The result follows by the Dirichelet convergence test and assumption L2.. □

A.1 Non-linear Least Squares

The estimation of the loadings and common factors is carried out by a non-linear least squares procedure in the wavelet domain. The optimization problem is expressed as follows,

$$\min_{\{\bar{\mathbf{\Lambda}}_{j,k}\}_{\forall j,k},\{\bar{\boldsymbol{F}}_k\}_{\forall j,k}} \quad (NT)^{-1}\sum_t\left[\boldsymbol{X}_{t;T}-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left(\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k\right)\psi_{j,k}(t)\right]'\left[\boldsymbol{X}_{t;T}-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\left(\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k\right)\psi_{j,k}(t)\right]$$

$$\text{s.t.} \quad \frac{2^{|j|}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})}{N}=\boldsymbol{I}_K \tag{A.1}$$

The constraint implies, asymptotically, that $\sum_{j=-\infty}^{-1}\frac{\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})}{N}=\boldsymbol{I}_K$. After distributing and using the definition of the empirical wavelet coefficients (4.1), one obtains,

$$(NT)^{-1}\Big[\sum_t\boldsymbol{X}_{t;T}'\boldsymbol{X}_{t;T}-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}_{j,k}'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\bar{\boldsymbol{F}}_k'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k}+\sum_t\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\sum_{l=-J}^{-1}\sum_{m=0}^{T-1}\psi_{j,k}(t)\psi_{l,m}(t)\bar{\boldsymbol{F}}_k'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_l(\frac{m}{T})\bar{\boldsymbol{F}}_m\Big]$$

Due to lemma 1, the latter objective function is asymptotically close to,

$$(NT)^{-1}\Big[\sum_t\boldsymbol{X}_{t;T}'\boldsymbol{X}_{t;T}-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}_{j,k}'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k-\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\bar{\boldsymbol{F}}_k'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k}+\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\bar{\boldsymbol{F}}_k'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\boldsymbol{F}}_k\Big]$$

and it's asymptotically equivalent if further restrictions are assumed on the factors (e.g. $\mathrm{E}\left[\boldsymbol{F}_k\boldsymbol{F}_{k'}'\right]=\mathbf{0}_K,\forall k\neq k'$), on the loadings (e.g. $\mathbf{\Lambda}_j(\frac{k}{T})'\mathbf{\Lambda}_l(\frac{m}{T})=\mathbf{0}_K,\forall j\neq l,k\neq m$) or a mixture of the two. Knowing that we will have to pursue some simulations to judge the convergence of our estimators, the two latter conditions impose too much structure on the loadings' functional forms, such that we will adopt factors that are serially uncorrelated.

Taking the first order conditions w.r.t the factors and remembering the constraint (A.1) yields,

$$\bar{\boldsymbol{F}}_k'\sum_{j=-J}^{-1}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})-\sum_{j=-J}^{-1}\boldsymbol{d}_{j,k}'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})=0 \quad ,\forall k$$

$$\bar{\boldsymbol{F}}_k'=\frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{j=-J}^{-1}\boldsymbol{d}_{j,k}'\bar{\mathbf{\Lambda}}_j(\frac{k}{T}) \tag{A.2}$$

Substituting this last expression in the asymptotically close objective function gives,

$$\begin{aligned}\min_{\{\bar{\mathbf{\Lambda}}_{j,k}\}\forall j,k} \quad (NT)^{-1}\Big[\sum_t \boldsymbol{X}'_{t;T}\boldsymbol{X}_{t;T} &- \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{j,k}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\\ &- \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{l,k}\bar{\mathbf{\Lambda}}_l(\frac{k}{T})\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k}\\ &+ \frac{1}{N^2}\frac{2^{2J}}{(2^J-1)^2}\sum_{l=-J}^{-1}\sum_{n=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{l,k}\bar{\mathbf{\Lambda}}_l(\frac{k}{T})\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\mathbf{\Lambda}}_n(\frac{k}{T})'\boldsymbol{d}_{n,k}\Big]\end{aligned}$$

$$\begin{aligned}\min_{\{\bar{\mathbf{\Lambda}}_{j,k}\}\forall j,k} \quad (NT)^{-1}\Big[\sum_t \boldsymbol{X}'_{t;T}\boldsymbol{X}_{t;T} &- \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{j,k}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\\ &- \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{l,k}\bar{\mathbf{\Lambda}}_l(\frac{k}{T})\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\boldsymbol{d}_{j,k}\\ &+ \frac{1}{N^2}\frac{2^{2J}}{(2^J-1)^2}N\frac{(2^J-1)}{2^J}\sum_{l=-J}^{-1}\sum_{n=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{l,k}\bar{\mathbf{\Lambda}}_l(\frac{k}{T})\bar{\mathbf{\Lambda}}_n(\frac{k}{T})'\boldsymbol{d}_{n,k}\Big] \qquad \text{from (A.1)}\end{aligned}$$

$$\min_{\{\bar{\mathbf{\Lambda}}_{j,k}\}\forall j,k} \quad (NT)^{-1}\Big[\sum_t \boldsymbol{X}'_{t;T}\boldsymbol{X}_{t;T} - \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\boldsymbol{d}'_{j,k}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\Big]$$

$$\Updownarrow$$

$$\max_{\{\bar{\mathbf{\Lambda}}_{j,k}\}\forall j,k} \quad \frac{1}{TN^2}\frac{2^J}{(2^J-1)}\sum_{l=-J}^{-1}\sum_{j=-J}^{-1}\sum_{k=0}^{T-1}\operatorname{tr}\left\{\bar{\mathbf{\Lambda}}_l(\frac{k}{T})'\boldsymbol{d}_{l,k}\boldsymbol{d}'_{j,k}\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\right\}$$

We recognize the raw wavelet periodogram $\boldsymbol{d}_{l,k}\boldsymbol{d}'_{j,k}$ developed by (Koch, 2015), which, as in the contemporaneously restricted case (4.2), has an unbiaised-consistent estimator $\widehat{\boldsymbol{S}}_{j,l}\,(k/T,k/T)$. To get complete details about the extension of the Evolutionary Wavelet Spectrum to the cross-scale dependence structures, one may refer to (Koch, 2015).

The least squares function can further be simplified by assumption 3c which yields

$$\max_{\{\bar{\mathbf{\Lambda}}_{j,k}\}\forall j,k} N^{-2}\operatorname{tr}\left\{\bar{\mathbf{\Lambda}}_j(\frac{k}{T})'\widehat{\boldsymbol{S}}_j(k/T)\bar{\mathbf{\Lambda}}_j(\frac{k}{T})\right\}$$

since the sums reach their optimal values when each term also attains it. If $\tilde{\mathbf{\Lambda}}_j(k/T)$ are the $N\times K$ matrices whose columns are the first K orthonormal eigenvectors of $N^{-1}\widehat{\boldsymbol{S}}_j(k/T)$, then the optimal solutions are given by $\widehat{\mathbf{\Lambda}}_j(\frac{k}{T}) := 2^{-\frac{|j|}{2}}\sqrt{N}\tilde{\mathbf{\Lambda}}_j(k/T)$, which implies the common factors estimator :

$$\widehat{\boldsymbol{F}}'_k = \frac{1}{N}\frac{2^J}{(2^J-1)}\sum_{j=-J}^{-1}\boldsymbol{d}'_{j,k}\widehat{\mathbf{\Lambda}}_j(\frac{k}{T}) \tag{A.3}$$

Appendix B Technical Lemmas

As in conventional factor modelling, the asymptotic theory hinges on an adequate expression of the processes' —time-varying —covariance matrix in terms of the factor components. We provide such a decomposition for our cross-evolutionary wavelet spectrum estimator at the beginning of this section.

To see that inside the CEWS estimator defined in (4.3) hides the factor structure, notice that by the expression of the raw wavelet periodogram (4.2), the raw wavelet coefficients (4.1), the MLSW processes' representation (2.1) and the factor structure (3.1), the estimator can be flattened into,

$$\widehat{\boldsymbol{S}}_j(k/T) = N^{-1}\frac{1}{2M+1}\sum_{s=-M}^{M}\sum_{l=-J}^{-1} A_{j,l}^{-1}\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}\sum_{m=0}^{T-1}\sum_{p=-J}^{-1}(\boldsymbol{\Lambda}_p(\frac{m}{T})\boldsymbol{F}_m + \boldsymbol{\varepsilon}_{p,m})\psi_{p,m}(t)$$
$$\sum_{m'=0}^{T-1}\sum_{p'=-J}^{-1}(\boldsymbol{\Lambda}_{p'}(\frac{m'}{T})\boldsymbol{F}_{m'} + \boldsymbol{\varepsilon}_{p',m'})'\psi_{p',m'}(t')\psi_{l,k+s}(t)\psi_{l,k+s}(t')$$

Anyone working with this expression would argue that it's not the most simple task to pursue. As it's usually the case when working with embricated and complex objects, one tries to reduce the complexity by breaking it into smaller manageable pieces. That's the stance we take with this estimator.

Most of the difficulties may be hidden by defining the $N \times 1$ vectors,

$$\boldsymbol{C}_t = \sum_{k=0}^{T-1}\sum_{j=-J}^{-1}\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{F}_k\psi_{j,k}(t) \tag{B.1}$$

$$\boldsymbol{E}_t = \sum_{k=0}^{T-1}\sum_{j=-J}^{-1}\boldsymbol{\varepsilon}_{j,k}\psi_{j,k}(t) \tag{B.2}$$

and the following $N \times N$ matrices,

$$\boldsymbol{Y}_{j,k}(s) = \sum_{l=-J}^{-1} A_{j,l}^{-1}\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}\boldsymbol{C}_t\boldsymbol{C}_{t'}'\psi_{l,k+s}(t)\psi_{l,k+s}(t') \tag{B.3}$$

$$\boldsymbol{W}_{j,k}(s) = \sum_{l=-J}^{-1} A_{j,l}^{-1}\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}\boldsymbol{C}_t\boldsymbol{E}_{t'}'\psi_{l,k+s}(t)\psi_{l,k+s}(t') \tag{B.4}$$

$$\boldsymbol{Z}_{j,k}(s) = \sum_{l=-J}^{-1} A_{j,l}^{-1}\sum_{t=0}^{T-1}\sum_{t'=0}^{T-1}\boldsymbol{E}_t\boldsymbol{E}_{t'}'\psi_{l,k+s}(t)\psi_{l,k+s}(t') \tag{B.5}$$

$$\boldsymbol{K}_{j,k} = \frac{1}{2M+1}\sum_{s=-M}^{M}\boldsymbol{Y}_{j,k}(s) \tag{B.6}$$

$$\boldsymbol{\Upsilon}_{j,k} = \frac{1}{2M+1}\sum_{s=-M}^{M}\boldsymbol{Z}_{j,k}(s) \tag{B.7}$$

$$\boldsymbol{\Theta}_{j,k} = \frac{1}{2M+1}\sum_{s=-M}^{M}\boldsymbol{W}_{j,k}(s) \tag{B.8}$$

By hiding all the complexity inside objects we can express the CEWS as,

$$\widehat{\boldsymbol{S}}_j(k/T) = \boldsymbol{K}_{j,k} + \boldsymbol{\Theta}_{j,k} + \boldsymbol{\Theta}'_{j,k} + \boldsymbol{\Upsilon}_{j,k} \tag{B.9}$$

Inspired by Dynamic Programming, such that, from properties of simpler objects one can derive more complex systems' behavior, the next technical lemmas lay the foundational work to —*in fine* —prove consistency of the CEWS estimator along with the factor structure.

This first lemma is straightforward but will be used extensively throughout the proofs.

Lemma 2. *Under the MLSW framework and assumptions LFE,*

$$\boldsymbol{\Lambda}_j(\frac{k}{T})\mathrm{E}\left[\boldsymbol{F}_k\boldsymbol{F}'_{k'}\right]\boldsymbol{\Lambda}_{j'}(\frac{k'}{T})' = \begin{cases} \boldsymbol{S}_j\left(k/T\right) - \mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}_{j,k}\right] & \textit{if } j = j' \textit{ and } k = k' \\ -\mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}_{j',k'}\right] & \textit{otherwise} \end{cases}$$

Proof. From the factor structure (3.1) and taking expectation,

$$\boldsymbol{W}_j\left(k/T\right)\mathrm{E}\left[\boldsymbol{\xi}_{j,k}\boldsymbol{\xi}'_{j',k'}\right]\boldsymbol{W}_{j'}\left(k'/T\right) = \boldsymbol{\Lambda}_j(\frac{k}{T})\mathrm{E}\left[\boldsymbol{F}_k\boldsymbol{F}_{k'}\right]\boldsymbol{\Lambda}_{j'}(\frac{k'}{T}) + \boldsymbol{\Lambda}_j(\frac{k}{T})\mathrm{E}\left[\boldsymbol{F}_k\boldsymbol{\varepsilon}'_{j',k'}\right] + \mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{F}_{k'}\right]\boldsymbol{\Lambda}_{j'}(\frac{k'}{T})' + \mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}'_{j',k'}\right]$$

By independence between factors and idiosyncratic errors —assumption E5. —we obtain,

$$\boldsymbol{W}_j\left(k/T\right)\mathrm{E}\left[\boldsymbol{\xi}_{j,k}\boldsymbol{\xi}'_{j',k'}\right]\boldsymbol{W}_{j'}\left(k'/T\right) = \boldsymbol{\Lambda}_j(\frac{k}{T})\mathrm{E}\left[\boldsymbol{F}_k\boldsymbol{F}_{k'}\right]\boldsymbol{\Lambda}_{j'}(\frac{k'}{T}) + \mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}'_{j',k'}\right]$$

Finally by the increments' orthonormality of MLSW processes (2) we get the result since

$$\mathrm{E}\left[\boldsymbol{\xi}_{j,k}\boldsymbol{\xi}'_{j',k'}\right] = \delta_{jj'}\delta_{kk'}\boldsymbol{I}_N$$

$\square$

Lemma 3. *Under assumptions E and the definition* (B.2)*,*

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{E}_t\right] &= 0 & \forall t \\ \mathrm{Cov}\left[\boldsymbol{E}_t, \boldsymbol{E}_{t'}\right] &< \infty & \forall t, t' \end{aligned}$$

Proof. The expectation property follows straightforwardly from the distributional assumption on the idiosyncratic error terms E1..

The covariance property can be flattened into,

$$\mathrm{E}\left[\boldsymbol{E}_t\boldsymbol{E}'_{t'}\right] \leq \sum_{m_0,m_1}\sum_{p_0,p_1}\left|\mathrm{E}\left[\boldsymbol{\varepsilon}_{p_0,m_0}\boldsymbol{\varepsilon}'_{p_1,m_1}\right]\right|\left|\psi_{p_0,m_0}(t)\right|\left|\psi_{p_1,m_1}(t')\right|$$

Then by definition of child wavelets, applying Cauchy on the expectation operator and taking the (α,β)-element of the matrix,

$$\begin{aligned} &= \sum_{m_0,m_1}\sum_{p_0,p_1}\mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2}\mathrm{E}\left[{\varepsilon^{(\beta)}_{p_1,m_1}}^2\right]^{1/2} 2^{\frac{p_0}{2}}\left|\psi(2^{p_0}t - m_0)\right| 2^{\frac{p_1}{2}}\left|\psi(2^{p_1}t - m_1)\right| \\ &= \left(\sum_{m_0}\sum_{p_0}2^{\frac{p_0}{2}}\mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2}\left|\psi(2^{p_0}t - m_0)\right|\right)\left(\sum_{m_1}\sum_{p_1}2^{\frac{p_1}{2}}\mathrm{E}\left[{\varepsilon^{(\beta)}_{p_1,m_1}}^2\right]^{1/2}\left|\psi(2^{p_1}t - m_1)\right|\right) \end{aligned}$$

Given that $\psi(x) \in \mathcal{L}^1$, then $\forall \delta > 0, \exists \bar{N} \in \mathbb{N} : \forall x > \bar{N} \Rightarrow |\psi(x)| < \delta$. Then $\forall \alpha$,

$$\begin{aligned}\sum_{m_0}\sum_{p_0} 2^{\frac{p_0}{2}} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2} |\psi(2^{p_0}t - m_0)| &= \sum_{m_0}^{\bar{N}}\sum_{p_0} 2^{\frac{p_0}{2}} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2} |\psi(2^{p_0}t - m_0)| \\ &\quad + \sum_{m_0=\bar{N}+1}\sum_{p_0} 2^{\frac{p_0}{2}} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2} |\psi(2^{p_0}t - m_0)| \\ &\leq \bar{N} \sup_{m_0} \sum_{p_0=-J}^{-1} 2^{\frac{p_0}{2}} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2} + \delta T \sup_{m_0} \sum_{p_0=-J}^{-1} 2^{\frac{p_0}{2}} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]^{1/2}\end{aligned}$$

This last term is bounded thanks to Dirichlet's convergence test and,

$$\begin{aligned}\sum_{p_0=-J}^{-1} \left(\sup_{m_0} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]\right)^{1/2} &\leq \left(\sum_{p_0=-J}^{-1} \sup_{m_0} \mathrm{E}\left[{\varepsilon^{(\alpha)}_{p_0,m_0}}^2\right]\right)^{1/2} \\ &< \infty \qquad\qquad \text{from assumption E4.}\end{aligned}$$

□

Lemma 4. *Under assumption LFE, the MLSW framework and the definition* (B.1),

$$\text{(I)} \qquad \mathrm{E}\left[\boldsymbol{C}_t\right] = 0 \qquad \forall t$$

$$\text{(II)} \qquad \mathrm{Cov}\left[\boldsymbol{C}_t, \boldsymbol{C}_{t'}\right] < \infty \qquad \forall t, t'$$

$$\text{(III)} \qquad \lim_{i\to\infty} \mathrm{Cov}\left[\boldsymbol{C}_{t_0}\boldsymbol{C}'_{t_1}, \boldsymbol{C}_{t_2}\boldsymbol{C}'_{t_3}\right] = 0 \qquad \forall t_0, t_1, t_2, t_3$$

$$\text{(IV)} \qquad \sup_t \mathrm{E}\left[[\boldsymbol{C}^{(\alpha)}_t]^4\right] < \infty \qquad \forall \alpha$$

$$\text{(V)} \qquad \sum_{\substack{t_0,t_1\\t_2,t_3}}^{T-1} \left|\mathrm{E}\left[[\boldsymbol{C}_{t_0}\boldsymbol{C}'_{t_1}\boldsymbol{C}'_{t_2}\boldsymbol{C}_{t_3}]^{(\alpha,\beta)}\right]\right| |\psi(t_0)|\,|\psi(t_1)|\,|\psi(t_2)|\,|\psi(t_3)| = O(N) \qquad \forall t_0, t_1, t_2, t_3, (\alpha, \beta)$$

Proof. The first expectation property follows from the distributional assumption F1. on the common factors.

The covariance property can be expanded as,

$$\mathrm{E}\left[\boldsymbol{C}_t\boldsymbol{C}'_{t'}\right] = \sum_{m_0,m_1}\sum_{p_0,p_1} \boldsymbol{\Lambda}_{p_0}(\frac{m_0}{T})\mathrm{E}\left[\boldsymbol{F}_{m_0}\boldsymbol{F}'_{m_1}\right]\boldsymbol{\Lambda}_{p_1}(\frac{m_1}{T})'\psi_{p_0,m_0}(t)\psi_{p_1,m_1}(t')$$

Given lemma 2,

$$= \sum_{m_0}\sum_{p_0} \boldsymbol{S}_{p_0}\left(m_0/T\right)\psi_{p_0,m_0}(t)\psi_{p_0,m_0}(t') + \sum_{m_0,m_1}\sum_{p_0,p_1} \mathrm{E}\left[\boldsymbol{\varepsilon}_{p_0,m_0}\boldsymbol{\varepsilon}'_{p_1,m_1}\right]\psi_{p_0,m_0}(t)\psi_{p_1,m_1}(t')$$

The lipschitz behavior of the CEWS along with the definition of the autocorrelation wavelet,

$$\begin{aligned}&= \sum_{p_0} \boldsymbol{S}_{p_0}\left(t/T\right)\Psi_{p_0}(t - t') + \mathrm{Cov}\left[E_t, E_{t'}\right] + O(T^{-1}) \\ &< \infty\end{aligned}$$

by lemma 3 and assumption 3a of section 2.

Property (III) is equivalent to

$$\lim_{i\to\infty} \mathrm{E}\left[\boldsymbol{C}_{t_0}\boldsymbol{C}'_{t_1}\boldsymbol{C}_{t_2+i}\boldsymbol{C}'_{t_3+i}\right] = \lim_{i\to\infty} \mathrm{E}\left[\boldsymbol{C}_{t_0}\boldsymbol{C}'_{t_1}\right] \mathrm{E}\left[\boldsymbol{C}_{t_2+i}\boldsymbol{C}'_{t_3+i}\right]$$

The (α,β)-element of the LHS can be flattened as,

$$\sum_{\substack{m_0,m_1\\m_2,m_3}} \sum_{\substack{p_0,p_1\\p_2,p_3}} \sum_{l=1}^{N} \sum_{\substack{i_0,i_1\\i_2,i_3\\=1}}^{K} \boldsymbol{\Lambda}_{p_0}(\frac{m_0}{T})^{(\alpha,i_0)} \boldsymbol{\Lambda}_{p_1}(\frac{m_1}{T})^{(l,i_1)} \boldsymbol{\Lambda}_{p_2}(\frac{m_2+i}{T})^{(l,i_2)} \boldsymbol{\Lambda}_{p_3}(\frac{m_3+i}{T})^{(\beta,i_3)} \mathrm{E}\left[\boldsymbol{F}^{(i_0)}_{m_0}\boldsymbol{F}^{(i_1)}_{m_1}\boldsymbol{F}^{(i_2)}_{m_2+i}\boldsymbol{F}^{(i_3)}_{m_3+i}\right]$$
$$\psi_{p_0,m_0}(t_0)\psi_{p_1,m_1}(t_1)\psi_{p_2,m_2}(t_2)\psi_{p_3,m_3}(t_3)$$
$$\longrightarrow \sum_{\substack{m_0,m_1\\m_2,m_3}} \sum_{\substack{p_0,p_1\\p_2,p_3}} \sum_{l=1}^{N} \sum_{\substack{i_0,i_1\\i_2,i_3\\=1}}^{K} \boldsymbol{\Lambda}_{p_0}(\frac{m_0}{T})^{(\alpha,i_0)} \boldsymbol{\Lambda}_{p_1}(\frac{m_1}{T})^{(l,i_1)} \boldsymbol{\Lambda}_{p_2}(\frac{m_2+i}{T})^{(l,i_2)} \boldsymbol{\Lambda}_{p_3}(\frac{m_3+i}{T})^{(\beta,i_3)} \mathrm{E}\left[\boldsymbol{F}^{(i_0)}_{m_0}\boldsymbol{F}^{(i_1)}_{m_1}\right] \mathrm{E}\left[\boldsymbol{F}^{(i_2)}_{m_2+i}\boldsymbol{F}^{(i_3)}_{m_3+i}\right]$$
$$\psi_{p_0,m_0}(t_0)\psi_{p_1,m_1}(t_1)\psi_{p_2,m_2}(t_2)\psi_{p_3,m_3}(t_3)$$

by assumption F4.. After splitting the sum, we obtain

$$= \mathrm{E}\left[\boldsymbol{C}_{t_0}\boldsymbol{C}'_{t_1}\right] \mathrm{E}\left[\boldsymbol{C}_{t_2+i}\boldsymbol{C}'_{t_3+i}\right]$$

To see that property (IV) holds, notice that,

$$\mathrm{E}\left[[C^{(\alpha)}_t]^4\right] \le \mathrm{E}\left[\left(\sum_m \sum_p \sum_{i=1}^{K} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right| 2^{\frac{p}{2}} \left|\psi(2^p t - m)\right|\right)^4\right]$$

and given $\psi(x) \in \mathcal{L}^1$then $\forall\delta > 0, \exists \bar{N} \in \mathbb{N} : \forall x > \bar{N} \Rightarrow |\psi(x)| < \delta$, thus we can decompose the inner sums,

$$\sum_{m=0}^{\bar{N}} \sum_p \sum_{i=1}^{K} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right| 2^{\frac{p}{2}} \left|\psi(2^p t - m)\right| + \sum_{m=\bar{N}+1}^{T-1} \sum_p \sum_{i=1}^{K} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right| 2^{\frac{p}{2}} \left|\psi(2^p t - m)\right|$$
$$\le \sum_{m=0}^{\bar{N}} \sum_p \sum_{i=1}^{K} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right| 2^{\frac{p}{2}} \left|\psi(2^p t - m)\right| + \delta \sum_{m=\bar{N}+1}^{T-1} \sum_p \sum_{i=1}^{K} 2^{\frac{p}{2}} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right|$$
$$\longrightarrow \sum_{m=0}^{\bar{N}} \sum_p \sum_{i=1}^{K} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right| 2^{\frac{p}{2}} \left|\psi(2^p t - m)\right|$$
$$\le \left(\sup_m \left|\psi(2^p t - m)\right|\right) \sum_{m=0}^{\bar{N}} \sum_p \sum_{i=1}^{K} 2^{\frac{p}{2}} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right|$$

Therefore,

$$\mathrm{E}\left[{C^{(\alpha)}_t}^4\right] \le \left(\sup_m \left|\psi(2^p t - m)\right|\right)^4 \mathrm{E}\left[\left(\sum_{m=0}^{\bar{N}} \sum_p \sum_{i=1}^{K} 2^{\frac{p}{2}} \left|\Lambda^{(\alpha,i)}_p(\frac{m}{T})\right| \left|F^{(i)}_m\right|\right)^4\right]$$

which by expanding the exponent, applying iteratively Cauchy-Schwarz and assumption F2. becomes,

$$\begin{aligned}
\mathrm{E}\left[C_t^{(\alpha)^4}\right] &\leq \left(\sup_m |\psi(2^p t - m)|\right)^4 M_1^4 \bar{N}^4 \left[\sup_m \sum_{i=1}^{K} \sum_{p=-J}^{-1} 2^{\frac{p}{2}} \left|\Lambda_p^{(\alpha,i)}(\frac{m}{T})\right|\right]^4 \\
&\leq \left(\sup_m |\psi(2^p t - m)|\right)^4 M_1^4 \bar{N}^4 K^2 \left[\sup_m \sum_{p=-J}^{-1} 2^{\frac{p}{2}} \left\|\boldsymbol{\lambda}_{m,p}^{(\alpha)}\right\|\right]^4 \quad \text{by Jensen's} \\
&< \infty
\end{aligned}$$

The result is obtained due to assumption L2. and the Dirichelet convergence test.

The last property —i.e. (V) —follows from Cauchy, $\psi(x) \in \mathcal{L}^1$ and property (IV). □

Lemma 5. *Under assumption E and definitions* (B.1), (B.2)*,*

$$\begin{aligned}
\boldsymbol{C}_t &\perp \boldsymbol{E}_{t'} && \forall t, t' \\
\lim_{i\to\infty} \operatorname{Cov}\left[\boldsymbol{C}_{t_0}\boldsymbol{E}'_{t_1}, \boldsymbol{C}_{t_2}\boldsymbol{E}'_{t_3}\right] &= 0 && \forall t_0, t_1, t_2, t_3
\end{aligned}$$

Proof. The first property holds since $\boldsymbol{C}_t \in span\{\{\boldsymbol{F}_m\}_{m\in\mathbb{Z}}\}$, $\boldsymbol{E}_t \in span\{\{\boldsymbol{\varepsilon}_{p,m}\}_{m\in\mathbb{Z},-p\in\mathbb{N}}\}$ and assumption E5..

The second property is analoguous to the proof of property (III) in lemma 4 by using assumption E7.. □

Lemma 6. *Under the MLSW framework, assumption LFE and definition* (B.3)*,*

$$\begin{aligned}
\mathrm{E}\left[Y_{j,k}^{(\alpha,\beta)}(s)\right] &= \left[\boldsymbol{S}_j\left((k+s)/T\right)\right]^{(\alpha,\beta)} - \mathrm{E}\left[Z_{j,k}^{(\alpha,\beta)}(s)\right] + O\left(T^{-1}\right) && \forall(\alpha,\beta) - elements \\
\operatorname{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s)\right] &= O(2^j N) && \forall(\alpha,\beta) - elements \\
\lim_{i\to\infty} \operatorname{Cov}\left[\boldsymbol{Y}_{j,k}(s), \boldsymbol{Y}_{j,k}(s+i)\right] &= \boldsymbol{0}_N && \forall j,k,s
\end{aligned}$$

Proof. By using lemma 2 and definition (B.5), one obtain

$$\mathrm{E}\left[\boldsymbol{Y}_{j,k}(s)\right] = \sum_{l=-J}^{-1} A_{l,j}^{-1} \sum_{t,t',m=0}^{T-1} \sum_{p=-J}^{-1} \boldsymbol{S}_p\left(m/T\right) \psi_{p,m}(t)\psi_{p,m}(t')\psi_{l,k+s}(t)\psi_{l,k+s}(t') - \mathrm{E}\left[\boldsymbol{Z}_{j,k}(s)\right]$$

Remembering that the CEWS is lipschitz continuous through time, one obtains, by a change of variable $m = n + k + s$,

$$\sum_{l=-J}^{-1} A_{l,j}^{-1} \sum_{p=-J}^{-1} \boldsymbol{S}_p\left((k+s)/T\right) \sum_{n=0}^{T-1} \left[\sum_{t=0}^{T-1} \psi_{p,n+k+s}(t)\psi_{l,k+s}(t)\right]^2 - \mathrm{E}\left[\boldsymbol{Z}_{j,k}(s)\right] + O(T^{-1})$$

The result then straightforwardly follows if one notice that $\sum_{n=0}^{T-1} \left[\sum_{t=0}^{T-1} \psi_{p,n+k+s}(t)\psi_{l,k+s}(t)\right]^2 = A_{p,l}$ which, thanks to the linear independence of autocorrelation wavelets, gives,

$$\sum_{p=-J}^{-1} \boldsymbol{S}_p\left((k+s)/T\right)\delta_{pj} - \mathrm{E}\left[Z_{j,k}(s)\right] + O(T^{-1})$$

Next, the variance can be bounded since,

$$\begin{aligned}
\operatorname{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s)\right] &\leq \sum_{l_0,l_1}\left|A_{j,l_0}^{-1}\right|\left|A_{j,l_1}^{-1}\right|\sum_{\substack{t_0,t_1\\t_2,t_3}}\left|\mathrm{E}\left[[C_{t_0}C'_{t_1}C'_{t_2}C_{t_3}]^{(\alpha,\beta)}\right]\right|\left|\psi_{l_0,k+s}(t_0)\right|\left|\psi_{l_0,k+s}(t_1)\right|\left|\psi_{l_1,k+s}(t_2)\right|\left|\psi_{l_1,k+s}(t_3)\right| \\
&= \sum_{l_0,l_1}\left|A_{j,l_0}^{-1}\right|\left|A_{j,l_1}^{-1}\right|\sum_{\substack{t_0,t_1\\t_2,t_3}}\left|\mathrm{E}\left[[C_{t_0}C'_{t_1}C'_{t_2}C_{t_3}]^{(\alpha,\beta)}\right]\right| 2^{\frac{l_0}{2}}\left|\psi(2^{l_0}t_0-k-s)\right| 2^{\frac{l_0}{2}}\left|\psi(2^{l_0}t_1-k-s)\right| \\
&\qquad 2^{\frac{l_1}{2}}\left|\psi(2^{l_1}t_2-k-s)\right| 2^{\frac{l_1}{2}}\left|\psi(2^{l_1}t_3-k-s)\right| \\
&= O(N)\left(\sum_l 2^l\left|A_{j,l}^{-1}\right|\right)^2 \qquad \text{from lemma 4} \\
&= O(2^jN)
\end{aligned}$$

where the last line uses an analoguous argument as in (Van Bellegem & von Sachs, 2008).

Finally, the asymptotic covariance property is implied by (III) in lemma 4. □

Lemma 7. *Under the MLSW framework, assumption LFE and definition* (B.4),

$$\begin{aligned}
\mathrm{E}\left[W_{j,k}^{(\alpha,\beta)}(s)\right] &= \mathbf{0}_N && \forall(\alpha,\beta)-elements \\
\operatorname{Var}\left[W_{j,k}^{(\alpha,\beta)}(s)\right] &= O(2^jN) && \forall(\alpha,\beta)-elements \\
\lim_{i\to\infty}\operatorname{Cov}\left[\boldsymbol{W}_{j,k}(s),\boldsymbol{W}_{j,k}(s+i)\right] &= \mathbf{0}_N && \forall j,k,s
\end{aligned}$$

Proof. The proofs are strictly analoguous to lemma 6 given lemma 5. □

Lemma 8. *Under the MLSW framework, the assumptions LFE and definitions* (B.6),(B.7) *and* (B.8),

$$\begin{aligned}
\mathrm{E}\left[\boldsymbol{K}_{j,k}\right] &= \boldsymbol{S}_j\left(k/T\right) - \mathrm{E}\left[\boldsymbol{\Upsilon}_{j,k}\right] + O\left(\frac{M}{T}\right) \\
\mathrm{E}\left[\boldsymbol{\Theta}_{j,k}\right] &= \mathbf{0}_N
\end{aligned}$$

Proof.

$$
\begin{aligned}
\mathrm{E}\left[\boldsymbol{K}_{j,k}\right] &= \frac{1}{2M+1}\sum_{s=-M}^{M}\mathrm{E}\left[\boldsymbol{Y}_{j,k}(s)\right] \\
&= \frac{1}{2M+1}\sum_{s=-M}^{M}\left[\boldsymbol{S}_j\left((k+s)/T\right) - \mathrm{E}\left[\boldsymbol{Z}_{j,k}(s)\right] + O(T^{-1})\right] && \text{from lemma 6} \\
&= \frac{1}{2M+1}\sum_{s=-M}^{M}\left[\boldsymbol{S}_j\left(k/T\right) + O\left(\frac{s}{T}\right)\right] - \frac{1}{2M+1}\sum_{s=-M}^{M}\mathrm{E}\left[\boldsymbol{Z}_{j,k}(s)\right] + O(T^{-1})
\end{aligned}
$$

by Lipschitz continuity and since $\sum_{s=-M}^{M}|s| = (M+1)M$.

$$
\begin{aligned}
&= \boldsymbol{S}_j\left(k/T\right) - \frac{1}{2M+1}\sum_{s=-M}^{M}\mathrm{E}\left[\boldsymbol{Z}_{j,k}(s)\right] + O\left(\frac{M}{T}\right) + O(T^{-1}) \\
&= \boldsymbol{S}_j\left(k/T\right) - \mathrm{E}\left[\Upsilon_{j,k}\right] + O\left(\frac{M}{T}\right)
\end{aligned}
$$

The second expectation follows directly from lemma 5. □

Lemma 9. *Under the MLSW framework, the assumption LFE and definitions* (B.6) *and* (B.8)*,*

$$
\left\|\boldsymbol{K}_{j,k} - \mathrm{E}\left[\boldsymbol{K}_{j,k}\right]\right\| = O_p\left(\sqrt{\frac{2^jN^3}{M} + \frac{N^2M^2}{T^2}}\right)
$$

$$
\left\|\boldsymbol{\Theta}_{j,k}\right\| = O_p\left(\sqrt{\frac{2^jN^3}{M}}\right)
$$

if one further assumes $M = O(T^{\eta^*})$ *and* $N = O(T^{\nu})$ *where* $0 < \nu < \frac{1}{4}$ *and* $\eta^* = \frac{1}{3}(\nu + \frac{j}{\log_2(T)} + 2)$*, then*

$$
\left\|\boldsymbol{K}_{j,k} - \mathrm{E}\left[\boldsymbol{K}_{j,k}\right]\right\| = O_p\left(2^{\frac{j}{3}}T^{\frac{1}{3}(4\nu-1)}\right)
$$

$$
\left\|\boldsymbol{\Theta}_{j,k}\right\| = O_p\left(2^{\frac{j}{3}}T^{\frac{1}{3}(4\nu-1)}\right)
$$

Proof. In finite sample $\boldsymbol{K}_{j,k}$ is a biased estimator, it's thus not sufficient, in order to find its convergence rate, to show the consistency of the variance. The idea of the proof lies in providing a vanishing bound for its *Mean Squared Error* (MSE), which combined with Tchebyshev's, implies the convergence in probability to its expectation.
For each (α,β)-element, we get

$$
\begin{aligned}
\mathbb{P}\left(\left|K_{j,k}^{(\alpha,\beta)} - \mathrm{E}\left[K_{j,k}^{(\alpha,\beta)}\right]\right| > \delta\right) &\leq \frac{\mathrm{Var}\left[K_{j,k}^{(\alpha,\beta)}\right] + Bias(K_{j,k}^{(\alpha,\beta)})^2}{\delta^2} \\
&= \frac{\sum_{s_0=-M}^{M}\sum_{s_1=-M}^{M}\left|\mathrm{Cov}\left[Y_{j,k}(s_0), Y_{j,k}(s_1)\right]^{(\alpha,\beta)}\right|}{(2M+1)^2\delta^2} + O\left(\frac{M^2}{T^2}\right)
\end{aligned}
$$

Let's focus on the numerator,

$$\sum_{s_0=-M}^{M}\sum_{s_1=-M}^{M}\left|\mathrm{Cov}\left[Y_{j,k}(s_0),Y_{j,k}(s_1)\right]^{(\alpha,\beta)}\right| = \sum_{s=-M}^{M}\mathrm{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s)\right] + 2\sum_{s_0=0}^{2M}\sum_{s_1=s_0+1}^{2M}\left|\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0-M),\boldsymbol{Y}_{j,k}(s_1-M)\right]^{(\alpha,\beta)}\right|$$

From lemma 6 and the definition of limits, $\forall\epsilon>0,\exists\bar{N}:\forall\,|s_0-s_1|>\bar{N}\Rightarrow\left|\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0),\boldsymbol{Y}_{j,k}(s_1)\right]^{(\alpha,\beta)}\right|<\epsilon$ for every (α,β)-elements of the matrix $\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0),\boldsymbol{Y}_{j,k}(s_1)\right]$. Therefore,

$$\begin{aligned}\sum_{s_0=0}^{2M}\sum_{s_1=s_0+1}^{2M}\left|\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0-M),\boldsymbol{Y}_{j,k}(s_1-M)\right]^{(\alpha,\beta)}\right| &= \sum_{s_0=0}^{2M}\sum_{s_1=s_0+1}^{s_0+\bar{N}}\left|\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0),\boldsymbol{Y}_{j,k}(s_1)\right]^{(\alpha,\beta)}\right| \\ &\quad+\sum_{s_0=0}^{2M}\sum_{s_1=s_0+1+\bar{N}}^{2M}\left|\mathrm{Cov}\left[\boldsymbol{Y}_{j,k}(s_0),\boldsymbol{Y}_{j,k}(s_1)\right]^{(\alpha,\beta)}\right| \\ &\le \sum_{s=0}^{2M}\bar{N}\mathrm{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s-M)\right]+\sum_{s_0=0}^{2M}\sum_{s_1=s_0+1+\bar{N}}^{2M}\epsilon \\ &\le \bar{N}(2M+1)\sup_s\mathrm{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s-M)\right]+(2M+1)^2\epsilon\end{aligned}$$

The initial probability can be bounded by,

$$\begin{aligned}\mathbb{P}\left(\left|K_{j,k}^{(\alpha,\beta)}-\mathrm{E}\left[K_{j,k}^{(\alpha,\beta)}\right]\right|>\delta\right) &\le O\left(\frac{1}{M}\right)\sup_s\mathrm{Var}\left[Y_{j,k}^{(\alpha,\beta)}(s)\right]+O\left(\frac{M^2}{T^2}\right) \\ &= O\left(\frac{2^jN}{M}+\frac{M^2}{T^2}\right)\end{aligned}$$

Since the Frobenius norm of a $N\times N$ matrix grows at $O(N)$, the first result follows.

The second result, thanks to lemma 7, is proven in a similar fashion, except that $\boldsymbol{\Theta}_{j,k}$ is unbiased for any sample size, we thus don't necessarily require its MSE —which is equivalent to its variance, in this case —to converge.

If $M=O(T^\eta)$ and $N=O(T^\nu)$ where $0<\nu<\frac{1}{4}$, then we can obtain the optimal growth of the smoothing window by minimizing $\sqrt{2^jT^{3\nu-\eta}+T^{2(\nu+\eta-1)}}$ with respect to η, such that, the solution has the following expression : $\eta^*=\frac{1}{3}(\nu+\frac{j}{\log_2(T)}+2)$. Substituting the latter in the general convergence rate gives the desired result. The same method applies to the optimal rate of $\|\boldsymbol{\Theta}_{j,k}\|$. $\square$

Appendix C Main Theorems

Theorem 1. *Under the MLSW framework and assumptions LFE,*

$$N^{-1}\left\|\widehat{\boldsymbol{S}}_j(k/T)-\boldsymbol{S}_j\left(k/T\right)\right\| = O_p\left(\sqrt{\frac{N}{2^{|j|}M}+\frac{M^2}{T^2}}\right)$$

where $\widehat{\boldsymbol{S}}_j(k/T)$ *is defined in* (4.3) *and* $\boldsymbol{S}_j(z)=\boldsymbol{\Lambda}_j(z)\boldsymbol{\Sigma}_F\boldsymbol{\Lambda}_j(z)+\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)$.
If one further assumes $N=O(T^\nu)$ *and* $M=O(T^\eta)$ *where* $0\leq\nu<\eta<1$, *then the optimal growth of the smoothing parameter is given by* $\eta^*=\frac{1}{3}(\nu-\frac{|j|+1}{\log_2 T}+2)$ *and thus,*

$$N^{-1}\left\|\widehat{\boldsymbol{S}}_j(k/T)-\boldsymbol{S}_j\left(k/T\right)\right\| = O_p\left(\frac{2^{\frac{|j|}{6}}}{T^{\frac{1}{3}(1-\nu)}}\right)$$

Proof. This theorem follows directly from lemma 9 by decomposing the norm as such,

$$\begin{aligned}\left\|\widehat{\boldsymbol{S}}_j(k/T)-\boldsymbol{S}_j\left(k/T\right)\right\| &= \left\|\boldsymbol{K}_{j,k}+\boldsymbol{\Theta}_{j,k}+\boldsymbol{\Theta}'_{j,k}+\boldsymbol{\Upsilon}_{j,k}-\boldsymbol{S}_j\left(k/T\right)\right\| \\ &\leq \left\|\boldsymbol{K}_{j,k}+\boldsymbol{\Upsilon}_{j,k}-\boldsymbol{S}_j\left(k/T\right)\right\|+2\left\|\boldsymbol{\Theta}_{j,k}\right\|\end{aligned}$$

The second result simply follows from the minimization of the rate of convergence with η as the decision variable. $\square$

Lemma 10. *Under the MLSW framework and assumptions LFE,*

$$\left\|\widehat{\boldsymbol{V}}_j(k/T)-\boldsymbol{V}_j(k/T)\right\| = o\left(\frac{1}{|j|}\right)O_p\left(\sqrt{\frac{N}{2^{|j|}M}+\frac{M^2}{T^2}}\right)$$

where $\widehat{\boldsymbol{V}}_j(k/T)$ *is defined in* (4.8) *and* $\boldsymbol{V}_j(k/T)$ *contains the eigenvalues of* $\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)\boldsymbol{\Sigma}_F$ *on its diagonal.*

If one further assumes $N=O(T^\nu)$ *and* $M=O(T^\eta)$ *where* $0\leq\nu<\eta<1$, *then the optimal growth of the smoothing parameter is given by* $\eta^*=\frac{1}{3}(\nu-\frac{|j|+1}{\log_2 T}+2)$ *and thus,*

$$\left\|\widehat{\boldsymbol{V}}_j(k/T)-\boldsymbol{V}_j(k/T)\right\| = o\left(\frac{1}{|j|}\right)O_p\left(\frac{1}{2^{\frac{|j|}{3}}T^{\frac{1}{3}(1-\nu)}}\right)$$

Proof. First of all take the classical covariance matrix estimator —i.e. $\widehat{\boldsymbol{\Sigma}}_F=\frac{1}{T}\sum_t\boldsymbol{F}_t\boldsymbol{F}'_t$ —and recall that,

$$\begin{aligned}\left\|\widehat{\boldsymbol{\Sigma}}_F\right\| &= O_p(1)\text{ by assumption F3.}\\ \left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\right\| &= o\left(\frac{1}{|j|}\right)\text{ by assumption L2.}\\ 2^{\frac{|j|}{2}}\left\|\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right\| &= O_p(1)\text{ by constraint (4.8)}\end{aligned}$$

Due to the decomposition of the estimated CEWS (B.9), its eigendecomposition can be written as

$$N^{-1}\left[\boldsymbol{K}_{j,k}+\boldsymbol{\Theta}_{j,k}+\boldsymbol{\Theta}'_{j,k}+\boldsymbol{\Upsilon}_{j,k}\right]\widehat{\boldsymbol{\Lambda}}_j(z)=\widehat{\boldsymbol{\Lambda}}_j(z)\widehat{\boldsymbol{V}}_j(z)\tag{C.1}$$

which is non-trivial only for finitely many scales —see assumption 3a.

Pre-multiply (C.1) by

$$\frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'$$

to obtain

$$\frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\left[N^{-1}\boldsymbol{K}_{j,k}+N^{-1}\boldsymbol{\Theta}_{j,k}+N^{-1}\boldsymbol{\Theta}'_{j,k}+N^{-1}\boldsymbol{\Upsilon}_{j,k}\right]\widehat{\boldsymbol{\Lambda}}_j(z)=\frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(z)\widehat{\boldsymbol{V}}_j(z)$$

Then by defining,

$$\boldsymbol{Q}_{j,k}=\frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T}) \tag{C.2}$$

$$\boldsymbol{P}_{j,k}=\frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}} \tag{C.3}$$

$$\boldsymbol{D}_{j,k}=\frac{1}{\sqrt{N}}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\left[N^{-1}\boldsymbol{\Theta}_{j,k}+N^{-1}\boldsymbol{\Theta}'_{j,k}\right]\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}} \tag{C.4}$$

The eigendecomposition we can show that the eigendecomposition may be expressed as

$$\boldsymbol{P}_{j,k}\boldsymbol{Q}_{j,k}+\boldsymbol{D}_{j,k}=\boldsymbol{Q}_{j,k}\widehat{\boldsymbol{V}}_j(k/T)$$

To see this, first note,

$$\begin{aligned}&\left\|\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\left[N^{-1}\boldsymbol{K}_{j,k}+N^{-1}\boldsymbol{\Upsilon}_{j,k}\right]\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}-\boldsymbol{P}_{j,k}\boldsymbol{Q}_{j,k}\right\|\leq\\
&\qquad\left\|\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\left[N^{-1}\boldsymbol{K}_{j,k}+N^{-1}\boldsymbol{\Upsilon}_{j,k}\right]\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}-\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}N^{-1}\boldsymbol{S}_j\left(k/T\right)\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right\|\\
&\qquad+\left\|\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}N^{-1}\boldsymbol{S}_j\left(k/T\right)\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}-\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})}{N}\widehat{\boldsymbol{\Sigma}}_F\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})2^{\frac{|j|}{2}}}{N}\right\|\end{aligned}$$

The first term converges asymptotically to zero, since, thanks to lemma 9, it's bounded by

$$\left\|\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\right\|\left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\right\|\left\|\left[N^{-1}\boldsymbol{K}_{j,k}+N^{-1}\boldsymbol{\Upsilon}_{j,k}-N^{-1}\boldsymbol{S}_j\left(k/T\right)\right]\right\|\left\|\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right\|=O_p(1)o\left(\frac{1}{|j|}\right)O_p\left(\sqrt{\frac{N}{2^{|j|}M}+\frac{M^2}{T^2}}\right)O_p(1)$$

The growth of the second term can be restricted in the same manner by,

$$\left\|\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\right\|\left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\right\|\left\|N^{-1}\boldsymbol{S}_j\left(k/T\right)-\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\widehat{\boldsymbol{\Sigma}}_F\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\right\|\left\|\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right\|$$

where, with lemma 2,

$$\begin{aligned}\left\|N^{-1}\boldsymbol{S}_j\left(k/T\right)-\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\widehat{\boldsymbol{\Sigma}}_F\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}\right\|&\leq\left\|\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\left[\mathrm{E}\left[\boldsymbol{F}_m\boldsymbol{F}'_m\right]-\widehat{\boldsymbol{\Sigma}}_F\right]\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'}{\sqrt{N}}+N^{-1}\mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k+s}\boldsymbol{\varepsilon}'_{j,k+s}\right]\right\|\\
&=o\left(\frac{1}{|j|}\right)O\left(\frac{1}{\sqrt{T}}\right)o\left(\frac{1}{|j|}\right)+o\left(\frac{1}{|j|\sqrt{N}}\right)\end{aligned}$$

The last line follows from assumption E4. —i.e. $\left\|N^{-1}\mathrm{E}\left[\boldsymbol{\varepsilon}_{j,k}\boldsymbol{\varepsilon}'_{j,k}\right]\right\| = o\left(\frac{1}{|j|\sqrt{N}}\right)$.

By following the classical argument initiated by Bai (2003), the matrix $\boldsymbol{Q}_{j,k}$ is invertible since it has full rank. We can thus write,

$$\left[\boldsymbol{P}_{j,k} + \boldsymbol{D}_{j,k}\boldsymbol{Q}_{j,k}^{-1}\right]\boldsymbol{Q}_{j,k} = \boldsymbol{Q}_{j,k}\widehat{\boldsymbol{V}}_j(k/T)$$

The matrix $\boldsymbol{Q}_{j,k}$ could be seen as columnwise-stacked orthogonal eigenvectors of the matrix $\boldsymbol{P}_{j,k} + \boldsymbol{D}_{j,k}\boldsymbol{Q}_{j,k}^{-1}$. Define $V^*_{j,k}$ to be the diagonal matrix of the diagonal elements of $\boldsymbol{Q}'_{j,k}\boldsymbol{Q}_{j,k}$. Let $\widehat{\Omega}_{j,k} = \boldsymbol{Q}_{j,k}\left[V^*_{j,k}\right]^{-\frac{1}{2}}$ be the matrix of unit length eigenvectors of $\boldsymbol{P}_{j,k} + \boldsymbol{D}_{j,k}\boldsymbol{Q}_{j,k}^{-1}$. Thus,

$$\left[\boldsymbol{P}_{j,k} + \boldsymbol{D}_{j,k}\boldsymbol{Q}_{j,k}^{-1}\right]\widehat{\Omega}_{j,k} = \widehat{\Omega}_{j,k}\widehat{\boldsymbol{V}}_j(k/T)$$

Note that, $\boldsymbol{P}_{j,k} + \boldsymbol{D}_{j,k}\boldsymbol{Q}_{j,k}^{-1}$ converges to $\boldsymbol{P}_{j,k} = \boldsymbol{\Sigma}_F^{\frac{1}{2}}\boldsymbol{\Sigma}_{j,k;\Lambda}\boldsymbol{\Sigma}_F^{\frac{1}{2}}$ since $\|\boldsymbol{D}_{j,k}\| = o_p\left(\frac{1}{|j|}\sqrt{\frac{2^jN}{M}}\right)$, by lemma 9. Then the matrix $\widehat{\boldsymbol{V}}_j(k/T)$ asymptotically contains the eigenvalues of $\boldsymbol{P}_{j,k}$. □

Lemma 11. *Under the MLSW framework and assumptions LFE,*

$$\left\|\boldsymbol{V}_j(k/T) - \boldsymbol{Q}'_{j,k}\boldsymbol{Q}_{j,k}\right\| = O_p\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}}\right)$$

where $\boldsymbol{Q}_{j,k}$ *is defined in* (C.2).

Proof. The eigendecomposition of the CEWS estimator defined in (4.8) allows us to write,

$$\widehat{\boldsymbol{V}}_j(k/T) = \frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}' N^{-1}\widehat{\boldsymbol{S}}_j(k/T)\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}$$

By Theorem 1 and lemma 2,

$$\widehat{\boldsymbol{V}}_j(k/T) = \frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}'\left[\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}}\boldsymbol{\Sigma}_F\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})}{\sqrt{N}} + \frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N} + O\left(\sqrt{\frac{1}{2^{|j|}NM} + \frac{M^2}{N^2T^2}}\right)\right]\frac{2^{\frac{|j|}{2}}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}$$

and thus,

$$\begin{aligned}\left\|\widehat{\boldsymbol{V}}_j(k/T) - \boldsymbol{Q}'_{j,k}\boldsymbol{Q}_{j,k}\right\| &\leq \left\|\frac{2^{|j|}\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N}\right\|\left(\left\|\frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N}\right\| + O\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}}\right)\right)\\ &= O_p(1)\left(o\left(\frac{1}{|j|\sqrt{N}}\right) + O\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}}\right)\right)\end{aligned}$$

□

Lemma 12. *Under the MLSW framework and assumptions LFE,*

$$\left\|\frac{2^{\frac{|j|}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N} - \boldsymbol{\Sigma}_F^{-\frac{1}{2}}\Omega_{j,k}\boldsymbol{V}_j(k/T)^{\frac{1}{2}}\right\| = O_p\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}}\right)$$

where $\Omega_{j,k}$ *contains the orthonormal eigenvectors of* $\boldsymbol{\Sigma}_F\boldsymbol{\Sigma}_{\boldsymbol{\Lambda};j}(z)$.

Proof. Recall that, if $V^*_{j,k}$ is the diagonal matrix of the diagonal elements of $\boldsymbol{Q}'_{j,k}\boldsymbol{Q}_{j,k}$, then

$$\begin{aligned}\boldsymbol{Q}_{j,k} &= \frac{2^{\frac{|j|}{2}}}{N}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{\frac{1}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})\\ \widehat{\Omega}_{j,k} &= \boldsymbol{Q}_{j,k}\left[V^*_{j,k}\right]^{-\frac{1}{2}}\end{aligned}$$

We can write,

$$\frac{2^{\frac{|j|}{2}}\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N} = \left[\widehat{\boldsymbol{\Sigma}}_F\right]^{-\frac{1}{2}}\boldsymbol{Q}_{j,k} = \left[\widehat{\boldsymbol{\Sigma}}_F\right]^{-\frac{1}{2}}\widehat{\Omega}_{j,k}\left[V^*_{j,k}\right]^{\frac{1}{2}}$$

Finally, the result is derived from assumption F3., lemma 11 and lemma 10,

$$\begin{aligned}\left[\widehat{\boldsymbol{\Sigma}}_F\right]^{-\frac{1}{2}} &\xrightarrow{p} \left[\boldsymbol{\Sigma}_F\right]^{-\frac{1}{2}}\\ \widehat{\Omega}_{j,k} &\xrightarrow{p} \Omega_{j,k}\\ \left[V^*_{j,k}\right]^{\frac{1}{2}} &\xrightarrow{p} \boldsymbol{V}_j(k/T)\end{aligned}$$

□

Theorem 2. *Under the MLSW framework and assumptions LFE, define*

$$\begin{aligned}\widehat{\boldsymbol{R}}_{j,k} &= \widehat{\boldsymbol{\Sigma}}_F\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})'\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{2^{\frac{|j|}{2}}N}\widehat{\boldsymbol{V}}_j^{-1}(k/T)\\ \boldsymbol{R}_{j,k} &= 2^{-\frac{|j|}{2}}\boldsymbol{\Sigma}_F^{\frac{1}{2}}\Omega_{j,k}\boldsymbol{V}_j(k/T)^{-\frac{1}{2}}\end{aligned}$$

then,

$$\left\|\widehat{\boldsymbol{R}}_{j,k} - \boldsymbol{R}_{j,k}\right\| = O_p\left(\sqrt{\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}}\right)$$

Proof. This theorem is a straightforward implication of assumption F3., lemma 12 and lemma 10. □

Theorem 3. *Under the MLSW framework and assumptions LFE,*

$$\frac{2^{\frac{|j|}{2}}}{\sqrt{N}}\left\|\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T}) - \boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{R}_{j,k}\right\| = o_p\left(\frac{|j|^2}{2^{|j|}}\sqrt{\frac{N}{M} + \frac{2^{|j|}M^2}{T^2}}\right)$$

where $\boldsymbol{R}_{j,k}$ is defined in lemma 2.

Proof. From the eigendecomposition (4.8), one can write,

$$\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T}) = N^{-1}\left[\boldsymbol{K}_{j,k} + \boldsymbol{\Theta}_{j,k} + \boldsymbol{\Theta}'_{j,k} + \boldsymbol{\Upsilon}_{j,k}\right]\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})\widehat{\boldsymbol{V}}_j^{-1}(k/T)$$

then, with $\widehat{\boldsymbol{R}}_{j,k}$ defined in lemma 2,

$$\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T}) - \boldsymbol{\Lambda}_j(\frac{k}{T})\widehat{\boldsymbol{R}}_{j,k} = \left[N^{-1}\boldsymbol{K}_{j,k} + N^{-1}\boldsymbol{\Theta}_{j,k} + N^{-1}\boldsymbol{\Theta}'_{j,k} + N^{-1}\boldsymbol{\Upsilon}_{j,k} - N^{-1}\boldsymbol{\Lambda}_j(\frac{k}{T})\boldsymbol{\Sigma}_F\boldsymbol{\Lambda}_j(\frac{k}{T})'\right]\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{2^{\frac{|j|}{2}}}\widehat{\boldsymbol{V}}_j^{-1}(k/T)$$

Taking the norm and by considering lemma 2, lemma 9 and lemma 10,

$$\frac{2^{\frac{|j|}{2}}}{\sqrt{N}}\left\|\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})-\boldsymbol{\Lambda}_j(\frac{k}{T})\widehat{\boldsymbol{R}}_{j,k}\right\| \leq \left[\left\|N^{-1}\boldsymbol{K}_{j,k}+N^{-1}\boldsymbol{\Upsilon}_{j,k}-N^{-1}\boldsymbol{S}_j\left(k/T\right)\right\|+\left\|\frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N}\right\|+2\left\|\frac{\boldsymbol{\Theta}_{j,k}}{N}\right\|\right]\left\|\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right\|\left\|\widehat{\boldsymbol{V}}_j^{-1}(k/T)\right\|$$

$$= \left[o_p\left(\sqrt{\frac{N}{2^{|j|}M}+\frac{M^2}{T^2}}\right)+o\left(\frac{1}{|j|\sqrt{N}}\right)+O\left(\sqrt{\frac{N}{2^{|j|}M}}\right)\right]O_p(2^{-\frac{|j|}{2}})o_p(|j|^2)$$

Given lemma 2, the results follows since the first term has the slowest rate of convergence. □

Theorem 4. *Under the MLSW framework and assumptions LFE,*

$$\left\|\widehat{\boldsymbol{F}}_k-2^{-\frac{|j|}{2}}\boldsymbol{R}_{j,k}^{-1}\boldsymbol{F}_k\right\|=o_p\left(\frac{|j|}{2^{|j|}}\sqrt{\frac{N}{M}+\frac{2^{|j|}M^2}{T^2}}\right)$$

Proof. The estimated factors can be expanded by using (4.1), (2.1) and (3.1) as to obtain

$$\widehat{\boldsymbol{F}}_k=\frac{2^J}{(2^J-1)}\sum_{p,l=-J}^{-1}\sum_{t'=0}^{T-1}\frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{N}\boldsymbol{F}_{t'}\sum_{t=0}^{T-1}\psi_{l,t'}(t)\psi_{p,k}(t)$$
$$+\frac{2^J}{(2^J-1)}\sum_{p,l=-J}^{-1}\sum_{t'=0}^{T-1}\frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\varepsilon}_{l,t'}}{N}\sum_{t=0}^{T-1}\psi_{l,t'}(t)\psi_{p,k}(t)$$

Analoguously to (Motta et al., 2011) and due to lemma 11, the inverse of rotation matrix $\boldsymbol{R}_{j,k}$, defined in lemma 2, can be expressed as

$$\boldsymbol{R}_{j,k}^{-1}=2^{\frac{|j|}{2}}\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})}{N}+2^{\frac{|j|}{2}}\left[\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}}\frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N}\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}+\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}}O\left(\frac{1}{2^{|j|}NM}+\frac{M^2}{N^2T^2}\right)\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right]\left(\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_F^-$$

Therefore,

$$\widehat{\boldsymbol{F}}_k-2^{-\frac{|j|}{2}}\boldsymbol{R}_{j,k}^{-1}\boldsymbol{F}_k=\frac{2^J}{(2^J-1)}\sum_{\substack{p,l=-J\\p\neq j\\l\neq j}}^{-1}\sum_{\substack{t'=0\\t'\neq k}}^{T-1}\frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{N}\boldsymbol{F}_{t'}\sum_{t=0}^{T-1}\psi_{l,t'}(t)\psi_{p,k}(t)$$
$$+\frac{2^J}{(2^J-1)}\sum_{p,l=-J}^{-1}\sum_{t'=0}^{T-1}\frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\varepsilon}_{l,t'}}{N}\sum_{t=0}^{T-1}\psi_{l,t'}(t)\psi_{p,k}(t)$$
$$+\left(\frac{2^J}{(2^J-1)}-1\right)\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})}{N}\boldsymbol{F}_k$$
$$+\left[\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}}\frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N}\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}+\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}}O\left(\frac{1}{2^{|j|}NM}+\frac{M^2}{N^2T^2}\right)\frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}}\right]\left(\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N}\right)^{-1}\widehat{\boldsymbol{\Sigma}}_F^-$$

and thus,

$$\left\|\widehat{\boldsymbol{F}}_k - 2^{-\frac{|j|}{2}}\boldsymbol{R}_{j,k}^{-1}\boldsymbol{F}_k\right\| \leq \frac{2^J}{(2^J-1)} \sum_{\substack{p,l=-J \\ p\neq j \\ l\neq j}}^{-1} \sum_{\substack{t'=0 \\ t'\neq k}}^{T-1} \left\| \frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{N} \right\| \|\boldsymbol{F}_{t'}\| \sum_{t=0}^{T-1} |\psi_{l,t'}(t)|\,|\psi_{p,k}(t)|$$

$$+ \frac{2^J}{(2^J-1)} \sum_{p,l=-J}^{-1} \sum_{t'=0}^{T-1} \left\| \frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\varepsilon}_{l,t'}}{N} \right\| \sum_{t=0}^{T-1} |\psi_{l,t'}(t)|\,|\psi_{p,k}(t)|$$

$$+ \left|\frac{2^J}{(2^J-1)} - 1\right| \left\| \frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'\boldsymbol{\Lambda}_j(\frac{k}{T})}{N} \right\| \|\boldsymbol{F}_k\|$$

$$+ \left[\left\| \frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}} \right\| \left\| \frac{\boldsymbol{\Sigma}_{j;\varepsilon}(k/T)}{N} \right\| \left\| \frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}} \right\| + \left\| \frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})'}{\sqrt{N}} \right\| O\left(\frac{N}{2^{|j|}M} + \frac{M^2}{T^2}\right) \left\| \frac{\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{\sqrt{N}} \right\| \right] \left\| \left(\frac{\boldsymbol{\Lambda}_j(\frac{k}{T})\widehat{\boldsymbol{\Lambda}}_j(\frac{k}{T})}{N} \right)^{-1} \right\| \left\| \widehat{\boldsymbol{\Sigma}}_F^{-1} \right\| \|\boldsymbol{F}_k\|$$

The first term is $O(\frac{1}{\sqrt{N}})$ since, it can be bounded by

$$\frac{2^J}{(2^J-1)} \bar{N}^2 \left(\sup_{t'} |\psi(t')|\right)^2 \left(\sup_{t'} \|\boldsymbol{F}_{t'}\|\right) \left(\sup_{t',p,l} \left\| \frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{N} \right\|\right) \sum_{\substack{p,l=-J \\ p\neq j \\ l\neq j}}^{-1} 2^{-\frac{|l|}{2}} 2^{-\frac{|p|}{2}}$$

and notice that due to Assumption D2, $\left\| \frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{\sqrt{N}} \right\|$ converges faster than $\left\| \frac{\boldsymbol{\Lambda}_p(\frac{k}{T})'\boldsymbol{\Lambda}_l(\frac{t'}{T})}{\sqrt{N}} \right\|$ since $\left\| \frac{\widehat{\boldsymbol{\Lambda}}_p(\frac{k}{T})}{\sqrt{N}} \right\| = O_p(2^{-\frac{|p|}{2}})$ (constraint A.1) and $\left\| \frac{\boldsymbol{\Lambda}_p(\frac{k}{T})}{\sqrt{N}} \right\| = o(|p|^{-1})$ (assumption L2.). Consequently, the latter upper bound is $o_p\left(\frac{1}{\sqrt{N}}\right)$.

The second term can be treated equivalently by using assumption 5.

Finally,

$$\left\|\widehat{\boldsymbol{F}}_k - 2^{-\frac{|j|}{2}}\boldsymbol{R}_{j,k}^{-1}\boldsymbol{F}_k\right\| = o_p\left(\frac{1}{\sqrt{N}}\right) + o_p\left(\frac{1}{\sqrt{N}}\right) + o_p\left(\frac{1}{2^J 2^{\frac{|j|}{2}}|j|}\right) + o_p\left(\frac{|j|}{2^{|j|}}\sqrt{\frac{N}{M} + \frac{2^{|j|}M^2}{T^2}}\right)$$

□

Theorem 5. *Under the MLSW framework and assumptions LFE-S-D1-D2,*

$$\frac{2^{\frac{|j|}{2}}}{\sqrt{N}} \left\|\widehat{\boldsymbol{C}}_{j,k} - 2^{-\frac{|j|}{2}}\boldsymbol{C}_{j,k}\right\| = o_p\left(\frac{|j|}{2^{|j|}}\sqrt{\frac{N}{M} + \frac{2^{|j|}M^2}{T^2}}\right)$$

Proof. The proof is straightforwardly implied by theorems 3 and 4. □