Economics School of Louvain - ESL

Economics School of Namur - ESN

Multiscale Factor Modelling of High Dimensional Locally Stationary Time Series

Author : LAFONTAINE Guillaume

Thesis Director : VAN BELLEGEM Sébastien

Thesis Reader : HAFNER Christian

Academic Year 2020-2021

Master in Economics – 120 credits – Focus : Research – Econometrics Orientation

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

Abstract

Extracting relevant information on the immense amount of time series data is paramount in contemporary multivariate analysis. Furthermore, those time series are usually non-stationary, serially and cross-correlated. Economic shocks and earnings announcements generally induce such series behaviours. In this paper, we provide a factor model based on the wavelet spectral representation of time series that allow us to handle both high dimensionality data and non-stationarity. On the one hand, the very essence of factor models is to encapsulate the variability of many variables in just a few common factors. On the other hand, Wavelets have been extensively used to capture abrupt changes and non-stationarities while allowing a sparser representation of data. We first define the model before developing an estimator for the time-varying factor loadings. We also report an empirical application.

Contents

Acknowledgments

This first master thesis concludes 9 months of research and study of factor models and wavelet theory.
First of all, I would like to thank my thesis supervisor, Professor Van Bellegem, for introducing me to fundamental research and more particularly to factor models and wavelet theory.
I would also like to thank Professor Hafner for his proofreading and encouragement.
I would like to acknowledge Yann Loux, Head of Asset and Liabilities Management (ALM) at Banque Internationale à Luxembourg (BIL), for his valuable comments and insight into the current evolution of interbank benchmark rates.
In addition, I would like to thank my parents and my brother for their unconditional love and support during this period of great uncertainty. They have been and will remain a great source of motivation.

I would like to address a more personal note on what this work has brought me. This thesis was my first contact with research work. I started out enthusiastic about this work, but I soon realised that it was not in this field that I felt most fulfilled. I have encountered many more disappointments than joys in this long journey, which is also part of a very peculiar period, the covid19 crisis.
Nevertheless, this work has highlighted great values, one of which is perseverance which is essential in research. Its teachings will resonate with me for a long time. I did not arrive at the destination I planned but the tools and insights I got by trying something new is what matter in the end.

1 Introduction

1.1 Factor Models

In this paper, we introduce the new factor model based on the evolutionary wavelet spectrum (EWS) developed by Koch (2015) in his thesis and apply it to the EURIBOR inter-bank benchmark rates. The essence of Factor models is to encapsulate the variability of many time series in a fewer number of constituents called common factors. This fits well in the analysis of EURIBOR benchmark rates which come in several tenors, —i.e. overnight, 1 week, 1 month, ... up to 12 months. Usually in order to analyse the dependencies of several time series, one uses a VAR or VARMA models. However such models cannot simply be estimated when the number of series (N) is larger than the number of time series observations (T) since there are only so many parameters that can be estimated with the observations given. Factor models constitute one way of dealing with this shortcoming of VAR models. A plethora of factor models is present in the literature. Most of the current models are based on the *approximate factor model* proposed by Chamberlain and Rothschild (1983) which contrast with the *static factor model* used by Ross (1976). The former is a generalization of the latter since it allows idiosyncratic components to be weakly correlated whereas the latter does not. Sims and Sargent (1977) provide the factor model with a dynamic form by way of spectral analysis. In other words they allow factors to be serially dependent. Yet Forni et al. (2000) noted that it is possible to adopt a static view to model a dynamic structure but this requires an intractable number of factors. The dynamic form is further developed by Forni et al. (2015) who contributed to the literature with a *generalized dynamic-factor model* which renders the *approximate factor model* dynamic, namely by having a dynamic structure and non-orthogonal idiosyncratic components.

Factors models are mostly used to tackle economic issues, whether macroeconomic, microeconomic or financial. This fact is well illustrated by Sims and Sargent (1977) which pointed out that under fairly general conditions on the behavior of macroeconomic variables, common factors spontaneously arise as a statistical framework to study macroeconomic issues. For instance, Forni and Lippi (2001) constructed a coincident indicator for the EURO zone with their *generalized dynamic-factor model* by way of aggregating several macroeconomic variables. Moreover, it is straightforward to see that the dynamic setting is a natural choice for dealing with business cycles. Aware of the deficiencies of VAR models when handling sparse data, Bernanke et al. (2005) combined factor models with VAR models to create their *factor-augmented vector autoregressive* model. The idea is to used common factors as additional information to analyse the effects of monetary policy shocks on macroeconomic variables. For instance, factors are thought to capture unobservable concepts such as "economic activity" or "credit condition". Another part of the literature is also attached to the idea of using factor models to enhance forecasting of time series. For instance, Stock and Watson (2002) showed that it is possible to consistently estimate the factor structure with principal components and apply their findings to improve forecasting, even in the case of small and idiosyncratic changes in factor loadings. Forni et al. (2015) further extended the *generalized dynamic-factor model*, that relies on two-sided filters, to the case of one-sided filters which better behave at the end of the observation period or for forecasting purposes.

Most primitive factor models assume underlying second-order stationarity of the data analysed. This assumption constitutes the fundamental support to develop any kind of asymptotic theory, in the sense that an increasing amount of similar information about a process is gathered with increasing observations. One of the first to admit the possibility for factor models to accommodate non-constant factor loadings, hence enabeling non-stationarities in the factor structure were Stock and Watson (2002). Time-varying factor loadings are nonetheless a recent development in the literature. The main reason for adopting loadings that change in time is to give them some of the dynamics, hence potentially reducing the number of common factors even further. However, it may complicate the analysis of factor models since the dynamics would be scatter across more dimensions. In this paper, we take the same stance as Eichler et al. (2011) and Motta et al. (2011) regarding the behavior of the covariance structure. We assume this structure to be smoothly changing over time. Motta et al. (2011) extended the *static* factor model to manage non-stationarities while Eichler et al. (2011) developed a non-stationary *dynamic factor model*. They both assumed the covariance matrix, or spectral density matrix in the case of dynamic factor model, to be differentiable with respect to time. However in this paper and as suggested by Motta et al. (2011), we use a weaker version of continuity —i.e. Lipschitz continuity. This continuity assumption assumes that the covariance structure to be differentiable *almost everywhere*, hence encompassing a broader class of functions. The theoretical framework we use to investigate non-stationarities is the commonly accepted "*rescaled time*" of Dahlhaus (1997). This re-scaling in time allows us to locally gather more and more information with approximately the same dynamics and thus to construct an appropriate asymptotic theory.
This paper is organized as follows. Section 1 we introduce the wavelet theory necessary for understanding the factor model. Section 2 we remind the Locally Stationary Wavelet model of Nason et al. (2000) and the extension made by Koch (2015) in his thesis. Section 3 deals with the construction of the factor model. We illustrate the factor model with the EURIBOR benchmark rates and discuss the importance of those rates in section 4. We finally conclude in section 5. Appendix A also contains simulation results and some convergence issues. Appendix B display out python implementation with some code examples.

1.2 Wavelets

The factor model analysed in this paper is based on the *locally stationary wavelet model* (LSW model) of Nason et al. (2000) which is itself based on wavelet functions. Their class of non-stationary random processes build upon a wavelet decomposition of a signal : the non-decimated wavelet decomposition [1]. This peculiar decomposition is extensively studied in Nason and Silverman (1995) from a statistical point of view. The basic idea is to allow the usual *discrete wavelet transform* (DWT) (Mallat (1989)) to be shifted at any locations and not just on dyadic ones. Note that the aforementioned "locations" are defined by the first scale namely $j = -1$ which provides the highest number of possible shifts.

Like the Fourier decomposition of a signal into different sine and cosine functions, orthonormal wavelets are functions that allow us to also build orthonormal basis in Hilbert spaces $L^2(\mathbb{R}^n)$. In other words we can represent arbitrary functions as a linear combination of those wavelets, which are usually supposed to be simpler and with desirable properties. Yet orthonormal wavelets are quite different from sine and cosine waves. First, the former have better localization properties. They allow a better trade-off between time and *frequency*[2] resolutions. Therefore they offer a natural way of dealing with non-stationarities by allowing parameters to depend both on time and frequency. Second, it is possible to use non-orthonormal wavelets to decompose a signal at a cost of redundancies in the representation, hence the need for some regularizations to guarantee identification of the parameters. Note that wavelets are part of a bigger literature on *Multiresolution Analysis*. For an introduction to multiresolution analysis and formal construction of wavelet functions we refer to Walnut (2004).
We can define any wavelet orthonormal basis $\{\psi_{j,k}(x)\}_{j,k\in\mathbb{Z}}$ from the mother wavelet $\psi(x)$ and father wavelet $\phi(x)$ with the *two-scale dilation equations* :

$$\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} g(k)\phi(2x-k) \tag{1.1}$$

$$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h(k)\phi(2x-k) \tag{1.2}$$

where $\{h(k)\}_{k\in\mathbb{Z}}$ and $\{g(k)\}_{k\in\mathbb{Z}}$ are respectively the high-pass and low-pass filter used in the pyramid algorithm of Mallat (1989). Those two filters constitute a *Quadrature Mirror Filter pair* since they satisfy the following condition : $|m_0(\gamma)|^2 + |m_1(\gamma)|^2 = 1$ where $m_0(\gamma)$ and $m_1(\gamma)$ are the frequency domain representations of the low-pass and high-pass filters respectively. This condition sums up all the necessary conditions on the mother and father wavelet to obtain an invertible transform.
For instance, the oldest known wavelet is the Haar wavelet. Given the indicator function $\chi_A(x)$, the father wavelet function $\phi_h(x) = \chi_{[0,1)}(x)$ and the mother wavelet has the following form : $\psi_h(x) = \chi_{[-\frac{1}{2},0)}(x) - \chi_{[0,\frac{1}{2})}(x)$. The corresponding quadrature mirror filters are : $g = \frac{1}{\sqrt{2}}(1,1)$ and

[1] This decomposition has several different names such as *stationary wavelet transform*, the *undecimated wavelet transform*, the *translation invariant wavelet transform* or the *maximal overlap wavelet transform*.

[2] Note that we cannot strictly speak about frequencies in the wavelet domain but rather about *scales* and *locations*

$h = \frac{1}{\sqrt{2}}(1, -1)$.

The *child wavelet* is a dilated and translated version of the mother wavelet and is defined as follows :

$$\bar{\psi}_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k) \tag{1.3}$$

where $j, k \in \mathbb{Z}$ are the scale and location parameter respectively. Note that the child wavelet is only defined for *dyadic* locations as indicated by the factor 2^j. It is customary in the literature to respect the so called *Meyer-Mallat* numbering of scales —i.e. $j = -1, \ldots, -\log_2(T)$. We thus follow that same practice in this paper.

In the LSW model, Nason et al. (2000) used non-decimated wavelets to decompose a random variable. We can define the *non-decimated child wavelet* with the formula :

$$\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - 2^{-(j+1)} k) \tag{1.4}$$

Contrarily to the decimated wavelet, this specification allows the shift to be independent of the scale. Therefore we get equally the same number of coefficients for every scale. The price to pay for this convenient property is the non-orthogonality of the wavelet functions. As a consequence, several coefficients carry identical information about the original signal.

The orthonormal wavelet representation of a signal can be summarized in the following definition.

Definition 1.1. *Given a mother wavelet $\psi(x)$, a father wavelet $\phi(x)$ and a child wavelet $\psi_{j,k}(x)$ as defined in (1.3), we can expand any function $f \in L^2$ in terms of the wavelet orthonormal basis $\{\psi_{j,k}(x)\}_{j,k \in \mathbb{Z}}$ such that*

$$f(x) = \sum_{j=-1}^{-J} \sum_{k=0}^{2^{-j}-1} \langle f, \psi_{j,k} \rangle \, \psi_{j,k}(x) + \sum_{k=0}^{2^{-J}-1} \langle f, \phi_{-J,k} \rangle \, \phi_{-J,k}(x) \tag{1.5}$$

where $J = \log_2(T)$ is the coarsest scale, $\phi_{j,k}(x)$ is the child wavelet associated with the father wavelet defined analogously to (1.3) and $\langle \cdot, \cdot \rangle$ is the inner product operator.

From this definition we get the wavelet representation of a function f with the wavelet coefficients $\langle f, \psi_{j,k} \rangle$ for all j, k and $\langle f, \phi_{-J,k} \rangle$ for all k and the coarsest scale J. We can see that the mother wavelet allows us to obtain the information contained in each separate scales whereas the father wavelet encapsulates all the remaining information. The uniqueness of this decomposition relies upon the orthogonality of the wavelet basis formed by the mother's child wavelets. However in case of non-decimated wavelets, we cannot simply use a linear combination of the inner products and the "basis" functions to represent our original function, considering that the inner product does not only contain information of a given basis function.

2 The Locally Stationary Wavelet Process

The factor model we develop in this paper is based on the Locally Stationary Wavelet model of Nason et al. (2000) which was extended by Koch (2015) in his thesis. The main idea is to expand the Cramér's Representation of stationary stochastic process to the wavelet domain. The Cramér's Representation has the following form :

$$X_t = \int_{-\pi}^{\pi} A(\omega) e^{-it\omega} d\xi(\omega) \tag{2.1}$$

where $A(\omega)$ is the frequency representation of the process, $e^{-it\omega}$ is the orthonormal basis in which we project X_t and $d\xi(\omega)$ are i.i.d zero-mean orthonormal increments with unit variance.
This form allows us to write the autocovariance of X_t in terms of the spectrum :

$$c_X(\tau) = \int_{-\pi}^{\pi} |A(\omega)|^2 e^{i\tau\omega} d\omega$$

where $|A(\omega)|^2$ is the spectrum of X_t.

2.1 The univariate LSW model

Nason et al. (2000) obtained a similar representation with non-decimated wavelet basis. The definition of the Locally Stationary Wavelet process is central for the development of the factor model, hence we recall it here entirely.

Definition 2.1. *The univariate LSW model is a sequence of doubly-indexed stochastic process $\{X_{t,T}\}_{t=1}^{T}$, having the representation in the mean-square sense :*

$$X_{t;T} = \sum_{j=-J}^{-1} \sum_{k} w_{j,k;T} \psi_{jk}(t) \xi_{jk} \tag{2.2}$$

where $\xi_{j,k}$ is a random orthonormal increment sequence and where $\{\psi_{j,k}(t)\}_{j,k}$ is a discrete non-decimated family of wavelets for $j = -1, \dots, -J$, $k = 0, \dots, T-1$, based on a mother wavelet $\psi(t)$ of compact support.
The quantities in representation (2.2) have the following properties.

(i)

$$\mathbb{E}(\xi_{j,k}) = 0. \tag{2.3}$$

(ii)

$$\mathbb{C}ov(\xi_{j,k}, \xi_{l,m}) = \delta_{j,l}\delta_{k,m}. \tag{2.4}$$

where $\delta_{j,l}$ is the Kronecker delta.

(iii) There exists for each $j \leq -1$ a Lipschitz continuous function $W_j(z)$ in rescaled time $z \in (0,1)$ which fulfils the following properties :

$$\sum_{j=-\infty}^{-1} |W_j(z)|^2 < \infty \text{ uniformly in } z \in (0,1) \tag{2.5}$$

the Lipschitz constants L_j are uniformly bounded in j and

$$\sum_{j=-\infty}^{-1} 2^{-j} L_j < \infty \tag{2.6}$$

there exists a sequence of constants C_j such that for each T

$$\sup_k \left| w_{j,k;T} - W_j(\frac{k}{T}) \right| \leq \frac{C_j}{T} \tag{2.7}$$

where for each $j = -1, \ldots, -J$ the supremum is over $k = 0, \ldots, T-1$ and where $\{C_j\}$ fulfils

$$\sum_{j=-\infty}^{-1} C_j < \infty \tag{2.8}$$

With this wavelet decomposition of the stochastic process $X_{t,T}$ we can define the *evolutionary wavelet spectrum* as $S_j(z) = |W_j(z)|^2$, which is finite and Lipschitz continuous by assumption. And similarly to the Fourier representation, we can specify the *local* autocovariance of the process $X_{t,T}$ with the formula :

$$c(z,\tau) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau) \qquad\qquad \tau \in \mathbb{Z},\ z \in (0,1) \tag{2.9}$$

where $\Psi_j(\tau)$ is the autocorrelation wavelet function characterized by $\Psi_j(\tau) := \sum_k \psi_{j,k}\psi_{j,k-\tau}$ for all $j < 0, \tau \in \mathbb{Z}$. This autocovariance is *local* in the sense that it depends on rescale time $z \in (0,1)$ and on the autocorrelation wavelet function which is compactly supported.

In empirical situation we don't have access to the increments ξ_{jk} and the wavelet coefficients $w_{j,k;T}$ separately. We are only able to observe the empirical wavelet coefficients $d_{j,k;T}$ from the

Stationary Wavelet Transform (SWT) defined by :

$$d_{j,k;T} := \sum_{t=1}^{T} X_{t;T}\psi_{j,k}(t) \tag{2.10}$$

However the raw periodogram $I_{j,k;T} = |d_{j,k;T}|^2$ is a biased estimator of the evolutionary wavelet spectrum $S_j(z)$, hence it needs to be corrected. This correction is made possible with the autocorrelation wavelet functions $\Psi_j := \{\Psi_j(\tau)\}_{\tau\in\mathbb{Z}}$ which we stack in a matrix $\Psi := (\Psi_0^\top, \ldots, \Psi_J^\top)$ and form the corresponding Gramian matrix $A_J := \Psi^\top\Psi$. Note that the matrix Ψ is an infinite dimensional object which needs to be truncated based on the compact support of the autocorrelation wavelet of the coarsest scale $-J$. It can be showed that the operator A_J is invertible since the columns of Ψ are linearly independent (see Nason et al. (2000)).
If we define $\bar{A}_J := A_J^{-1}$. We can finally get the corrected wavelet periodogram with the following formula :

$$\bar{I}_{j,k;T} = \sum_{l=-1}^{-J} \bar{A}_{l,k} I_{j,k;T} \tag{2.11}$$

which is an unbiased estimator of the evolutionary wavelet spectrum. In order to get consistency of the estimator we have to smooth it with some kind of smoothing procedure —i.e. kernel smoothing, translation invariant de-noising (Coifman and Donoho (1995)).

2.2 Extension of the LSW model

Building upon that univariate LSW model, Koch (2015) relaxed the assumption (2.4) by allowing increments to be cross-correlated. However he limited the degree of correlation among increments by the *order* of the LSW model. This order of the LSW denotes by how many neighbouring scales a particular scale j is correlated to. Formally, this is achieved by giving assumption (2.4) the new following form :

$$\mathbb{C}ov(\xi_{j,k}, \xi_{j+i,l}) = \begin{cases} \Gamma_{i,j}\left(\frac{k}{T}\right) & \text{if } -q \geq i \geq -1 \text{ and } k = l \\ 0 & otherwise \end{cases} \tag{2.12}$$

Note that the covariance of the increments is also assumed to be time-varying and dependent on scale and order of the model. However we only allow increments to be contemporaneously dependent, hence not serially correlated.
We can redefine the EWS functions as $S_{i,j}(z) = W_j(z)W_{j+i}(z)\Gamma_{i,j}(z)$, which is also assumed to be finite and Lipschitz continuous in rescaled time $z \in (0, 1)$ for every scale j and order i.

Koch (2015) also develop the *multivariate LSW model* which allows the EWS to be correlated between different processes. Each elements of the multivariate case is modelled as a univariate LSW model. Idiosyncratic increments of one process are now allowed to be correlated with idiosyncratic increments of other processes. Koch (2015) adapted to the multivariate framework the following :

$$\mathbb{C}ov(\xi_{j,k}^{(u)}, \xi_{j+i,l}^{(v)}) = \begin{cases} \Gamma_{i,j}^{(u,v)}\left(\frac{k}{T}\right) & \text{if } -q \geq i \geq -1 \text{ and } k = l \\ 0 & otherwise \end{cases} \tag{2.13}$$

$$S_{i,j}^{(u,v)}(z) = W_j^{(u)}(z) W_{j+i}^{(v)}(z) \Gamma_{i,j}^{(u,v)}(z) \tag{2.14}$$

Note that even though $S_{i,j}^{(u,v)}(z)$ and $W_{j+i}^{(u)}$ are assumed Lipschitz continuous for all j, i, u, v, this property does not directly carry over the covariance term $\Gamma_{i,j}^{(u,v)}(z)$. Lemma 1 shows that $W_{j+i}^{(v)}(z) \neq 0$ is a necessary condition to have $\Gamma_{i,j}^{(u,v)}(z)$ Lipschitz continuous. However this condition is not always fulfilled in case of the LSW model, which may result in non-Lipschitz continuous behavior of the covariance. Each and every theorems and findings of Nason et al. (2000) can be transposed to this multivariate LSW model. We do not, however, state them here since they are straightforward extensions of the univariate case. This extension is essential for developing a factor model since we are now able to state the cross-correlation between the processes (u) and (v) in terms of the Evolutionary Wavelet Spectrum, which gives us the possibility to model the covariance matrix of the processes. The empirical periodogram also needs to be corrected by some kind of Gramian matrix and then smoothed to achieve consistency, all the details about the estimation procedure can be found in Koch (2015).

There is however one drawback from using this representation of multivariate processes. Unlike most factor models which use the concept of double asymptotics —i.e. T and N go to infinity simultaneously without restrictions, the consistency of the EWS requires the cross-sectional dimension to not grow faster than T^α with $0 \leq \alpha < \frac{1}{3}$, hence the factor model will inherit this restriction.

3 The Factor model

3.1 Classical factor models

In this section we go over the main theory of stationary factor models. Factor models allow the characterization of the variability of many variables in fewer constituents called *common factors*. Let N be the number of cross-sectional processes and T the number of observations. Formally, the factor model takes the following form :

$$X_t = \Lambda F_t + \epsilon_t \qquad \text{for } t = 1, \ldots, T \tag{3.1}$$

where X_t is a $(N \times 1)$ vector of observations with $\mathbb{E}(X_t) = 0$ and covariance matrix Σ_X, F_t is a $(K \times 1)$ vector of common factors with covariance matrix Σ_F, Λ is a $(N \times K)$ matrix containing the factor loadings —i.e. the mapping from the factor space to the original space of processes. And ϵ_t is a $(N \times 1)$ vector of idiosyncratic components independent of the factors with covariance matrix Σ_ϵ. It is customary to define the term $C_t = \Lambda F_t$ the $(N \times 1)$ vector of *common components*. It is assumed that the number of factors $K < p$ in order to have fewer constituents describing the variability of X_t. Having $K = p$ would only produce factors that are a reorganization of the information contained in X_t without reducing the dimensionality. Idiosyncratic components are information not explained by the common factors and are usually assumed to affect a limited number of processes if not only one. Note that in this form the factor loadings don't depend on time which renders difficult the modeling of non-stationarities.

The major benefit of using such a decomposition (3.1) is a simpler and sparser characterization of the second moments of the observations in terms of the common factors and factor loadings :

$$\Sigma_X = \Lambda \Sigma_F \Lambda^\top + \Sigma_\epsilon \tag{3.2}$$

The *static* representation assumes this matrix to be diagonal —i.e. the idiosyncratic errors are independent of one another. However, this assumption seems too restrictive when dealing with large cross-dimensions, where shocks are likely to be correlated to some extend. The *approximate* factor model of Chamberlain and Rothschild (1983) relaxed this assumption by imposing a bound on the largest eigenvalue of Σ_ϵ. In this way idiosyncratic errors are allow to be "weakly" correlated. By "weakly" we mean correlated to a finite number of other idiosyncratic errors. This assumption formalizes the intuition of Sims and Sargent (1977) who mentioned the fact that for the factor model decomposition to make sense the variance of the idiosyncratic must be small compared to the variance of the process X_t.

However, the decomposition of the covariance (3.2) is not unique, hence non-identified. To see this we can construct new factors $F_t^* = R^{-1} F_t$ and loadings $\Lambda^* = \Lambda R$, where R is an invertible matrix, without changing the representations (3.1) and (3.2). To solve this inherent indeterminacy, several methods have been studied int he literature, see Anderson and Rubin (1956). The most common solution is to make additional assumptions on the behavior of factors and/or factor loadings. For instance, assuming normalized and orthogonal factors such that $\mathbb{E}\left(F_t F_t^\top\right) = I_K$, where I_K is the identity matrix of rank K, allows us to restate the second moment of X_t as :

$$\Sigma_X = \Lambda\Lambda^\top + \Sigma_\epsilon$$

which limits the indeterminacy to an orthonormal transformation such that $RR^\top = I_K$. By using the eigendecomposition of the matrix $\Lambda\Lambda^\top$ and taking $\Lambda^\top\Lambda$ to be a diagonal matrix, it is possible to show that we can further reduce the indeterminacy up to a multiplication by ± 1 (see Anderson and Rubin (1956)).

3.2 Non-stationary Factor models

In order to cope with non-stationarities in the data, we can expand the representation (3.1) by making the loadings time-varying and more specifically dependent on *rescaled time* as proposed by Dahlhaus (1997) :

$$X_{t;T} = \Lambda\left(\frac{t}{T}\right) F_t + \epsilon_t \qquad t = 1, \ldots, T \tag{3.3}$$

which gives a time dimension to the second-order moments of the observations,

$$\Sigma_X\left(z\right) = \Lambda(z)\Sigma_F\Lambda^\top(z) + \Sigma_\epsilon \tag{3.4}$$

where $z = \frac{t}{T} \in (0, 1)$. This "rescaling" in time is key to construct an appropriate asymptotic theory in the non-stationary framework. The very nature of non-stationary phenomena is to change their structures over time, thus rendering future observations meaningless for analysing the beginning of the process. Rescaled time allows us to locally gather more and more information with approximately the same dynamics, hence permitting an asymptotic reasoning.

4 LSW Factor Model

We are now in position to describe the factor model introduced by Koch (2015) based on the Locally Stationary Wavelet decomposition of processes of Nason et al. (2000). We first state the factor model in its stacked form before making the necessary assumptions on the factors, loadings and error terms. Observe that we assume that there exists a factor structure in the population. However more theoretical conditions should be given to ensure such a structure. We believe that it could be possible to develop a representation theory as in Forni and Lippi (2001).

The next assumptions recall the LSW decomposition to remind the reader that we are working in the wavelet domain, hence on the empirical wavelet coefficients and not on the processes $X_{t;T}^{(u)}$ directly.

ASSUMPTIONS P :

1. Let $X_T(t) = \left(X_T^{(1)}(t), \dots, X_T^{(N)}(t)\right)^\top$ be a multivariate process where each component follows a locally stationary wavelet process of order q for each $t = 1, \dots, T > 0$ and $N = O(T^\alpha)$ with $0 \leq \alpha < \frac{1}{3}$:

$$X_T^{(u)}(t) = \sum_{j=-J}^{-1} \sum_k W_j^{(u)}(z) \xi_{jk}^{(u)} \psi_{jk}(t)$$

   which was introduced in definition (2.1).

2. The increments $\xi_{jk}^{(u)}$ have the following properties for each $k = 1, \dots, T$, $j = -1, \dots, -J = -\log_2(T)$, $u, v = 1, \dots, N$:

   (a) $\mathbb{E}(\xi_{jk}^{(u)}) = 0$

   (b) $Var(\xi_{jk}^{(u)}) = 1$

   (c) The covariance of the increments $\xi_{jk}^{(u)}$ is defined as :

$$\mathbb{C}ov(\xi_{j,k}^{(u)}, \xi_{j+i,l}^{(v)}) = \begin{cases} \Gamma_{i,j}^{(u,v)}\left(\frac{k}{T}\right) & \text{if } -q \geq i \geq -1 \text{ and } k = l \\ 0 & otherwise \end{cases} \tag{4.1}$$

   As noted previously the order of the LSW model is reflected in this covariance since this latter allows increments to be correlated between scales.

Note that the cross-section is constrained by the number of time series observations. The fact is inherited from the consistency of the wavelet periodogram. The asymptotic theory will thus only depend on T.

The LSW factor model is entirely developed on the empirical wavelet coefficients coming from the Stationary Wavelet Transform (SWT)(Nason and Silverman (1995)). It has the following form :

$$d_{j,k}^{(u)} = \sum_{m=1}^{K} \Lambda_{j,m}^{(u)}(z) F_{m,k} + \epsilon_{j,k}^{(u)} \tag{4.2}$$

for each $j = -, \dots, -J$, $k = 0, \dots, T$ and $u = 1, \dots, N$.

4.1 Packing of the model

We can create a succinct notation for the whole factor model by grouping similar objects:

1. Let $d_{j,k} = (d_{j,k}^{(1)}, \dots, d_{j,k}^{(N)})^\top$ and define $d_{JN,k} = (d_{-1,k}^\top, \dots, d_{-J,k}^\top)^\top$, the $(JN \times 1)$ vector that stacks $d_{j,k}$ for every $j = -1, \dots, -J$.
2. Let $F_k = (F_{1,k}, \dots, F_{K,k})^\top$ be the $(K \times 1)$ vector of factors at a given time k.
3. Let $\Lambda_{j,m}(z) = (\Lambda_{j,m}^{(1)}(z), \dots, \Lambda_{j,m}^{(N)}(z))^\top$ and define $\Lambda_{JN,m}(z) = (\Lambda_{-1,m}(z)^\top, \dots, \Lambda_{-J,m}(z)^\top)^\top$ to be the $(JN \times 1)$ vector that stacks $\Lambda_{j,m}(z)$ for every $j = -1, \dots, -J$. Finally set $\Lambda_{JN}(z) = (\Lambda_{JN,1}(z), \dots, \Lambda_{JN,K}(z))$ as the $(JN \times K)$ matrix that regroups the loadings for every factors.
4. Let $\epsilon_{j,k} = (\epsilon_{j,k}^{(1)}, \dots, \epsilon_{j,k}^{(N)})^\top$ and define $\epsilon_{JN,k} = (\epsilon_{-1,k}^\top, \dots, \epsilon_{-J,k}^\top)^\top$, the $J(N \times 1)$ vector that stacks $\epsilon_{j,k}$ for every $j = -1, \dots, -J$.

With those notations we can rewrite the factor model (4.2) as :

$$d_{JN,k} = \Lambda_{JN}(z)F_k + \epsilon_{JN,k} \qquad \forall k = 1, \dots, T. \tag{4.3}$$

and the evolutionary wavelet spectrum matrix as :

$$S_{JN}(z) = \Lambda_{JN}(z)\Lambda_{JN}^\top(z) + \Sigma_\epsilon(z) \tag{4.4}$$

Recall that we assume this second-order structure to be Lipschitz continuous which may not constitute a restrictive assumption if the modelled dynamics are smoothly changing. However, in case of "jump" process —i.e. financial time series, this assumption is not adequate. For instance, Van Bellegem and von Sachs (2008) showed that it would still be possible to define the EWS with bounded variation instead of Lipschitz continuity.

4.2 Assumptions on the LSW factor model

The assumptions used for the decomposition (4.3) are classical in the factor model literature in the sense that they allow to identify the model and to make it well-behaved.

ASSUMPTIONS F :

1. F_k is a zero-mean white noise process with covariance matrix $\Sigma_F = I_K$ where I_K is the identity matrix of rank K.

ASSUMPTIONS L :

1. $\Lambda_{JN}(z)^\top \Lambda_{JN}(z)$ is a $(K \times K)$ diagonal matrix.
2. The Euclidean norm of the $(JN \times 1)$ vector $\Lambda_{JN,m}(z)$ is such that $\|\Lambda_{JN,m}(z)\| = O(N^\kappa)$ with $0 < \kappa \leq 1$.

ASSUMPTIONS E :

1. $\epsilon_{JN,k}$ is a zero-mean white noise ($JN \times 1$) vector with covariance matrix $\Sigma_\epsilon(z)$ which is assumed to have bounded largest eigenvalue λ_m for all $z \in (0,1)$ such that $\lambda_{max}(\Sigma_\epsilon(z)) = O(1)$.

2. $F_{m,k} \perp\!\!\!\perp \epsilon_{j,k}^{(u)}$ for all k, m, j and u.

Assumptions F regroup assumptions on the factors. In particular assumption F1 allows the identification of the model along with assumption L1 as previously discussed in section 3. Moreover given that the covariance matrix of the common factors is identity, factors are assumed to have a normalized variance of 1 and to be independent of one another. Each Factor is thus responsible to encapsulate different information about several processes at multiple scales. Assumption L1 about factor loadings also allows identification of the model. Unlike classical factor models where loadings are assumed to be bounded, assumption L2 only restrict the growth of the euclidean norm of the loadings. In other words, loadings cannot grow too fast with respect to the number of cross-sections. This condition also underlines the fact that each factor must have a non-vanishing importance in explaining the original processes. Note that we do not assume some kind of continuity on the factor loadings. We could have assumed loadings to be Lipschitz continuous, although proofs of convergence would have been a mere reproduction of the proofs of Motta et al. (2011), at the exception of adding *almost everywhere* differentiability. We will discuss the implication of continuity of factor loadings in the subsequent section. Most of the Assumptions E on the idiosyncratic error terms are in accordance with the *approximate factor model* of Chamberlain and Rothschild (1983). The covariance matrix of errors is assumed to be time-varying and with bounded largest eigenvalue which makes the idiosyncratic components correlated to at most a finite set of other idiosyncratic terms as the cross-section grows. The idiosyncrasy of those error terms comes from the fact that they are assumed to not be correlated to any past, present or future factors. Assumption E3 formalizes this aspect. $\epsilon_{JN,k}$ is thus a vector of information that is not explained by the factors but is still present in one or a finite number of processes.

4.3 Discussion

One important question that does not arise in stationary factor model but does in non-stationary ones is the fact that time-varying eigenvalues of the time-varying covariance matrix may intersect at some rescaled time $z_0 \in (0,1)$. This poses an important problem concerning the continuity of eigenvectors —i.e. the continuity of factor loadings. It is commonly known that eigenvalues of a continuous matrix function are continuous (see Katō (1995)). In other words, a matrix where each of its entries are continuous also produces continuous eigenvalues. However we cannot say the same thing for eigenvectors, especially when the multiplicity of eigenvalues is greater than 1. To illustrate that point, take the identity matrix of rank 2 with a small perturbation ϵ on the main diagonal and compute the two corresponding eigenvectors v_1 and v_2,

$$\begin{pmatrix} 1+\epsilon & 0 \\ 0 & 1+\epsilon \end{pmatrix} \qquad v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

However now if we consider a small perturbation on the ascending diagonal of the identity matrix we obtain,

$$\begin{pmatrix} 1 & \epsilon \\ \epsilon & 1 \end{pmatrix} \qquad \tilde{v_1} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \qquad \tilde{v_2} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

which gives us completely different eigenvectors $\tilde{v_1}$ and $\tilde{v_2}$. One way to circumvent this issue is to assume a strictly positive eigengap —i.e. eigenvalues must have multiplicities of 1, hence be different. Another way to avoid "jumps" of eigenvectors is to make other restrictions on the eigenvalues. For instance, Motta et al. (2011) allowed eigenvalues to be equal at some points but prevented those eigenvalues to have equal derivatives at those points. This issue was not pointed out by Koch (2015) in his thesis. Koch (2015) used the Davis-Kahan theorem (Davis and Kahan (1970)) in order to prove convergence of the empirical eigenvectors to the true eigenvectors of the EWS matrix. However the Davis-Kahan theorem requires a strictly positive eigengap, which was not assumed by Koch (2015). Fortunately for Koch (2015) who was not cautious in its use of this theorem, Eldridge et al. (2017) proved that the Davis-Kahan theorem also holds in case of eigenvalues multiplicities.

Following the factor model of Koch (2015), we don't assume that eigenvalues have multiplicities of 1, thus the corresponding eigenvectors may not be continuous. This will imply that factor loadings will not always show continuous behaviour. This does not mean, however, that they do not evolve over time.

4.4 LSW factor model and principal component analysis

In classical factor model, we can develop estimators for the common factors and loadings with the principal component analysis. This method amounts to solving a non-linear least square problem and take the optimum parameters as estimators :

$$\begin{aligned} \left(\hat{F}_t, \hat{\Lambda}_N\right) := \arg \min_{F_t, \Lambda_N} &= (NT)^{-1} \sum_{t=1}^{T} (X_t - \Lambda_N F_t)^\top (X_t - \Lambda_N F_t) \\ s.t. & \qquad \frac{\Lambda_N^\top \Lambda_N}{N} = I_K \end{aligned}$$

We can straightforwardly solve this problem with known methods and find the optimal factors and loadings that obey :

$$\begin{aligned} N^{-1}\hat{\Sigma}_X\bar{\Lambda}_N &= \bar{\Lambda}_N\hat{V}_N \\ \hat{\Lambda}_N &= \sqrt{N}\bar{\Lambda}_N \\ \hat{F}_t &= \hat{\Lambda}_N^\top X_t/N \end{aligned}$$

where $\hat{\Sigma}_X$ is the consistent estimator of the covariance between the processes given by $T^{-1}\sum_{t=1}^T X_tX_t^\top$, $\hat{V}_N$ and $\bar{\Lambda}_N$ are the empirical matrix of eigenvalues and eigenvectors of $\hat{\Sigma}_X$ respectively. The consistent estimator of factors loadings is equal to the empirical eigenvectors rescaled by $\sqrt{N}$. Note that if the model has K common factors, we only take the first K columns of the matrix $\bar{\Lambda}_N$ as estimated factor loadings. Note also that $\hat{\Sigma}_X$ is a covariance matrix, it is a symmetric positive semi-definite matrix which implies that its eigenvectors will be of unit length (see Lütkepohl (1996)).

In the case of non-stationary factor models, the minimization problem (**??**) becomes time-varying. For instance, Motta et al. (2011) used a kernel in the time domain to make this optimization problem time dependent. In the present case, we address this issue in the wavelet domain by way of minimizing the time-varying non-linear least square problem with the empirical wavelet coefficients for all $k = 1, \dots, T$:

$$L_{JN;T}(k) = \min_{F_k,\Lambda_k} (JN)^{-1}\Big(d_{JN,k} - \Lambda_{JN}\big(\frac{k}{T}\big)F_k\Big)^\top\Big(d_{JN,k} - \Lambda_{JN}\big(\frac{k}{T}\big)F_k\Big) \tag{4.5}$$

$$s.t. \qquad \frac{\Lambda_{JN}\big(\frac{k}{T}\big)^\top\Lambda_{JN}\big(\frac{k}{T}\big)}{JN} = I_K \tag{4.6}$$

Note that we do not have to have a summation term going through the time index k. As we have seen in previous section, the empirical wavelet coefficients alone are sufficient to get an estimator of the time-domain covariance of the processes —i.e. by virtue of (2.9) and (2.11).
By taking the first order condition with respect to F_k, we obtain,

$$\Lambda_{JN}\left(\frac{k}{T}\right)^\top\Lambda_{JN}\left(\frac{k}{T}\right)F_k = \Lambda_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k}$$

with the constraint (4.6), we can further isolate the factor,

$$F_k = (JN)^{-1}\Lambda_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k} \tag{4.7}$$

Next, in order to find an estimator for the factor loadings, we can replace (4.7) in the initial minimization problem (4.5) and obtain,

$$\tilde{L}_{JN;T}(k) = \min_{\Lambda_k} (JN)^{-1} \left[d_{JN,k}^\top d_{JN,k} - (JN)^{-1} d_{JN,k}^\top \Lambda_{JN}\left(\frac{k}{T}\right) \Lambda_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k} \right]$$

which is equivalent to maximizing the following,

$$\max_{\Lambda_k} (JN)^{-1} \left[d_{JN,k}^\top \Lambda_{JN}\left(\frac{k}{T}\right) \Lambda_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k} \right]$$

Note that the term that needs to be maximized is a scalar —i.e. a (1×1) matrix. By taking the trace of that matrix and using the property of the trace, we have,

$$\max_{\Lambda_k} (JN)^{-1} tr \left\{ \Lambda_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k} d_{JN,k}^\top \Lambda_{JN}\left(\frac{k}{T}\right) \right\} \tag{4.8}$$

we obviously recognize the raw wavelet periodogram in its stacked form $I_{JN,k} = d_{JN,k} d_{JN,k}^\top$. As already mentioned in section 2, $I_{JN,k}$ is a biaised estimator of the Evolutionary Wavelet Spectrum $S_{JN}(z)$. If we do not replace this raw periodogram by its smoothed and corrected version $\tilde{I}_{JN,k}$, we will not be able to express the covariance of the processes $X_{t;T}$ in terms of the factors and factor loadings as given by (4.4). The right maximization problem would then be ,

$$\max_{\Lambda_k} (JN)^{-1} tr \left\{ \Lambda_{JN}\left(\frac{k}{T}\right)^\top \tilde{I}_{JN,k} \Lambda_{JN}\left(\frac{k}{T}\right) \right\} \tag{4.9}$$

The estimator of the factor loadings is then defined as the scaled eigenvectors of the smoothed and corrected wavelet periodogram.

$$\hat{\Lambda}_{JN}\left(\frac{k}{T}\right) = \sqrt{JN}\, \bar{\Lambda}_{JN}\left(\frac{k}{T}\right) \qquad \forall k = 1, \ldots, T \tag{4.10}$$

$$(JN)^{-1} \tilde{I}_{JN,k}\, \bar{\Lambda}_{JN}\left(\frac{k}{T}\right) = \bar{\Lambda}_{JN}\left(\frac{k}{T}\right) \hat{V}_{JN,k} \qquad \forall k = 1, \ldots, T \tag{4.11}$$

With that estimator of the loadings we get the estimator of the factors via (4.7) :

$$\hat{F}_k = (JN)^{-1} \hat{\Lambda}_{JN}\left(\frac{k}{T}\right)^\top d_{JN,k} \qquad \forall k = 1, \ldots, T \tag{4.12}$$

Note that this estimator is not directly identified since we decompose the signals using non-orthogonal wavelets. The empirical wavelet coefficients $d_{JN,k}$ are thus not unique. However, using the *algorithme à trous* (Holschneider et al. (1990)) to compute the *Stationary Wavelet Transform*

(SWT), we get a decomposition which is the minimal L_2 norm solution. This comes from the fact that the SWT contains the coefficients of all possible shifted *Discrete Wavelet Transform* (DWT) (see Nason and Silverman (1995)). Moreover Coifman and Donoho (1995) pointed out that this transform exactly corresponds to the minimum L_2 norm solution.

4.4.1 Asymptotic estimation theory

Most of the asymptotic theory was developed by Koch (2015). However we are going to extend it by providing consistency of the common factors. As already mentioned in a previous section (see Section 2), the cross-section N cannot grow faster than T^α with $0 \leq \alpha < \frac{1}{3}$ in order to have consistency of the smoothed and corrected wavelet periodogram $\tilde{I}_{JN,k}$. Our asymptotic reasoning is thus solely governed by $T \to \infty$. The number of decomposition J used in the LSW process is considered fixed. We only display it in order to remind the reader that we are working in the wavelet domain. Recall that we cannot make $J \to \infty$ since it is then not possible to consistently estimate the EWS (Koch (2015)).
From now on, we consider $\|\cdot\|$ to be the Frobenius norm, or the L_2 norm in case of vectors.

We first remind the theorem of Koch's thesis concerning consistency of factor loadings.

Theorem 1. *Under Assumptions PFLE, and given the rescaled time* $z = \frac{k}{T}$*, the degree of high-dimensionality* $0 \leq \alpha < \frac{1}{3}$ *and the degree of persistence* $0 < \kappa \leq 1$*,*

$$\left\| \hat{\Lambda}_{JN}(z) - \frac{\Lambda_{JN}(z)}{\|\Lambda_{JN}(z)\|} \right\| = O\left(\frac{1}{T^{\alpha\kappa}}\right) \tag{4.13}$$

for all $z \in (0, 1)$*.*

Proof. See appendix. □

The estimated factor loadings converge to the normalized true loadings. This normalization is due to assumption L1, which states that loadings are orthogonal and not orthonormal. However given that the smoothed and corrected periodogram $\tilde{I}_{JN,k}$ is a real and symmetric matrix, its eigenvectors $\hat{\Lambda}_{JN}(z)$ are orthonormal (see Lütkepohl (1996)). We therefore need that normalization. However we believe that it would be possible to change the assumptions and get rid of that normalization. More particularly, we could allow factors to have a general diagonal covariance matrix but also factor loadings could have an arbitrary positive semi-definite covariance matrix. However we let this generalization for future work.
Recall that the proof given by Koch (2015) is based on the Davis-Kahan theorem (1970) which requires a positive eigengap. However we do not have any assumptions on the behavior of eigenvalues of the EWS matrix. Even though Koch (2015) implicitly assume multiplicity of 1 of eigenvalues,

its convergence theorem also apply in case of multiplicity strictly greater than 1 thanks to Eldridge et al. (2017).

The next original theorem states the convergence of the common factors to the true normalized factors. This normalization is inherited from theorem 1. To prove this convergence we need an extra assumption, which is natural in the literature (Motta et al. (2011), Bai (2003)).

ASSUMPTION X :

$$\left\| \frac{\Lambda_{JN}(z)^{\top} \epsilon_{JN,k}}{\sqrt{JN}} \right\| = O(1) \qquad \text{for all } t \text{ and for all } T \text{ as } N \longrightarrow \infty. \tag{4.14}$$

This additional assumption states that the loadings and the error terms are asymptotically uncorrelated.

Theorem 2. *Under Assumptions PFLE and X, and given the rescaled time* $z = \frac{k}{T}$*, the degree of high-dimensionality* $0 \leq \alpha < \frac{1}{3}$ *and the degree of persistence* $0 < \kappa \leq 1$*,*

$$\left\| \hat{F}_k - \frac{F_k}{\|\Lambda_{JN}(z)\|} \right\| = O\left(\frac{1}{T^{\frac{\alpha\kappa}{2}}} \right) \tag{4.15}$$

for all $z \in (0, 1)$ *and* $k = 1, \ldots, T$*.*

Proof. The proof lies in Appendix A. □

The LSW factor model is comparable to a factor model based on a double-sided filter. This comes from the *local* structure of the LSW model. Therefore any point of the factors is an average of surrounding neighbours. This renders forecasting with our model difficult.

5 Application : Interbank Benchmark Rates

At the beginning of this section we comment on the economic importance and the construction of the EURIBOR rates. Then we illustrate the factor model developed by Koch (2015) with the inter-bank benchmark rates —i.e. EURIBOR, before pointing out to recent reforms of those rates.

5.1 Importance of the Benchmark rates

The Interbank benchmark rates occupy an substantial place in the economic activity. Those rates, among which the EURIBOR, provide standard rates on which different kind of financial products are based on. For example, when a household want to take on a mortgage to buy a new house, she can decide with her bank to pay a yearly interest rate of the EURIBOR rate plus a premium. This kind of mixture between a benchmark rate and a fixed rate produces a *floating rate* for the contract. The major consequence of such a contract would be to benefit from future potential decrease in interest rates. On the contrary, if it is believe that interest rates will increase in the future, household may not want to agree on such a contract. However, households could also choose a fixed rate, in which case the interest rate will remain constant throughout the life of the contract. That was only one example of the use of an interbank benchmark rate. Other financial products also base their rates on a mixture of benchmark and premium rates —i.e. savings accounts, interest rate swaps, ... The EURIBOR rate comes in 5 different tenors namely 1 week, 1 month, 3 months, 6 months and 1 year. The chosen rate for a particular contract will depend on its life-time. Usually the 12 months rate is used for the longest contracts —e.g. mortgages. Overall, the European Commission (2016) estimated that more than 180 000 billion worth of contracts are under-pined by the EURIBOR making it one of the most important interest rate in the world. The panel of banks surveyed is composed of European national banks with the highest volume of business in the EU. At the time of writing this paper, 18 banks[3] are surveyed daily on their lending and borrowing rates in the interbank market, namely Belfius (Belgium), Société Générale, BNP-Paribas (France), Deutsche Bank (Germany) just to name a few. The EURIBOR benchmark rate also constitutes an important instrument for the monetary policy of the European Central Bank (ECB). The rates on the interbank market is representative of the actual policy rate of the ECB. As a consequence, after implementing a monetary policy, the ECB gauges the effectiveness of its policy by analysing the benchmark rates. It is thus desirable to keep this rate as close as possible to the real sentiment of the market. This fact will hold an important position for the current reforms (see section 5.4).

5.2 Construction and failure of the benchmarks

One focal point of those rates is their construction. How to construct a representative interest rate of the financial lending and borrowing health that will be used throughout the economy ? Initially, the way to create the benchmark was to ask banks about the price of borrowing from one another and aggregate their answer in a single index via a trimmed mean —i.e. an average where the top and bottom 15% sampled rates are dropped. The idea was that if enough banks are surveyed we can obtain a decent estimator. However this estimator could be biased if the panel of banks do not make independent choices. For instance, some banks in the panel could agree to report higher rates in order to push the benchmark rate higher, hence pocketing the difference. On the contrary, banks could have incentives to under-report their borrowing costs in order to hide their riskiness, hence avoiding a negative market reaction. Attracted by the lure of money, several banks from the panel manipulated in that way the benchmark rates between 2005 and 2012. For instance, the Commodities and Future Trading Commission (CFTC), which regulates the US derivatives markets, charged several banks for attempted manipulations and false reports

[3] see EMMI - European Money Markets Institute—PANEL BANKS. Last Visited 2021-04-27,from `https://www.emmi-benchmarks.eu/euribor-org/panel-banks.html` for the full list.

regarding the interbank markets. The penalties ranged from several hundreds million to a billion dollars. We refer to Hou and Skeie (2014) for a complete exposition on the different scandals. Following those events threatening the integrity of such an important economic rate, a flowering of papers emerged to understand and quantify the collusion that took place but more importantly to find alternatives less reliant on subjective judgements of banks. Eisl et al. (2013) analysed the individual daily reports of the panel of banks to quantify the possible manipulation. Snider and Youle (2010) modelled the effect of misreporting incentives on the behavior of an individual bank with LIBOR rates, the financial cousin of the EURIBOR. Snider and Youle (2012) evaluated the extend of misreporting incentives related to portfolio position of banks. Duffie and Dworczak (2014) theoretically studied and proposed a more robust approach for computing a benchmark rate namely they recommended the *volume-weighted average price* (VWAP) as the best linear unbiased method. This method is currently used to settle the price of the EURIBOR rate.

Since 2019, the methodology used to determine the EURIBOR tries to rely less on subjective judgments of banks and more on observable transaction data for the specified tenor. However, in some cases, actual transaction data may not be available. To circumvent this issue, a 3-level methodology is proposed by the European Money Markets (2019). This methodology imposes a progressive calculation hierarchy to estimate the cost of lending and borrowing. In hierarchical order, the methodology is as follow :

1. Only use transaction data from the previous day and from the defined tenor. Only use computation methods provided by the EMMI namely via a VWAP.

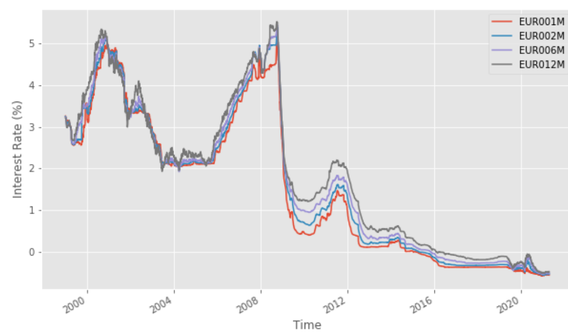
2. Only use transaction data from previous days and from the whole tenor spectrum. Only use computation methods provided by the EMMI —i.e. an adjusted linear interpolation from adjacent tenors.

3. Use transaction data from the whole tenor spectrum but also from markets closely related to the interbank market. Use a combination of modelling techniques and own bank's judgments.

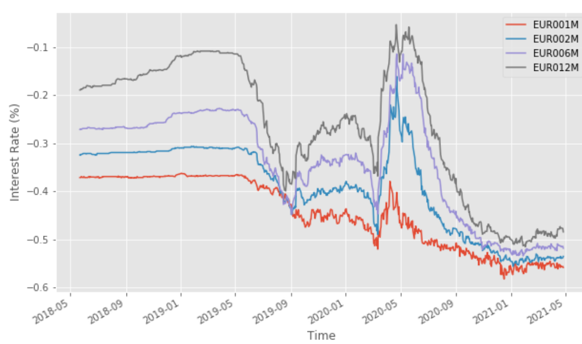
As it can clearly be seen, this 3-level methodology does not advocate the use of bank's opinion until the third level. This methodology is part of a broader set of reforms than we mention in section 5.4.

5.3 Factor modelling of the Benchmarks

It may sometimes be useful to get an overall sentiment for the borrowing and lending health of the economy. Factor models are well suited for such a summarizing task. It is well-known that economic variables usually showcase some kind of non-stationarities over time. For instance, some variables behave differently under economic expansion or contraction —e.g. unemployment rate, production, investments, ... The interbank benchmark rates are no exception to the rule. During economic turmoil, the second-order structure of those rates changes. As depicted in figure 1(a), before 2008, the EURIBOR rates display a constant variance. At the height of the crisis, all tenors saw a dramatic increase in their volatility and plummeted from 5% down to 1% in a few weeks. After the crisis, the benchmark rates recovered a somewhat constant variance. Figure 1(b) also manifests the same picture for the covid-crisis in march 2020. The rates and their variances were rather constant before the crisis. When the covid hit the financial markets volatility increased considerably. Therefore using a non-stationary method for analysis such dynamic structures seem natural.



(a) EURIBOR benchmark rates time series since foundation of the rate.



(b) EURIBOR benchmark rates time series during the Covid-crisis.

Figure 1: EURIBOR rates time series

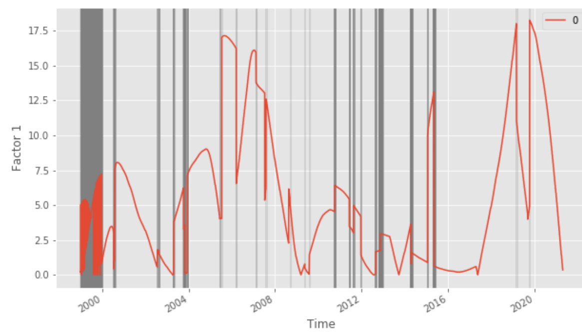
5.3.1 Calibration of the LSW factor model

The model uses the time series in level and not in returns. We do not require to have stationary series for fitting the model since the detail coefficients obtained via the stationary wavelet transform are more stationary than the original series.
The characteristics of the factor model we use to analyse the interbank benchmark rates are as follows :

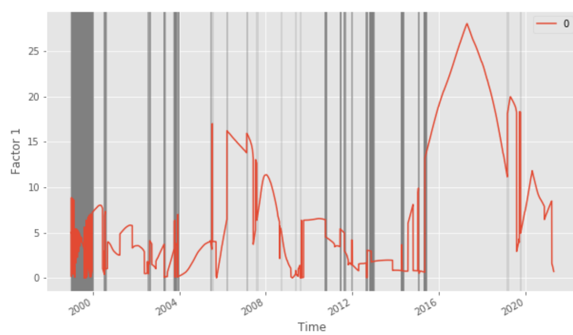
1. Number of factors : 2
2. Order of the LSW model : 0
3. Maximum level of decomposition : 12
4. Wavelet for signal decomposition : Haar
5. Wavelet for smoothing : Daubechie 6

The number of factors we use is quite arbitrary since we didn't use any selection methods. However we believe that a 2 factors for extracting the relevant information of 48 time series is greatly enough given the high correlation between rates. Note that the 4 initial time series transform into 48 times series with the stationary wavelet transform with 12 levels of decomposition. The order of the model is zero, hence we don't assume idiosyncratic increments of the LSW process to be correlated between different scales. This assumption is restrictive since benchmark rates have roughly the same underlying data generating process —i.e. an average of transaction data on the money market. However, as noted in Appendix A, our python implementation only converges to the true evolutionary wavelet spectrum when the order of the LSW is zero. Recall that we pointed out different possible explanations for this lack of convergence.
The number of decomposition levels is determined by the number of time series observations available. In our case we have 5700 data points for each benchmark rate tenor (1 month, 3 months, 6 months and 12 months). Twelve levels of decomposition is thus available to us, since the inequality $2^j \leq 5700 < 2^{j+1}$ is satisfied for $j = 12$. In line of the theory developed by Nason et al. (2000) who showed convergence of the empirical EWS to the true EWS for the Haar and Shannon wavelet, we decide to take the Haar wavelet to perform the *stationary wavelet transform* of our time series. To smooth the corrected raw periodogram we apply the translation invariant de-noising of Coifman and Donoho (1995) with the Daubechie 6 wavelet. As noted by Coifman and Donoho (1995) it is customary to use a smoothing wavelet with a higher number of vanishing moments compared to the wavelet used for decomposing the time-series. The reason is that higher vanishing moments will mean a stronger smoothing.

The factors we get after applying the theory of section 4 are given in the following figures (see Figure 2(a) and 2(b)). Those factors experience jumps since they are estimated from the empirical



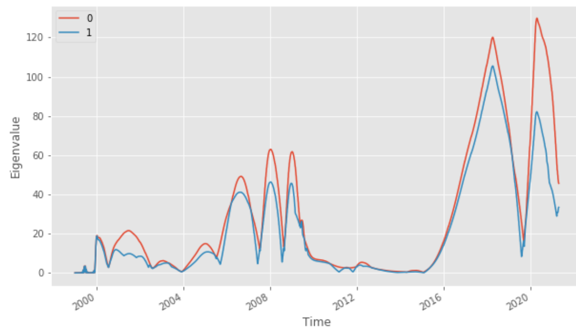
(a) First factor



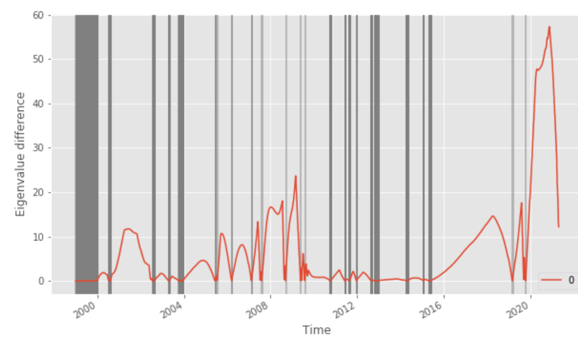
(b) Second factor

Figure 2: Common Factors

factor loadings, which are not continuous. We have highlighted some of the jump regions in grey. In truth the grey regions are points where the continuous eigenvalues intersect, as mentioned in section 4.3. The grey regions were computed from figure 3(b) which displays the difference between the first and second eigenvalue. When this difference is zero, eigenvalues obviously intersect. In figure 3, We have plotted the two eigenvalues corresponding to the two first eigenvectors, and by extension also corresponding to factor loadings. As expected those 2 eigenvalues are continuous function of time. Recall that rescaled time is only used for theoretical development and not empirically, hence the figures exposed in this section are all expressed in original time. The fact that factors are not continuous and jump sharply sometimes renders clean interpretation of the factors difficult. However by comparison with the initial benchmark rates, we can still state a few assertions regarding their behavior. It seems that both factors are sensitive to periods of crisis. They both reacted



(a) The first two eigenvalues corresponding to the 2 eigenvectors



(b) Difference of the two first eigenvalues

Figure 3: Eigenvalues

sharply during the 2008 crisis and covid-crisis in 2020. Notice that both figures 2(a) and 2(b) have different y-scale. You will also note that although the covid-crisis had a smaller impact that the 2008 crisis in figure 1(a), both factors still showcase a dramatic increase when covid hit the market. This comes from the fact that the LSW model of Nason et al. (2000) is used to represent *local* processes. Therefore, most of the information contained at one time in a factor is a summary of information around that given point. This allows us to compare the market sentiment with respect to an event. For instance, the 2 factors seem to agree that the covid-crisis was perceived by the market as serious as the 2008 crisis, even thought in figure 1(a) the covid crisis had a much smaller impact on the rates compared to 2008. Before 2014, the second factor was positively correlated to the rates. However, after 2014, it became negatively correlated with the rates. This can visually be seen by observing that, before 2014, a decrease in the rates implies an decrease of the second factor and the converse after 2014. The peak seen towards the end of 2017 indicates when the most important rates namely the 3 months and 6 months tenors plateaued before slowly reverting.

It is difficult to give a univocal meaning to the common factors since their relation to the initial time series seems to change over time. To make to matter more difficult, the sudden and sharp jumps observed are irregular and don't seem to relate to any particular event. They may only be artefacts of the method used.

5.4 Current Reforms

As already mentioned the interbank benchmark rates are prone to manipulation from market participants. This manipulation renders the rates far less representative of the current borrowing and lending situation, hence threatening the desirable efficiency of markets. Ideally no parties in the contract should be able to influence the interest rate of the contract. Currently those rates are thus under profound reforms. The "Working group on euro risk-free rates" is a private sector working group commissioned by the European Central Bank (ECB) to identify alternatives to current benchmarks. In 2018, this group has recommended the use of the *euro short-term rate* (€STR) rather than the EONIA, the EURIBOR overnight rate. The main difference between these two rates is that the former is administered by the ECB itself whereas the latter is the responsibility of the EMMI, a private non-profit association. The ECB has a strong interest in the smoothness of the money market since its monetary policy effectiveness depends on it. Directly supervising the benchmark rate will render that task easier. Moreover, the ECB is the organisation with the most complete and reliable datasets for creating such an index. The computational aspect of the €STR is similar to the current EURIBOR since it consists in a volume-weighted trimmed mean. However the €STR is more robust than its predecessor because it has a wider scope of observable data. The EONIA only looked at trades between banks whereas the €STR also takes into account investment and pension funds, central banks, money market funds, ... It is currently scheduled that by the end of 2021, the EONIA will be discontinued as a benchmark rate, in favor of the €STR.

Analogously to European benchmarks, other important international financial benchmark rates are also been reformed. For instance, in the United-States, the USD LIBOR was replaced in 2017 by the *Secured Overnight Financing Rate* (SOFR) calculated by the Federal Reserve (Fed). However even the SOFR is being questioned in light of several shortcomings, namely the possibility for the Fed to manipulate it, and its fragility under economic distress. Some may see the *Bloomberg's Short-Term Bank Yield Index* (BSBY) as an alternative to SOFR. We do not go into greater details here so broad and changing the subject is, and let the comparison of the different proposed rates for future work.

As you can see reforming such important indices is particularly challenging so great is the number and diversity of agents involved. All administrators seem however to agree on the fact that benchmark rates should be computed from sufficient, transparent and publicly available data —e.g. transaction data, as to insure its representativeness. Factor models could help in that regards given their resilience and flexibility in application.

6 Conclusion

In conclusion, in this work, we have exposed the necessary theory underlying the LSW factor model developed by Koch (2015). His model is based on the Locally Stationary Wavelet process of Nason et al. (2000) which he extended to the multivariate case. Our contribution consists mainly in laying the groundwork for the development of a rigorous and general asymptotic theory for this factor model. We have also pointed that Koch (2015) based the consistency of his factor model on the Davis-Kahan theorem (Davis and Kahan (1970)) without assuming a positive eigengap, which is required by this theorem. We showed that his consistency results still hold thanks to Eldridge et al. (2017). By virtue of this generalization, eigenvalue multiplicities is not an issue anymore. We thus don't require factor loadings to be continuous. The price to pay for this new flexibility is more time spent trying to understand the behavior of factors and their loadings, since non-smooth behavior is more challenging to fully grasp.
Another important contribution is the extensive python implementation of the LSW model of Nason et al. (2000), the extension of that model by Koch (2015) along with his LSW factor model. You will also observe that in order to implement those, we had to also implement the *Stationary Wavelet Transform* and its inverse, and the *translation invariant de-noising* of Coifman and Donoho (1995). As pointed out in the appendix A, our implementation only converges with the model of Nason et al. (2000). Even though we have started from scratch several times, the implementation doesn't converge when we apply the model of Koch (2015). We therefore suspect the thesis of Koch (2015) to have several inconsistencies which should be further investigated. This fact did not prevent us to use the factor model based on LSW model that do converge (Nason et al. (2000)).
We applied the factor model to the European Interbank Benchmark Rates namely the EURIBOR. Those benchmark represents the state of the borrowing and lending health in the European Union. We mentioned that common factors are *local* in the sense that at each time point they summarize information of original processes in the neighbourhood of that point. This was illustrated by comparing the crisis of 2008 and march 2020. The EURIBOR and other benchmark rates are currently under profound reform following different manipulations and abuses.

Going further, work still needs to be done regarding asymptotic theory. Koch (2015) assumed identification restrictions that may be too stringent —i.e. the covariance matrix of factors is the identity matrix and diagonal covariance matrix of loadings. We believe it would be possible to develop an asymptotic theory following the example of Motta et al. (2011), namely they proved general consistency of the factor model and let the choice of identification restrictions to the researcher.

References

Anderson, T. W., & Rubin, H. (1956). Statistical Inference in Factor Analysis. , *5*, 111–150.

Bai, J. (2003, January). Inferential Theory for Factor Models of Large Dimensions. *Econometrica*, *71*(1), 135–171. Retrieved 2020-12-05, from `http://doi.wiley.com/10.1111/1468-0262.00392` doi: 10.1111/1468-0262.00392

Bernanke, B. S., Boivin, J., & Eliasz, P. (2005, February). Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach. *The Quarterly Journal of Economics*, *120*(1), 387–422. Retrieved 2021-04-08, from `https://academic.oup.com/qje/article-lookup/doi/10.1162/0033553053327452` doi: 10.1162/0033553053327452

Chamberlain, G., & Rothschild, M. (1983, September). Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets. *Econometrica*, *51*(5), 1281. Retrieved 2020-12-05, from `https://www.jstor.org/stable/1912275?origin=crossref` doi: 10.2307/1912275

Coifman, R. R., & Donoho, D. L. (1995). Translation-Invariant De-Noising. In P. Bickel et al. (Eds.), *Wavelets and Statistics* (Vol. 103, pp. 125–150). New York, NY: Springer New York. Retrieved 2021-03-17, from `http://link.springer.com/10.1007/978-1-4612-2544-7`$_9$ (Series Title: Lecture Notes in Statistics) doi: 10.1007/978-1-4612-2544-7$_9$

Commission, E. (2016, August). Commission implementing regulation (EU) 2016/1368. *Official Journal of the European Union*. Retrieved 2021-04-27, from `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R1368`

Dahlhaus, R. (1997, February). Fitting time series models to nonstationary processes. *The Annals of Statistics*, *25*(1), 1–37. Retrieved 2021-02-13, from `http://projecteuclid.org/euclid.aos/1034276620` doi: 10.1214/aos/1034276620

Davis, C., & Kahan, W. M. (1970, March). The Rotation of Eigenvectors by a Perturbation. III. *SIAM Journal on Numerical Analysis*, *7*(1), 1–46. Retrieved 2021-04-21, from `http://epubs.siam.org/doi/10.1137/0707001` doi: 10.1137/0707001

Duffie, D., & Dworczak, P. (2014, October). *Robust Benchmark Design* (Tech. Rep. No. w20540). Cambridge, MA: National Bureau of Economic Research. Retrieved 2021-04-25, from `http://www.nber.org/papers/w20540.pdf` doi: 10.3386/w20540

Eichler, M., Motta, G., & von Sachs, R. (2011, July). Fitting dynamic factor models to non-stationary time series. *Journal of Econometrics*, *163*(1), 51–70. Retrieved 2021-04-09, from `https://linkinghub.elsevier.com/retrieve/pii/S0304407610002113` doi: 10.1016/j.jeconom.2010.11.007

Eisl, A., Jankowitsch, R., & Subrahmanyam, M. G. (2013). Are Interest Rate Fixings Fixed? An Analysis of Libor and Euribor. *SSRN Electronic Journal*. Retrieved 2021-04-24, from `http://www.ssrn.com/abstract=2201013` doi: 10.2139/ssrn.2201013

Eldridge, J., Belkin, M., & Wang, Y. (2017, June). Unperturbed: spectral analysis beyond Davis-Kahan. *arXiv:1706.06516 [cs, stat]*. Retrieved 2021-04-21, from http://arxiv.org/abs/1706.06516 (arXiv: 1706.06516)

Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000, November). The Generalized Dynamic-Factor Model: Identification and Estimation. *Review of Economics and Statistics*, *82*(4), 540–554. Retrieved 2020-12-05, from http://www.mitpressjournals.org/doi/10.1162/003465300559037 doi: 10.1162/003465300559037

Forni, M., Hallin, M., Lippi, M., & Zaffaroni, P. (2015, April). Dynamic factor models with infinite-dimensional factor spaces: One-sided representations. *Journal of Econometrics*, *185*(2), 359–371. Retrieved 2020-12-05, from https://linkinghub.elsevier.com/retrieve/pii/S0304407614002693 doi: 10.1016/j.jeconom.2013.10.017

Forni, M., & Lippi, M. (2001, December). The Generalized Dynamic Factor Model : Representation Theory. *Econometric Theory*, *17*(6), 1113–1141. Retrieved 2020-12-05, from https://www.cambridge.org/core/product/identifier/S0266466601176048/type/journal$_article$ doi: 10.1017/S0266466601176048

Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamitchian, P. (1990). A Real-Time Algorithm for Signal Analysis with the Help of the Wavelet Transform. In J.-M. Combes, A. Grossmann, & P. Tchamitchian (Eds.), *Wavelets* (pp. 286–297). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved 2021-05-03, from http://link.springer.com/10.1007/978-3-642-75988-8$_2$8 (Series Title: inverse problems and theoretical imaging) doi: 10.1007/978-3-642-75988-8$_2$8

Hou, D., & Skeie, D. R. (2014). LIBOR: Origins, Economics, Crisis, Scandal, and Reform. *SSRN Electronic Journal*. Retrieved 2021-04-25, from http://www.ssrn.com/abstract=2423387 doi: 10.2139/ssrn.2423387

Katō, T. (1995). *Perturbation theory for linear operators.* Berlin: Springer.

Koch, D. (2015). *Multiscale methods for the analysis of high-dimensional locally stationary time series* (Doctoral dissertation, Universite Catholique de Louvain, Louvain-la-Neuve). Retrieved from http://hdl.handle.net.proxy.bib.ucl.ac.be/2078.1/162316 (Supervisor : Van Bellegem, Sébastien)

Lütkepohl, H. (1996). *Handbook of matrices.* Chichester ; New York: Wiley.

Mallat, S. (1989, July). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693. Retrieved 2021-04-10, from http://ieeexplore.ieee.org/document/192463/ doi: 10.1109/34.192463

Motta, G., Hafner, C. M., & von Sachs, R. (2011, December). Locally Stationary Factor Models : Identification and Nonparametric Estimation. *Econometric Theory*, *27*(6), 1279–1319. Retrieved 2020-12-11, from https://www.cambridge.org/core/product/identifier/S0266466611000053/type/journal$_{a}rticle$ doi: 10.1017/S0266466611000053

Nason, G. P., & Silverman, B. W. (1995). The Stationary Wavelet Transform and some Statistical Applications. In P. Bickel et al. (Eds.), *Wavelets and Statistics* (Vol. 103, pp. 281–299). New York, NY: Springer New York. Retrieved 2021-03-17, from http://link.springer.com/10.1007/978-1-4612-2544-7$_{1}$7 (Series Title: Lecture Notes in Statistics) doi: 10.1007/978-1-4612-2544-7$_{1}$7

Nason, G. P., von Sachs, R., & Kroisandt, G. (2000, May). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(2), 271–292. Retrieved 2020-12-13, from http://doi.wiley.com/10.1111/1467-9868.00231 doi: 10.1111/1467-9868.00231

Ross, S. A. (1976, December). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, *13*(3), 341–360. Retrieved 2021-04-07, from https://linkinghub.elsevier.com/retrieve/pii/0022053176900466 doi: 10.1016/0022-0531(76)90046-6

Sims, C. A., & Sargent, T. J. (1977, January). *Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory* (preprint). Working Paper. Retrieved 2021-04-07, from https://researchdatabase.minneapolisfed.org/concern/publications/7d278t13n doi: 10.21034/wp.55

Snider, C. A., & Youle, T. (2010). Does the LIBOR Reflect Banks' Borrowing Costs? *SSRN Electronic Journal*. Retrieved 2021-04-24, from http://www.ssrn.com/abstract=1569603 doi: 10.2139/ssrn.1569603

Snider, C. A., & Youle, T. (2012). The Fix Is in: Detecting Portfolio Driven Manipulation of the Libor. *SSRN Electronic Journal*. Retrieved 2021-04-24, from http://www.ssrn.com/abstract=2189015 doi: 10.2139/ssrn.2189015

Stock, J. H., & Watson, M. W. (2002, December). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. Retrieved 2020-12-05, from http://www.tandfonline.com/doi/abs/10.1198/016214502388618960 doi: 10.1198/016214502388618960

Van Bellegem, S., & von Sachs, R. (2008, August). Locally adaptive estimation of evolutionary wavelet spectra. *The Annals of Statistics*, *36*(4), 1879–1924. Retrieved 2020-12-10, from https://projecteuclid.org/euclid.aos/1216237303 doi: 10.1214/07-AOS524

Walnut, D. F. (2004). *An Introduction to Wavelet Analysis*. Retrieved 2020-12-10, from https://doi.org/10.1007/978-1-4612-0001-7 (OCLC: 1113459008)

Appendix

A Proofs

A.1 Proof of Lemma 1

Lemma 1. *Let f, g be real bounded Lipschitz continuous functions with Lipschitz constants L_f and L_g respectively. Set $f := gh$, where h is a real bounded function, then $|g|$ being bounded from below by a constant $m_g > 0$ is a necessary condition for h to be Lipschitz continuous.*

Proof. This proof is based on the Lipschitz continuity definition and some inequality manipulations. From the definition of f and the reverse triangle inequality, we get :

$$\begin{aligned}
|f(x) - f(y)| &= |g(x)h(x) - g(y)h(y)| \\
&= \left|g(x)h(x) - g(y)h(x) - \big(g(y)h(y) - g(y)h(x)\big)\right| \\
&\geq \Big|\, |g(x)h(x) - g(y)h(x)| - |g(y)h(y) - g(y)h(x)| \,\Big| \\
&\geq |g(y)|\,|h(x) - h(y)| - |h(x)|\,|g(x) - g(y)|
\end{aligned}$$

We can bound from above f with the Lipschitz continuity assumption,

$$|g(y)|\,|h(x) - h(y)| - |h(x)|\,|g(x) - g(y)| \leq |f(x) - f(y)| \leq L_f\,|x - y|$$

Now by virtue of Lipschitz continuity of g and the fact that h is bounded by a constant, say M_h,

$$|g(y)|\,|h(x) - h(y)| \leq \big(L_f + M_h L_g\big)\,|x - y| \tag{A.1}$$

Finally, if $0 < m_g \leq |g|$, then

$$|h(x) - h(y)| \leq \frac{\big(L_f + M_h L_g\big)}{m_g}\,|x - y|$$

which proves the results.
Note that if $g(y) = 0$, h is not restricted to be Lipschitz continuous since the inequality (A.1) is trivially fulfilled. $\square$

A.2 Proof of theorem 1

Theorem. *Under Assumptions PFLE, and given the rescaled time* $z = \frac{k}{T}$*, the degree of high-dimensionality* $0 \leq \alpha < \frac{1}{3}$ *and the degree of persistence* $0 < \kappa \leq 1$*,*

$$\left\| \hat{\Lambda}_{JN,m}(z) - \frac{\Lambda_{JN,m}(z)}{\|\Lambda_{JN,m}(z)\|} \right\| = O\left(\frac{1}{T^{\alpha\kappa}}\right) \tag{A.2}$$

for all $z \in (0,1)$ *and where* $\Lambda_{JN,m}(z)$ *is the m-th column of the loading matrix* $\Lambda_{JN}(z)$*.*

Proof. This proof comes from Koch (2015).
The Davis-Kahan theorem (Davis and Kahan (1970)) allows us to write

$$\left\| \hat{\Lambda}_{JN,m}(z) - \frac{\Lambda_{JN,m}(z)}{\|\Lambda_{JN,m}(z)\|} \right\| \leq \frac{\sqrt{2}\,\|\Sigma_\epsilon(z)\|}{\min\left(\left|\lambda_{m-1}(S_{JN}(z)) - \|\Lambda_{JN,m}(z)\|^2\right|, \left|\|\Lambda_{JN,m}(z)\|^2 - \lambda_{m+1}(S_{JN}(z))\right| \right)}$$

The denominator can be further decomposed using the triangle inequality,

$$\left|\lambda_{m-1}(S_{JN}(z)) - \|\Lambda_{JN,m}(z)\|^2\right| \geq \left|\|\Lambda_{JN,m-1}(z)\|^2 - \|\Lambda_{JN,m}(z)\|^2)\right| - \left|\|\Lambda_{JN,m-1}(z)\|^2 - \lambda_{m-1}(S_{JN}(z))\right|$$
$$\left|\|\Lambda_{JN,m}(z)\|^2 - \lambda_{m+1}(S_{JN}(z))\right| \geq \left|\|\Lambda_{JN,m}(z)\|^2 - \|\Lambda_{JN,m+1}(z)\|^2\right| - \left|\lambda_{m+1}(S_{JN}(z) - \|\Lambda_{JN,m+1}(z)\|^2\right|$$

where $\lambda_m(S_{JN}(z))$ is the m-th eigenvalue of the matrix $S_{JN}(z)$.
Using the Weyl's inequalities we can bound $\left|\|\Lambda_{JN,m-1}(z)\|^2 - \lambda_{m-1}(S_{JN}(z))\right|$ and $\left|\lambda_{m+1}(S_{JN}(z) - \|\Lambda_{JN,m+1}(z)\|^2\right|$ from above by $\lambda_m(\Sigma_\epsilon(z))$ for all m. Note that assumption E1 implies $\lambda_{max}(\Sigma_\epsilon(z)) = O(1)$. Finally, given assumption L2 and $N = O(T^\alpha)$, we have

$$\left|\|\Lambda_{JN,m-1}(z)\|^2 - \|\Lambda_{JN,m}(z)\|^2)\right| = O(N^\kappa)$$
$$\left|\|\Lambda_{JN,m}(z)\|^2 - \|\Lambda_{JN,m+1}(z)\|^2\right| = O(N^\kappa)$$

for all m. Putting everything together we obtain the result. □

A.3 Proof of theorem 2

Theorem. *Under Assumptions PFLE, and given the rescaled time* $z = \frac{k}{T}$*, the degree of high-dimensionality* $\alpha > 0$ *and the degree of persistence* $\kappa > 0$*,*

$$\left\| \hat{F}_k - \frac{F_k}{\|\Lambda_{JN}(z)\|} \right\| = O\left(\frac{1}{T^{\frac{\alpha\kappa}{2}}}\right) \tag{A.3}$$

for all $z \in (0,1)$ *and* $k = 1, \ldots, T$*.*

Proof. From the definition of the estimator of the factors 4.12 and the factor model 4.3, it is possible to write the estimator as,

$$\hat{F}_k = \frac{\hat{\Lambda}_{JN}(z)^\top \Lambda_{JN}(z)}{JN} F_k + \frac{\hat{\Lambda}_{JN}(z)^\top \epsilon_{JN,k}}{JN}$$

which gives us,

$$\left\| \hat{F}_k - \frac{\hat{\Lambda}_{JN}(z)^\top \Lambda_{JN}(z)}{JN} F_k \right\| = \left\| \frac{\hat{\Lambda}_{JN}(z)^\top \epsilon_{JN,k}}{JN} \right\| \tag{A.4}$$

If we manage to prove that $\frac{\hat{\Lambda}_{JN}(z)^\top \Lambda_{JN}(z)}{JN} = I_K$, where I_k is the identity matrix of order K, then from assumption X (4.14) the convergence follows. However, from theorem **??** the estimated loadings converge to the normalized true ones. As a consequence we can only state that

$$\left\| \frac{\hat{\Lambda}_{JN}(z)^\top \frac{\Lambda_{JN}(z)}{\|\Lambda_{JN}(z)\|}}{JN} - I_K \right\| = O\left(\frac{1}{T^{\alpha\kappa}}\right) \tag{A.5}$$

since the loadings are the orthonormal eigenvectors of the EWS matrix. Formally, this is expressed in the constraint of the non-linear least square —i.e. condition 4.6.
The right-hand side of A.4 converges at a rate of $O\left(\frac{1}{T^{\frac{\alpha\kappa}{2}}}\right)$. To see this, from assumption X, we have $\left\| \frac{\Lambda_{JN}(z)^\top \epsilon_{JN,k}}{JN} \right\| = O\left(\frac{1}{\sqrt{N}}\right)$. Then, given that $N = O\left(T^\alpha\right)$,

$$\begin{aligned}
\left\| \frac{\hat{\Lambda}_{JN}(z)^\top \epsilon_{JN,k}}{JN} \right\| &= \left\| \frac{\frac{\Lambda_{JN}(z)}{\|\Lambda_{JN}(z)\|} \epsilon_{JN,k}}{JN} \right\| + O\left(T^{-\alpha\kappa}\right) \\
&= \|\Lambda_{JN}(z)\| \left\| \frac{\Lambda_{JN}(z)^\top \epsilon_{JN,k}}{JN} \right\| + O\left(T^{-\alpha\kappa}\right) \\
&= O\left(N^\kappa\right) O\left(\frac{1}{\sqrt{N}}\right) + O\left(T^{-\alpha\kappa}\right) \\
&= O\left(\frac{1}{T^{\frac{\alpha\kappa}{2}}}\right)
\end{aligned}$$

The result then follows,

$$\left\| \hat{F}_k - \frac{F_k}{\|\Lambda_{JN}(z)\|} \right\| = O\left(\frac{1}{T^{\frac{\alpha\kappa}{2}}}\right)$$

□

B Convergence issues

This section demonstrates the converge of the corrected wavelet periodogram to the true evolutionary wavelet spectrum. It serves as a sanity checks before applying the model to real world situations. Having a good implementation is key in order to obtain meaning full results. Figures depicted here come from the code example at the end of the "Python implementation" section (see section **??**). We first begin by illustrating the theory of Nason et al. (2000), before demonstrating the results of the thesis of Koch (2015).

B.0.1 Illustrations of Nason et al. (2000)

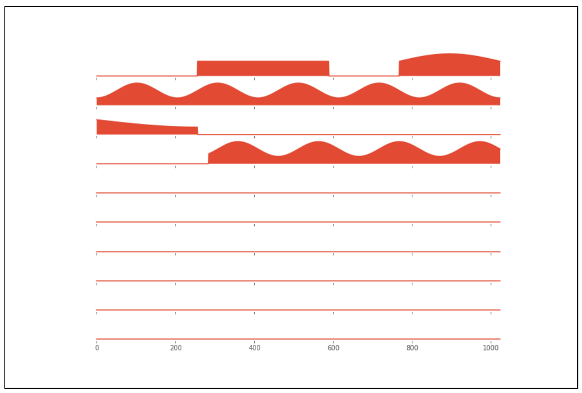
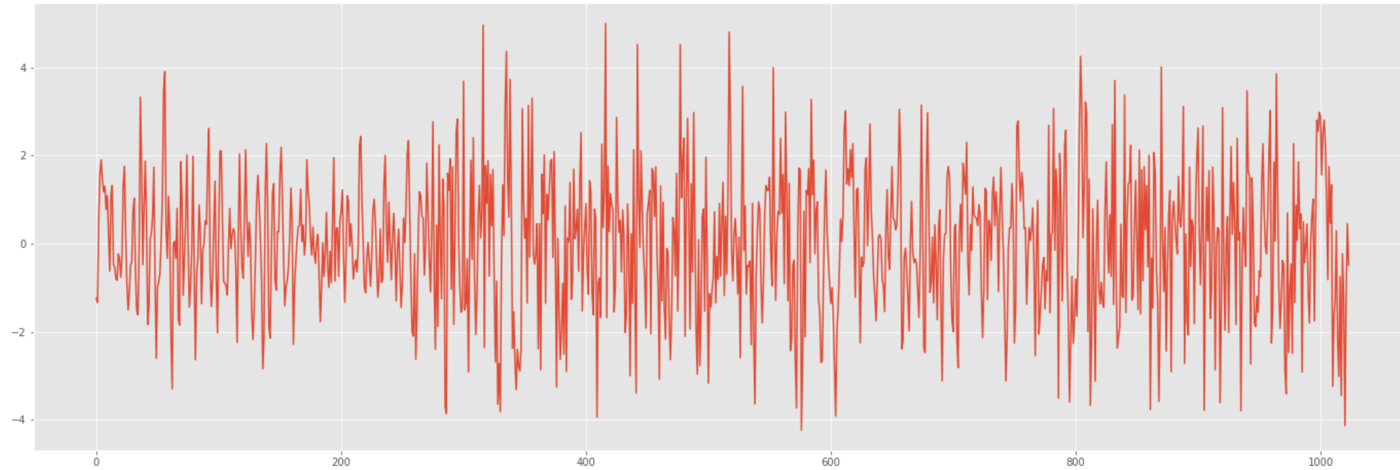


Figure 4: True Evolutionary Wavelet Spectrum. The scales are organized from top to bottom —i.e. the scale $j = -1$ is the first on top. All levels are on the same y-axis scale, between 0 and 1.

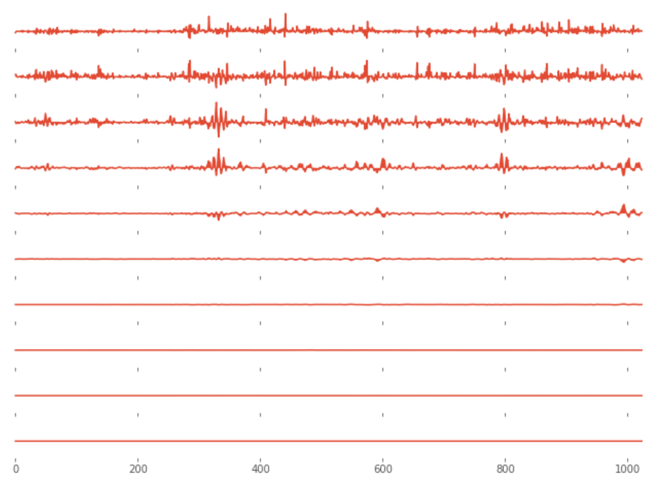
Figure 4 show the true evolutionary wavelet spectrum we use for the following simulations.

From this theoretical EWS it is possible to compute a sample path of the Locally Stationary Wavelet process (see figure 5(a)). We can then decompose this sample signal as if it were a true signal and get the corrected wavelet periodogram (see figure 5(b)). However we have to smooth that corrected periodogram in order to get a consistent estimator of the EWS (see figure 5(c)). We used the translation invariant de-noising of Coifman and Donoho (1995) for smoothing.

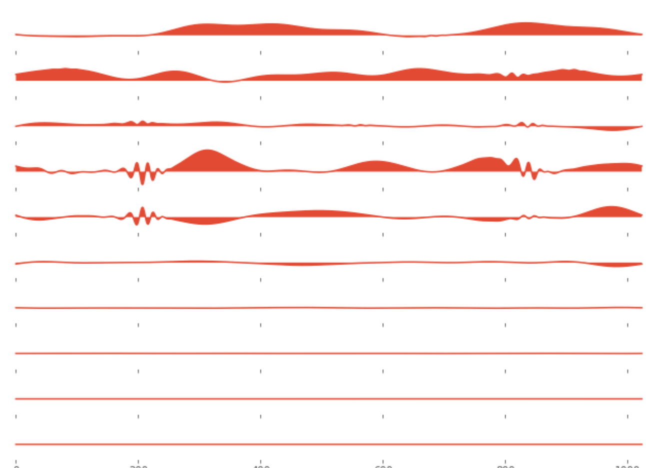
Finally, to show that the corrected wavelet periodogram tends to the true EWS, we averaged 250 realizations of the corrected periodogram. We display the averaged periodogram in figure 5. From those figures we can claim that we were able to replicate the findings of Nason et al. (2000).



(a) Sample path of the Locally Stationary Wavelet Process given the theoretical EWS depicted in figure 4.



(b) Sample corrected EWS from the decomposition of the sample signal shown in figure 5(a)



(c) Smoothed periodogram from the corrected periodogram in figure 5(b)

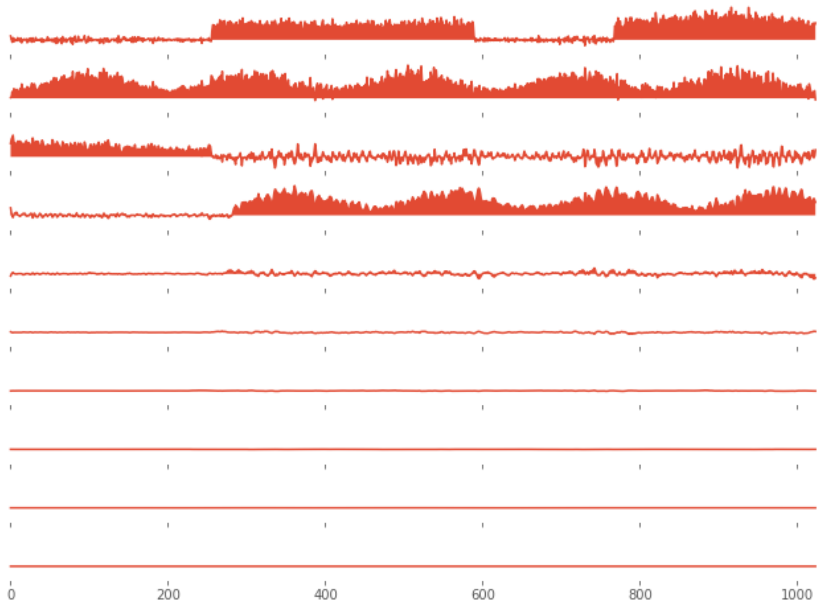


Figure 5: Average of 250 samples of corrected wavelet periodogram. This clearly shows the convergence of the sample corrected periodogram to the true EWS.

B.0.2 Illustration of Koch (2015)

We know illustrate the extension of the Locally Stationary Wavelet decomposition realized by Koch (2015) in his thesis. Recall that this extension allows the increments of the LSW model (see Definition 2.1) to be correlated between scales. Note that it does not, however, allows those increments to be serially dependent. We refer to section 2.2 for a complete exposition on the extensions.

We take the same true Evolutionary Wavelet Spectrum as the previous section (see figure 4). And we assume the correlation matrix for the increments to be constant over time and to have the following form :

$$\Gamma(z) = \begin{pmatrix} 1.0 & 0.89 & 0.59 & 0.22 & 0.04 & 0 & 0 & 0 & 0 & 0 \\ 0.89 & 1.0 & 0.86 & 0.51 & 0.19 & 0.04 & 0 & 0 & 0 & 0 \\ 0.59 & 0.86 & 1.0 & 0.84 & 0.5 & 0.19 & 0.04 & 0 & 0 & 0 \\ 0.22 & 0.51 & 0.84 & 1.0 & 0.84 & 0.50 & 0.19 & 0.04 & 0 & 0 \\ 0.04 & 0.19 & 0.5 & 0.84 & 1.0 & 0.84 & 0.50 & 0.19 & 0.04 & 0 \\ 0 & 0.04 & 0.19 & 0.5 & 0.84 & 1.0 & 0.84 & 0.5 & 0.19 & 0.04 \\ 0 & 0 & 0.04 & 0.19 & 0.5 & 0.84 & 1.0 & 0.84 & 0.51 & 0.22 \\ 0 & 0 & 0 & 0.04 & 0.19 & 0.5 & 0.84 & 1.0 & 0.86 & 0.59 \\ 0 & 0 & 0 & 0 & 0.04 & 0.19 & 0.51 & 0.86 & 1.0 & 0.89 \\ 0 & 0 & 0 & 0 & 0 & 0.04 & 0.22 & 0.59 & 0.89 & 1.0 \end{pmatrix}$$

The (n, m)th entry of that matrix reads as "the correlation between the scale n and the scale m. Therefore the (n, m)th element $\Gamma_{n,m}(z)$ of that matrix correspond to the element $\Gamma_{-|m-n|,-n}(z)$ in the definition of the EWS (see section 2.2). This matrix leads us to use a LSW model of order 4 since scales are dependent on up to 4 neighbouring scales. For instance, the EWS of order 1 —i.e. the EWS where the first scale represents the EWS between scale $j = -1$ and $j = -2$, is shown below :

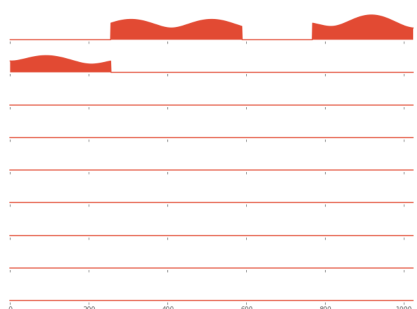
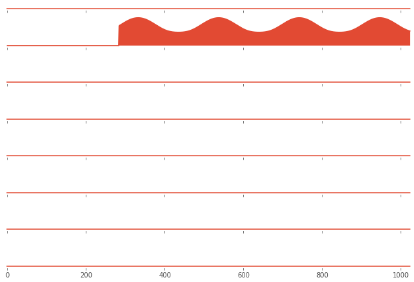


Figure 6: EWS of order 1 —between scale $j = -1$ and $j = -2$.

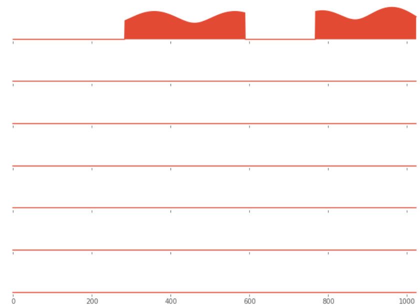
we can analogously define the remaining orders of the periodogram.

You will note that the 4th order is empty since either the scale or the 4th neighbouring scale is zero everywhere. You also observe that when the order of the EWS increases, the number of scales decreases.

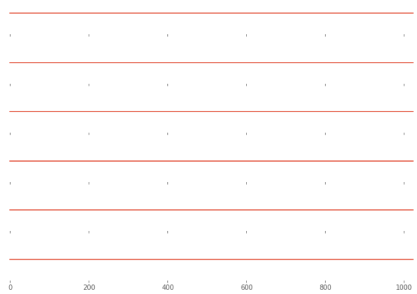
By following the same procedure as in the previous section, we can obtain a sample path of the



(a) EWS of Order 2



(b) EWS of Order 3



(c) EWS of Order 4

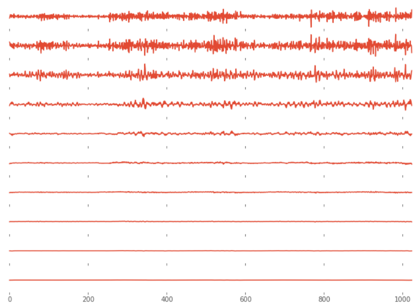
Figure 7: Remaining orders of the LSW model of order 4.

LSW(q) model, its corrected and smoothed wavelet periodogram.
It is also possible to check the convergence of the average periodogram to the true EWS. However, if we precisely follow the theory developed by Koch (2015) in his thesis, we don't obtain that convergence. The following figure (Figure 8(a)) gives us the average of 250 realizations of the corrected periodogram. This average clearly does not converge.

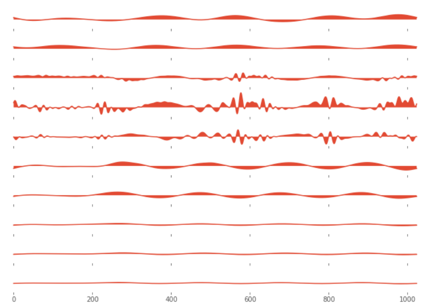
After changing empirically the correction procedure of the EWS we recover the convergence of the average corrected EWS, although only for order zero (see Figure 8(c) and Figure 8(d)). We modified the theory of Koch (2015) empirically in the sense that in its correction of the wavelet periodogram we only take non-negative orders. Particularly, the summation term with index "i" goes from 0 to q and not from $-q$ to q (see section 4.2.3, page 144 of Koch (2015)). we suspect our correction is appropriate considering the fact that, following the notation of Koch's thesis, the correction term $A^{-1}_{i,r}(j,l)$ is the same for positive or negative orders "i". Allowing "i" to be negative

would thus have the effect of correcting the spectrum twice.
We see several possible explanations for this lack of convergence. First, even after having checked several times our code, our python implementation may still contain bugs. Second, we may have a more profound misunderstanding of the model at hand, since we suspect the thesis of Koch (2015) to contain several inconsistencies, namely about the indices of summation terms, which may have misled us. Third, the theory developed by Koch (2015) may be somewhat erroneous concerning the correction of the raw wavelet periodogram. This last suggestion is the least plausible but still deserves to be mentioned. This convergence issue still needs to be further investigated.

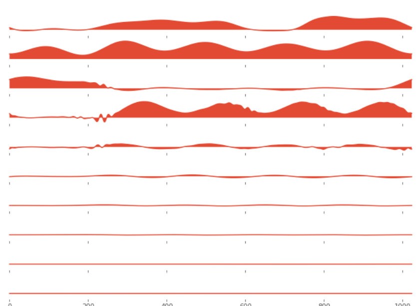
Until now our simulations were only concerned with the univariate LSW. Our implementation also takes into account the multivariate case. However we don't show simulations results here since they are identical to the univariate ones. We also encounter the same convergence issue as before, namely the periodogram between different scales does not converge to the true EWS.



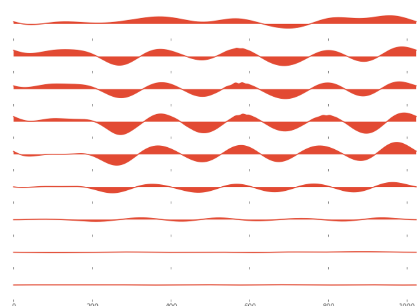
(a) This figure display the average of 250 realizations of the corrected periodogram (at order 0). This periodogram should converge to the true EWS of order 0 depicted by Figure 4.



(b) Smoothed and corrected non-convergent wavelet periodogram of order 0 of the 250 realizations of the process. This periodogram should converge to the true EWS of order 0 depicted by Figure 4.



(c) Smoothed and corrected "empirically modified" wavelet periodogram of order 0 from averaging 250 realizations. This periodogram converges to the true evolutionary wavelet spectrum (Figure 4).



(d) Smoothed and corrected "empirically modified" wavelet periodogram of order 1 from averaging 250 realizations. This periodogram should converge to the EWS shown in Figure 6.

Figure 8: Non-convergent and empirically modified periodogram