Université catholique de Louvain

Institute for Multidisciplinary Research in Quantitative
Modelling and Analysis (IMMAQ)

*Thèse de doctorat présentée en vue de l'obtention du grade de
Docteur en Sciences*

Multiscale methods for the analysis of high-dimensional locally stationary time series

Membres du jury

Prof. Pierre Devolder *(Président)*

Prof. Idris Eckley

Prof. Hernando Ombao

Prof. Sébastien Van Bellegem *(Promoteur)*

Prof. Rainer von Sachs (*Secrétaire)*

Acknowledgments

This thesis concludes four years of research, which could never have been successfully completed without the support of several people. First and foremost I would like to express my deep gratitude towards my supervisor Prof. Van Bellegem for introducing me to the field of multisale modelling and sharing his insights with me during extensive discussions over these four years. Without his invaluable scientific advice, this work would not have been realized. I would further like to address special thanks to the remaining members of my supervisory committee, Prof. von Sachs and Prof. Devolder, for the time they invested in keeping this thesis on track and for their availability and help during crucial moments of my PhD period. I want to thank all the members of the jury for having invested time and effort in reading the first version of this thesis. Their comments and suggestions have been of great value to me and helped me turn the thesis into what it is today. I also want to thank all my colleagues at the CORE and the ISBA, who made my PhD years an unforgettable experience that I will cherish forever. Last but not least, I would like to express my deepest thanks to my parents and my sister for their encouragement, love and unconditional support on this long journey.

*

This research was supported by the Fonds National de la Recherche Luxembourg under the AFR (Aides à la Formation Recherche) scheme (Project Code: 4091499).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis deals with the multiscale modelling of the covariance pattern of discrete time series with time-varying autocovariance function.

Discrete time series are sequences of random variables with a natural temporal ordering. A random variable is a variable whose possible values are numerical outcomes of a random phenomenon. We limit our analysis to real-valued random variables and hence in this work a random variable X is defined as a measurable function from the sample space Ω, that is the set of possible outcomes of the random phenomenon, to the set of real numbers $\mathbb{R}$. To any random variable corresponds a probability distribution, which assigns a measure of likeliness to each measurable subset of the sample space. If the image of a random variable is finite or countably infinite, we refer to the latter as a discrete random variable and the corresponding probability distribution can be described by a probability mass function which assigns a probability to each value in the image of the random variable. For example, the possible outcomes of a coin toss can be described by the sample space $\Omega = \{\text{heads}, \text{tails}\}$. We can introduce a real-valued random variable X such that

$$X_\omega = \begin{cases} 1 & \text{if } \omega = \text{heads} \\ 0 & \text{if } \omega = \text{tails} \end{cases}$$

If the coin is a fair coin, the probability mass function, denoted as $f_X(x)$ is given by

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{2} & \text{if } x = 0 \end{cases}$$

If the image of a random variable is uncountably infinite, we refer to the latter as a continuous random variable and the probability distribution can be described by a cumulative distribution function which describes the probability that the random variable X will be found to have a value less than or equal to x, which we denote as $f_X(x) = P(X \leq x)$. Examples of continuous random variables are the price in Euros of a given stock in a weeks time, the total power output in megawatt of a wind farm over the coming month or the accumulated liquid precipitation in liters within a given area during the coming year.

As it has been mentioned above, a time series is a sequence of random variables in which the constituent elements have an inherent order according to a time index. We denote as X_n a time series of length n, such that

$$X_n = \{X(t)\}_{t=0}^{n-1}. \tag{1.1}$$

From a probability theoretical viewpoint, X_n is a vector-valued function from the sample space of its constituent elements to $\mathbb{R}^n$. Consequently, any given time series gives rise to a joint probability distribution. In this thesis we focus on time series that consist of jointly continuous random variables, in which case the probability distribution function takes the form of a joint cumulative distribution, that is

$$F_{X_n}(x_n) = P\left(X(0) \leq x(0), \ldots, X(n-1) \leq x(n-1)\right). \tag{1.2}$$

One of the main objectives of the statistical analysis of time series is to design estimators of the joint cumulative distribution. The approaches designed to do this can be divided into two big classes, parametric approaches on the one hand and and non-parametric approaches on the other hand. Parametric approaches assume that the time series has a certain underlying structure which can be described using a small number of parameters. In that sense parametric approaches are based on the assumption that the family of the function in (1.2) is known. For example if we assume that the elements of X_n follow a joint normal distribution and that they are zero mean, the probability distribution of X_n is entirely defined given the set of covariances of all the pairs of random variables.

The covariance is a measure of linear dependence of two random variables. More precisely, consider a pair of random variables of the sequence in (1.1), let us say $X(t)$ and $X(t+\tau)$. Under the assumption of zero-mean random variables the covariance between $X(t)$ and $X(t+\tau)$ is given by the expected value of their product and hence if we denote as $f(x(t),x(t+\tau))$ the joint normal density of $X_n(t)$ and $X(t+\tau)$, we have that

$$\begin{aligned}
&\operatorname{Cov}(X(t),X(t+\tau)) \\
&= E\,(X(t)X(t+\tau)) \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x(t)x(t+\tau)f\,(x(t),x(t+\tau))\,dx(t)dx(t+\tau)
\end{aligned}$$

If positive values of $X(t)$ mainly correspond with positive values of $X(t+\tau)$, and the same holds for the negative values, which means that the variables tend to show similar behavior, the covariance is positive. On the contrary, when positive values of one variable mainly correspond to the negative values of the other, and hence if the variables tend to show opposite behavior, the covariance is negative. The sign of the covariance therefore shows the direction in the linear relationship between the variables. The set of covariances of the elements of a given time series forms the autocovariance function of the time series

$$\begin{gathered}
C_n\,(t,\tau) = E\,(X_n(t)X_n(t+\tau)) \\
t = 0,\ldots,n-1;\tau = -t,\ldots,n-1-t
\end{gathered} \tag{1.3}$$

Time-varying autocovariance function

Many well-established models in the field of time series analysis are based on the assumption that the autocovariance function is stationary in time and hence

$$C_n\,(t,\tau) = C_n\,(\tau)$$

for $t = 0,\ldots,n-1$ and $\tau = -t,\ldots,n-1-t$ with $C_n(\tau) = C_n(-\tau)$. The hypothesis of stationarity is an appealing concept in the sense that the autocovariance function can be estimated by computing the corresponding sample averages, which then allow to build an asymptotic theory based mainly on some version of the Central limit theorem. If we denote as $\hat{C}_n(\tau)$ the sample estimator of $C_n(\tau)$, we have that

$$\hat{C}_n\,(\tau) = \frac{1}{n-\tau}\sum_{k=0}^{n-1-\tau} X(k)X(k+\tau) \tag{1.4}$$

for $0 \leq \tau \leq n-1$. In most cases, the assumption of stationarity represents a mere mathematical idealization which leads to an oversimplified viewpoint of the underlying second-order properties of a time series. However, to build an asymptotic theory in the nonstationary case, we cannot simply drop the assumption of constant cross-covariances of the increment terms without replacing it with something else. Priestley (1965) [101] provides a characterization of stochastic processes with non-stationary second-order structure by introducing a form of time series that are globally non-stationary but have a locally stationary behavior relying on the theory of evolutionary spectra. In this sense, the first step allowing to move from stationary to nonstationary time series modelling, is to replace the assumption of constant covariances of the elements of $X_n(t)$ by a more flexible regularity assumption. One example of a more flexible regularity assumption is to suppose that the cross-covariance function is Lipschitz continuous, in which case $|C(\tau,t+k)-C(\tau,t)| \leq Ck$ with $C \geq 0$. However, the regularity assumption alone does not suffice to build a rigorous asymptotic framework. Dahlhaus (1997) [27] defines an adequate asymptotic theory for the estimation of locally stationary time series, based on the concept of rescaled time. Rescaled time, denoted as z, is defined as the ratio of the real time index by the length of the random sequence, and hence $z = \frac{t}{n}$. Based on the rescaled time notation, we can rewrite the covariance function in (1.13) as $C(\tau,z)$. Analogously to nonparametric regression analysis, the notion of rescaled time makes use of the idea that as the number of data points increases the time-varying parameters of the locally stationary process are observed on the same interval but on a finer grid. If we combine the Lipschitz continuity assumption of the covariance function and the rescaled time notation, we get the upper bound

$$\left| C\left(\tau, \frac{t+k}{n}\right) - C\left(\tau, \frac{t}{n}\right) \right| \leq \frac{Ck}{n} \tag{1.5}$$

which means that as the length of the random sequence increases, more information about the local behavior of the time-varying parameters can be collected.

Recall from the representation in (1.4) that the estimator of the autocovariance function in the stationary case takes the form of an uniformly weighted average of all the random variables in (1.1). In the nonstationary case, the estimation procedure needs to take into account the fact that the second-order pattern of the time series is locally constant but globally nonstationary. Hence, as a uniform weighting of the elements of (1.1) does not allow to capture the local behavior of the time series, we need to replace the latter with a weighted averaging scheme. More precisely, in the nonstationary case the sequence of uniform weights is replaced by a sequence in which the $t-$th component and the components close to the latter take on larger values than the ones far away from component t. For example, we can consider the identity function

$$\mathbb{I}(t-M \leq k \leq t+M) = \begin{cases} 1 & \text{if } t-M \leq k \leq t+M \\ 0 & \text{otherwise} \end{cases}$$

where $0 < M < n-1$ and to define the estimator

$$\hat{C}_n(\tau,t) = \frac{1}{2M+1} \sum_{k=0}^{n-1-\tau} \mathbb{I}(t-M \leq k \leq t+M) X(k) X(k+\tau). \tag{1.6}$$

If the covariance function of a given time series was time invariant, the estimator in (1.6) should be approximately constant in time. On the contrary, if the covariance pattern of the time series is characterized by a nonstationary behavior we should observe a time-varying behavior in the estimator. To illustrate this, we consider the daily log-returns of the German stock index between November 2004 and April 2015 plotted in Figure 1.1. Figure 1.1 contains the estimated variance, that is $\hat{C}_n^{(u,v)}(0,t)$, of the German stock index using the estimator in (1.6) with $M = 30$. Note that from the third quarter 2008 until summer 2009 which roughly delimits the period of the Global financial crisis and during the height of the Euro crisis in 2011, the variance of the log-returns takes on very large values compared to the remaining intervals of the period of analysis and hence indicating the presence of a time-varying second-order pattern in the log-return series.

Mandelbrot (1963) [87] points out the insufficiency of the normal distribution for modelling the marginal distribution of log-returns because of the apparent fat-tailed character of the latter. Since then, the non-Gaussian character of the distribution of log-returns has been observed on many occasions in various studies (see for example Fama (1965) [46] and Cont (2001) [26]. A fat-tailed distribution is a probability distribution that has the property that it exhibits large skewness or kurtosis, where the skewness is a measure of the asymmetry and the kurtosis is a measure of the ”peakedness” of the distribution. Hence, a common way of quantifying the deviation from the normal distribution is by computing the empirical versions of the kurtosis and the skewness of a given sample and compare its values to the corresponding theoretical values of the skewness and kurtosis in the case of a normal distribution through the Jarque-Bera test statistics (see Jarque and Bera, 1980 [73]), which is

$$JB = \frac{n}{6}\left(\gamma^2 + \frac{1}{4}(\kappa - 3)^2\right)$$

where γ and κ are the estimators of skewness and kurtosis respectively. We use the Jarque-Bera test to verify the normality assumption on the log-returns of the German, French and Belgian stock indices using a sample which stretches out from November 2004 to April 2015 containing 2500 data points. For the testing procedure we split the sample into K subsamples of size $\frac{n}{K}$ and carry out a Jarque-Bera test for each one of these subsamples. Table 1.1 contains the proportion of subsamples for which the $p-$value of the respective Jarque-Bera test is smaller than $\alpha = 0.05$ and this for $\frac{n}{K} = 250, 125, 100, 50, 25$. The idea behind this approach is that if the observed log-returns within each of the subsamples are generated from a normal distribution with variance remaining constant throughout the entire subsample the proportion of rejections should be close to the value of α. Table 1.2 contains the results when applying the same testing procedure using the Shapiro-Wilk test (see Shapiro and Wilk, 1965 [111]) instead of the Jarque-Bera test. For both testing procedures, the number of subsamples, for which the normality testing procedure leads to a rejection of the null hypothesis, tends towards the chosen value for α as the size of the subsamples tends to zero.

The null hypothesis of both, the Jarque-Bera and the Shapiro-Wilk testing procedure, is that all the random variables within the same subsample follow the same normal distribution. However, if we suppose that the underlying time series is locally stationary and hence that the variance function is expressed in rescaled time notation and is Lipschitz continuous, the absolute difference between the variances of any given pair of elements of the subsample is bounded from above by $\frac{CK}{n}$. Hence, if K is large the absolute difference between the variances of some elements within the subsample can be potentially very large, ultimately resulting in a rejection of the null hypothesis of the normality testing procedures. Hence, even though each of the random variables within a given subsample follows a normal distribution the strong differences in terms of their respective variances leads to a rejection of the assumption of a single underlying normal distribution. However, if K is small the distribution of the elements within a given subsample can be well-approximated by a single normal distribution under the null hypothesis, leading to a non-rejection of the latter in most subsamples. Hence, the detection of non-Gaussianity in the log-return series of financial assets may, in some cases, merely be an artifact of an underlying nonstationary behavior of the time series' second-order structure. For a more detailed discussion on the link between nonstationary covariance pattern and non-Gaussianity see Mikosch and Stărică (1999) [92] and Fryzlewicz (2005) [57].

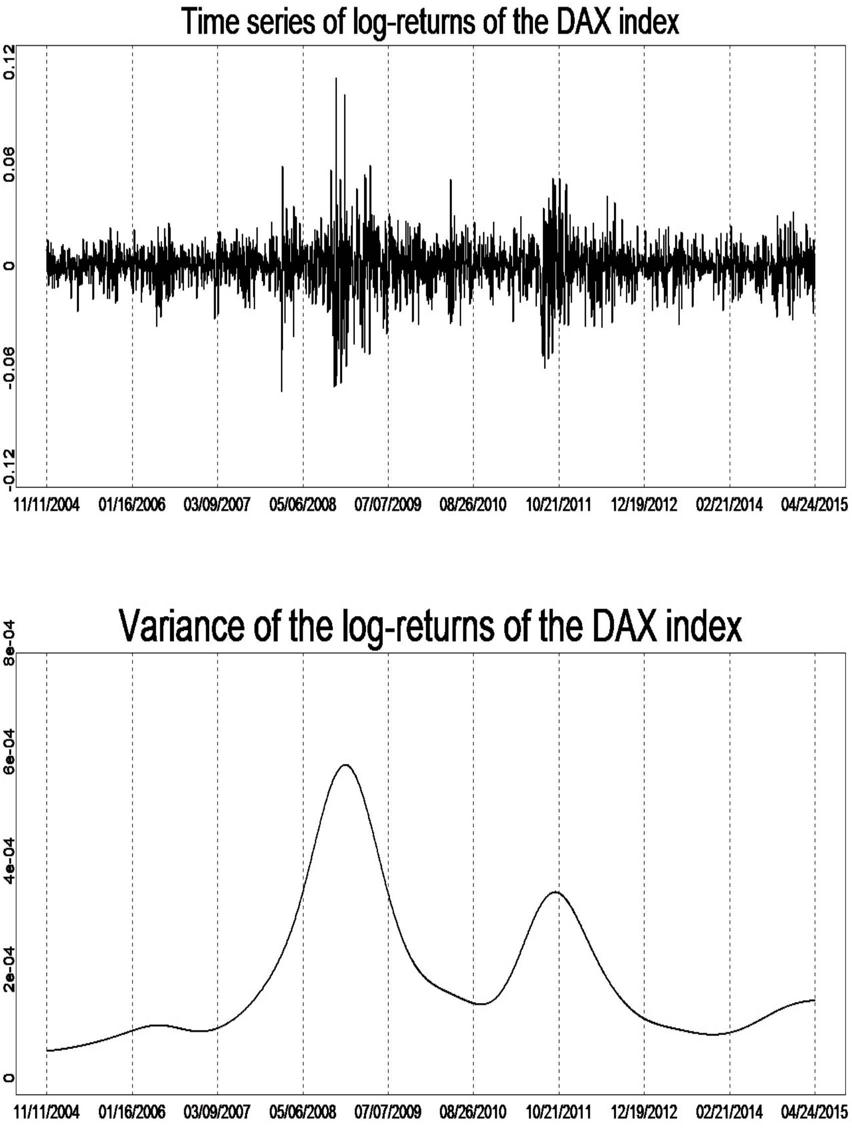


Figure 1.1: ***Time series and estimated variance of the log-returns of the DAX stock index from November 2004 to April 2015***

Sample Size	Proportion of rejections
250	0.8
125	0.35
100	0.2
50	0.16
25	0.07

Sample Size	Proportion of rejections
250	0.8
125	0.35
100	0.2
50	0.06
25	0.04

Sample Size	Proportion of rejections
250	0.9
125	0.4
100	0.4
50	0.18
25	0.07

Table 1.1: ***Jarque Bera-testing of the log-returns of the German, the French and the Belgian stock index from November 2004 to April 2015***

Sample Size	Proportion of rejections
250	0.8
125	0.35
100	0.2
50	0.06
25	0.04

Sample Size	Proportion of rejections
250	0.8
125	0.35
100	0.24
50	0.12
25	0.06

Sample Size	Proportion of rejections
250	0.7
125	0.35
100	0.24
50	0.1
25	0.078

Table 1.2: ***Shapiro-Wilk testing of the log-returns of the German, the French and the Belgian stock index from November 2004 to April 2015***

Multiscale time series modelling

We can push the parametric modelling paradigm further by assuming that the elements of the time series in (1.1) are generated by an underlying time series model. As for the assumption of joint normality, supposing the existence of an underlying time series model, aims at reducing the numbers of parameters of the joint probability distribution of the time series X_n.

A common approach in parametric time series modelling is the class of stationary moving average (MA) models (see Brockwell and Davis, 2009 [20]) in which each random variable is represented as a convolution of a sequence of independent normal random variables, denoted as $\{\varepsilon_k\}_k$, and referred to as increment terms, which means that for some l such that $0 \leq l < +\infty$, we have

$$X(t) = \sum_{k=0}^{l} \psi(k)\varepsilon_{t-k} \tag{1.7}$$

where the increment terms ε_{t-k} are such that $E\left(\varepsilon_{t-k}\right) = 0$ and $\mathrm{Var}\left(\varepsilon_{t-k}\right) = S$. According to the MA representation of the time series, the corresponding autocovariance function takes the form

$$C_n\left(\tau\right) = S\Psi(\tau) \tag{1.8}$$

with $\Psi(\tau)$ taking the form of an autocorrelation of the sequence of coefficients of the linear filter representation, that is $\Psi(\tau) = \sum_{k=0}^{l} \psi(k)\psi(k+\tau)$. Hence, the autocovariance function is entirely defined given the l filter coefficients of the MA and the variance of the increment terms, that is S. Thus, the main advantage of the moving average decomposition of the time series is the highly sparse representation of its autocovariance function in terms of a small set of parameters. However, the number of lags of the moving average model has to be fixed a priori or needs to be derived using data driven model selection procedures. In addition to that, the sparse representation of the decomposed time series does not allow for a profound analysis of the data generating process.

One way of solving these issues is to drop the idea of decomposing the time series with respect to a single MA model representation and to adopt a decomposition based on multiple models. Multiscale modelling refers to a mathematical approach in which multiple models at different scales of resolution are used simultaneously to describe a given system. The multiscale approach presents several advantages compared to classic single-scale modelling approaches. First, note that a given system might be the aggregated outcome of a multitude of different phenomena. If these phenomena are homogeneous in their nature and behavior, modelling their combined outcome at an aggregated level might be a good compromise. However, if the sub-systems are characterized by strong heterogeneity, finding an appropriate model specification at the macro-level is not always possible. In that case, it might be easier to define individual models for the different sub-systems at different scales of resolution and model the system at the macro-level by aggregating the individual modelling approaches. Secondly, if the system at the macroscale is a result of various types of heterogeneous phenomena, the aggregation of the latter might hide a lot of valuable information. Given that multiscale approaches provide insights from different scales of resolution, they allow to reveal patterns within a given system that are hidden at the macroscale representation. Finally, if only a few of the constituent sub-systems render good approximation to the aggregated system, multiscale methods allow to distinguish the informative from the non-informative building blocks and to obtain significant compression for the representation of the original system.

Multiscale time series modelling approaches use a set of MA models to represent one single time series. More precisely, each random variable of the time series is expressed as a sum of random variables as

$$X(t) = \sum_{j=-J}^{-1} X_j(t) \tag{1.9}$$

where each of the random variables $X_{-1}(t), \ldots, X_{-J}(t)$ has a stationary MA representation as

$$X_j(t) = \sum_{k=0}^{l(j)} \psi_j(k) \varepsilon_{j,t-k}. \tag{1.10}$$

where the increment terms $\varepsilon_{j,t-k}$ are such that $E\left(\varepsilon_{j,t-k}\right) = 0$ and $\mathrm{Var}\left(\varepsilon_{j,t-k}\right) = S_j$. If, in addition to that, we assume that the increment terms belonging to different MA representations in 1.10 are independent, the covariance function of the time series is given by

$$\begin{aligned} C(\tau) &= \sum_{j=-J}^{-1} \operatorname{Cov}\left(X_j(t), X_j(t+\tau)\right) \\ &= \sum_{j=-J}^{-1} S_j \Psi_j(\tau) \end{aligned} \tag{1.11}$$

with $\Psi_j(\tau) = \sum_{k=-\infty}^{+\infty} \psi_j(k)\psi_j(k+\tau)$.

High-dimensional time series modelling

In most cases, statisticians are not only interested in modelling and estimating the autocovariance function of a single time series but of a collection of p different series in the context of a multivariate time series model, that is

$$\begin{aligned} &X_n^{(u)} = \left\{X^{(u)}(t)\right\}_{t=0}^{n-1} \\ &u = 1, \ldots, p. \end{aligned} \tag{1.12}$$

The autocovariance function of a multivariate time series

$$\begin{gathered} C_n^{(u,v)}(t,\tau) = E\left(X^{(u)}(t)X^{(v)}(t+\tau)\right) \\ t = 0, \ldots, n-1; \tau = -t, \ldots, n-1-t \\ u, v = 1, \ldots, p. \end{gathered} \tag{1.13}$$

In classic multivariate time series modelling, the number of constituent time series is supposed to be much smaller than the length of the latter and hence $p << n$. The inherent assumption on which all methods of classic multivariate analysis are based, is that statisticians can use some a priori knowledge to specify a small subset of variables which are relevant for the analysis. However, in most statistical data analysis procedures the subset of relevant variables is not known beforehand. Hence in most cases, we face a large set of possibly relevant variables instead of a small set of effectively relevant variables. This potentially leads to a high number of variables within the model framework and hence to a high-dimensional dataset. This shift in the modelling paradigm of multivariate time series goes hand in hand with a change in the type of asymptotic reasoning. Asymptotic theory in classic multivariate modelling is based on the idea that the number of variables p is fixed with the sample size n going to infinity. In high-dimensional settings the number of variables is supposed to be such that $p = O(n^{\alpha})$ with $\alpha > 0$, which leads to a high-dimensional asymptotic reasoning in which the number of variables increases along with the sample size. However, in most cases the assumption that α is merely larger than zero is not enough. These kind of settings require another type of increasing dimension asymptotics, in which the number of variables increases along with the sample size such that the ratio $\frac{p}{n}$ tends to a constant, which means that α is supposed to be larger than or equal to one.

Contributions of the thesis

Locally stationary wavelet (LSW) processes, introduced in the seminal paper of Nason et al. (2000) [94], are a specific type of multiscale time series models, which combine the multiscale modelling paradigm with the idea of nonstationary autocovariance functions which vary smoothly in time. LSW processes decompose the time series into a sum of random variables as in (1.9) with each of its elements admitting a MA representation as in (1.10). However, the sequence of coefficients of each of the MA representations is set to be equal to a discrete wavelet function in the sense that the sequence of the representation at scale j is given by the discrete wavelet function at scale j, with $j = -1, \ldots, -J$. The wavelet functions at the different scales stem from the same multiresolution analysis which is supposed to be known in advance. The variances of the increment terms are assumed to be Lipschitz continuous which reflects in the slowly changing autocovariance function of the time series at the macro-scale. Sanderson et al. (2010)[108] expand the LSW model framework to the bivariate setting. A multivariate LSW model is introduced in Park et al. (2014) [98] and Cho and Fryzlewicz (2015) [24]. Note that these models use Lipschitz continuity as a regularity assumption on the variance and cross-covariance functions of the increment terms. Van Bellegem and von Sachs (2008) [129] introduce a univariate LSW process in which the Lipschitz continuity assumption is replaced by the more general notion of bounded total variation. Fryzlewicz et al. (2003) [60] addresses the issue of forecasting univariate LSW processes.

The contributions of this thesis are build around one major drawback of the current approach to LSW process modelling. The vectors of coefficients of the multiscale moving average filter in the LSW representation are time-varying vertical stretchings or compressions of discrete wavelet functions at different scales of resolution, which puts some serious limitations on the form of the autocovariance function of the time series. This becomes apparent in the multivariate LSW framework, which uses the same collection of wavelet functions in each of the constituent LSW processes of the multivariate model framework. Note that even though the form of the time-varying stretchings or compressions of the wavelet functions is different for each of the constituent time series, the assumption of identical wavelet functions across the constituent LSW processes leads to an autocovariance function which is such that for $u, v = 1, \ldots, p$

$$C_n^{(u,v)}(t,z) = C_n^{(v,u)}(\tau,z) \quad \text{for } \tau \in \mathbb{Z} \text{ and } t \in \mathbb{Z}.$$

In this thesis we propose a novel approach to LSW modelling. To do this, we start from the representation in 1.9. As it is the case for most multiscale modelling approaches, LSW processes are based on the assumption that the components of the sum of random variables are independent from each other. We expand the LSW model framework by allowing for correlation between these random variables by introducing a correlation pattern within the multiscale increment process such that for $j = -J+1, \ldots, -1$ and $-J - j \leq i \leq -1$

$$\text{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,l}\right) = \begin{cases} S_{i,j}(z) & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases} . \tag{1.14}$$

with $z = \frac{k}{n}$.

To see the impact of the more flexible covariance pattern among the increment terms, denote as $S(z)$ the covariance matrix matrix of the sequence of increments $\{\varepsilon_{j,k}\}_{j=-J}^{-1}$. $U(z)$ is a matrix composed of the eigenvectors of $S(z)$ and $\Lambda(z)$ is a diagonal matrix containing the eigenvalues of $S(z)$. Hence, we can write each random variable of the time series as

$$X_n(t) = \sum_{m=1}^{J} \sum_{k=0}^{n-1} \Delta_m(k) Z_{m,k} \tag{1.15}$$

where $\Delta_m(k) = \lambda_m(z) \sum_{j=-J}^{-1} u_{j,m}(z)\psi_j(k)$ and $Z_{m,k}$ are independent increment terms with zero mean. Thus, we obtain a representation as in (1.9) and (1.10).

However, the vectors of coefficients of the multiscale MA representation are time varying linear combinations of the set of discrete wavelet functions instead of mere time-varying stretchings or compressions of the latter.

Outline of the thesis

The thesis is divided into five parts, covering different fields of statistical analysis

Chapter 2: Wavelet analysis

Chapter 3: Univariate time series modelling

Chapter 4: Multivariate time series modelling

Chapter 5: Factor modelling

Chapter 6: High-dimensional statistics

We now provide a short overview of the main results derived in each of these chapters.

Chapter 2
Cross-correlations of discrete wavelet functions

We focus on the problem of measuring the level of redundancy which exists within a given collection of discrete non-decimated wavelet functions, denoted as $\psi_J = \{\psi_{j,k}\}_{k=0}^{+\infty}$ with $j = -1, \ldots, -J$, where j represents the dilation index and k is the translation index. For that purpose we introduce the novel concept of classes of cross-correlation wavelet functions (CCWF) of non-decimated wavelets, which are defined as inner products of the elements of ψ_J. We denote as $\Psi_J = \{\Psi_{i,j}\}_{i=-J-j}^{J+j}$ with $j = -1, \ldots, -J$ the set of CCWF where

$$\Psi_{i,j}(\tau) = \sum_{t=-\infty}^{+\infty} \psi_{j,0}(t)\psi_{j+i,\tau}(t) \quad \text{with } \tau \in \mathbb{Z}$$

We derive several properties of the class of cross-correlation wavelet functions and show that the elements of Ψ_J are independent for any integer $J > 0$.

Chapter 3
Generalized locally stationary wavelet processes

We introduce the class of LSW(q) processes, consisting of doubly-indexed sequences of random variables, denoted as $X_n = \{X_n(t)\}_{t=0}^{n-1}$, $n > 0$, such that

$$X_n(t) = \sum_{j=-\infty}^{-1} \sum_{k=0}^{n-1} W_j\left(\frac{k}{n}\right) \varepsilon_{j,k} \psi_{j,k}(t)$$

with the correlation pattern among the increment terms being given by

$$\mathrm{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,l}\right) = \begin{cases} \Gamma_{i,j}\left(\frac{k}{n}\right) & \text{if } -q \leq i \leq 0 \text{ and } k = l \\ 0 & \text{if } i < -q \text{ or } k \neq l \end{cases} .$$

for some integer $q \geq 0$. We show that the autocovariance function of X_n converges towards the local autocovariance function

$$C_q(\tau, z) = \sum_{j=-\infty}^{-1} S_{0,j}(z) \Psi_{0,j}(\tau) + \sum_{i=-q}^{-1} \sum_{j=-\infty}^{-1} S_{i,j}(z) \left(\Psi_{i,j}(\tau) + \Psi_{-i,j+i}(\tau)\right),$$

$$z \in (0,1), \quad \tau \in \mathbb{Z}$$

where $\Psi = \{\Psi_{0,j}, \{\Psi_{i,j}, \Psi_{-i,j+i}\}_{i=-q}^{-1}\}_{j=-\infty}^{-1}$ are a set of CCWF introduced in Chapter 2. We prove that in the case of the Haar basis there exists $\delta > 0$ such that the smallest eigenvalue of the inner product operator A of the set Ψ behaves as $\lambda_{\min}(A) > \delta$. We derive a consistent estimator of the vector of coefficients $\{\{S_{i,j}\}_{i=-q}^{0}\}_{j=-\infty}^{-1}$ using a smoothed and corrected wavelet periodogram approach. In the applied section, we use LSW(q) processes to define a coherence measure for time series of financial assets. We compute the coherence measure for the German, the Greek and the Irish stock indices using log-return data from 2003 to 2015.

Chapter 4
Multivariate locally stationary wavelet processes

We introduce the class of multivariate LSW(q) processes, consisting of doubly-indexed sequences of random variables, denoted as $X_n = \{X_n^{(u)}(t)\}_{t=0}^{n-1}$, $n > 0$, $u = 1, \ldots, p$, such that

$$X_n^{(u)}(t) = \sum_{j=-\infty}^{-1} \sum_{k=0}^{n-1} W_j\left(\frac{k}{n}\right) \varepsilon_{j,k}^{(u)} \psi_{j,k}(t)$$

with the correlation pattern among the increment terms being given by

$$\mathrm{Cov}\left(\varepsilon_{j,k}^{(u)},\varepsilon_{j+i,l}^{(v)}\right)=\begin{cases}\Gamma_{i,j}^{(u,v)}\left(\frac{k}{n}\right) & \text{if } -q\le i\le 0 \text{ and } k=l\\ 0 & \text{if } i<-q \text{ or } k\neq l\end{cases}.$$

for $u,v=1,\ldots,p$ and some integer $q\geq 0$. We show that the autocovariance function of X_n converges towards the local autocovariance function

$$\begin{aligned}C^{(u,v)}(\tau,z) &= \sum_{j=-\infty}^{-1} S_{0,j}^{(u,v)}(z)\,\Psi_{0,j}(\tau)+\sum_{i=-q}^{-1}\sum_{j=-\infty}^{-1} S_{i,j}^{(u,v)}(z)\,\Psi_{i,j}(\tau)\\ &\quad+\sum_{i=-q}^{-1}\sum_{j=-\infty}^{-1} S_{i,j}^{(v,u)}(z)\,\Psi_{-i,j+i}(\tau)\end{aligned}$$

$$z\in(0,1);\quad \tau\in\mathbb{Z};\quad u=1,\ldots,p;\quad v=1,\ldots,p$$

where $\Psi=\{\Psi_{0,j}\{\Psi_{i,j},\Psi_{-i,j+i}\}_{i=-q}^{-1}\}_{j=-\infty}^{-1}$ are a set of CCWF introduced in Chapter 3. We derive consistent estimators of the set of coefficients $\{\{S_{i,j}^{(u,v)}\}_{i=-q}^{0}\}_{j=-\infty}^{-1}$ with $u,v=1,\ldots,p$, using a smoothed and corrected wavelet periodogram approach. We show that the consistency result remains valid if $p=O(n^{\alpha})$ with $0\leq\alpha<\frac{1}{3}$. In the applied section, we derive the impulse response functions of the MLSW(q) process model. We compute the cosine similarity of the impulse response functions of Euro interbank offered rates (Euribors) of different maturities on the one hand and the Euro currency exchange rate on the other hand using log-return data from 2003 to 2015.

Chapter 5
Locally stationary wavelet factor model

Consider the vectors of increments of the MLSW(q) model framework $\varepsilon_{j,k}=(\varepsilon_{j,k}^{(1)},\ldots,\varepsilon_{j,k}^{(p)})^t$ with $j=-1,\ldots,-\log_2(n)$ and $\varepsilon_t=(\varepsilon_{-1,k},\ldots,\varepsilon_{-\log_2(n),k})^t$ and denote as $S(z)=E(\varepsilon_k\varepsilon_k^t)$ the EWS matrix.

We decompose ε_k as

$$\varepsilon_k=B_n(z)F_k+\xi_k \tag{1.16}$$

where $B_n(z)$ is a $Jp\times K$ matrix of factor loadings, F_k is a $K\times 1-$dimensional random vector of common factors and ξ_k is a $Jp\times 1-$dimensional random vector of idiosyncratic components. We derive a consistent estimation procedure of the number of common factors and the standardized factor loadings as $n \to +\infty$ with $p \to +\infty$ as $p = O(n^{\alpha})$ with $0 < \alpha < \frac{1}{3}$, given persistence assumptions for the common factors and the idiosyncratic components, using a singular value decomposition of the estimated evolutionary wavelet spectrum matrix. In the applied section, we use LSW factor processes to model the degree of economic integration of the founding members of the European currency area.

Chapter 6
Wavelet-based estimator of high-dimensional covariance matrices

We denote as $\{X(t)\}_{t=0}^{n-1}$ a $p-$dimensional random process with independent and identically distributed random vectors $X(t) \in \mathbb{R}^p$ such that $E(X(t)) = \mu_p$ and $\mathrm{Var}(X(t)) = \Sigma_p$. We propose a multiscale estimator of Σ_p using an unbalanced Haar wavelet decomposition. We prove that the multiscale estimator is consistent in operator norm under mild moment conditions of the elements of $X(t)$ and given conditions on the degree of sparsity of the wavelet transform of the population covariance matrix. In the applied section, we use the proposed estimator within mean-variance portfolio optimization settings. We show the good performance of the latter compared to shrinkage-type estimators in simulation studies and real-data examples using the log-returns of the stock prices of thirty US companies from 1990 to 2014.

Chapter 2

Cross-correlations of discrete wavelet functions

Multiscale analysis is based on the idea that a given function can be approximated, or under certain conditions be fully represented, as a weighted sum of a collection of functions whose elements are obtained through a set of transformations of a single basis function. Fourier (1822)[56] provides the first mention of the underlying idea of multiscale analysis by stating that the elements of the class of periodic functions can be expressed as a weighted sum of dilations of the complex exponential function, that is $e^{i2\pi x\omega} = \cos(2\pi\omega t) + i\sin(2\pi\omega t)$ with $x \in \mathbb{R}$, $\omega \in [0,1)$ and $i = \sqrt{-1}$.

The Fourier decomposition of a given function allows to reveal the frequency content of the latter. However, given that the elements of $\{e^{i2\pi t\omega}\}_{\omega}$ are mere dilations of the complex exponential, the Fourier decomposition does not provide any information on the variations of the frequency content on the standard domain. In contrast to Fourier analysis, wavelet multiscale analysis decomposes the frequency domain into discrete scales. The functions used in wavelet analysis are not only dilated versions of the underlying basis function but are generated upon the combined use of dilations and translations. Hence, at each scale of the wavelet decomposition there exists a collection of functions with elements being shifted versions of a single function which itself is obtained through a dilation of the basis function. The existence of a collection of shifted functions within each of the scales is the key element for studying the changing pattern of the frequency content within each of the scales.

A common starting point of wavelet analysis can be found in Grossman and Morlet (1984) [64] which uses window functions that were generated by dilation or compression of a prototype Gaussian to decompose seismic signals . Meyer (1986) [91] discovered that the basis functions proposed by Morlet and Grossman contained a great deal of redundancy and hence started working on developing wavelets with better localization properties resulting ultimately in the construction of orthogonal wavelet basis functions with very good time and frequency localization. However, the honor of having discovered the first orthonormal wavelet basis functions goes way back to Alfréd Haar (1910) [65]. Meyer and Mallat (1986) [91] developed the concept of multiresolution analysis for compactly supported wavelet functions which is the underlying design scheme for most discrete wavelet transforms (DWT) and the theoretical justification for the algorithm of the fast wavelet transform (FWT). Daubechies (1990) [31] and Mallat (1989a,b,c) [83][84][85] are credited with developing the transition from continuous to discrete signal analysis.

The first part of this chapter aims at providing an overview of the field of wavelet analysis by introducing some of its fundamental concepts such as multiresolution analysis (see Section 2.1.1), the dilation equation (see Section 2.1.2) and the construction algorithm for discrete wavelet functions (see Section 2.1.4). In the second part of this chapter we focus on the problem of measuring the level of redundancy within a given collection of wavelet functions. Saito and Beylkin (1992) [107] study the notion of autocorrelation wavelet functions which allow to capture the redundancy of sets of wavelet functions belonging to the same scale of resolution. However, given that autocorrelation wavelet functions do not take into account any potential redundancy between two sets of functions belonging to different wavelet scales, the latter are not sufficient to capture the degree of redundancy of an entire system of wavelet functions. In this chapter we introduce the class of cross-correlation wavelet functions which allows to capture both, intra-scale and inter-scale redundancies and study the properties of the latter.

2.1 Wavelet Analysis

In this section we summarize some results on orthogonal wavelet systems. To begin with, we recall the concept of multiresolution analysis. We then show how the basic properties of the multiresolution analysis lead to the dilation equation of the scaling function. In the same context we introduce the notion of wavelet function. An exhaustive introduction to these concepts can be found in Daubechies (1992) [32] and Mallat and Mallat (2008) [86].

2.1.1 Multiresolution analysis of $L^2(\mathbb{R})$

Multiresolution analysis (see Meyer, 1986 [91] and Mallat, 1989b [84]) constitutes the conceptual framework for the construction of orthogonal wavelet bases and provides the design method for the discrete wavelet transform introduced in Section 2.1.4.

Definition 2.1. *Let $V_j,\ j \in \mathbb{Z}$ be a sequence of subspaces of $L^2(\mathbb{R})$. We say that $\{V_j\}_{j\in\mathbb{Z}}$ is a multiresolution analysis (MRA) of $L^2(\mathbb{R})$ if*

$$V_j \subset V_{j+1} \qquad \textit{(nestedness)}$$

$$\bigcup_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R}) \qquad \textit{(density)}$$

$$\bigcap_{j\in\mathbb{Z}} V_j = \emptyset \qquad \textit{(separation)}$$

$$f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1} \qquad \textit{(scaling)}$$

$$f(x) \in V_j \Leftrightarrow f(x-k) \in V_j \quad \text{for all } k \in \mathbb{Z} \qquad \textit{(translation invariance)}$$

and there exists a function $\phi(x) \in V_0$, with $\int_{\mathbb{R}} \phi(x)dx \neq 0$, called a scaling function such that the set $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis for V_0.

We will assume hereafter that all scaling functions $\phi(x)$ are real-valued functions with $\int_{\mathbb{R}} \phi(x)dx = 1$. Note however, that all the results mentioned in this section could be generalized to multiresolution analyses with complex-valued scaling functions.

Given the scaling property and the translation invariance property of the MRA we define the functions $\phi_{j,k}(x) \in V_j$ as

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k). \tag{2.1.1}$$

The multiplication by 2^j with corresponds to the scaling of the initial function. Suppose for example that the support of $\phi(x)$ was given by the interval $[0,1]$. The support of $\phi(2^j x)$ is then given by $[0,2^{-j}]$. Subtracting k leads to a shifting of the initial function and hence the support of $\phi(x-k)$ is $[k,1+k]$. Combining dilation and translation produces the function $\phi_{j,k}(x)$ with support $[2^{-j}k, 2^{-j}(k+1)]$.

We now show that given the properties of the MRA in Definition 2.1, the set of functions $\left\{\phi_{j,k}(x)\right\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_j. To begin with, using the scaling property and the translation invariance property of the MRA, the inner product in $L^2(\mathbb{R})$ of two functions in V_j is such that

$$
\begin{aligned}
< \phi_{j,k}(x), \phi_{j,l}(x) > &= \int_{\mathbb{R}} \phi_{j,k}(x)\phi_{j,l}(x)dx \\
&= 2^j \int_{\mathbb{R}} \phi(2^j x - k)\phi(2^j x - l)dx \\
&= \int_{\mathbb{R}} \phi(u)\phi(u-(l-k))du.
\end{aligned}
$$

for every scale $j \in \mathbb{Z}$ and each shift $k \in \mathbb{Z}$. Given that $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ is a collection of orthonormal functions we thus have that $< \phi_{j,k}(x), \phi_{j,l}(x) >$ is equal to 1 if $k = l$ and equal to 0 if $k \neq l$. Hence, $\{\phi_{j,k}(x)\}_{k\in\mathbb{Z}}$ is an orthonormal set of functions on V_j. Moreover, let $f(x)$ be any element in V_j. Then by the scaling property of the MRA we know that $f(2^{-j}x) \in V_0$ for every $j \in \mathbb{Z}$. Using again that $\{\phi(x-k)\}_{k\in\mathbb{Z}}$ represents an orthonormal basis for V_0, we thus have the decomposition $f(2^{-j}x) = \sum_{k\in\mathbb{Z}} a_k \phi(x-k)$ where $a_k \in \mathbb{R}$. Replacing x by $2^j x$ we can write

$$
f(x) = \sum_{k\in\mathbb{Z}} \left(\frac{a_k}{2^{\frac{j}{2}}}\right) 2^{\frac{j}{2}} \phi(2^j x - k) = \sum_{k\in\mathbb{Z}} \left(\frac{a_k}{2^{\frac{j}{2}}}\right) \phi_{j,k}(x).
$$

Hence, we can write any given element $f(x)$ of V_j as a linear combination of $\{\phi_{j,k}(x)\}_{k\in\mathbb{Z}}$. Moreover, given that the elements of $\{\phi_{j,k}(x)\}_{k\in\mathbb{Z}}$ are orthonormal to each other for fixed j, the latter represents an orthonormal basis of V_j.

A multiresolution analysis provides an interesting way to decompose $L^2(\mathbb{R})$. First, the approximation spaces V_j are nested which gives us the ability to zoom from one to another using the scaling property in Definition 2.1. Secondly, we have an orthonormal basis for V_j generated from one single basis scaling function $\phi(x) \in V_0$ using the dilation and translation properties.

Moreover, the density property tells us that the multiresolution analysis covers all of $L^2(\mathbb{R})$.

2.1.2 Dilation equation of $\phi(x)$

In this section we are going to derive the dilation equation of the function $\phi(x)$. The dilation equation plays a major role in wavelet theory given that it allows to build the function $\phi_{j,k}(x)$ from the set of functions $\left\{\phi_{j+1,l}(x)\right\}_{l\in\mathbb{Z}}$ and hence to design a recursive scheme to build the basis functions of all the elements of the sequence of spaces V_j, $j \in \mathbb{Z}$.

Using the nestedness property of the MRA we know that $\phi(x) \in V_1$. Given that $\left\{\phi_{1,k}(x)\right\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of V_1, we can represent $\phi(x)$ as

$$\begin{aligned} \phi(x) &= \sum_{l\in\mathbb{Z}} h_l \phi_{1,l}(x) \\ &= \sqrt{2} \sum_{l\in\mathbb{Z}} h_l \phi(2x-l). \end{aligned} \tag{2.1.2}$$

where $h_l \in \mathbb{R}$. Using that $< \phi(x), \phi_{1,k}(x) >= \sum_{l\in\mathbb{Z}} h_l < \phi_{1,l}(x), \phi_{1,k}(x) >$ and that $< \phi_{1,l}(x), \phi_{1,k}(x) >$ is equal to 1 if $k = l$ and equal to 0 if $k \neq l$, we have that $h_l =< \phi(x), \phi_{1,l}(x) >$ for every $l \in \mathbb{Z}$.

The coefficients $\dots, h_{-1}, h_0, h_1, \dots$ form what is called the *scaling filter* of the MRA. In general we can write for every scale j and every shift k

$$\phi_{j,k}(x) = \sum_{l\in\mathbb{Z}} a_{k,l} \phi_{j+1,l}(x) \tag{2.1.3}$$

where $a_{k,l} =< \phi_{j,k}(x), \phi_{j+1,l}(x) >$. The coefficients $a_{k,l}$ can be expressed in terms of the elements of the scaling filter

$$\begin{aligned} a_{k,l} &=< 2^{\frac{j}{2}} \phi(2^j x - k), 2^{\frac{j+1}{2}} \phi(2^{j+1} x - l) > \\ &= 2^j \sqrt{2} \int_{\mathbb{R}} \phi(2^j x - k) \phi(2^{j+1} x - l) dx \\ &= \int_{\mathbb{R}} \phi(u) \sqrt{2} \phi(2u - (l - 2k)) du \\ &= \int_{\mathbb{R}} \phi(u) \phi_{1,l-2k}(u) du \\ &= h_{l-2k}. \end{aligned}$$

Hence, the linear combination in (2.1.3) becomes

$$\phi_{j,k}(x) = \sum_{l\in\mathbb{Z}} h_{l-2k}\phi_{j+1,l}(x). \tag{2.1.4}$$

Hence, the basis functions of V_j can be defined as a linear combination of the basis functions of V_{j+1} which themselves are linear combinations of the basis functions of V_{j+2}. This leads to a recursive scheme which allows to link the basis functions of any element of the sequence $\{V_j\}_{j\in\mathbb{Z}}$ to the basis functions of the remaining elements. We can express the relation in (2.1.4) in terms of the scaling function, leading to the dilation equation of $\phi(x)$

$$\phi(2^j x - k) = \sqrt{2}\sum_{l\in\mathbb{Z}} h_{l-2k}\phi(2^{j+1}x - l). \tag{2.1.5}$$

for every $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$.

2.1.3 The wavelet function $\psi(x)$

Consider any function $f(x)$ in V_j. Given the nestedness property of the MRA we have that $f(x) \in V_{j+i}$ with $i \geq 0$ and hence we can express $f(x)$ as a linear combination of the basis functions of V_{j+i} that is $\left\{\phi_{j+i,k}(x)\right\}_{k\in\mathbb{Z}}$. To express $f(x)$ in terms of $\left\{\phi_{j+i,k}(x)\right\}_{k\in\mathbb{Z}}$ when $i < 0$, we need an additional type of functions.

Definition 2.2. *Suppose that $\{V_j\}_{j\in\mathbb{Z}}$ forms a multiresolution analysis for $L^2(\mathbb{R})$ with scaling function $\phi(x)$ that satisfies the dilation equation in* (2.1.5). *The wavelet function $\psi(x) \in V_1$ is defined as*

$$\psi(x) = \sqrt{2}\sum_{l\in\mathbb{Z}} g_l\phi(2x - l) \tag{2.1.6}$$

and more generally the functions $\psi_{j,k}(x) \in V_{j+1}$ by

$$\psi_{j,k}(x) = \sum_{l\in\mathbb{Z}} g_{l-2k}\phi_{j+1,l}(x) \tag{2.1.7}$$

for all integers $j,k \in \mathbb{Z}$. The wavelet filter is given by the coefficients $\ldots, g_{-1}, g_0, g_1, \ldots$ where

$$g_k = (-1)^k h_{1-k}. \tag{2.1.8}$$

Mallat (1989a [83], 1989b [84]) and Meyer (1986) [91] show that if $W_j = \text{span}\left\{\psi_{j,k}(x)\right\}_{k\in\mathbb{Z}}$ then W_j is the orthogonal complement of V_j in V_{j+1} which we denote as

$$V_{j+1} = V_j \otimes W_j \tag{2.1.9}$$

and that the set $\left\{\psi_{j,k}(x)\right\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of W_j.

Given (2.1.9) we can write each function $f(x) \in V_{j+1}$ as

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \phi_{j,k}(t) + \sum_{k\in\mathbb{Z}} b_k \psi_{j,k}(t). \tag{2.1.10}$$

The right-hand side of (2.1.10) contains two parts, where $\sum\limits_{k\in\mathbb{Z}} a_k \phi_{j,k}(t)$ represents the approximation of $f(x)$ in V_{j-1} and where $\sum_{k\in\mathbb{Z}} b_k \psi_{j,k}(t)$ is an element of W_j. Moreover, given that $\phi_{j,k}(t) \in V_j$, we can write the latter as $\phi_{j,k}(t) = \sum_{l\in\mathbb{Z}} a_{k,l}\phi_{j-1,k}(t) + \sum\limits_{l\in\mathbb{Z}} b_{k,l}\psi_{j-1,k}(t)$ and hence

$$f(x) = \sum_{k\in\mathbb{Z}} a_k \left(\sum_{l\in\mathbb{Z}} a_{k,l}\phi_{j-1,l}(t) + \sum_{l\in\mathbb{Z}} b_{k,l}\psi_{j-1,l}(t) \right) + \sum_{k\in\mathbb{Z}} b_k \psi_{j,k}(t).$$

Again we can write $\phi_{j-1,l}(t)$ as a linear combination of $\{\phi_{j-2,k}(t)\}_{k=-\infty}^{+\infty}$ and $\{\psi_{j-2,k}(t)\}_{k=-\infty}^{+\infty}$ with $\{\phi_{j-2,l}(t)\}_{k=-\infty}^{+\infty}$ being a linear combination of $\{\phi_{j-3,k}(t)\}_{k=-\infty}^{+\infty}$ and $\{\psi_{j-3,k}(t)\}_{k=-\infty}^{+\infty}$ and so on. Repetitive use of this recursive scheme then allows to write $f(x)$ as an infinite series of elements of $W_j, W_{j-1}, W_{j-2}, \ldots$ respectively.

Example 2.1 (Haar basis). *Haar (1910) [65] provides the first example of an MRA. The scaling filter of the Haar basis is given by the sequence of coefficients $\{h_k\}_{k=-\infty}^{+\infty}$ such that*

$$h_k = \begin{cases} \frac{\sqrt{2}}{2} & \text{for } k = 0,1 \\ 0 & \text{otherwise} \end{cases} \tag{2.1.11}$$

The wavelet functions have an analytic expression and take the form of *stepwise constant functions*

$$\psi_{j,k}(x) = \begin{cases} 2^{\frac{j}{2}} & \textit{if } x \in \left[-2^{-j}(1-k), 2^{-j}k\right] \\ \\ -2^{\frac{j}{2}} & \textit{if } x \in \left]2^{-j}k, 2^{-j}(k+1)\right] \\ \\ 0 & \textit{otherwise} \end{cases} \quad . \tag{2.1.12}$$

Example 2.2 (**Daubechies basis**). *The Haar MRA in Example 2.1 is part of the larger class of the so-called Daubechies MRA (see Daubechies, 1990 [31]). There are several types of Daubechies MRA, each of them using a different scaling filter. Table 2.1 below contains the scaling filter for the D2 to D8 MRA. Note that the D2 scaling filter corresponds to the Haar scaling filter. The wavelet functions of the D2 MRA has an analytic expression given by* (2.1.12)*. For the remaining Daubechies MRA, there exists no analytic expression of the corresponding wavelet functions.*

k	D2	D4	D6	D8
0	0.7071	0.4830	0.3328	0.2304
1	0.7071	0.8365	0.8069	0.7148
2	0	0.2241	0.4599	0.6309
3	0	-0.1294	-0.1350	-0.0280
4	0	0	-0.0854	-0.1870
5	0	0	0.0352	0.0308
6	0	0	0	0.0329
7	0	0	0	-0.0106

Table 2.1: ***Daubechies scaling filters (D2-D8)***

2.1.4 The discrete wavelet transform

Consider any deterministic sequence $s_0, \ldots, s_{n-1}$ in l_2 where $n = 2^J$ with $J \in \mathbb{N}$ and define the continuous function

$$f(x) = \sum_{k=-\infty}^{+\infty} c_{0,k}\phi(x-k) \quad x \in \mathbb{R}$$

with $c_{0,k} = s_k$ for $k = 0, \ldots, n-1$.

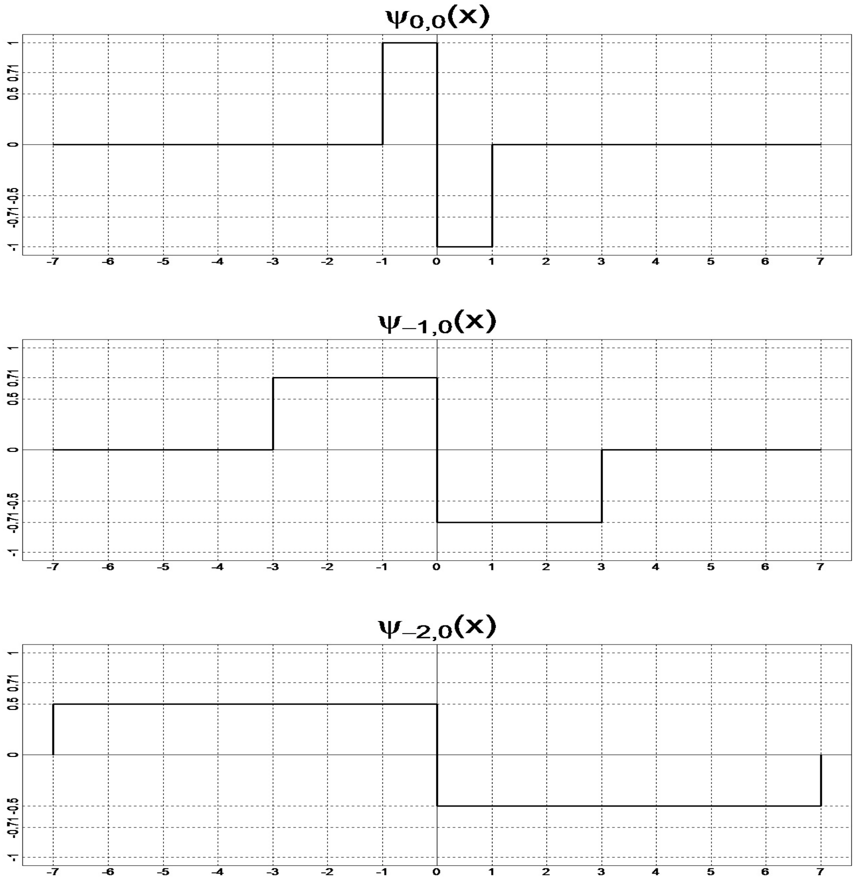


*Figure 2.1: **(Haar wavelet functions)** $\psi_{0,0}, \psi_{-1,0}$ **and** $\psi_{-2,0}$*

The function $f(x)$ is in the space V_0. Using (2.1.9), the function $f(x)$ can be decomposed as

$$f(x) = \sum_{k=-\infty}^{+\infty} c_{-1,k}\phi_{-1,k}(x) + \sum_{k=-\infty}^{+\infty} d_{-1,k}\psi_{-1,k}(x)$$

where $c_{-1,k} =< f(x), \phi_{-1,k}(x) >$ denote the scaling coefficients and $d_{-1,k} =< f(x), \psi_{-1,k}(x) >$ represent the wavelet coefficients. Using (2.1.5) and (2.1.7) we have that $\phi_{-1,k}(x) = \sum_l h_{l-2k}\phi(x-l)$ and $\psi_{-1,k}(x) = \sum_l g_{l-2k}\phi(x-l)$ respectively. Hence, the coefficients $c_{-1,k}$ and $d_{-1,k}$ take

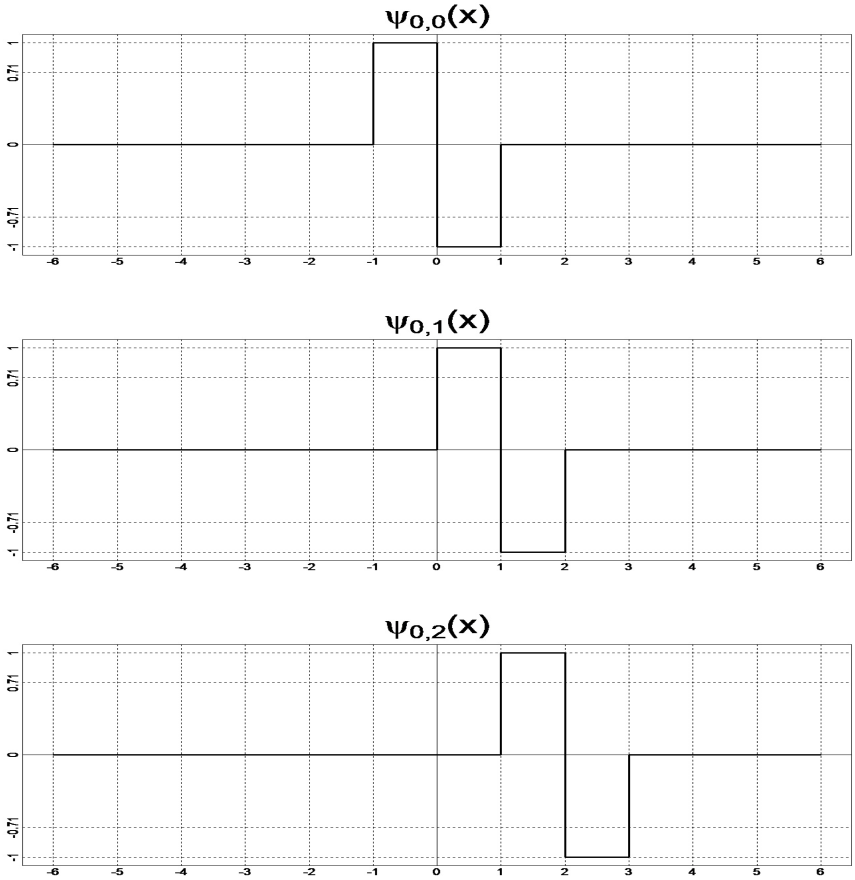


*Figure 2.2: **(Haar wavelet functions)** $\psi_{0,0}, \psi_{0,1}$ **and** $\psi_{0,2}$*

the form

$$c_{-1,k} = \sum_l h_{l-2k} c_{0,l} \quad \text{and} \quad d_{-1,k} = \sum_l g_{l-2k} c_{0,l}. \tag{2.1.13}$$

with $k = 0, \dots, \frac{n}{2} - 1$. Note that the number of scaling coefficients at level -1 is half of the number of scaling coefficients at level 0. The discrete wavelet function at level -1 and shift 0 is identical to the scaling filter vector, that is $\psi_{-1,0}(t) = g_t$ for $t \in \mathbb{Z}$ with $g_t \neq 0$ for $t = 0, \dots, L[h] - 1$ where $L[h] < +\infty$ denotes the length of the support of the scaling filter. From this follows that the support of the discrete wavelet function at scale -1 is $0 \le t \le L[h] - 1$.

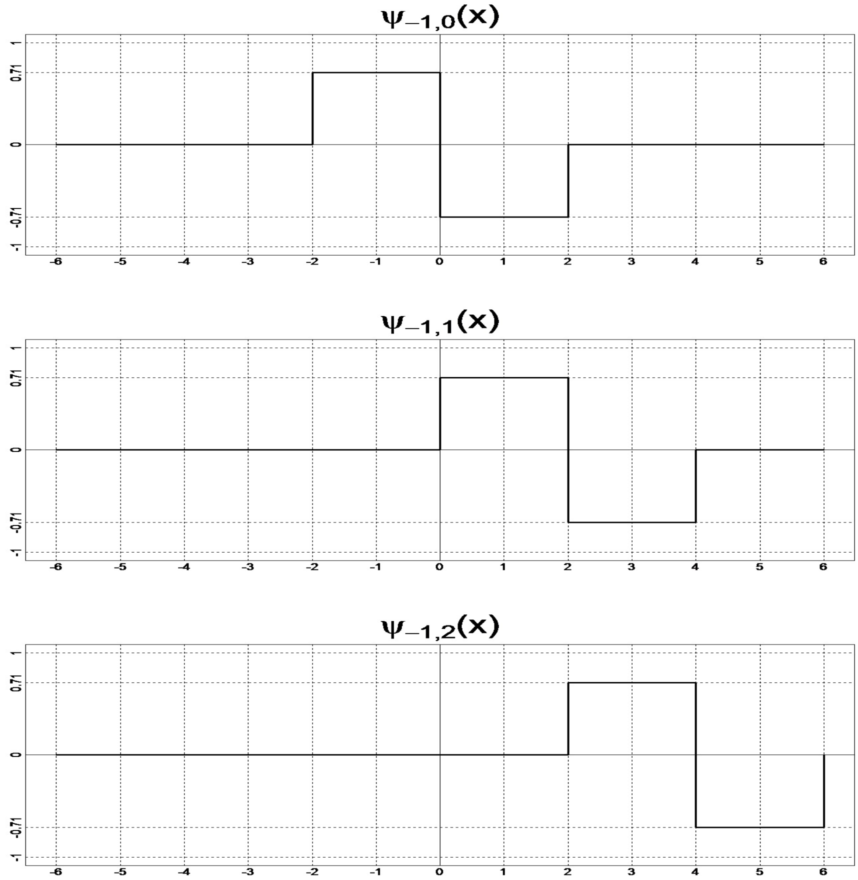


Figure 2.3: ***(Haar wavelet functions)*** $\psi_{-1,0}, \psi_{-1,1}$ ***and*** $\psi_{-1,2}$

Next consider the sequence of coefficients $c_{-1,0}, \ldots, c_{-1,\frac{n}{2}-1}$ and define the continuous function

$$f_{-1}(x) = \sum_k c_{-1,k}\phi_{1,k}(x) \quad x \in \mathbb{R}$$

which again can be decomposed as

$$f_{-1}(x) = \sum_k c_{-2,k}\phi_{-2,k}(x) + \sum_k d_{-2,k}\psi_{-2,k}(x)$$

with $c_{-2,k} =< f_{-1}(x), \phi_{-2,k}(x) >$ and $d_{-2,k} =< f_{-1}(x), \psi_{-2,k}(x) >$. Using (2.1.5) and (2.1.7) we have that $\phi_{-2,k}(x) = \sum_l h_{l-2k}\phi_{-1,l}(x)$ and $\psi_{-2,k}(x) =$

$\sum_l g_{l-2k}\phi_{-1,l}(x)$ respectively. Hence, the coefficients $c_{-2,k}$ and $d_{-2,k}$ take the form

$$c_{-2,k} = \sum_l h_{l-2k} c_{-1,l} \quad \text{and} \quad d_{-2,k} = \sum_l g_{l-2k} c_{-1,l}. \tag{2.1.14}$$

Using the expressions for $c_{-1,l}$ in (2.1.14) we can write $c_{-2,k}$ and $d_{-2,k}$ as

$$c_{-2,k} = \sum_m c_{0,m} \sum_l h_{m-2l} h_{l-2k} \qquad d_{-2,k} = \sum_m c_{0,m} \sum_l h_{m-2l} g_{l-2k} \tag{2.1.15}$$

with $k = 0, \ldots, \frac{n}{4} - 1$. The discrete wavelet function at level -2 and shift 0 then takes the form $\psi_{-2,0}(t) = \sum_l h_{t-2l}\psi_{-1,0}(l) = \sum_l h_{t-2l} g_l$ with $l \in \mathbb{Z}$. Recall that $h_{t-2l} = 0$ for $t < 2l$ and $t > L[h] - 1 + 2l$ and that $g_l = 0$ for $l < 0$ and $l > L[h] - 1$. If we combine these constraints on the support of the scaling filter and the wavelet filter we get that the support of the discrete wavelet function at scale -2 is given by $0 \leq t \leq 3\,(L[h] - 1)$.

This procedure can be iterated leading to a recursive definition of the discrete wavelet functions

$$\begin{array}{ll} \psi_{-1,0}(t) = g_t & \text{for } t \in \mathbb{Z} \\ \psi_{j-1,0}(t) = \sum\limits_k h_{t-2k}\psi_{j,0}(k) & \text{for } t \in \mathbb{Z} \text{ and } j = -2, \ldots, -J. \end{array} \tag{2.1.16}$$

Hence, the wavelet functions at each scale of the discrete wavelet transform are defined as repeated convolutions of the scaling and wavelet filters. Moreover, the length of the support of the wavelet functions L_j is a multiple of the length of the support of h

$$L_j = \left(2^j - 1\right)(L[h] - 1) + 1. \tag{2.1.17}$$

Figures 2.4 and 2.5 contain the discrete D4 wavelet functions for $j = -1, \ldots, -2$ and $k = 0$ and $j = -3, \ldots, -4$ and $k = 0$ respectively.

2.1.5 Non-decimated discrete wavelet system

In the decimated discrete wavelet system, the functions $\psi_{j,k}(t)$ are shifted versions of the function $\psi_{j,0}(t)$ in (2.1.16) as

$$\psi_{j,k}(t) = \psi_{j,0}\left(t - 2^{-j}k\right). \tag{2.1.18}$$

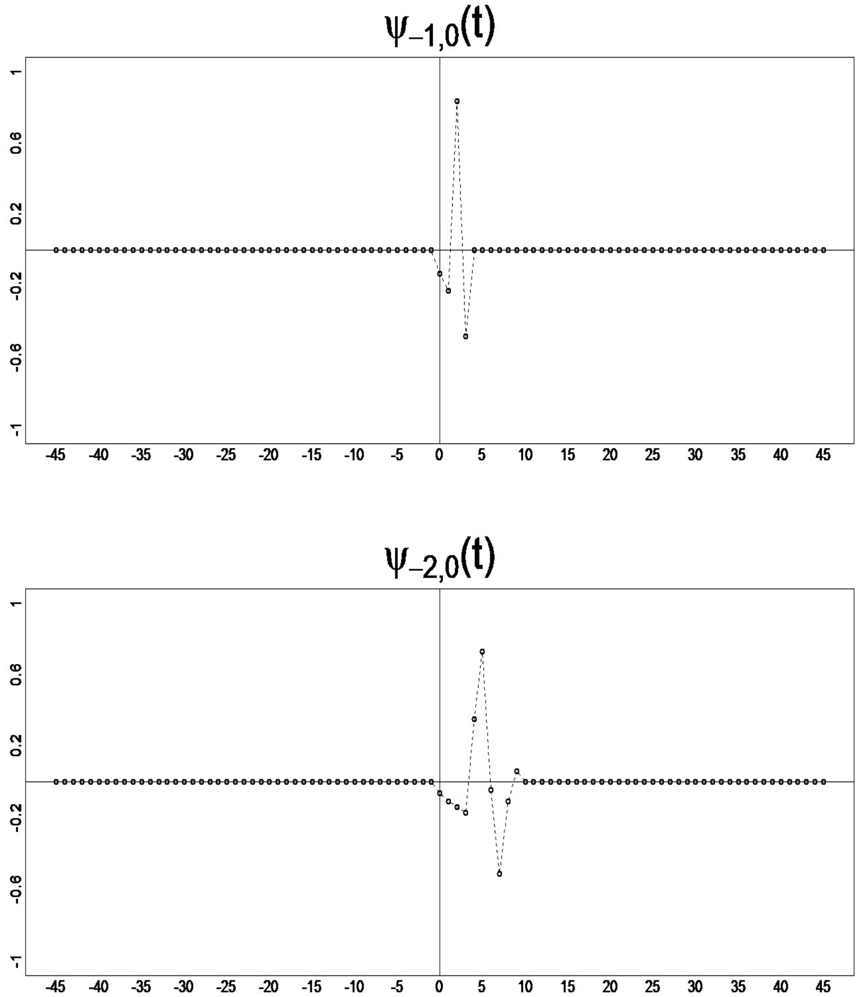


Figure 2.4: ***Discrete wavelet functions of the D4 basis at shift for*** $j=-1,\ldots,-2$ ***and*** $k=0$

Thus, the decimated wavelet system only allows for dyadic shifts of the wavelet functions with the size of the shifts being dependent upon the respective wavelet scale. As we have seen in Section 2.1.4, this scale-variance of the shifting operation leads to the decimation phenomenon with the number of coefficients at a given scale of resolution being half of the number of wavelet coefficients at the previous scale.

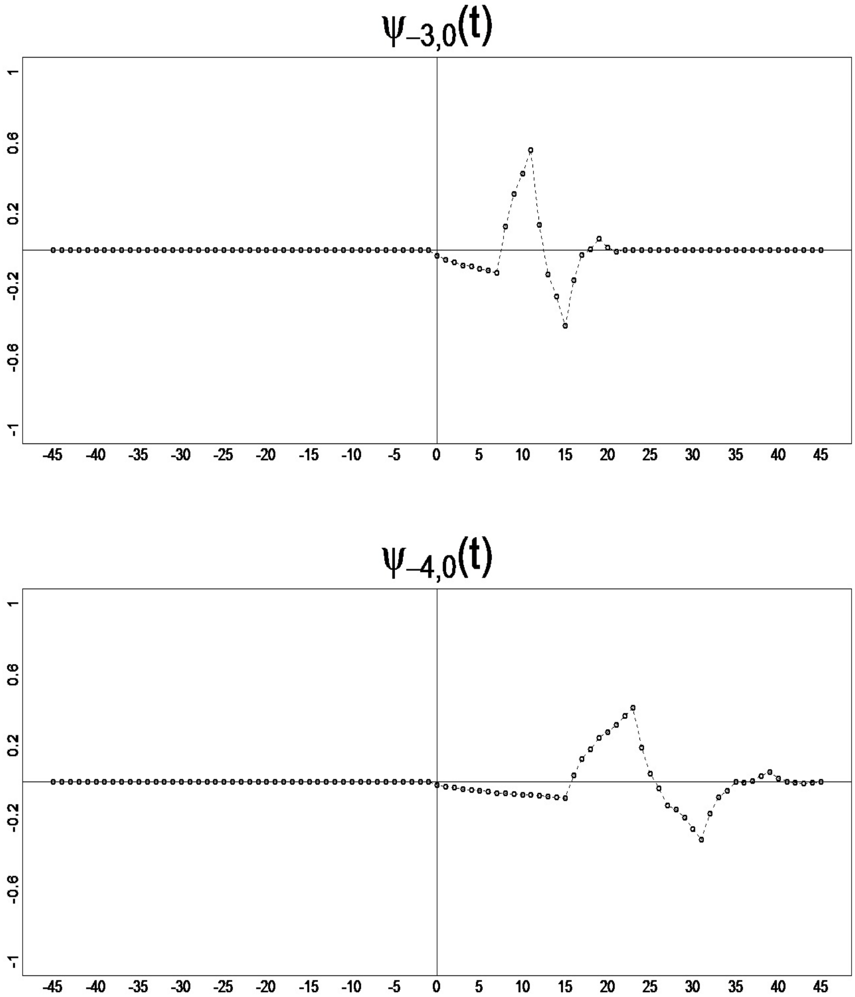


Figure 2.5: ***Discrete wavelet functions of the D4 basis at shift for*** $j = -3, \ldots, -4$ ***and*** $k = 0$

Non-decimated wavelet systems (see Berkner and Wells, 2002 [15]) refer to wavelet systems which are constructed such that the decimation phenomenon does not occur. More precisely, non-decimated wavelet systems allow for the wavelet functions to be shifted to any location and not just by shift of 2^j. Hence, in the non-decimated wavelet system, the wavelet function at scale j

and shift τ is given by

$$\psi_{j,\tau}(t) = \psi_{j,0}(t-\tau). \quad (2.1.19)$$

Hence, the decimated wavelet function at scale j and shift k is identical to the non-decimated wavelet function at scale j and shift $2^{-j}k$.

In contrast to the decimated wavelet system, the number of wavelet functions at each scale of resolution does not depend on the wavelet scale. Hence, the number of wavelet functions in the non-decimated wavelet system is always larger than in the decimated system. Note however, that the larger number of functions comes at a price in the sense that the constituent functions of the non-decimated wavelet system are not linearly independent. The linear dependence among the elements of the non-decimated wavelet system leads to a redundant representation of a given signal in the wavelet domain. In the next section we introduce the concept of cross-correlation wavelet functions allowing to measure the degree of redundancy of given systems of non-decimated wavelet functions.

2.2 Cross-correlations of discrete wavelet functions

Consider two real sequences $x(t)$ and $y(t)$ indexed by $t \in \mathbb{Z}$. The cross-correlation of these two sequences, denoted as $C_{x,y}(\tau)$, takes the form

$$C_{x,y}(\tau) = \sum_{t=-\infty}^{+\infty} x(t)y(t+\tau) \quad \text{for } \tau \in \mathbb{Z}. \quad (2.2.1)$$

The cross-correlation is a sliding sliding inner product of two signals and hence represents a measure of similarity of $x(t)$ and $y(t)$ as a function of the lag of one relative to the other. The autocorrelation is defined as the cross-correlation of a series with itself. We denote as $\{C_{x,y}(\tau)\}_{\tau=-\infty}^{+\infty}$ the cross-correlation sequence of $x(t)$ and $y(t)$. Consequently, we refer to $\{C_{x,x}(\tau)\}_{\tau=-\infty}^{+\infty}$ and $\{C_{y,y}(\tau)\}_{\tau=-\infty}^{+\infty}$ as the autocovariance sequence of $x(t)$ and $y(t)$ respectively. If both, $x(t)$ and $y(t)$ are discrete wavelet functions originating from a common recursive scheme as in (2.1.16), the corresponding sequence of cross-correlations is referred to as cross-correlation wavelet sequences or cross-correlation wavelet functions (see Definition 2.3). Definition 2.3 below formally introduces the notion of CCWF for of collections non-decimated discrete wavelet functions.

Definition 2.3. *Denote as* $\{\psi_{j,\tau}\}_{\tau\in\mathbb{Z}}$ *with* $j=-1,\ldots,-J$ *and* $J\in\mathbb{N}^+$ *a collection of non-decimated wavelet functions. We denote as* $\Psi_J = \{\{\Psi_{i,j}\}_{i=-j-1}^{-J-j}\}_{j=-J}^{-1}$ *the set of cross-correlation functions of the collection of non-decimated wavelet functions, such that*

$$\Psi_{i,j}(\tau) = \sum_{t=-\infty}^{+\infty} \psi_{j,0}(t)\psi_{j+i,\tau}(t) \quad \textit{with } \tau\in\mathbb{Z} \tag{2.2.2}$$

for $i=-j-1\ldots,-J$ *and* $j=-1,\ldots,-J$.

Note that if the wavelet functions are standardized as $\sum_{t=-\infty}^{+\infty}\psi_{j,0}^2(t)=1$ and $\sum_{t=-\infty}^{+\infty}\psi_{j+i,0}^2(t)=1$, the CCWF $\Psi_{i,j}(\tau)$ is identical to the cosine of the angle between the vectors $\psi_{j,0}$ and $\psi_{j+i,\tau}$.

The function $\Psi_{0,j}$ is identical to the autocovariance wavelet function (ACWF) studied in Saito and Beylkin (1992) [107]. The continuous father wavelet autocorrelation function is the fundamental function of the Deslauriers and Dubuc (1989) [34] interpolation scheme. Discrete autocovariance wavelets functions play a crucial role in the LSW process model framework which will be introduced in Chapter 3.

The ACWF at each scale are symmetric about zero, that is $\Psi_j(\tau)=\Psi_j(-\tau)$ for all j, for any wavelet basis. To show this, set $k'=k+\tau$ which leads to $\psi_{j,0}(k)\psi_{j,0}(k+\tau)=\psi_{j,0}(k'-\tau)\psi_{j,0}(k')$ and hence

$$\sum_{k=-\infty}^{+\infty}\psi_{j,0}(k)\psi_{j,0}(k+\tau) = \sum_{k'=-\infty}^{+\infty}\psi_{j,0}(k'-\tau)\psi_{j,0}(k')$$

for all $\tau\in\mathbb{Z}$.

To derive the upper and the lower bounds of the support of the autocorrelation wavelets recall from (2.1.17) that the support of $\psi_{j,0}(k)$ is given by $\left[0,\left(2^{-j}-1\right)(L[h]-1)\right]$ and that the support of $\psi_{j,0}(k+\tau)$ is given by $\left[-\tau,\left(2^{-j}-1\right)(L[h]-1)-\tau\right]$. Thus, if $\tau\geq\left(2^{-j}-1\right)(L[h]-1)+1$ or $\tau\leq-\left(2^{-j}-1\right)(L[h]-1)-1$ we have that

$$\left[0,\left(2^{-j}-1\right)(L[h]-1)\right]\cap\left[-\tau,\left(2^{-j}-1\right)(L[h]-1)-\tau\right]=\emptyset$$

and as a consequence $\sum_{k=-\infty}^{+\infty}\psi_{j,0}(k)\psi_{j,0}(k+\tau)=0$. Thus, we have that $\Psi_j(\tau)=0$ if $|\tau|\geq\left(2^{-j}-1\right)(L[h]+1)$.

Example 2.3 (Haar autocorrelation wavelet functions). *The autocorrelation functions of discrete Haar wavelets have an analytic representation given by*

$$\Psi_j(\tau)=\begin{cases}1-2^j3|\tau| & \text{if } \tau\in\left[0,2^{-j-1}\right]\cap\mathbb{Z}\\ -\frac{1}{2}+2^j|\tau| & \text{if } \tau\in\left[2^{-j-1},2^{-j}-1\right]\cap\mathbb{Z}\\ 0 & \text{otherwise}\end{cases}.$$

Given that most discrete wavelet functions, as for example the members of the Daubechies wavelet class (apart from the D2 basis), do not have an analytic representation, their respective ACWF does not have an analytic representation either. However, as Lemma 2.1 below states, there exists a set of points given by $\tau=2^{-j-1}m$ with $m\in\mathbb{Z}$, for which the ACWF at level j can be expressed as an autocorrelation of the wavelet filter and hence admits a closed form expression.

Lemma 2.1. *Denote as* $\{\psi_{j,\tau}\}_{\tau\in\mathbb{Z}}$ *with* $j=-1,\ldots,-J$ *and* $J\in\mathbb{N}^+$ *a collection of non-decimated wavelet functions and suppose that the associated scaling filter fulfills*

$$\sum_{k=-\infty}^{+\infty}h_kh_{k+2\delta}=\begin{cases}1 & \text{if } \delta=0\\ 0 & \text{if } \delta\in\mathbb{Z}_{\neq 0}\end{cases}. \tag{2.2.3}$$

If we denote by Ψ_j *the ACWF at the* $j-$*th wavelet scale, we have that for* $j=-1,\ldots,-J$ *and* $m\in\mathbb{Z}$

$$\Psi_j\left(2^{-j-1}m\right)=\sum_u g_ug_{u+m}.$$

Based on the statement of Lemma 2.1 we can construct Table 2.2 containing the values of the ACWF for $\tau=2^{-j-1}m$ with $m=0,\ldots,5$ for D2-D12 wavelet functions obtained by a direct computation of the autocorrelation function of the wavelet filter.

As pointed out before, ACWF are only a special case of the more general notion of cross-correlation wavelet functions. Unlike ACWF, the properties

	$\Psi_j(0)$	$\Psi_j\left(2^{-j-1}\right)$	$\Psi_j\left(2^{-j}\right)$	$\Psi_j\left(2^{-j-1}3\right)$	$\Psi_j\left(2^{-j+2}\right)$	$\Psi_j\left(2^{-j-1}5\right)$
D2	1	-0.5	0	0	0	0
D4	1	-0.5625	0	0.0625	0	0
D6	1	-0.5859	0	0.0977	0	-0.0117
D8	1	-0.5981	0	0.1196	0	-0.0239
D10	1	-0.6056	0	0.1346	0	-0.0346
D12	1	-0.6107	0	0.1454	0	-0.0436

Table 2.2: ***Point values for Daubechies-type ACWF***

of cross-correlations of discrete wavelet functions have not been studied in the literature so far. Hence, in the remaining part of this chapter we provide an overview of some of the main properties of the class of CCWF.

To start with, Lemma 2.2 states that unlike ACWF, cross-correlation wavelet functions are not symmetric about zero.

Lemma 2.2. *Denote as* $\Psi_J = \{\{\Psi_{i,j}\}_{i=-j-1}^{-J-j}\}_{j=-J}^{-1}$ *with* $J \in \mathbb{N}^+$ *a set of cross-correlation wavelet functions as defined in Definition 2.3. The elements of* Ψ_J *then have the following properties*

(a) $\Psi_{i,j}(\tau) = \Psi_{-i,j+i}(-\tau)$

(b) $\Psi_{i,j}(\tau) = 0$ *if* $\begin{cases} \tau \leq -\left(2^{-j}-1\right)\left(L[h]-1\right) \\ \tau > \left(2^{-j-i}-1\right)\left(L[h]-1\right)+1 \end{cases}$

(c) *The ACWF are given by* $\Psi_{0,j}(\tau) = \sum_{t=-\infty}^{+\infty} \psi_{j,0}(t)\psi_{j,\tau}(t)$

Example 2.4 (Haar cross-correlation wavelet functions). *The cross-correlation wavelet functions of discrete Haar wavelet functions for* $j = -1,-2,\ldots$ *and* $i = -1,-2,\ldots$ *take the following form*

$$\Psi_{i,j}(\tau) = \begin{cases} 2^{j+\frac{i}{2}}\left(2^{-j}-2\left|2^{-i}+2^{-j-1}-2-\tau\right|\right) & \text{if } \tau \in I_1 \\ -2^{j+\frac{i}{2}}\left(2^{-j-1}-\left|2^{-i+1}+2^{-j}+2^{-j-1}-4-\tau\right|\right) & \text{if } \tau \in I_2 \\ -2^{j+\frac{i}{2}}\left(2^{-j-1}-\left|2^{-j-1}-\tau\right|\right) & \text{if } \tau \in I_3 \\ 0 & \text{otherwise} \end{cases}$$

where $I_1 = \left[2^{-i}-1, 2^{-i}+2^{-j}-3\right] \cap \mathbb{Z}$, $I_2 = \left[2^{-i+1}+2^{-j}-3, 2^{-i+1}+2^{-j+1}-5\right] \cap \mathbb{Z}$ and $I_3 = \left[0, 2^{-j}-1\right] \cap \mathbb{Z}$.

Lemma 2.3 below states that for each element within a given collection of cross-correlation functions there exists a non-empty set of values of τ between $-\left(2^{-j-i}-1\right)(L[h]-1)+1$ and $\left(2^{-j-i}-1\right)(L[h]-1)-1$ for which $\Psi_{i,j}(\tau)=0$.

Lemma 2.3. *Denote as* $\{\psi_{j,\tau}\}_{\tau\in\mathbb{Z}}$ *with* $j=-1,\ldots,-J$ *and* $J\in\mathbb{N}^+$ *a collection of non-decimated wavelet functions and suppose that the associated scaling filter fulfills the assumption in* (2.2.3). *Suppose moreover that*

$$\sum_{k=-\infty}^{+\infty} h_{k+2\delta} g_k = 0 \tag{2.2.4}$$

for any $\delta\in\mathbb{Z}$. *If we denote as* $\Psi_{i,j}$ *the CCWF of order i at scale j, we have that for* $j,l=-1,\ldots,-J$, $i,s=-1\ldots,-J-j$ *and* $m\in\mathbb{Z}$

$$\Psi_{i,j}\left(2^{l-1}\left(2^{-s+1}-1\right)m\right)=0 \quad \text{if } l+s=j+i \text{ and } s>i.$$

Lemma 2.4 states that for each CCWF there exists a non-empty set of values of τ for which $\Psi_{i,j}(\tau)$ has a closed form expression as a product of a subset of elements of the scaling filter and wavelet filter.

Lemma 2.4. *Denote as* $\{\psi_{j,\tau}\}_{\tau\in\mathbb{Z}}$ *with* $j=-1,\ldots,-J$ *and* $J\in\mathbb{N}^+$ *a collection of non-decimated wavelet functions with scaling and wavelet filters fulfilling the assumptions in* (2.2.3) *and* (2.2.4). *If we denote as* $\Psi_{i,j}$ *the CCWF of order i at scale j and as* $L[h]$ *the length of the respective wavelet filter, we have that for* $j=-1,\ldots,-J$ *and* $i=-1\ldots,-J-j$

$$\Psi_{i,j}\left(2^{j-1}\left(2^{-i+1}-1\right)(L[h]-1)\right)=g_0 h^i_{L[h]-1} g_{L[h]-1}.$$

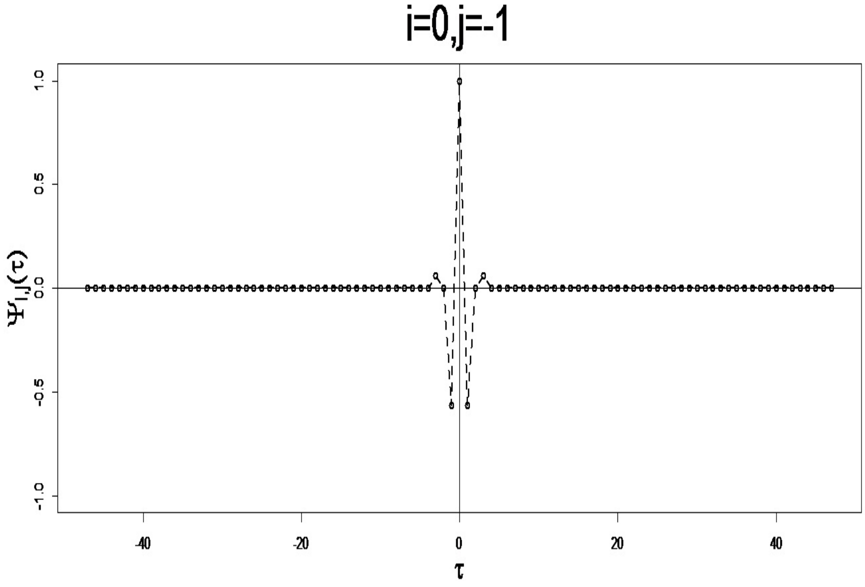
Nason et al. (2000) [94] show that for any given value of J, the elements of the collection of ACWF $\{\Psi_j(\tau)\}_{j=-J}^{-1}$ are linearly independent regardless of the choice of the scaling filter. Theorem 2.1 below provides a generalization of this statement to the larger class of CCWF.

Theorem 2.1. *If* $\{\psi_{j,\tau}\}_{\tau\in\mathbb{Z}}$ *with* $j=-1,\ldots,-J$ *and* $J\in\mathbb{N}^+$ *denotes a collection of non-decimated wavelet functions with scaling and wavelet filters fulfilling the assumptions in* (2.2.3) *and* (2.2.4), *then the family of cross-correlation wavelet functions*

$$\Psi_J=\left\{\Psi_{0,j},\left\{\Psi_{i,j},\Psi_{-i,j+i}\right\}_{i=-J-j}^{-1}\right\}_{j=-J}^{-1}$$

is linearly independent.

The notion of CCWF will reappear in Sections 3.4.1 and 4.1.1 when dealing with the covariance representation of univariate and multivariate $\mathrm{LSW}(q)$ processes. In Sections 3.4.2 and 4.1.2 we will use the statement of Theorem 2.1 to prove asymptotic uniqueness of the inverse representation of the local autocovariance functions of both, univariate and multivariate $\mathrm{LSW}(q)$ processes.



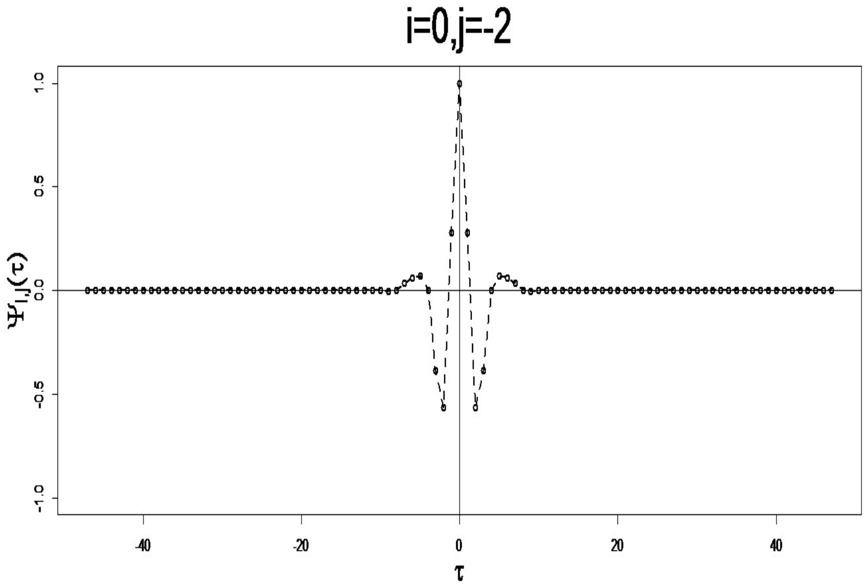


Figure 2.6: ***Set of CCWF of discrete D4 wavelet functions for*** $J = 4$

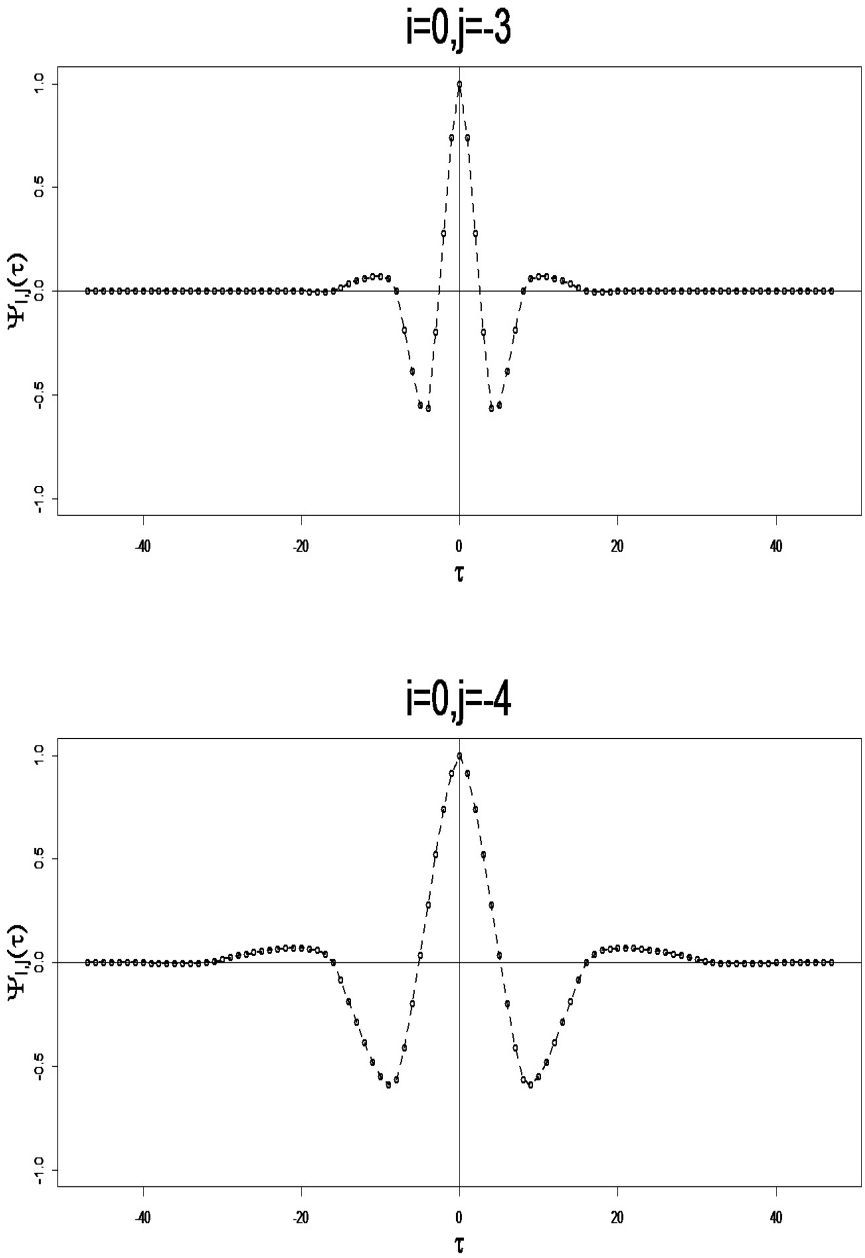


Figure 2.7: ***Set of CCWF of discrete D4 wavelet functions for*** $J = 4$

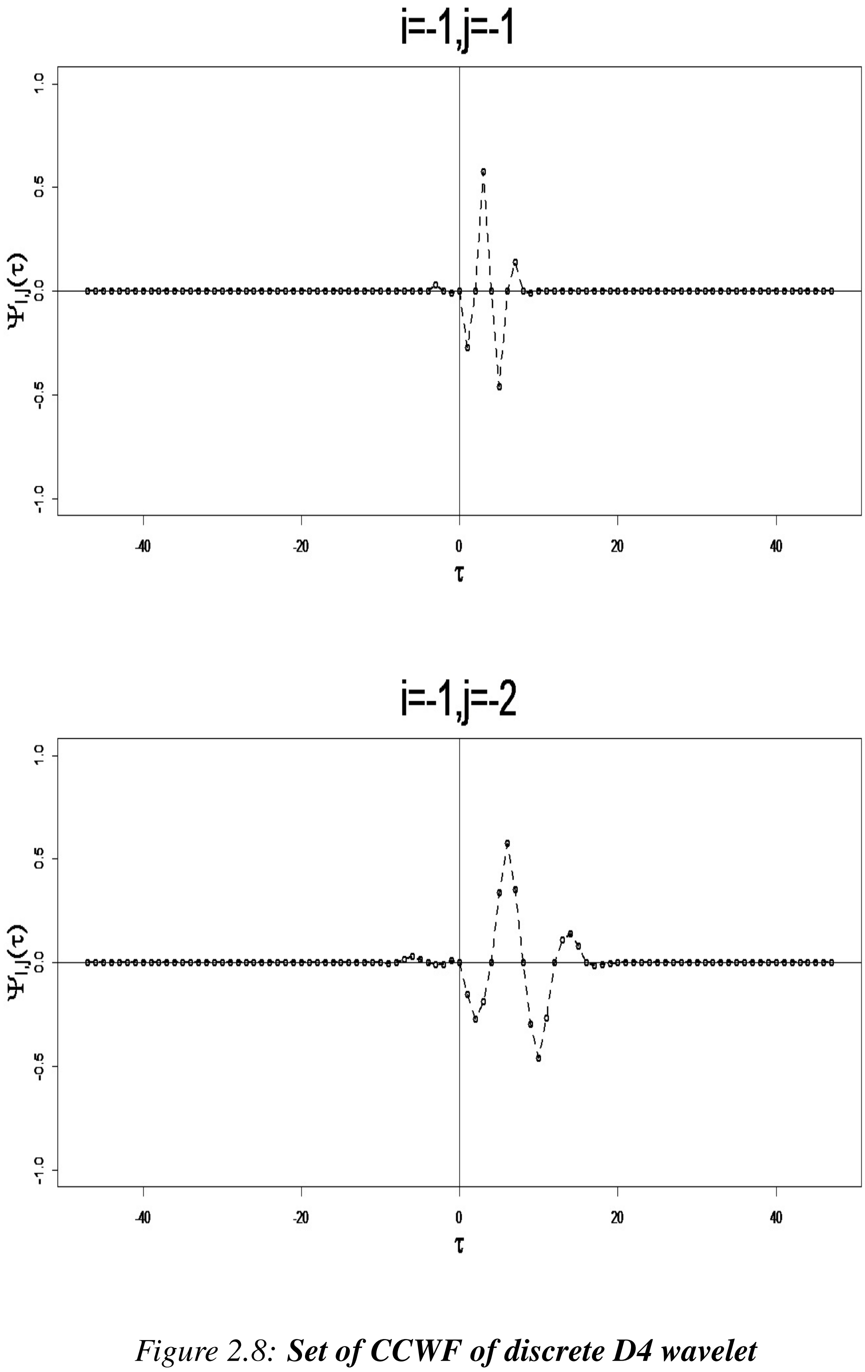


Figure 2.8: ***Set of CCWF of discrete D4 wavelet functions for*** $J = 4$

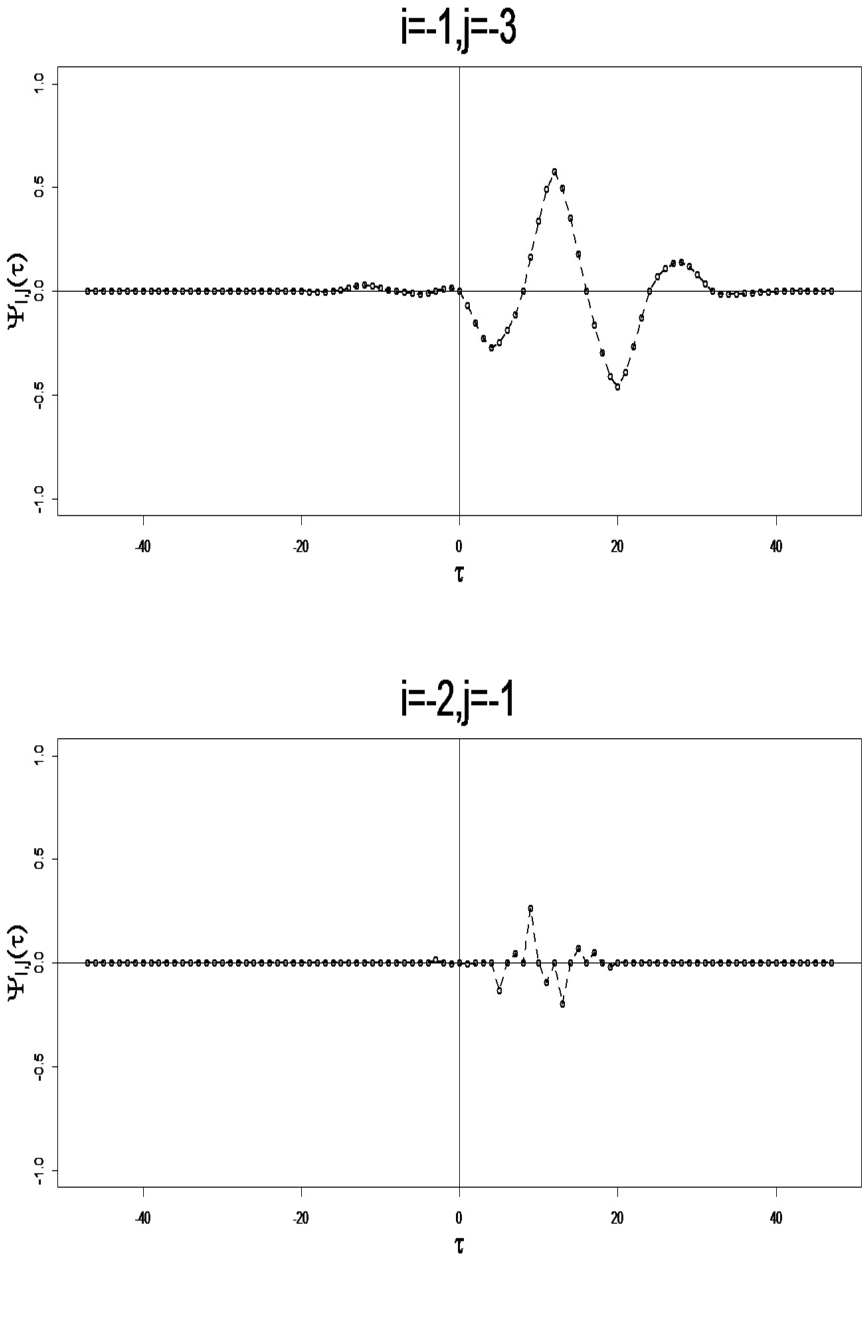
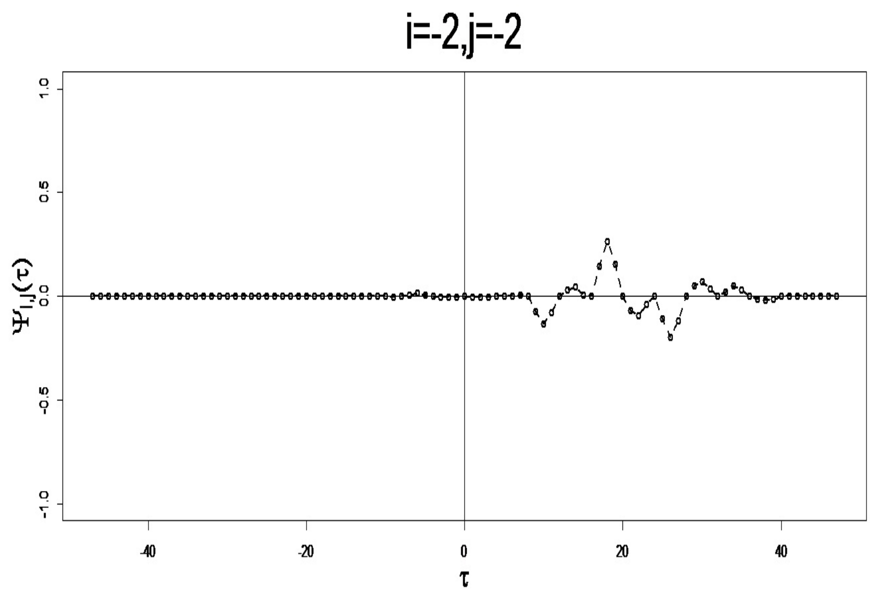


Figure 2.9: ***Set of CCWF of discrete D4 wavelet functions for*** $J = 4$



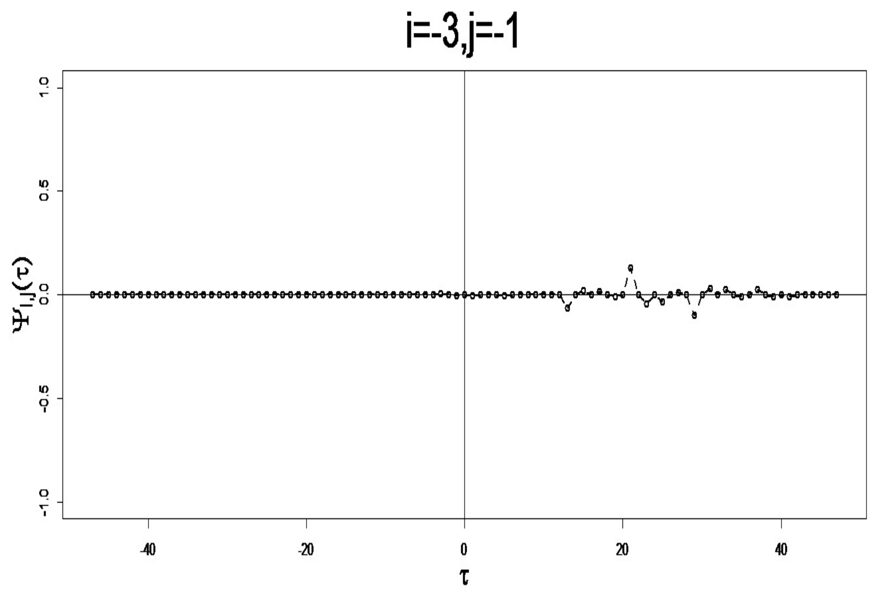


Figure 2.10: ***Set of CCWF of discrete D4 wavelet functions for** $J = 4$*

2.3 Proofs

2.3.1 Proof of Lemma 2.1

If $j=-1$ we obtain immediately $\Psi_{-1}(m)=\sum_{k=-\infty}^{+\infty}\psi_{-1,0}(k)\psi_{-1,0}(k+m)$ which can be rewritten as $\Psi_{-1}(m)=\sum_{k}g_k g_{k+m}$ using the recursive relation in (2.1.16). For $j=-2$ we have that $\Psi_{-2}(2m)=\sum_{k=-\infty}^{+\infty}\psi_{-2,0}(k)\psi_{-2,0}(k+2m)$. Using again the recursive definition in (2.1.16) we can write this as

$$\Psi_{-2}(2m)=\sum_{u}g_u\sum_{u'}g_{u'}\sum_{k=-\infty}^{+\infty}h_{k-2u}h_{k-2u'+2m}$$

Given the assumption in (2.2.3), we have that $\sum_{k=-\infty}^{+\infty}h_{k-2u}h_{k-2u'+2m}=1$ if $u'=u+m$ and 0 otherwise. This leads to $\Psi_{-2}(2m)=\sum_u g_u g_{u+m}$. Similarly, if $j=-3$ we can write

$$\Psi_{-3}(4m)=\sum_{u_2}g_{u_2}\sum_{u_1}h_{u_1-2u_2}\sum_{u_2'}g_{u_2'}\sum_{u_1'}h_{u_1'-2u_2'}\sum_{k=-\infty}^{+\infty}h_{k-2u_1}h_{k-2u_1'+4m}$$

Using that $\sum_{k=-\infty}^{+\infty}h_{k-2u_1}h_{k-2u_1'+4m}$ is equal to 1 if $u_1'=u_1+2m$ and 0 otherwise, this becomes $\Psi_{-3}(4m)=\sum_{u_2}g_{u_2}\sum_{u_2'}g_{u_2'}\sum_{u_1}h_{u_1-2u_2}h_{u_1-2u_2'+2m}$ where $\sum_{u_1}h_{u_1-2u_2}h_{u_1-2u_2'+2m}$ is equal to 1 if $u_2'=u_2+m$ and is equal to 0 otherwise and hence $\Psi_{-3}(4m)=\sum_{u_2}g_{u_2}g_{u_2+m}$.

In general we have that

$$\begin{aligned}\Psi_j\left(2^{-j-1}m\right)= & \sum_{u_{-j}}g_{u_{-j}}\sum_{u_{-j-1}}h_{u_{-j-1}-2u_{-j}}\cdots\sum_{u_2}h_{u_2-2u_3}\\ & \sum_{u'_{-j}}g_{u'_{-j}}\sum_{u'_{-j-1}}h_{u'_{-j-1}-2u'_{-j}}\cdots\sum_{u'_2}h_{u'_2-2u'_3}\sum_{u_1}h_{u_1-2u_2}h_{u_1-2u'_2+2^{-j-1}m}\end{aligned}$$

where (2.2.3) allows to write

$$\sum_{u_l=-\infty}^{+\infty}h_{u_l-2u_{l+1}}h_{u_l-2u'_{l+1}-2^{-j-l}}=\begin{cases}1 & \text{if } u'_{l+1}=u_{l+1}+2^{-j-1-l}m\\ 0 & \text{otherwise}\end{cases}$$

for $l = 1, \dots, -j$, and hence $\Psi_j\left(2^{j-1}m\right) = \sum_{u_{-j}} g_{u_{-j}} g_{u_{-j}+m}$.

2.3.2 Proof of Lemma 2.2

Set $k' = k + \tau$ which allows to write $\psi_{j,0}(k)\psi_{j+i,0}(k+\tau) = \psi_{j,0}(k' - \tau)\psi_{j+i,0}(k')$ and hence

$$\sum_{k=-\infty}^{+\infty} \psi_{j,0}(k)\psi_{j+i,0}(k+\tau) = \sum_{k'=-\infty}^{+\infty} \psi_{j,0}(k'-\tau)\psi_{j+i,0}(k')$$

for all $\tau \in \mathbb{Z}$ which means that $\Psi_{i,j}(\tau) = \Psi_{-i,j+i}(-\tau)$.

Recall that the support of $\psi_{j,0}(k)$ is given by $\left[0, \left(2^{-j}-1\right)(L[h]-1)\right]$ and that the support of $\psi_{j+i,0}(k+\tau)$ is $\left[-\tau, \left(2^{-j-i}-1\right)(L[h]-1)-\tau\right]$. Thus, if $\tau \geq \left(2^{-j-i}-1\right)(L[h]-1)+1$ or $\tau \leq -\left(2^{-j}-1\right)(L[h]-1)$ we have that

$$\left[0, \left(2^{-j}-1\right)(L[h]-1)\right] \cap \left[-\tau, \left(2^{-j-i}-1\right)(L[h]-1)-\tau\right] = \emptyset$$

and hence $\sum_{k=-\infty}^{+\infty} \psi_{j,0}(k)\psi_{j+i,0}(k+\tau) = 0$. This means that $\Psi_{i,j}(\tau) = 0$ if $\tau \leq -\left(2^{-j}-1\right)(L[h]-1)$ or $\tau > \left(2^{-j-i}-1\right)(L[h]-1)+1$.

The cross-correlation wavelets of order 0 are given by

$$\Psi_{0,j}(\tau) = \sum_{k=-\infty}^{+\infty} \psi_{j,0}(-k)\psi_{j,0}(-k+\tau)$$

and hence are equal to the ACWF.

2.3.3 Proof of Lemma 2.3

Using the recursive definition of the discrete wavelet functions in (2.2.3) we can write

$$\begin{aligned}\Psi_{i,j}(\tau) = \quad & \sum_{u'_{-j-i}} g_{u'_{-j-i}} \sum_{u'_{-j-i-1}} h_{u'_{-j-i-1}-2u'_{-j-i}} \cdots \sum_{u'_2} h_{u'_2-2u'_3} \\ & \sum_{u_{-j}} g_{u_{-j}} \sum_{u_{-j-1}} h_{u_{-j-1}-2u_{-j}} \cdots \sum_{u_2} h_{u_2-2u_3} \sum_{u_1} h_{u_1-2u'_2+\tau} h_{u_1-2u_2}\end{aligned}$$

Given that for any $m \in \mathbb{Z}$ we have that $\sum_q h_{q-2u} h_{q-2u'+2^\kappa m}$ with $\kappa \geq 1$ is equal to 1 if $u' = u + 2^{\kappa-1}m$ and equal to 0 if $u' \neq u - 2^{\kappa-1}m$ this reduces to

$$\Psi_{i,j}(\tau) = \sum_{u'_{-j-i}} g_{u'_{-j+1}} \sum_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}}$$

if $i = -1$ and to

$$\Psi_{i,j}(\tau) = \sum_{u'_{-j-i}} g_{u'_{-j-i}} \sum_{u'_{-j-i-1}} h_{u'_{-j-i-1}-2u'_{-j-i}} \cdots \sum_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}}$$

if $i < -1$.

Set $\tau = m\sum_{r=0}^{-s} 2^{-l-s-1-r}$ and hence $2^{j+1}\tau = m\sum_{r=0}^{-s} 2^{j-l-s-r} = 2^{j-l}\left(2^{-s+1}-1\right)m$. Thus, if $l < j$ then $2^{j+1}\tau$ is even for any value of $m \in \mathbb{Z}$ which means that $-2u'_{-j+1} + 2^{j+1}\tau$ is always even. Hence given the assumption in (19), we get that $\sum\limits_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}} = 0$ for every possible value of u'_{-j+1}.

2.3.4 Proof of Lemma 2.4

Using again the recursive definition of the discrete wavelet functions in (2.1.16) and that for any $m \in \mathbb{Z}$ we have that $\sum_q h_{q-2u} h_{q-2u'+2^\kappa m}$ with $\kappa \geq 1$ is equal to 1 if $u' = u + 2^{\kappa-1}m$ and equal to 0 if $u' \neq u - 2^{\kappa-1}m$, we can write

$$\Psi_{i,j}(\tau) = \sum_{u'_{-j-i}} g_{u'_{-j+1}} \sum_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}}$$

if $i = -1$ and to

$$\Psi_{i,j}(\tau) = \sum_{u'_{-j-i}} g_{u'_{-j-i}} \sum_{u'_{-j-i-1}} h_{u'_{-j-i-1}-2u'_{-j-i}} \cdots \sum_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}}$$

if $i < -1$. Set $m = L[h] - 1$ and put $\tau = m\sum_{r=0}^{i} 2^{-j-i-1-r}$. In that case $g_{u'_{-j-i}} \neq 0$ if and only if $0 \leq u'_{-j-i} \leq m$. Similarly, $h_{u'_{-j-i-1}-2u'_{-j-i}} \neq 0$ if and only if $2u'_{-j-i} \leq u'_{-j-i-1} \leq 2u'_{-j-i} + m$. In general we have that $h_{u'_{-j-i-v}-2u'_{-j-i-v+1}} \neq 0$ if and only if $2u'_{-j-i-v+1} \leq u'_{-j-i-v} \leq 2u'_{-j-i-v+1} + m$. By solving these inequalities recursively we obtain that for $v = 1, \ldots, -i$

the coefficients $h_{u'_{-j-i-v+1}}$ are equal to zero if

$$u'_{-j-i-v+1} > (2^{v}-1)\,m$$

Thus, if $u'_{-j+1} > \left(2^{-i}-1\right)m$ we have that $h_{u'_{-j+1}} = 0$. Moreover, $h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} \quad g_{u_{-j}} = 0$ if $-2u'_{-j+1} + 2^{j+1}\tau > m$. Using that $2^{j+1}\tau = m\sum_{r=0}^{i} 2^{-i-r} = \left(2^{-i+1}-1\right)m$ this becomes $u'_{-j+1} < \left(2^{-i}-1\right)m$. If we set $u'_{-j+1} = \left(2^{-i}-1\right)m$ we obtain that $\sum_{u_{-j}} h_{u_{-j}-2u'_{-j+1}+2^{j+1}\tau} g_{u_{-j}} = \sum_{u_{-j}} h_{u_{-j}+m} g_{u_{-j}}$ which reduces to $h_m g_0$.

Similarly, if $u'_{-j+2} > \left(2^{-i-1}-1\right)m$ we have that $h_{u'_{-j+1}} = 0$. Moreover, $h_{u_{-j+1}-2u'_{-j+2}} = 0$ if $u'_{-j+1} - 2u'_{-j+2} > m$. Given that $u'_{-j+1} = \left(2^{-i}-1\right)m$ this means that $u'_{-j+2} < \left(2^{-i-1}-1\right)m$. If we set $u'_{-j+2} = \left(2^{-i-1}-1\right)m$ we obtain that $\sum_{u_{-j+1}} h_{u_{-j+1}-2u'_{-j+2}} = h_m$.

This procedure can be iterated to obtain that $g_0 h_m^{-i} g_m$.

2.3.5 Proof of Theorem 2.1

Denote as $\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j} S_{i,j}^{(1)}\Psi_{i,j}(\tau)$ and $\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j} S_{i,j}^{(2)}\Psi_{i,j}(\tau)$ two linear combinations of the family of autocorrelation wavelets. To prove linear independence of Ψ we have to show that

$$\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j} S_{i,j}^{(1)}\Psi_{i,j}(\tau) = \sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j} S_{i,j}^{(2)}\Psi_{i,j}(\tau) \quad \text{for } \tau \in \mathbb{Z}$$

implies that $\Delta_{i,j} = S_{i,j}^{(1)} - S_{i,j}^{(2)} = 0$ for $j = -1,\ldots,-J\ldots$ and $i = -j-1,\ldots,-J-j$.

To start with, split Ψ into three subsets defined as $\Psi^{(0,-J)} = \left\{\Psi_{0,-J}\right\}$, $\bar{\Psi}^{(0,-J)} = \left\{\Psi_{i,j} :\ i+j=-J,\ i<0\right\}$ and $\tilde{\Psi}^{(0,-J)} = \left\{\Psi_{i,j} :\ i+j > -J\right\}$. Recall that $\Psi_{i,j}(\tau) = 0$ if $\tau \le -\left(2^{-j}-1\right)\left(L[h]-1\right)$ or $\tau > \left(2^{-j-i}-1\right)\left(L[h]-1\right)+1$. Hence, $\Psi_{i,j}\left(2^{J-1}\left(L[h]-1\right)\right) = 0$ for $\Psi_{i,j}(\tau) \in \tilde{\Psi}^{(0,-J)}$. Moreover, using Lemma 2.3 we know that $\Psi_{i,j}\left(2^{J-1}\left(L[h]-1\right)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(0,-J)}$. Lemma 2.4 then tells us that $\Psi_{0,-J}\left(2^{J-1}\left(L[h]-1\right)\right) = g_{L[h]-1}g_0$ with

$g_{L[h]-1}g_0 \neq 0$. Thus, for $\tau = \left(2^{J-1}\right)(L[h]-1)$ we get that

$$\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j}\Delta_{i,j}\Psi_{i,j}(\tau) = \Delta_{0,-J}g_0g_m$$

Thus, for $\Delta_{0,-J}\Psi_{0,-J}\left(2^{J-1}(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{0,-J} = 0$.

Next, define the sets $\Psi^{(-1,-J)} = \left\{\Psi_{-1,-J}\right\}$ and $\bar{\Psi}^{(-1,-J)} = \left\{\Psi_{i,j} : \, i+j=-J, \, i<-1\right\}$. Using Lemma 2.3 we know that $\Psi_{i,j}\left(\left(2^{J-1}+2^{J-2}\right)(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(-1,-J)}$. Thus, for $\tau = \left(\left(2^{J-1}+2^{J-2}\right)(L[h]-1)\right)$ we get that

$$\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j}\Delta_{i,j}\Psi_{i,j}(\tau) = \Delta_{-1,-J}g_0g_m$$

Thus, for $\Delta_{-1,-J}\Psi_{-1,-J}\left(\left(2^{J-1}+2^{J-2}\right)(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{0,-J} = 0$.

This procedure can be generalized by defining the sets $\Psi^{(s,-J)} = \{\Psi_{-s,-J}\}$ and $\bar{\Psi}^{(s,-J)} = \left\{\Psi_{i,j} : \, i+j=-J, \, i<s\right\}$ with $s<-1$ with $\Psi_{i,j}\left(\left(\sum_{r=0}^{-s}2^{J-1-r}\right)(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(s,-J)}$ and $\Psi_{i,j}(\tau) \in \tilde{\Psi}^{(s,-J)}$. Thus, if we solve the system of equations iteratively for $s = -2,-3,-4,\ldots$ by setting at each step $\tau = \left(\sum_{r=0}^{-s}2^{J-1-r}\right)(L[h]-1)$ we get that

$$\sum_{j=-1}^{-J}\sum_{i=-j-1}^{-J-j}\Delta_{i,j}\Psi_{i,j}(\tau) = \Delta_{s,-J}g_0g_m$$

Thus, for $\Delta_{s,-J}\Psi_{s,-J}\left(\left(\sum_{r=0}^{-s}2^{J-1-r}\right)(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{s,-J} = 0$.

Similarly, we define the sets $\Psi^{(0,-J+1)} = \left\{\Psi_{0,-J+1}\right\}$, $\bar{\Psi}^{(0,-J+1)} = \left\{\Psi_{i,j} : \, i+j=-J, \, i<0\right\}$, $\hat{\Psi}^{(0,-J+1)} = \left\{\Psi_{i,j} : \, i+j=-J, \, i>0\right\}$ and $\tilde{\Psi}^{(0,-J+1)} = \left\{\Psi_{i,j} : \, i+j>-J+1\right\}$. Given the supports of the CCWF we know that $\Psi_{i,j}\left(2^{J-2}(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \hat{\Psi}^{(0,-J+1)}$ and $\Psi_{i,j}(\tau) \in \tilde{\Psi}^{(0,-J+1)}$. Moreover, given Lemma 2.3 we have that $\Psi_{i,j}\left(2^{J-2}(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(0,-J+1)}$. Thus, for

$\tau = \left(2^{J-2}\right)(L[h]-1)$ we get that

$$\sum_{j=-1}^{-J} \sum_{i=-j-1}^{-J-j} \Delta_{i,j} \Psi_{i,j}(\tau) = \Delta_{0,-J+1} g_0 g_m$$

Thus, for $\Delta_{0,-J+1} \Psi_{0,-J+1}\left(2^{J-2}(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{0,-J+1} = 0$.

In general we define the sets $\Psi^{(s,-J+1)} = \left\{\Psi_{s,-J+1}\right\}$, $\bar{\Psi}^{(s,-J+1)} = \left\{\Psi_{i,j} : \; i+j = -J, \; i < s\right\}$, $\hat{\Psi}^{(0,-J+1)} = \left\{\Psi_{i,j} : \; i+j = -J, \; i > -s\right\}$ and $\tilde{\Psi}^{(s,-J+1)} = \left\{\Psi_{i,j} : \; i+j > -J+1\right\}$ with $\Psi_{i,j}\left(\left(\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(s,-J+1)}$, $\Psi_{i,j}(\tau) \in \hat{\Psi}^{(s,-J+1)}$ and $\Psi_{i,j}(\tau) \in \tilde{\Psi}^{(s,-J+1)}$. Thus, if we solve the system of equations iteratively for $s = -1, -2, -3, \ldots$ by setting at each step $\tau = \left(\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)$ we get that

$$\sum_{j=-1}^{-J} \sum_{i=-j-1}^{-J-j} \Delta_{i,j} \Psi_{i,j}(\tau) = \Delta_{s,-J+1} g_0 g_m$$

Thus, for $\Delta_{s,-J+1} \Psi_{s,-J+1}\left(\left(\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{s,-J+1} = 0$. Similarly, for $\Psi_{i,j}\left(\left(\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)\right) = 0$ for $\Psi_{i,j}(\tau) \in \bar{\Psi}^{(s,-J+1)}$, $\Psi_{i,j}(\tau) \in \hat{\Psi}^{(s,-J+1)}$ and $\Psi_{i,j}(\tau) \in \tilde{\Psi}^{(s,-J+1)}$. Thus, if we solve the system of equations iteratively for $s = 1, 2, 3, \ldots$ by setting at each step $\tau = \left(-\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)$ we get that

$$\sum_{j=-1}^{-J} \sum_{i=-j-1}^{-J-j} \Delta_{i,j} \Psi_{i,j}(\tau) = \Delta_{s,-J+1} g_0 g_m$$

Thus, for $\Delta_{s,-J+1} \Psi_{s,-J+1}\left(\left(\sum_{r=0}^{-s} 2^{J-2-r}\right)(L[h]-1)\right) = 0$ to hold we need to have $\Delta_{s,-J+1} = 0$.

Chapter 3

Generalized locally stationary wavelet processes

This chapter deals with the analysis of univariate discrete time series of zero-mean random variables. Many well-established statistical models are based on the assumption that the time series is second-order stationary and hence that its autocovariance function is invariant over time. The hypothesis of stationarity is an appealing notion since it allows to estimate the autocovariances and autocorrelations by computing the corresponding sample averages and hence to build a sound asymptotic theory for the estimation procedure based mainly on some version of the Central limit theorem. However, in most cases relevant for practical purposes, the assumption of stationarity represents a mere mathematical idealization which leads to an oversimplified viewpoint of the underlying second-order properties of a time series.

If we drop the stationarity assumption we need to replace the latter with a more general notion which allows to carry out meaningful statistical analysis and to develop consistent estimation procedures. Priestley (1965) [101] provides a characterization of stochastic processes with non-stationary second-order structure by introducing a form of time series that are globally non-stationary but have a locally stationary behavior relying on the theory of evolutionary spectra. Dahlhaus (1997) [27] introduces a modelling framework which allows to build an adequate asymptotic theory for the estimation of locally stationary time series, based on the concept of rescaled time. Analogously to nonparametric regression analysis, the notion of rescaled time makes use of the idea that as the number of data points increases the time-varying parameters of the locally stationary process are observed on the same interval but on a finer grid. Thus, as the sample size increases more information about the local behavior of the time-varying parameters can be collected.

A specific form of nonstationary stochastic processes are the so-called locally stationary wavelet (LSW) processes which have been defined first in Nason et al. (2000) [94]. LSW processes provide a representation of nonstationary time series with smoothly changing autocovariance functions based on discrete non-decimated wavelets and an orthogonal Gaussian increment process. Van Bellegem and von Sachs (2008) [129] expand the class of LSW processes to processes whose autocovariance function changes abruptly in time. Fryzlewicz and Nason (2006) [59] suggest the use of a Haar-Fisz method for the estimation of the evolutionary wavelet spectrum by combining the Haar wavelets and the variance stabilizing Fisz transform. Fryzlewicz et al. (2003) [60] address the issue of forecasting univariate LSW processes.

Note that in all the LSW model specifications introduced in the literature so far, each element of the univariate time series at the macroscale is represented as a linear combination of an orthonormal, multiscale increment process. The objective of this chapter is to introduce a new class of univariate LSW processes that allows for correlations among the increment terms. Nason et al. (2000) [94] show that for the case of orthonormal increment terms the covariance representation of the LSW model is asymptotically unique and derive a consistent estimation procedure of the evolutionary wavelet spectrum (EWS). We show that the asymptotic uniqueness of the autocovariance representation remains true in the presence of correlated increment processes and derive a consistent estimation procedure of the corresponding EWS.

The chapter is divided into seven sections. Section 3.1 contains a general introduction to the use of multiscale representation methods in the modelling of time series with non-stationary autocovariance function. In that context, we will have a closer look at locally stationary stochastic processes and the concept of evolutionary spectrum of a time series (Priestley, 1965 [101]). Section 3.2 discusses the notion of rescaled time (Dalhaus, 1997 [27]). Section 3.3 introduces the locally stationary wavelet process model as it is defined in Nason et al. (2000) [94]. The remaining sections of this chapter contain original work. In Section 3.4 we propose an extension of the standard LSW processes by allowing for more flexible correlation patterns in the increment process. In that context we define LSW processes of order q with q representing the degree of complexity of the correlation structure within the increment process. We show that the autocovariance representation of the extended model framework is asymptotically unique given that the model complexity q does not increase along with the sample size. Section 3.5 deals with the estimation of the generalized LSW model framework. We present a consistent estimation procedure of the evolutionary wavelet spectrum based on a generalized version of the smoothed and corrected wavelet periodogram (see Nason et al, 2000 [94]). In Sections 3.6 and 3.7 we study the performance of the proposed estimator using simulation studies and a real data analysis respectively.

3.1 Multiscale modelling of locally stationary time series

Consider a stochastic process $X = \{X(t)\}_{t=-\infty}^{+\infty}$ such that $E(X(t)) = 0$ for all $t \in \mathbb{Z}$ and denote as

$$C(\tau,t) = Cov(X(t), X(t+\tau)); \quad t \in \mathbb{Z}; \quad \tau \in \mathbb{Z}$$

its autocovariance function.

Denote as $\{\phi_\omega(t)\}_{\omega\in\mathbb{R}}$ a collection of functions whose elements are obtained through a set of transformations of a single basis function. The idea of multiscale time series analysis is to decompose the process $\{X(t)\}_{t=-\infty}^{+\infty}$ with respect to $\{\phi_\omega(t)\}_\omega$ and a zero-mean process of potentially complex-valued increments, denoted as $\{d\varepsilon(\omega)\}_\omega$, which leads to

$$X(t) = \int_{-\infty}^{+\infty} \phi_\omega(t) d\varepsilon(\omega), \quad t \in \mathbb{Z} \tag{3.1.1}$$

If the process can be represented as in (3.1.1), its autocovariance function also admits a multiscale decomposition

$$C(\tau,t) = \int_{\omega}\int_{\omega'} \phi_t(\omega)\phi_t(\omega')S(\omega,\omega')d\omega d\omega', \quad t \in \mathbb{Z}, \quad \tau \in \mathbb{Z} \tag{3.1.2}$$

where $S(\omega,\omega') = E\left(d\varepsilon(\omega)\overline{d\varepsilon(\omega')}\right)$ with $\overline{d\varepsilon(\omega')}$ being the complex conjugate of $d\varepsilon(\omega')$. Thus, the multiscale decomposition of $\{X(t)\}_{t=-\infty}^{+\infty}$ allows to represent its autocovariance structure as a quadratic form of the set of functions $\{\phi_\omega(t)\}_\omega$ with coefficients given by the components of the covariance kernel $\{S(\omega,\omega')\}_{\omega\mathbb{R},\omega'\mathbb{R}}$ of the increment process.

The multiscale representation of the covariance function allows to discover underlying patterns in the second-order properties of $\{X(t)\}_{t=-\infty}^{+\infty}$ which are hidden in the single-scale representation. Moreover, provided an appropriate choice of the set of functions $\{\phi_\omega(t)\}_{\omega\in\mathbb{R}}$, the covariance pattern of the increment process might have a higher degree of sparsity compared to the covariance structure of $\{X(t)\}_{t=-\infty}^{+\infty}$. In that case, the multiscale decomposition of $\{X(t)\}_{t=-\infty}^{+\infty}$ allows to represent its second-order properties in a compressed form which makes the analysis of the latter easier.

3.1.1 Spectral representation of second-order stationary processes

We say that $\{X(t)\}_{t=-\infty}^{+\infty}$ is covariance-stationary if its autocovariance function does not depend upon time and hence if

$$C(\tau) = Cov(X(t), X(t+\tau))$$

for $t \in \mathbb{Z}$ and $\tau \in \mathbb{Z}$. The spectral representation theorem (see Priestley, (1981a)[102] for example) states that each covariance stationary stochastic process can be written as an integral of mutually orthogonal complex exponential functions with random weights given by the elements of an orthonormal increment process. Thus we have that

$$X(t) = \int_{-\infty}^{+\infty} A(\omega) e^{i\omega t} dZ(\omega) \tag{3.1.3}$$

where $A(\omega)$ is a possibly complex-valued function, $\left\{e^{-i\omega t}\right\}_\omega$ denotes the set of complex exponential functions with varying frequencies ω and $dZ(\omega)$ denotes a zero-mean orthonormal process of complex-valued increments and hence $E(dZ(\omega)\overline{dZ(\omega')})$ is equal to 1 if $\omega = \omega'$ and equal to 0 if $\omega \neq \omega'$. Given the spectral representation of the stationary process, its autocovariance function can be decomposed as

$$C(\tau) = \int_{-\infty}^{+\infty} e^{-i\omega\tau} S(\omega) \tag{3.1.4}$$

where $S(\omega) = |A(\omega)|^2$ denotes the spectrum of the process with $|A(\omega)|$ being the modulus of $A(\omega)$.

The spectral representation is a good example to illustrate the advantages of having a multiscale viewpoint of $\{X(t)\}_{t=-\infty}^{+\infty}$. Indeed, given the orthonormality of the increment process, we can study the second-order properties of the initial process by a simple scale-wise analysis of the spectral coefficients $\{S(\omega)\}_\omega$ in the frequency domain.

3.1.2 Piecewise-stationary stochastic processes

The spectral decomposition described above is only valid for stationary processes. So the natural question at this point is if something similar to the spectral representation for stationary processes can be done in the non-stationary case?

Arguably the most basic form of second-order non-stationary processes are piecewise-stationary processes (see Priestley, (1981a) [102] for example). The most basic representation of a piecewise-stationary process $\{X(t)\}_{t=-\infty}^{+\infty}$ takes the form

$$X(t) = \begin{cases} X_1(t) & \text{if } t \leq t_0 \\ X_2(t) & \text{if } t > t_0 \end{cases}$$

where both, $\{X_1(t)\}_{t=-\infty}^{t_0}$ and $\{X_2(t)\}_{t=t_0+1}^{+\infty}$ are stationary processes but with different autocovariance functions. The process $\{X(t)\}_{t=-\infty}^{+\infty}$ is globally non-stationary since its second-order properties change within time. However, given that the two sub-processes $\{X_1(t)\}_{t=-\infty}^{t_0}$ and $\{X_2(t)\}_{t=t_0+1}^{+\infty}$ are both stationary on the intervals $[-\infty, t_0]$ and $[t_0+1, +\infty]$ respectively we say that $\{X(t)\}_{t=-\infty}^{+\infty}$ is piecewise stationary. Given that the process $\{X(t)\}_{t=-\infty}^{+\infty}$ is non-stationary, it has no spectral representation as in (3.1.3). However, if t_0 is known we can decompose both processes as $X_1(t) = \int_{-\infty}^{+\infty} A_1(\omega) e^{i\omega t} dZ(\omega)$ and $X_2(t) = \int_{-\infty}^{+\infty} A_2(\omega) e^{i\omega t} dZ(\omega)$.

The reasoning above can be extended to a more general form of piecewise stationary processes for which $X(t) = X_k(t)$ for $t_{k-1} < t \leq t_k$ and $k = 1, \ldots, N$, where the processes $\{X_k(t)\}_{t=t_{k-1}+1}^{t_k}$ are stationary processes with different autocovariance functions and with spectral representations given by $X_k(t) = \int_{-\infty}^{+\infty} A_k(\omega)e^{i\omega t}dZ(\omega)$ respectively. Thus, $\{X(t)\}_{t=-\infty}^{+\infty}$ admits a multiscale representation with

$$X(t) = \int_{-\infty}^{+\infty} A_t(\omega)e^{i\omega t}dZ(\omega). \tag{3.1.5}$$

with $A_t(\omega) = A_k(\omega)$ for $t_{k-1} < t \leq t_k$ and $k = 1,\ldots,N$.Thus, the time-varying second-order structure of piecewise stationary processes leads to a spectral representation of the latter with time-varying spectrum and Fourier representation of the covariance function given by

$$C(\tau,t) = \int_{-\infty}^{+\infty} S_t(\omega)e^{i\omega t}d\omega \tag{3.1.6}$$

with $S_t(\omega) = |A_k(\omega)|^2$ for $t_{k-1} < t \leq t_k$ and $k = 1,\ldots,N$.

3.1.3 Evolutionary spectrum

The simple representation in terms of time-varying spectrum of the specific class of piecewise-constant process in (3.1.5) has been generalized by Priestley (1965) [101] by introducing the concept of evolutionary spectrum. Priestley shows that if there exists a family of functions $\{\phi_\omega(t)\}_\omega$ defined on the real line such that the autocovariance function admits a representation of the form

$$C(t,\tau) = \int_{-\infty}^{+\infty} \phi_\omega(t)\overline{\phi_\omega(t+\tau)}S(\omega)d\omega \tag{3.1.7}$$

where $\overline{\phi_\omega(t+\tau)}$ denotes the complex conjugate of $\phi_\omega(t+\tau)$, the corresponding process can be represented as

$$X(t) = \int_{-\infty}^{+\infty} \phi_\omega(t)dZ(\omega); \quad t \in \mathbb{Z} \tag{3.1.8}$$

where $\{dZ(\omega)\}_\omega$ is an orthogonal process with $E(dZ(\omega)\overline{dZ(\omega)} = S(\omega)$. If $\{X(t)\}_t$ is in the class of oscillatory processes, the functions $\phi_\omega(t)$ can be decomposed as

$$\phi_\omega(t) = H_\omega(t)e^{i\omega t} \tag{3.1.9}$$

for all ω where $H_t(\omega) = \int_{-\infty}^{+\infty} e^{it\theta} dK_\omega(\theta)$ with $K_\omega(\theta)$ having an absolute maximum at $\theta = 1$. The evolutionary spectrum at time t is then defined as

$$S_t(\omega) = |H_t(\omega)|^2 S(\omega). \tag{3.1.10}$$

3.2 Locally stationary processes and asymptotic theory

We denote as $X_n = \{X(t)\}_{t=0}^{n-1}$ with $n > 0$ a doubly-indexed sequence of random variables. For an extensive discussion of the notion of doubly-indexed random sequences we refer to Serfling (2002) [110]. The autocovariance function of X_n is defined as

$$C_n(\tau,t) = Cov\left(X_n(t), X_n(t+\tau)\right); \quad 0 \le t \le n; \quad -t \le \tau \le n-1-t.$$

Given that in the case of second-order stationary time series, the autocovariance function is time invariant, we can gather more and more information about the autocovariance structure simply by increasing the sample size. In the non-stationary case, the autocovariance function is time-varying and hence extending the process into the future, will not give any information on the behavior of the process at the beginning of the time interval. Hence, in the non-stationary case we need a different asymptotic concept.

To set ideas straight consider an AR(1) model with time-varying lag coefficient

$$X(t) = \rho(t)X(t-1) + \xi(t) \quad \text{with } \xi(t) \sim N\left(0, \sigma^2\right)$$

for $t = 0, \ldots, n-1$. The first element that we need in order to design an appropriate asymptotic framework for non-stationary time series is some kind of regularity in the behavior of the autoregressive coefficient $\rho(t)$. A natural extension of the stationary case would be to assume that the absolute difference between the autoregressive coefficients at two different time points $\rho(t)$ and $\rho(t')$ is bounded by the absolute difference between t and t'

multiplied by a real-valued constant K such that $K > 0$

$$\left|\rho(t') - \rho(t)\right| \leq K\left|t' - t\right|. \tag{3.2.1}$$

The regularity conditions on $\rho(t)$ imply that the closer two random variables $X(t)$ and $X(t')$ are located in the time domain, the slower the value of the autoregressive coefficient can change. However, in order to be able to build a sound asymptotic theory, the upper bound in (3.2.1) needs to decrease as the sample size n increases.

In order to do this we need the concept of rescaled time introduced in Dahlhaus (1997) [27]. Rescaled time is defined as the ratio of the real time unit t over the sample size n

$$z = \frac{t}{n}. \tag{3.2.2}$$

The AR(1) model above can be rewritten in terms of rescaled time as

$$X(t) = \rho(\tfrac{t}{n})X(t-1) + \xi(t) \quad \text{with } \xi(t) \sim N\left(0, \sigma^2\right).$$

Note that according to the rescaled time notation, $\rho(z)$ is now a function on $(0,1)$ which means that as n increases, the process is observed on the same interval but on a finer grid. Hence, we obtain that

$$\left|\rho\left(\frac{t'}{n}\right) - \rho\left(\frac{t}{n}\right)\right| \leq \frac{K\left|t' - t\right|}{n}. \tag{3.2.3}$$

Thus, as n increases, we obtain more and more information about the local behavior of $\rho(t)$.

3.3 Locally stationary wavelet processes

Locally stationary wavelet (LSW) processes are a class of multiscale time series models defined first in Nason et al. (2000). Definition 3.1 specifies the class of LSW processes based on the characterizations in Nason et al. (2000) [94] and Van Bellegem and von Sachs (2008) [129].

Definition 3.1. *A sequence of doubly-indexed stochastic processes* $\{X_n(t)\}_{t=0}^{n-1}$, $n > 0$, *with mean zero elements is in the class of locally* *stationary wavelet (LSW) processes if there exists a representation*

$$X_n(t) = \sum_{j=-\infty}^{-1} \sum_{k=0}^{n-1} W_j(z)\varepsilon_{j,k}\psi_{j,k}(t). \tag{3.3.1}$$

The collection $\{\psi_{j,k}\}_{j,k}$, $j = -1, -2, \ldots$, $k = 0, \ldots, n-1$ *is a set discrete non-decimated wavelet functions. The ratio* $z = \frac{k}{n}$ *denotes rescaled time. For each* $j \leq -1$ *the evolutionary wavelet spectrum (EWS) functions, denoted as* $S_j(z) = W_j^2(z)$, *are assumed to be Lipschitz continuous on* $(0,1)$ *such that*

$$\sum_{j=-\infty}^{-1} S_j(z) < \infty \text{ uniformly in } z \in (0,1)$$

and with the Lipschitz constants L_j *satisfying* $\sum_{j=-\infty}^{-1} L_j^2 C_j < \infty$ *where* L_j *denotes the support of the wavelet functions at level* j. *The increment terms* $\varepsilon_{j,k}$ *are Gaussian random variables such that* $E\left(\varepsilon_{j,k}\right) = 0$ *and* $Cov\left(\varepsilon_{j,k}, \varepsilon_{i,l}\right) = \delta_{j,i}\delta_{k,l}$ *where* $\delta_{j,i} = 1$ *if* $j = i$ *and* 0 *otherwise.*

In the LSW model framework, each element of the doubly-indexed sequence $X_n(t)$ is defined as a linear combination of the increments $W_j(z)\,\varepsilon_{j,k}$ with the coefficients of the linear combination being given by the collection of non-decimated wavelet functions $\{\psi_{j,k}\}_{j,k}$ and hence

$$\text{Cov}\left(X_n(t), \varepsilon_{j,k}\right) = \psi_{j,k}(t) W_j(z).$$

Recall from Section 2.1.4 that the length of the support of the discrete wavelet functions at a given wavelet scale is given by $L_j = (2^{-j} - 1)(L[h] - 1) + 1$, where $L[h]$ denotes the length of the corresponding wavelet filter. Hence, for $k = 0, \ldots, n-1$ and $j = -1, -2, \ldots$ we have that $\text{Cov}\left(X_n(t), \varepsilon_{j,k}\right) = 0$ if $t \geq k + L_j$. Hence, the larger the support of the wavelet function $\psi_{j,k}(t)$, the larger the subset of random variables of $\{X_n(t)\}_{t=0}^{n-1}$ that are affected by the random fluctuations of $\varepsilon_{j,k}$. Thus, LSW processes allow for a multiscale representation of the sequence $\{X_n(t)\}_{n,t}$ with the increments at coarser and finer wavelet scales having an interpretation as long term and short term random shocks respectively.

The variance of the increments, e.g. $S_j(z) = W_j^2(z)$, depends on k which leads to the time-varying second-order structure of the time series. The Lipschitz continuity assumption on the EWS functions combined with the rescaled time notation leads to a smoothly changing autocovariance structure of $\{X_n(t)\}$. Finally, the assumption that for all value of j the Lipschitz constants behave such $\sum_{j=-1}^{+\infty} 2^{-2j} C_j < \infty$, implies that the rates of change of the EWS functions at finer scales are larger compared to the ones for the coarser scales.

Recall from Section 2.1.5 that the collection of non-decimated wavelet functions $\{\psi_{j,k}\}_{j,k}$, $j = -1, \ldots, -\infty$, $k = 0, \ldots, n-1$ is not linearly independent. Thus, the wavelet coefficients $W_j(z)$ are not uniquely defined given a fixed collection of wavelet functions. However, Nason et al. (2000) [94] and Van Bellegem and von Sachs (2008) [129] prove that, even though the multiscale time series representation of the LSW model is not unique, the multiscale representation of the autocovariance functions is asymptotically unique. For a more profound discussion of the asymptotic uniqueness of the autocovariance representation, we refer to Section 3.3.2.

Example 3.1. *We consider an LSW process for which*

$$W_j(z) = \begin{cases} \cos^2(0.5\pi z) & \text{if } j = -1 \\ \cos^2(0.25\pi z) & \text{if } j = -2 \\ \sin^2(0.5\pi z) & \text{if } j = -3 \end{cases} \tag{3.3.2}$$

The collection of discrete non-decimated wavelet functions is generated from a Haar basis. Figure 3.1 contains the functions in (3.3.2) *and one realization of the LSW process. Recall from Section 2.1.4 that the support of the discrete wavelet functions is given by* $(2^j - 1)(L[h] - 1)$ *and that for the Haar basis* $L[h] = 2$ *(see Example 2.1). Thus, we can write each random variable* $X(t) = X_{-1}(t) + X_{-2}(t) + X_{-3}(t)$ *where* $X_{-1}(t) = \sum_{k=0}^{n-1} W_{-1}(z)\varepsilon_{-1,k}\psi_{-1,k}(t)$ *represents an MA(1) process,* $X_{-2}(t) = \sum_{k=0}^{n-1} W_{-2}(z)\varepsilon_{-2,k}\psi_{-2,k}(t)$ *is an MA(3) process and* $X_{-3}(t) = \sum_{k=0}^{n-1} W_{-3}(z)\varepsilon_{-3,k}\psi_{-3,k}(t)$ *is an MA(7) process.*

3.3.1 Autocovariance representation

According to the LSW representation and using the rescaled time notation, the covariance between $X_n(t)$ and $X_n(t+\tau)$ can be written as $C_n(\tau, z) = \mathrm{Cov}[X_n(nz), X_n(nz+\tau)]$. Denote as $C(\tau, z)$ the local autocovariance function

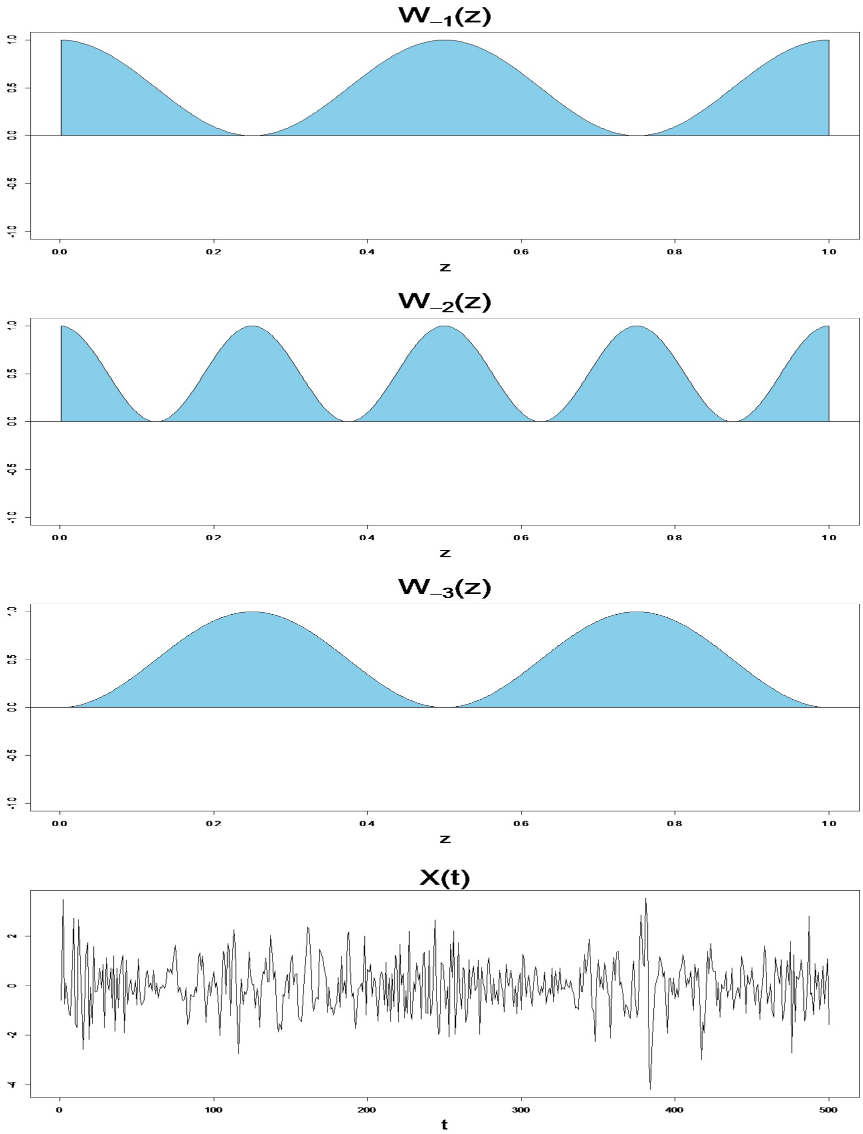


Figure 3.1: ***(Example 3.1) Wavelet coefficient functions and simulated time series***

defined as

$$C(\tau,z) = \sum_{j=-\infty}^{-1} S_j(z)\Psi_j(\tau) \qquad (3.3.3)$$

$$z \in (0,1) \text{ and } \tau \in \mathbb{Z}$$

where $\{\Psi_j\}_{j=-\infty}^{-1}$ is a family of autocorrelation wavelets functions (ACWF), which have been introduced in Section 2.2. Recall that the ACWF are symmetric about zero. Thus, the local autocovariance function, being a linear combination of symmetric functions, is also symmetric about zero. Proposition 3.1 can be found in Van Bellegem and von Sachs (2008) [129].

Proposition 3.1. *Given the assumptions in Definition 3.2 and if $n \to \infty$ we have that $C_n(\tau, z) = Cov(X_n(nz), X_n(nz+\tau))$ converges to the local autocovariance function* (3.3.3)*, given that*

$$\sum_{\tau=-\infty}^{+\infty} \int_0^1 |C_n(\tau, z) - C(\tau, z)| dz = O\left(\frac{1}{n}\right).$$

Thus, asymptotically the covariance between $X(t)$ and $X(t+\tau)$ is a linear combination of a family of autocorrelation wavelets with time-varying coefficients given by $S_j(z)$ with $j = -1, -2, \ldots$.

Note that if the time series is covariance stationary, we have that $W_j(z) = W_j$ for all $j = -1, -2, \ldots$ and hence each element of the doubly-indexed stochastic process can be written as

$$X_n(t) = \sum_{j=-\infty}^{-1} W_j \sum_{k=0}^{n-1} \varepsilon_{j,k} \psi_{j,k}(t).$$

The time-invariant wavelet coefficients lead to an EWS which is invariant in z, that is $S_j(z) = S_j$ for $j = -1, -2, \ldots$ and hence the local autocovariance function of covariance stationary time series is given by

$$C(\tau) = \sum_{j=-\infty}^{-1} S_j \Psi_j(\tau), \quad \tau \in \mathbb{Z}.$$

3.3.2 Uniqueness of the autocovariance representation

Define the operator $A := (A(j,l))_{j,l<0}$ such that

$$A(j,l) = \sum_{\tau \in \mathbb{Z}} \Psi_j(\tau) \Psi_l(\tau) \tag{3.3.4}$$

and the $J-$ dimensional matrix $A_J = (A(j,l))_{j,l=-1,\dots,-J}$. Recall from Section 2.2 that the family of ACWF $\{\Psi_j(\tau)\}_{j=-J}^{-1}$ is linearly independent (see Theorem 2.1). As a consequence, the operator norm of A_J^{-1} is bounded from above by some constant $0 < \delta_J < +\infty$ for all $J \geq 1$. Thus, we can write the evolutionary wavelet spectrum as the inverse representation of the local autocovariance function

$$S_j(z) = \sum_{l=-\infty}^{-1} A^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Psi_l(\tau) C(\tau, z) \tag{3.3.5}$$

Hence, even though the wavelet coefficients $W_j(z)$ are not uniquely defined given the corresponding LSW process, the local autocovariance function has a unique representation in terms of the evolutionary wavelet spectrum.

Denote as S_n the inverse representation of the autocovariance function $C_n(\tau, z)$

$$S_{n,j}(z) = \sum_{l=-\infty}^{-1} A^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Psi_l(\tau) C_n(\tau, z). \tag{3.3.6}$$

Proposition 3.2 can be found in Van Bellegem and von Sachs (2008) [129] and states that $C_n(\tau, z)$ converges towards (τ, z) as n increases.

Proposition 3.2. *Assume that the local autocovariance function is such that $\sum_{\tau \in \mathbb{Z}} |C(\tau, z)| < +\infty$ for each $z \in (0,1)$. If the Gramian matrix A_J is such that there exists an absolute lower bound $\delta \geq 0$ with $\lambda_{\min}(A_J) \geq \delta$ for $J \geq 1$ where $\lambda_{\min}(A_J)$ denotes the smallest eigenvalue of A_J and if $\sum_{j=-\infty}^{-1} |A^{-1}(j,l)| < +\infty$, we have that*

$$\lim_{n \to \infty} \int_0^1 \left| S_{n,j}(,z) - S_j(z) \right| = 0.$$

Hence, in order to have convergence of $C_n(\tau, z)$ towards (τ, z), there needs to exists an absolute lower bound $\delta \geq 0$ such that $\lambda_{\min}(A_J) \geq \delta$ for $J \geq 1$. Nason et al. (2000) show that $\delta > 0$ exists for both, the Haar wavelet basis and the Shannon wavelet basis. As the Shannon wavelet is the limiting wavelet in the Daubechies' compactly supported series with number of vanishing moments going to infinity, the authors conjecture that Proposition 3.2 is also valid for the rest of Daubechies' compactly supported wavelets. Frylzewicz et al. (2003) show that for Haar wavelet functions we have that $\sum_{j=-\infty}^{-1} |A_{i,r}^{-1}(j,l)| < C 2^{\frac{l}{2}}$ with $C > 0$.

3.4 Locally stationary wavelet processes of order q

There is a fundamental difference between the spectral representation of second-order stationary processes in (3.1.3) and the LSW model specification in (3.3.1). In the spectral representation, one starts from the class of second-order stationary processes. The form in (3.1.3) can then be derived using Szegö's limit theorems (see Szegö, 1915 [125]) describing the asymptotic behavior of the determinants of large Toeplitz matrices. Hence, the collection of functions, e.g. the set of complex exponentials, and the covariance pattern of the increment process is fixed by the class of covariance stationary processes defined, independently from its representation in the frequency domain. In the case of LSW processes things are different, in the sense that the classes of LSW processes are defined on the basis of their representation in the wavelet domain. Thus, in contrast to the class of covariance stationary processes, the definition of the class of LSW processes and its representation in the wavelet domain are identical, leading to a tautological relationship between the class and its representation in the wavelet domain.

In this section we are going to define new classes of LSW processes, containing the class defined in Nason et al. (2000) [94] as a subclass, using correlated increment processes. To start with, note that there are three different forms of correlation patterns that may exist within a multiscale increment process

1. **Inter-scale contemporaneous correlation:** $\mathrm{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,k}\right) = S_{i,j,k}$
2. **Intra-scale serial correlation:** $\mathrm{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j,k-l}\right) = S_{j,k-l}$
3. **Inter-scale serial correlation:** $\mathrm{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,k-l}\right) = S_{i,j,k-l}$

The advantage of having an orthogonal increment process is that it leads to a highly sparse representation of $\{X(t)\}_{t=0}^{n-1}$ in the wavelet domain. In order to maintain a high degree of sparsity, we only consider contemporaneous correlations between the increments terms belonging to different scales of resolution. The degree of inter-scale contemporaneous correlation determines the order of the multiscale increment process and hence we say that an increment process is of order q if for $j = -1, -2, \dots$ and $i = 0, \dots, -q$

$$\mathrm{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,l}\right) = \begin{cases} \Gamma_{i,j}\left(\frac{k}{n}\right) & \text{if } -q \leq i \leq 0 \text{ and } k = l \\ 0 & \text{if } i < -q \text{ or } k \neq l \end{cases}.$$

with $\Gamma_{0,j}\left(\frac{k}{n}\right) = 1$. The order of the increment process determines the order of the corresponding LSW process (see Definition 3.2).

Definition 3.2. *A sequence of doubly-indexed stochastic processes $\{X_n(t)\}_{t=0}^{n-1}$, $n > 0$, with zero-mean elements is in the class of locally stationary wavelet processes of order q if there exists a representation*

$$X_n(t) = \sum_{j=-\infty}^{-1} \sum_{k=0}^{n-1} W_j(z)\, \varepsilon_{j,k} \psi_{j,k}(t). \tag{3.4.1}$$

The collection $\{\psi_{j,k}\}_{j,k}$, $j = -1, -2, \dots$, $k = 0, \dots, n-1$ is a set of discrete non-decimated wavelet functions. The ratio $z = \frac{k}{n}$ denotes rescaled time. The increment terms $\varepsilon_{j,k}$ are Gaussian random variables such that $E(\varepsilon_{j,k}) = 0$, $Var\left(\varepsilon_{j,k}\right) = 1$, and

$$Cov\left(\varepsilon_{j,k}, \varepsilon_{j+i,l}\right) = \begin{cases} \Gamma_{i,j}\left(\frac{k}{n}\right) & \text{if } -q \leq i \leq -1 \text{ and } k = l \\ 0 & \text{if } i < -q \text{ or } k \neq l \end{cases}$$

with $j = -1, -2, \dots$, $i = -1, -2, \dots$. The EWS functions $S_{i,j}(z) = W_j(z)W_{j+i}(z)\Gamma_{i,j}(z)$ are assumed to be Lipschitz continuous on $(0,1)$ with Lipschitz constants $C_{i,j}$ satisfying

$$\sum_{i=-q}^{0} \sum_{j=-\infty}^{0} L_{j+i}^2 C_{i,j} < \infty$$

where L_{j+i} denotes the support of the wavelet functions at level $j+i$.

3.4.1 Autocovariance representation

The local autocovariance function of the LSW(q) model is given by

$$C_q(\tau, z) = \sum_{j=-\infty}^{-1} S_{0,j}(z)\, \Psi_{0,j}(\tau) + \sum_{i=-q}^{-1} \sum_{j=-\infty}^{-1} S_{i,j}(z) \left(\Psi_{i,j}(\tau) + \Psi_{-i,j+i}(\tau)\right), \tag{3.4.2}$$

$$z \in (0,1), \quad \tau \in \mathbb{Z}$$

where $\{\Psi_{0,j}\{\Psi_{i,j}, \Psi_{-i,j+i}\}_{i=-q}^{-1}\}_{j=-\infty}^{-1}$ is a family of cross-correlation wavelets which have been introduced in Section 2.2. For the sake of notation simplicity, we write the family of CCWF as

$$\Phi_q = \left\{ \Phi_{0,j}(\tau) = \Psi_{0,j}(\tau), \left\{ \Phi_{i,j}(\tau) = \Psi_{i,j}(\tau) + \Psi_{-i,j+i}(\tau) \right\}_{i=-q}^{-1} \right\}_{j=-\infty}^{-1} \tag{3.4.3}$$

Proposition 3.3 states that the autocovariance function of the LSW(q) process model converges asymptotically to the local autocovariance function given in (3.4.2) and hence generalizes the statements of Proposition 3.1 to LSW processes of order larger than zero.

Proposition 3.3. *Given the assumptions in Definition 3.2 and if $n \to \infty$ we have that $C_n(\tau, z) = Cov\,(X_n(nz), X_n(nz + \tau))$ converges to the local autocovariance function, given that*

$$\sum_{\tau=-\infty}^{+\infty} \int_0^1 |C_{n,q}(\tau, z) - C_q(\tau, z)| dz = O\left(\frac{1}{n}\right). \tag{3.4.4}$$

3.4.2 Uniqueness of the autocovariance representation

Define the operator $A_{i,r} := (A_{i,r}(j,l))_{j,l<0}$ such that

$$A_{i,r}(j,l) = \sum_{\tau \in \mathbb{Z}} \Phi_{i,j}(\tau) \Phi_{r,l}(\tau). \tag{3.4.5}$$

and define $A_q := \{A_{i,r}\}_{-q \le i,r \le 0}$ as the inner product operator of the family of functions in (3.4.3).

Denote as $A_{q,J}$ the Gramian matrix of the sub-family of functions

$$\Phi_q^{(J)} = \left\{ \left\{ \Phi_{i,j} \right\}_{i=\max(-J-j,-q)}^{0} \right\}_{j=-J}^{-1}. \tag{3.4.6}$$

such that

$$A_{q,J} = \left(\Phi_q^{(J)}\right)^t \Phi_q^{(J)} \tag{3.4.7}$$

According to Theorem 2.1, the family of CCWF Φ_q^J is linearly independent.

Thus, we have that the operator norm of $A_{q,J}^{-1}$ is bounded from above by some constant $\delta_{q,J} > 0$ for $J \geq 1$ and $q \geq 0$ which implies that A is an invertible operator. Hence, the EWS can be written as the inverse representation of the local autocovariance function that is

$$S_{q,i,j}(z) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{0} A_{i,r}^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Phi_{r,l}(\tau) C(\tau, z). \tag{3.4.8}$$

Denote as $S_{q,n}$ the inverse representation of the autocovariance function $C_{n,q}(\tau, z)$, that is

$$S_{q,n,i,j}(z) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{0} A_{(i,r)}^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Phi_{r,l}(\tau) C_n(\tau, z). \tag{3.4.9}$$

Proposition 3.4 below states that for $S_{q,n,i,j}(z)$ to converge to the EWS, there needs to exists an absolute lower bound $\delta > 0$ such that $\delta \leq \lambda_{\min}(A_{q,J})$ where $\lambda_{\min}(A_{q,J})$ denotes the smallest eigenvalue of the Gramian matrix $A_{q,J}$.

Proposition 3.4. *Consider the Gramian matrix in* (3.4.7) *and suppose that* $0 \leq q < +\infty$*. If there exists* $\delta > 0$ *such that for the smallest eigenvalue of* $A_{q,J}$ *, denoted as* $\lambda_{\min}(A_{q,J})$*, we have* $\lambda_{\min}(A_{q,J}) \geq \delta$ *for* $J \geq 1$ *and if* $\sum_{i=-q}^{0} \sum_{j=-\infty}^{-1} |A_{i,r}^{-1}(j,l)| < +\infty$*, we obtain that*

$$\int_0^1 \left| S_{q,n,i,j}(z) - S_{q,i,j}(z) \right| dz = O\left(\frac{1}{n}\right)$$

Note that Proposition 3.4 generalizes the statements of Proposition 3.2 to the class of LSW(q) processes.

3.4.3 Examples of LSW(q) processes

As seen in Section 3.4.1, the properties of the LSW(q) process model given in Definition 3.2, lead to an asymptotic representation of its autocovariance function in which for all value of z there exists a vector $S_q(z) = \{S_{i,j}(z)\}_{i,j}$ such that $C_q(z) = \{C_q(\tau, z)\}_{\tau \in \mathbb{Z}}$ can be expressed as $C_q(z) = \Phi_q S_q(z)$. Given that $A_q = \Phi_q^t \Phi_q$ is invertible we can write $S_q(z) = A_q^{-1} \Phi_q^t C(z)$. Note that if we denote as $||C_q(z) - \Phi_q S_q(z)||_2$ the L_2-norm of the vector

$C_q(z) - \Phi_q S_q(z)$, the order of a given LSW process can be obtained by solving

$$q^* = \underset{q}{\operatorname{argmin}} \; \min_{S_q(z)} \; ||C_{q^*}(z) - \Phi_q S_q(z)||_2 \tag{3.4.10}$$

with the set of functions Φ_q being fixed by the a priori choice of the collection of non-decimated wavelet functions. Consider $q \geq 0$ and $q' \geq 0$ such that $q > q'$. Given that the collection of functions $\Phi_{q'}$ is a subset of Φ_q we have that $||C_{q^*}(z) - \Phi_q S_q(z)||_2 \leq ||C_{q^*}(z) - \Phi_{q'} S_{q'}(z)||_2$. Note that due to the linear independence of the elements of Φ_q, we have that if $||C_{q^*}(z) - \Phi_q S_q(z)||_2 = ||C_{q^*}(z) - \Phi_{q'} S_{q'}(z)||_2$ the vectors $S_q(z)$ and $S_{q'}(z)$ are such that $S_q(z) = \{S_{q'}(z), 0, \ldots, 0\}$.

Note that for each value of q we have that $||C_{q^*}(z) - \Phi_q S_q(z)||_2 \geq ||C_{q^*}(z) - \Phi_q A_q^{-1} \Phi_q^t C(z)||_2$ which means that we can write the optimization setting in (3.4.10) as

$$q^* = \underset{q}{\operatorname{argmin}} \; ||C_{q^*}(z) - P_q C_{q^*}(z)||_2 \tag{3.4.11}$$

with $P_q = \Phi_q A_q^{-1} \Phi_q^t$.

Write $||C_{q^*}(z) - P_q C_{q^*}(z)||_2^2 = C_{q^*}^t(z) \left(I - P_q\right) C(z)$ and consider

$$R_q(z) = 1 - \frac{C_{q^*}^t(z) \left(I - P_q\right) C_{q^*}(z)}{C_{q^*}^t(z) C_{q^*}(z)}. \tag{3.4.12}$$

Using that $C_{q^*}^t(z) \left(I - P_q\right) C(z)$ is bounded by $C_{q^*}^t(z) C(z) ||I - P_q||_2$ and that $I - P_q$ is an idempotent matrix and hence that all its eigenvalues are either equal to 0 and 1, we have that $C_{q^*}^t(z) \left(I - P_q\right) C(z) \leq C_{q^*}^t(z) C_{q^*}(z)$ and hence that

$$0 \leq R_q(z) \leq 1$$

Consider $q \geq 0$ and $q' \geq 0$ such that $q > q'$. Given that the collection of functions $\Phi_{q'}$ is a subset of Φ_q we have that $R_q(z) < R_{q'}(z)$.

We are now going to study several classes of LSW processes in more details. For each one of them we will analyze the spectral behavior of the Gram matrix of the respective collection of second-order wavelets. At each time we will provide a specific example of the respective class of LSW processes and compute $R_q(z)$ for LSW processes of lower order.

3.4.4 LSW(∞) processes

The covariance pattern of LSW(∞) processes is given by

$$\mathrm{Cov}\left(\varepsilon_{j,k},\varepsilon_{j+i,l}\right)=\begin{cases}\Gamma_{i,j}(z) & \text{if } i\leq 0 \text{ and } k=l\\ 0 & \text{if } k\neq l\end{cases}$$

with $\Gamma_{0,j}(z)=1$ for $j=-1,-2,\ldots$, $i=-1,-2,\ldots$.

Theorem 3.1. *Consider the Gramian matrix in* (3.4.7). *If we let J depend on n as $J(n)=\log_2(n)$ and set $q(n)=\log_2(n)-1$, we have that for Haar wavelet functions $\lambda_{\min}(A_{q(n),J(n)})=O\left(\frac{1}{n}\right)$.*

According to Theorem 3.1 the smallest eigenvalue of $A_{q,J}$ does not a have an absolute lower bound and hence $S_{q,n}$ does not converge to the evolutionary wavelet spectrum as $n\to+\infty$. Thus, the autocovariance representation of LSW(∞) processes is not asymptotically unique.

If we look at the proof of Theorem 3.1 we see that the smallest eigenvalue of $A_{q,J}$ converges towards zero at rate $\frac{1}{n}$ because the norm of the function in $\Psi_{-log_2(n)-l,l}(\tau)+\Psi_{log_2(n)+l,-log_2(n)}(\tau)$ tends to zero as n increases for l fixed and $\log_2(n)$ growing along with n. From this we can conclude that, for the LSW processes to have a unique second-order representation, the order of the increment process cannot increase with the length of the time series n. In the next sections we will study the spectral behavior of the Gramian matrix of the collection of functions in (3.4.6) of LSW processes for which the order q is fixed in advance.

3.4.5 LSW(1) processes

The covariance pattern of the increment process of the LSW(1) model is given by

$$\mathrm{Cov}\left(\varepsilon_{j,k},\varepsilon_{j+i,l}\right)=\begin{cases}\Gamma_{i,j}(z) & \text{if } -1\leq i\leq 0 \text{ and } k=l\\ 0 & \text{if } i<-1 \text{ or } k\neq l\end{cases}.$$

with $\Gamma_{0,j}(z)=1$ for $j=-1,-2,\ldots$, $i=-1,-2,\ldots$.
The local autocovariance function of the LSW(1) model is given by

$$C(\tau,z)=\sum_{j=-\infty}^{-1}S_{0,j}(z)\Phi_{0,j}(\tau)+\sum_{j=-\infty}^{-1}S_{-1,-j}(z)\Phi_{-1,j}(\tau). \tag{3.4.13}$$

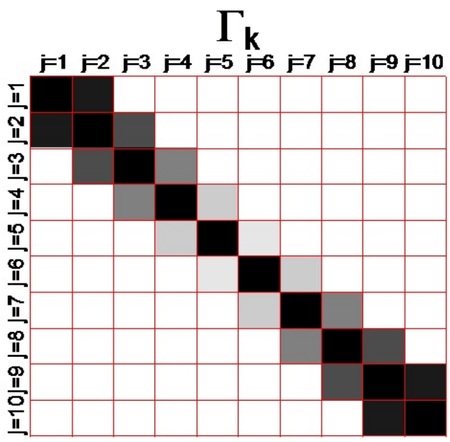


Figure 3.2: ***(LSW(1) model) Covariance pattern of the increment process***

Theorem 3.2 states that for the Haar wavelet basis, the absolute lower bound of the sequence of Gram matrices is finite and can be derived explicitly.

Theorem 3.2. *Denote as $A_{q,J}$ the Gramian matrix of the family of functions in* (3.4.6) *and consider the absolute lower bound δ such that $\delta \leq \lambda_{\min}(A_{q,J})$ for $J \geq 1$. If we set $q = 1$ and for Haar wavelet functions we have that $\delta = 0.0202$.*

Corollary 3.1. *Denote as $A_{q,J}$ the Gramian matrix of the family of functions in* (3.4.6)*. If we set $q = 1$ and for Haar wavelet functions we have that $\sum_{i=-q}^{0}\sum_{j=-\infty}^{-1}|A_{i,r}(j,l)^{-1}| \leq C2^{\frac{l}{2}}$ with $C > 0$ and $r, l < 0$.*

If we combine the statements of Proposition 3.4, Theorem 3.2 and Corollary 3.1, we can conclude that the inverse representation of the autocovariance function of the LSW(1) model converges asymptotically towards the corresponding EWS. Hence, the autocovariance representation of the LSW(1) model is asymptotically unique in the case of Haar wavelet bases.

Example 3.2. *We consider a stationary LSW(1) process with EWS given by*

$$S_{i,j}(z) = \begin{cases} 1 & \text{if } i = 0 \\ 0.9 & \text{if } -1 \leq i \leq -1 \end{cases}$$

Figure 3.3 shows the true local autocovariance function of the $LSW(1)$ process and the projection of the latter onto the space spanned by the CCWF of the $LSW(0)$ process model for the D6 wavelets basis. Tables 3.1 contains the $R(z)$ coefficients, defined in (3.4.12)*, for $LSW(0)$ processes for wavelet bases D2-D10.*

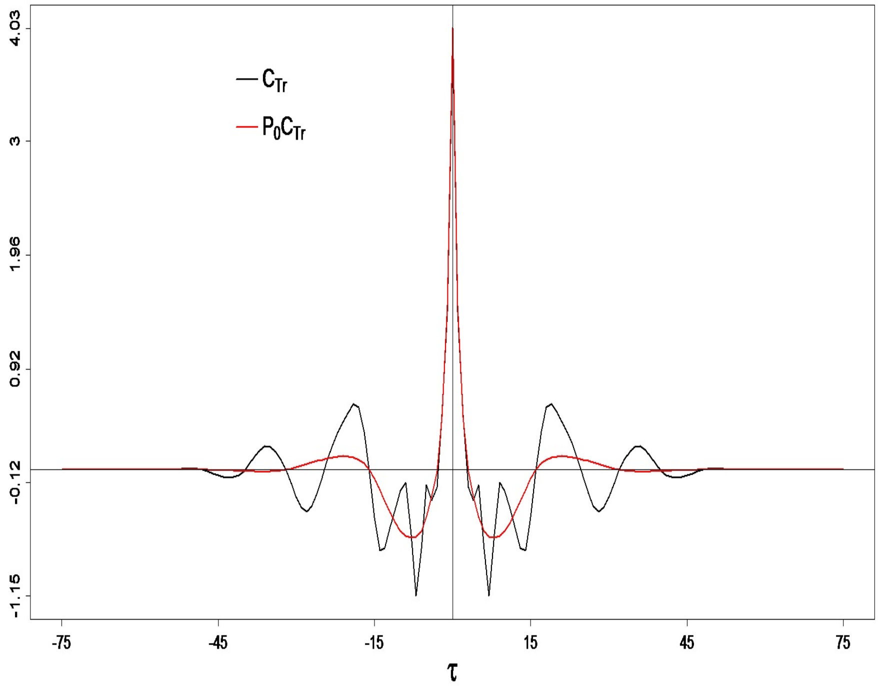


Figure 3.3: ***(Example 3.2) True and projected local autocovariance function***

Wavelet Basis	$R(z)$
D2	0.98
D4	0.85
D6	0.80
D8	0.84
D10	0.84

Table 3.1: ***(Example 3.2)*** $R_q(z)$ ***of*** $\boldsymbol{LSW}(0)$ ***processes***

3.4.6 LSW(2) processes

The covariance pattern of the increment process of the LSW(2) model is given by

$$\operatorname{Cov}\left(\varepsilon_{j,k}, \varepsilon_{j+i,l}\right) = \begin{cases} \Gamma_{i,j}(z) & \text{if } -2 \leq i \leq 0 \text{ and } k = l \\ 0 & \text{if } i < -2 \text{ or } k \neq l \end{cases}.$$

with $\Gamma_{0,j}(z) = 1$ for $j = -1, -2, \ldots$, $i = -1, -2, \ldots$.

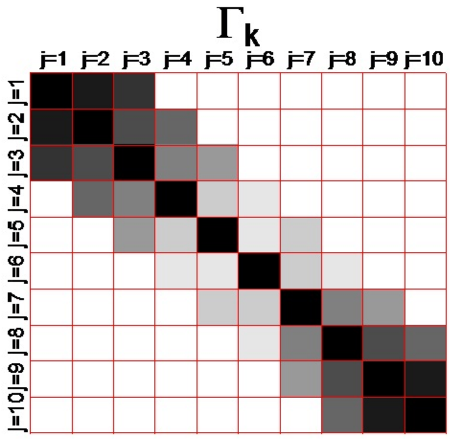


Figure 3.4: ***(LSW(2) model) Covariance pattern of the increment process***

The local autocovariance function of the LSW(2) model is given by

$$\begin{aligned} C(\tau,z) &= \sum_{j=-\infty}^{-1} S_{0,j}(z)\Phi_{0,j}(\tau) + \sum_{j=\infty}^{-1} S_{-1,-j}(z)\Phi_{-1,j}(\tau) \\ &\quad + \sum_{j=-\infty}^{-1} S_{-2,-j}(z)\Phi_{-2,j}(\tau). \end{aligned} \tag{3.4.14}$$

Theorem 3.3 states that for the Haar wavelet basis, the absolute lower bound of the Gram operator is finite and can be derived explicitly.

Theorem 3.3. *Denote as $A_{q,J}$ the Gramian matrix of the family of functions in* (3.4.6) *and consider the absolute lower bound δ such that $\delta \leq \lambda_{\min}(A_{q,J})$*

for $J \geq 1$. If we set $q = 2$ and for Haar wavelet functions we have that $\delta = 0.004$.

Corollary 3.2. *Denote as $A_{q,J}$ the Gramian matrix of the family of functions in* (3.4.6). *If we set $q = 2$ and for Haar wavelet functions we have that* $\sum_{i=-q}^{0}\sum_{j=-\infty}^{-1}|A_{i,r}^{-1}(j,l)| \leq C2^{\frac{l}{2}}$ *with $C > 0$ and $r, l < 0$.*

If we combine the statements of Proposition 3.4, Theorem 3.3 and Corollary 3.2, we can conclude that the inverse representation of the autocovariance function of the LSW(2) model converges asymptotically towards the corresponding EWS. Hence, the autocovariance representation of the LSW(2) model is asymptotically unique in the case of Haar wavelet bases.

Example 3.3. *We consider a stationary LSW(2) process with EWS given by*

$$S_{i,j}(z) = \begin{cases} 1 & \text{if } i = 0 \\ 0.9 & \text{if } -2 \leq i \leq -1 \end{cases}$$

Figure 3.4 shows the true local autocovariance function of the LSW process and the projection of the latter onto the space spanned by the CCWF of the LSW(0) and LSW(1) process models for the D6 wavelets basis. Tables 3.2 and 3.3 contain the $R(z)$ coefficients, defined in (3.4.12)*, for LSW(0) and LSW(1) processes for wavelet bases D2-D10.*

Wavelet Basis	$R(z)$
D2	0.97
D4	0.85
D6	0.83
D8	0.83
D10	0.82

Table 3.2: ***(Example 3.3)*** *$R_q(z)$* ***of LSW(0) processes***

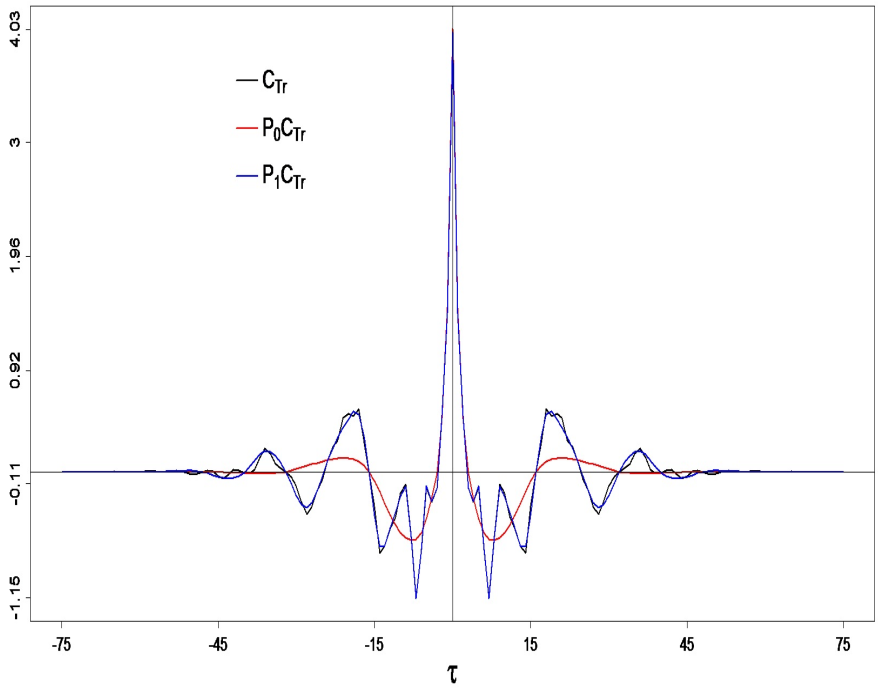


Figure 3.5: ***(Example 3.3) True and projected local autocovariance function***

Wavelet Basis	$R(z)$
D2	0.99
D4	0.98
D6	0.99
D8	0.99
D10	0.99

Table 3.3: ***(Example 3.3)*** $R_q(z)$ ***of*** $LSW(1)$ ***processes***

3.4.7 Discussion

From a conceptual point of view, the scheme of proofs of Theorems 3.2 and 3.3 could be used to prove the existence of an absolute lower bound for the smallest eigenvalue of A_q in cases where $q > 2$. However, the number of blocks within the matrices T and $\tilde{T}$ is equal to the square of the order of the corresponding LSW process plus one. Computing the analytic form of the components within each one of these blocks might is out of reach for large values of q. However, we can obtain the absolute lower bounds numerically. The table below contains the lower bounds for the smallest eigenvalues of A_q for the Haar wavelet basis with q=0,...,5.

q	0	1	2	3	4	5
δ	0.6926	0.1238	0.0248	0.0053	0.0012	0.0003

Table 3.4: ***Smallest eigenvalues of A_q for LSW models of different orders***

In general, we have that if the order of the LSW processes is fixed as $q < +\infty$, then the series of matrices $\left\{A_{q,J}\right\}_{J>0}$ is such that the corresponding series of lowest eigenvalues $\left\{\lambda_{\min}(A_{q,J})\right\}_{J>0}$ fulfills $\lambda_{\min}(A_{q,J}) \geq \delta$ with $\delta > 0$, leading to a unique autocovariance representation of the corresponding LSW process.

In the next section we will see that the existence of an absolute lower bound for the eigenspectrum of $A_{q,J}$ is also a necessary condition for the consistency of the periodogram-based estimator of the EWS of the LSW(q) processes.

3.5 Estimation of the LSW(q) process model

Consider the representation in (3.4.1). If the elements of the collection of discrete wavelet functions $\{\psi_{j,k}(t)\}_{j,k}$ were orthonormal to each other we could express the increment terms as a projection of the time series $\{X_n(t)\}_{t=0}^{n-1}$ on the set of wavelet functions, that is

$$W_j\left(\frac{k}{n}\right)\varepsilon_{j,k} = \sum_{t=0}^{n-1} \psi_{j,k}(t)X_n(t).$$

Given this unique representation of the random process in the wavelet domain, we can estimate the covariance pattern of the increment terms using the elements of the random sequence $W_j\left(\frac{k}{n}\right)\varepsilon_{j,k}$.

As it has been mentioned in Section 2.1.5, the elements of a collection of non-decimated wavelet functions are not linearly independent and which means that in the case of LSW processes the time series does not have a unique representation in terms of a sequence of random variables being projections of the time series on $\{\psi_{j,k}(t)\}_{j,k}$. Nevertheless, as for the case of orthonormal wavelet functions we start of by defining the random variables

$$d_{j,k,n} = \sum_{t=0}^{n-1} \psi_{j,k}(t)X(t) \tag{3.5.1}$$

for $j = -1,\dots,-J$ and $k = 0,\dots,n-1$, which we refer to as the set of empirical wavelet coefficients.

3.5.1 Raw wavelet periodogram

The raw wavelet periodogram of the LSW(q) model consists in a set of random variables, denoted as $I_{i,j,k,n}$ with $j = -1,\dots,-J$, $i = -1,\dots,\max(-J-j,-q)$ and $k = 0,\dots,n-1$, such that

$$I_{i,j,k,n} = \begin{cases} d_{j,k,n}^2 & \text{if } i = 0 \\ 2d_{j,k,n}d_{j+i,k,n} & \text{if } \max(-J-j,-q) \leq i \leq -1 \end{cases}. \tag{3.5.2}$$

We refer to the random sequence $\{I_{i,j,k,n}\}_{k=0}^{n-1}$ as a layer of the raw wavelet periodogram with $\{I_{0,j,k,n}\}_{k=0}^{n-1}$ being the intra-wavelet scale layers and $\{I_{i,j,k,n}\}_{k=0}^{n-1}$, $i < 0$, being the inter-wavelet scale layers.

Theorem 3.4 states that the raw wavelet periodogram is an asymptotically biased estimator of the EWS of the LSW(q) process model.

Theorem 3.4. *Given the assumptions in Definition 3.2, the expectation of the raw wavelet periodogram is*

$$E\left(I_{i,j,k,n}\right) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{0} S_{r,l}(z)A_{i,r}(j,l) + O\left(\frac{1}{n}\right) \tag{3.5.3}$$

for $j = -1,\dots,-J$, $i = -1,\dots,\max(-J-j,-q)$ *and* $k = 0,\dots,n-1$ *with* $z = \frac{k}{n}$, $J > 0$ *and* $q \geq 0$.

Thus, given the statement of Theorem 3.4, the expected value of the raw wavelet periodogram can be written asymptotically as a linear combination of the EWS functions with weights given by the elements of the inner product operator of the set of CCWF.

Theorem 3.5. *Given the assumptions in Definition 3.2, the variance of the raw wavelet periodogram is*

$$\begin{aligned} Var\left(I_{i,j,k}\right) &= \tfrac{1}{4}\left(\sum_r\sum_l S_{r,l}(z)A_{i,r}(j,l)\right)^2 \\ &+\sum_r\sum_l S_{r,l}(z)A_{0,r}(j,l)\sum_r\sum_l S_{r,l}(z)A_{0,r}(j+i,l)+O\left(\tfrac{1}{n}\right) \end{aligned}$$

for $j=-1,\ldots,-J$, $i=-1,\ldots,\max(-J-j,-q)$ *and* $k=0,\ldots,n-1$ *with* $z=\frac{k}{n}$, $J>0$ *and* $q\geq 0$.

Note that Theorem 3.4 states that the variance of the raw wavelet periodogram does not vanish as n increases.

3.5.2 Smoothed wavelet periodogram

An intuitive approach in order to get an estimator of the EWS with vanishing variance, would be to construct a smoothed version of the raw wavelet periodogram. Hence, we consider the smoothed wavelet periodogram, whose elements are defined as moving averages of the elements of the raw wavelet periodogram within each layer leading to the sequences of random variables

$$I_{i,j,k,n,M}=\frac{1}{2M+1}\sum_{s=-M}^{M}I_{i,j,k+s,n} \tag{3.5.4}$$

for $j=-1,\ldots,-J$, $i=-1,\ldots,\max(-J-j,-q)$ and $k=0,\ldots,n-1$ with $z=\frac{k}{n}$, $J>0$ and $0\leq q<+\infty$.

Theorem 3.6. *Given the assumptions in Definition 3.2 and if* $|S_{i,j}(z)|\leq\frac{K}{L_{j+i}}$ *for any* $i=0,\ldots,-q$ *with* $q<+\infty$ *and* $j=-1,-2,\ldots$*, the variance of the smoothed wavelet periodogram is*

$$Var\left(I_{i,j,k,n,M}\right)=O\left(\frac{2^{-j}j^2}{M}\right)$$

for $j=-1,\ldots,-J$, $i=-1,\ldots,\max(-J-j,-q)$ *and* $k=0,\ldots,n-1$ *with* $z=\frac{k}{n}$, $J>0$ *and* $0\leq q<+\infty$.

Note from the proof of Theorem 3.6 that, a necessary condition for the smoothed wavelet periodogram to have vanishing variance with increasing M, is that the elements of $\left\{I_{i,j,k+s,n}\right\}_{s=-M}^{M}$ are such that $\sum_{\kappa}|\mathrm{Cov}\left(I_{i,j,k+s,n}, I_{i,j,k+s+\kappa,n}\right)| < +\infty$. This condition of not verified however, given that $\sum_{\kappa}|\mathrm{Cov}\left(I_{i,j,k+s,n}, I_{i,j,k+s+\kappa,n}\right)| = O\left(2^{-j}\right)$ reflecting into the layer dependent variance of the smoothed wavelet periodogram. The latter has some serious consequence for the asymptotic al behavior of the variance of the smoothed wavelet periodogram at the coarser scales. Take for example the estimator at layer $j = -J$. If we let J grow along with n such that $J(n) = \log_2(n)$ we have that $\mathrm{Var}\left(I_{0,J,k,n,M}\right) = O\left(\frac{log_2(n)^2 n}{M}\right)$ and hence the variance at the coarsest scale of resolution does not vanish with increasing values of the smoothing parameter. Hence, we can conclude from this that the smoothing of the raw wavelet periodogram does not lead to an estimator with vanishing variance at all the layers of resolution.

Theorem 3.7 below states that the asymptotic bias of the raw and the smoothed wavelet periodogram are identical.

Theorem 3.7. *Given the assumptions in Definition 3.2, the expected value of the smoothed wavelet periodogram is*

$$E\left(I_{i,j,k,n,M}\right) = E\left(I_{i,j,k,n}\right) + O\left(\frac{M}{n}\right)$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$ *and* $k = 0, \ldots, n-1$ *with* $z = \frac{k}{n}$, $J > 0$ *and* $0 \leq q < +\infty$.

3.5.3 Smoothed and corrected wavelet periodogram

The smoothed and corrected wavelet periodogram is defined as

$$\tilde{I}_{i,j,k,n,M} = \sum_{r=-q}^{0} \sum_{l=-J}^{-1} A_{i,r}^{-1}(j,l) I_{r,l,k,n,M}.$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$ and $k = 0, \ldots, n-1$ with $z = \frac{k}{n}$, $J > 0$ and $0 \leq q < +\infty$ where A_q is the inner product operator of the collection of CCWF (see Section 3.4.2).

Theorem 3.8. *Given the assumptions in Definition 3.2 and if* $|S_{i,j}(z)| \leq \frac{K}{L_{j+i}}$ *for any* $i = 0, \ldots, -q$ *with* $q < +\infty$ *and* $j = -1, -2, \ldots,$ *and* *provided that there exists $\delta > 0$ such that $\lambda_{\min}(A_{q,J}) \geq \delta$ and that $\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}\left|A_{i,r}^{-1}(j,l)\right| \leq C2^{\frac{j}{2}}$, we have that*

$$E\left(\tilde{I}_{i,j,k,n,M}\right) = S_{i,j}(z) + O\left(\frac{M\log_2(n)}{n}\right).$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$ and $k = 0, \ldots, n-1$ with $z = \frac{k}{n}$, $J > 0$ and $0 \leq q < +\infty$.

Theorem 3.9. *Given the assumptions in Definition 3.2 and if $|S_{i,j}(z)| \leq \frac{K}{L_{j+i}}$ for any $i = 0, \ldots, -q$ with $q < +\infty$ and $j = -1, -2, \ldots$, and provided that there exists $\delta > 0$ such that $\lambda_{\min}(A_{q,J}) \geq \delta$ and that $\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}\left|A_{i,r}^{-1}(j,l)\right| \leq C2^{\frac{j}{2}}$, we have that*

$$Var\left(\tilde{I}_{i,j,k,n,M}\right) = O\left(\frac{2^j\log_2(n)}{M}\right).$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$ and $k = 0, \ldots, n-1$ with $z = \frac{k}{n}$, $J > 0$ and $0 \leq q < +\infty$.

If we set $M = O\left(n^{\beta}\right)$ and given the statements of Theorems 3.8 and 3.9, the mean squared error of the smoothed and corrected wavelet periodogram at each layer is given by

$$\text{MSE}\left(\tilde{I}_{i,j,k,n,M}\right) = O\left(\log_2^2(n)\left(\frac{1}{n^{2(1-\beta)}} + \frac{2^j}{n^{\beta}\log_2(n)}\right)\right). \tag{3.5.5}$$

Hence, the optimal convergence rate of the MSE can be obtained by solving $\max\left(\min\left(2(1-\beta), \beta\right)\right)$ with respect to β leading to $\beta = \frac{2}{3}$ in which case

$$\text{MSE}\left(\tilde{I}_{i,j,k,n,M}\right) = O\left(\frac{\log_2^2(n)}{n^{\frac{2}{3}}}\right)$$

Using Markov's inequality, we have that for all $\varepsilon > 0$

$$P\left(|\tilde{I}_{i,j,k,n,M} - S_{i,j}(z)| > \varepsilon\right) \leq \frac{\text{MSE}\left(\tilde{I}_{i,j,k,n,M}\right)}{\varepsilon}$$

which, given the result in (3.5.5), proves convergence in probability of the smoothed and corrected wavelet periodogram.

3.6 Simulation Studies

In this section we use Monte Carlo simulations to compare the performances of LSW models of different orders in small sample settings.

To start with, we denote as $\tilde{C}_{n,q}(z) = \Phi_{J,q}\tilde{I}_{M,t,n,q}$ the estimator of the local autocovariance function $C_q(z)$ of the LSW(q) process with $\tilde{I}_{M,t,n,q}$ being the vector of smoothed and corrected periodograms, that is $\tilde{I}_{M,t,n,q} = \{\tilde{I}_{i,j,t,n,M}\}$ for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J - j, -q)$. Define as $\theta(C_q(z), \tilde{C}_{n,q}(z))$ the angle between the two vectors with cosine of $\theta(C_q(z), \tilde{C}_{n,q}(z))$ being given by

$$\cos\left[\theta(C_q(z), \tilde{C}_{n,q}(z))\right] = \frac{\tilde{C}^t_{n,q}(z)C_q(z)}{\sqrt{\tilde{C}^t_{n,q}(z)\tilde{C}_{n,q}(z)}\sqrt{C^t_q(z)C_q(z)}} \tag{3.6.1}$$

Denote as F_q and $F_{q'}$ the cumulative distribution functions of the random variables $\cos\left[\theta(C(z), \tilde{C}_{n,q}(z))\right]$ and $\cos\left[\theta(C(z), \tilde{C}_{n,q'}(z))\right]$ respectively. We say that $\tilde{C}_{n,q}(z))$ has a first-order stochastic dominance over $\tilde{C}_{n,q'}(z))$ if for any value of p we have that

$$F_q^{-1}(p) \geq F_{q'}^{-1}(p)$$

Example 3.4. *We consider a stationary LSW(3) process with D8 wavelet basis and with EWS given by $S_{0,j}(z) = 1$ for $j = -1, \ldots, -3$ and $S_{i,j}(z) = 0.9$ for $j = -1, \ldots, -2$ and $i = -1, \ldots, -3 - j$.*

Figures 3.6, 3.7 and 3.8 plot the autocovariance function of the LSW(3) model in Example 3.4 and the projections of the latter on the space spanned by the CCWF of the LSW(0), LSW(1) and LSW(2) models respectively. Table 3.5 contains the cosine similarities between the autocovariance function of the LSW(3) model and the projections on the space spanned by the CCWF functions of the lower order LSW models.

q	Cosine Similarity
2	0.9999
1	0.9980
0	0.9434

Table 3.5: ***(D8-Wavelets)*** $n = 30$

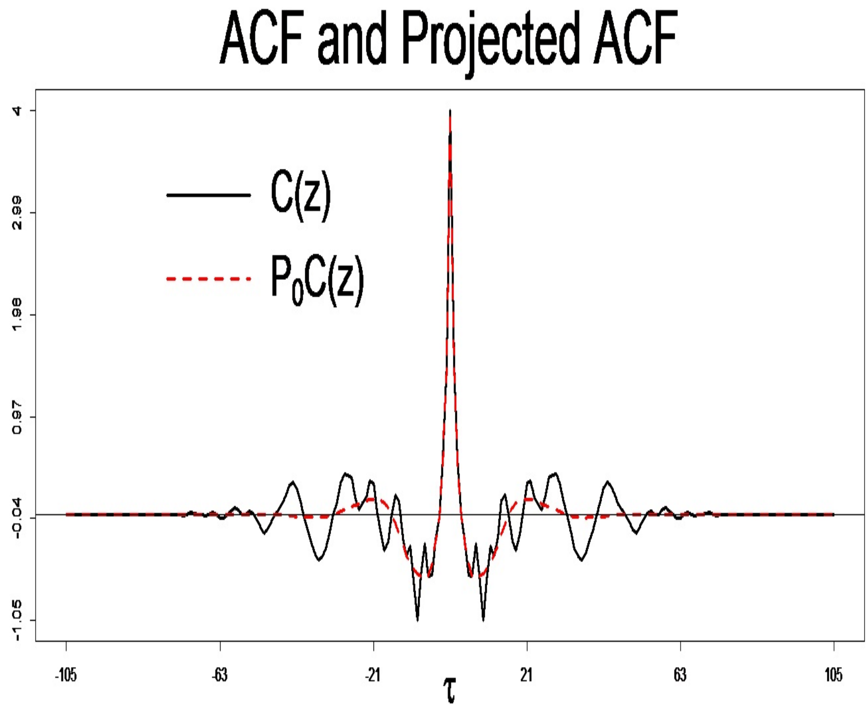


Figure 3.6: ***(Example 3.4)*** $C(z)$ ***and projection on the space spanned by the CCWF of the LSW***(0) ***model***

For the simulation studies we draw observations from the process described in Example 3.4 and compute the local autocovariance function estimators of four different LSW models with orders $q = 0, 1, 2, 3$ respectively. Thus, if we generate L scenarios we obtain L observations of the cosine similarity in (3.6.1), denoted as

$$\left\{ \cos \left[\theta(C_3(\tau, z), \tilde{C}_{n,q}^{(l)}(\tau, z)) \right] \right\}_{l=1}^{L}$$

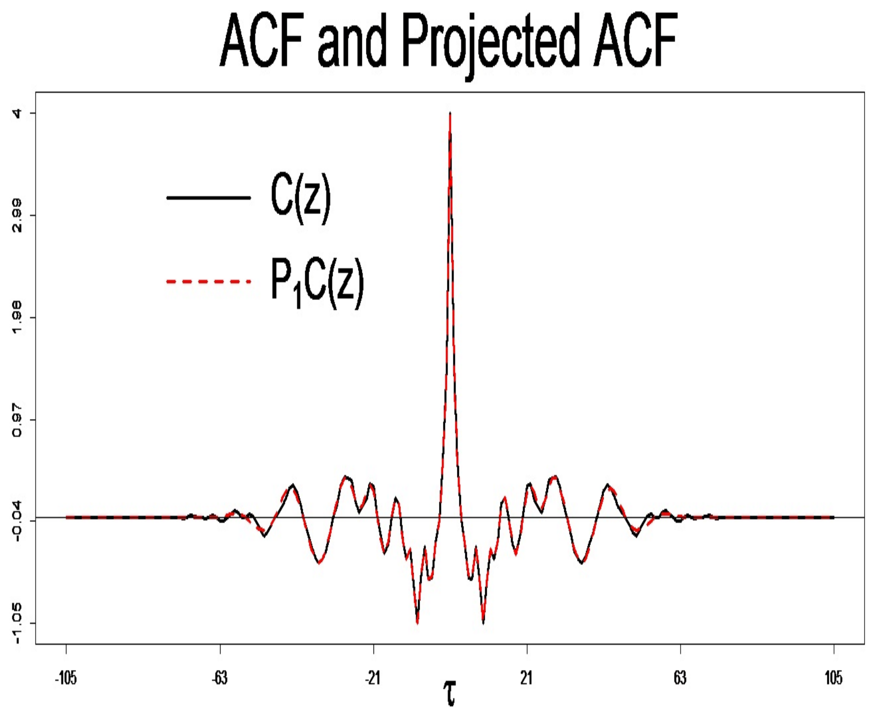


Figure 3.7: ***(Example 3.4)*** $C(z)$ ***and projection on the space spanned by the CCWF of the LSW***(1) ***model***

with $q = 0,1,2,3$. We use these observations to compute the empirical distribution functions $\hat{F}_q(x)$ of the random variables

$$\hat{F}_q(x) = \frac{\sum_{l=1}^{L} \mathbb{I}(\cos\left[\theta(C_3(\tau,z), \tilde{C}_{n,q}^{(l)}(\tau,z))\right] \leq x)}{L}$$

where $\mathbb{I}$ denotes the indicator function.

Tables 3.6, 3.7, 3.8 and 3.9 contain the quantiles $\hat{F}_q^{-1}(p)$ for $q = 0,1,2,3$ and $p = 0.1, 0.2, 0.3, 0.4, 0.5$ for different sample sizes.

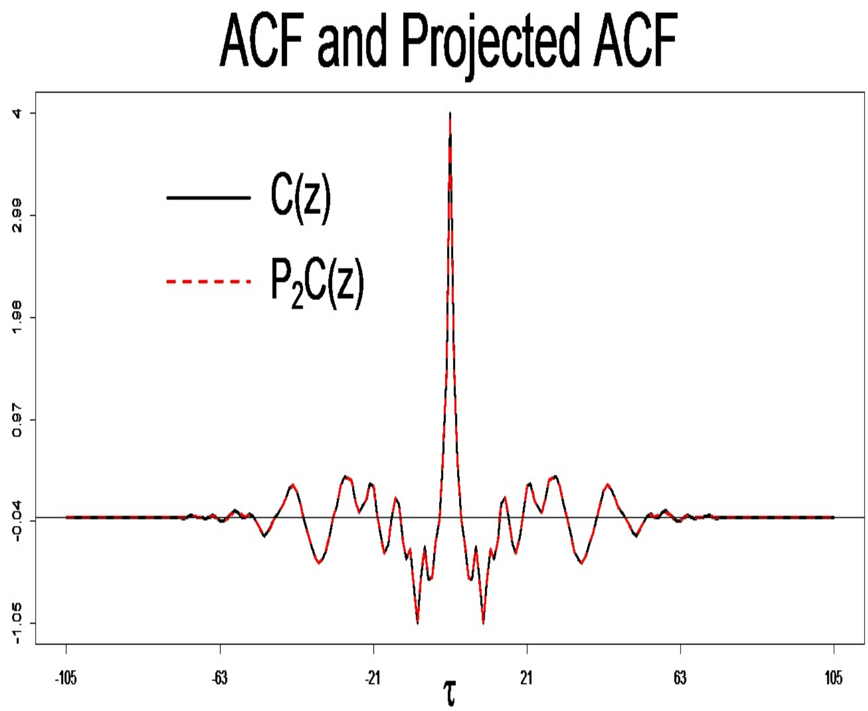


Figure 3.8: ***(Example 3.4)*** $C(z)$ ***and projection on the space spanned by the CCWF of the LSW***(2) ***model***

q	0.05	0.1	0.2	0.3	0.4	0.5
3	-	-	-	-	-	-
2	-	-	-	-	-	-
1	0.8840	0.9098	0.9358	0.9508	0.9611	0.9700
0	0.8738	0.8951	0.9220	0.9331	0.9411	0.9482

Table 3.6: ***D8-Wavelets:*** $n = 30$

q	0.05	0.1	0.2	0.3	0.4	0.5
3	-	-	-	-	-	-
2	0.8687	0.8873	0.9102	0.9268	0.9392	0.9497
1	0.8769	0.9004	0.9217	0.9375	0.9510	0.9594
0	0.8699	0.8872	0.9058	0.9180	0.9262	0.9323

Table 3.7: ***D8-Wavelets:*** $n = 60$

q	0.05	0.1	0.2	0.3	0.4	0.5
3	-	-	-	-	-	-
2	0.9043	0.9166	0.9341	0.9449	0.9545	0.9630
1	0.9145	0.9278	0.9432	0.9538	0.9631	0.9699
0	0.8941	0.9086	0.9199	0.9272	0.9332	0.9373

Table 3.8: ***D8-Wavelets:*** $n = 90$

q	0.05	0.1	0.2	0.3	0.4	0.5
3	0.8495	0.8697	0.8965	0.9107	0.9251	0.9382
2	0.8539	0.8774	0.9017	0.9161	0.9308	0.9429
1	0.8721	0.8937	0.9182	0.9312	0.9457	0.9546
0	0.8568	0.8749	0.8942	0.9061	0.9150	0.9220

Table 3.9: ***D8-Wavelets:*** $n = 120$

For each sample size, the local autocovariance function estimator of the LSW(1) model stochastically dominates the estimators provided by the three others LSW models. The LSW(2) stochastically dominates the LSW(0) model for $n = 90$ and $n = 120$. The LSW(3) does not manage to outperform the LSW(0) model. If we have a look at Figure 3.7 as well as Table 3.5, we see that the projection of the true autocovariance function on the space spanned by the CCWF of the LSW(1) model, which means that the model bias of the LSW(1) model is relatively low. Moreover, the correlation pattern of the increment process of order 1 is still characterized by a high-degree of sparsity resulting in a low variance of the periodogram-based estimator. The covariance structure of the increment processes of the LSW(2) and LSW(3) models are less sparse which translates into a larger variance of the corrected and smoothed wavelet periodogram, resulting in a larger variance of the corresponding cosine similarity relative to the LSW(1) model.

3.7 Real Data Analysis: Modelling the log-returns of European stock indexes

In this section, we use LSW(q) processes to study the second-order behavior of the time series of log-returns of the stock indexes of Germany, Ireland and Greece. The covariance patterns of returns and log-returns of financial time series play an important role in many portfolio optimization algorithms and are a major keystone in financial risk management models (see for example Markowitz (1952) [89], Fama and French (1993) [47] and Fan et al. [48]).

We denote as $P(t)$ the value of a given stock index at time point t and denote as $X(t)$ its log return at time t, such that

$$X(t) = \ln(P(t)) - \ln(P(t-1)).$$

Figures 3.9, 3.10 and 3.11 plots the German, the Irish and the Greek stock indexes and the corresponding log-returns for the period 04/01/2001-11/03/2014. Note that for all three series of stock indexes there has been an important increase between the beginning of the second quarter of 2003 up to end of the second quarter 2007, marking the beginning of the financial crisis of 2007-2008. Uncertainties regarding bank solvency, declines in credit availability and damaged investor confidence had a negative impact on economic activities in the US and most European countries. This economic downturn lead to a sharp decline of the three stock indexes until the third quarter of 2009. Governments and central banks responded with unprecedented financial stimulus programs and monetary policy expansion combined with bank bailouts which explains the recovery of the German and the Irish stock market from the end of 2009 on. However, we can observe a different dynamics in the Greek stock market which is mainly due to the Greek government-debt crisis which had been triggered at the aftermath of the financial crisis by a combination of structural weaknesses of the Greek economy along with a decade of high structural deficits and debt-to-GDP levels on public accounts.

We assume that the three stock indices are in the class of $\mathrm{LSW}(1)$ processes with non-decimated wavelet functions stemming from a D6 wavelet basis and the number of wavelet scales given by $J = 5$. Figures 3.12, 3.13 and 3.14 plot the smoothed and corrected periodogram estimators of the inter-scale EWS functions. As far as the German stock index is concerned, the most striking feature is the steady decline of the periodogram at scale $-1,-1$ from the beginning of 2005 on up to the second quarter of 2008 marking the starting point of the global financial crisis. Note moreover, that the periodograms at the remaining scales are very close to zero. As for the Irish stock index, there are two periods during which the periodogram at scale $-1,-1$ takes on strong negative values, that is around 2006 and 2011. Note that the year 2006 marks the height of the Irish housing market bubble and that 2011 is considered as having been the peak of the debt crisis which affected the Irish economy after the global financial crisis. As far as Greece is concerned, the periodogram at scale $-1,-1$, remains negative between the start of the analysis period until the end of the financial crisis to become positive between the second quarter of 2009 and the second quarter of 2013 with a peak around spring 2011. Note that this period is characterized by the first and the second bailout launched by the so-called Troika, that is the Eurozone countries, European Central Bank (ECB) and International Monetary Fund (IMF). From the second quarter of 2013 on, the periodogram at scale $-1,-1$ turns negative again.

The analysis of the wavelet periodogram of the $\mathrm{LSW}(1)$ model involves the study of $J+(J-1)$ different layers. However, it is not always obvious how to interpret the combined information provided by these layers. Hence, the aim of the last part of this section is to discuss a possible way of summing up the information contained in the different layers into one single coefficient.

To start with, consider the random vector of increment terms

$$\eta_{J,k} = \left(W_{-1}(z)\varepsilon_{-1,k},\dots,W_{-J}(z)\varepsilon_{-J,k}\right)^t$$

with $J>0$ and $k=0,\dots,n-1$. Denote as $S_{J,q}(z)=E(\eta_{J,k}\eta_{J,k}^t)$ the covariance matrix of $\eta_{J,t}$ and as $\tilde{I}_{J,M,q}(z)$ the estimator of $S_{J,q}(z)$ based on the corrected and smoothed wavelet periodogram. Denote as $\{\lambda_m\left(S_{J,q}(z)\right)\}_{m=1}^J$ the set of ordered eigenvalues of $S_{J,q}(z)$ with $\lambda_1\left(S_{J,q}(z)\right)$ being the largest eigenvalue and as $\{u_m\left(S_{J,q}(z)\right)\}_{m=1}^J$ the associated set of eigenvectors.

We write each increment $W_j(z)\varepsilon_{j,k}$ as a linear combination of independent zero-mean Gaussian random variables with unit variance, as

$$W_j(z)\varepsilon_{j,k} = \sum_{m=1}^{J}\Delta_{m,j,q}(z)Z_{m,k}. \tag{3.7.1}$$

where $\Delta_{m,j,q}(z)=\sqrt{\lambda_m\left(S_{J,q}(z)\right)}u_m\left(S_{J,q}(z)\right)$ and define the coherence measure

$$\mathscr{C}_q(z) = \sum_{m=0}^{J}\sum_{j=-J+1}^{-1}\sum_{i=-J+j}^{0}\left(\Delta_{m,j,q}(z)-\Delta_{m,j+i,q}(z)\right)^2. \tag{3.7.2}$$

Note that $\mathscr{C}_q(z)$ is equal to zero if and only if the correlation of all pairs of elements of the random vector $\eta_{J,k}$ is equal to one. In that case we say that the random vector η_k as a perfectly coherent.

Denote as $\tilde{I}_{J,M,q}(z)$ the estimator of $S_J(z)$ based on the corrected and smoothed periodogram. Moreover, denote as $\left\{u_m\left(\tilde{I}_{J,M,q}(z)\right)\right\}_{m=1}^{J}$ the set of eigenvectors associated with the set of ordered eigenvalues of $\tilde{I}_{J,M,q}(z)$, denoted as $\left\{\lambda_m\left(\tilde{I}_{J,M,q}(z)\right)\right\}_{m=1}^{J}$. We define the estimator of the form in (3.7.2) as

$$\hat{\mathscr{C}}(z) = \sum_{m=0}^{J} \lambda_m\left(\tilde{I}_{J,M,q}(z)\right) \sum_{j=-J+1}^{-1} \sum_{i=-J+j}^{0} \left(u_{m,j}\left(\tilde{I}_{J,M,q}(z)\right) - u_{m,j+i}\left(\tilde{I}_{J,M,q}(z)\right)\right)^2. \tag{3.7.3}$$

Figure 3.15 contains the estimates given in (3.7.3) for the DAX, the ISEQ and the ATHEX stock indices based on the corresponding log-return data. Note that for all three log-return series, the coherence measure is characterized by a strong increase from the beginning of the period of analysis up to the aftermath of the global financial crisis. However, from the second quarter of 2009 on, the coherence measures of the German and the Irish stock index start decreasing again, while the coherence measure of the Greek stock index continues to increase up to the second quarter of 2015.

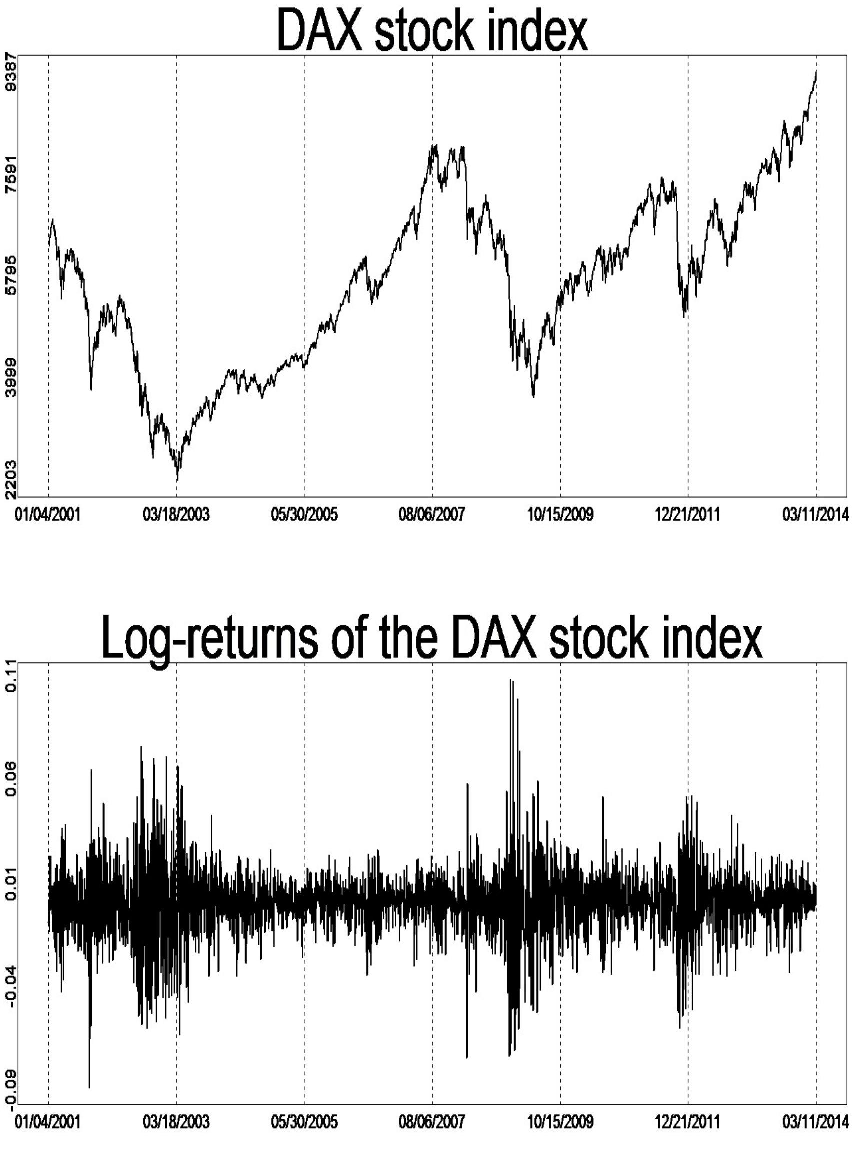


Figure 3.9: ***German stock index and its log-returns from 04/01/2001 to 11/03/2014***

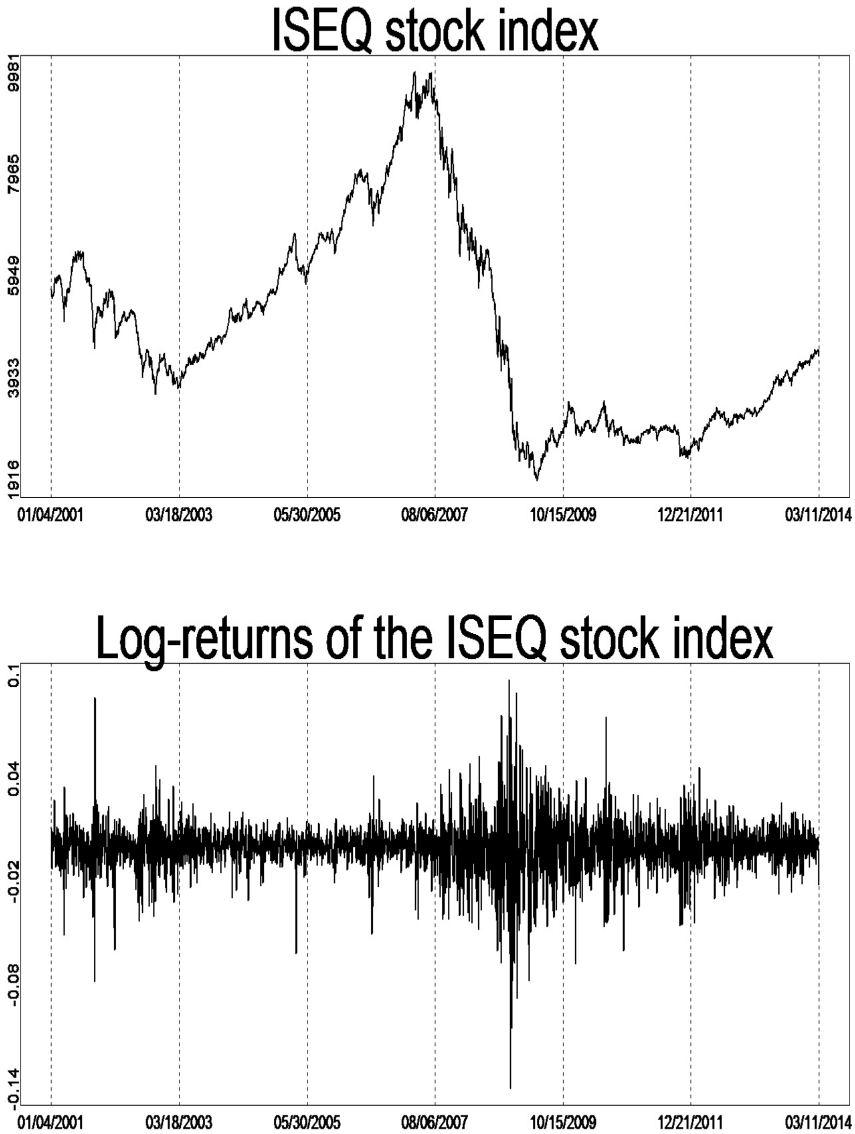


Figure 3.10: ***Irish stock index and its log-returns from 04/01/2001 to 11/03/2014***

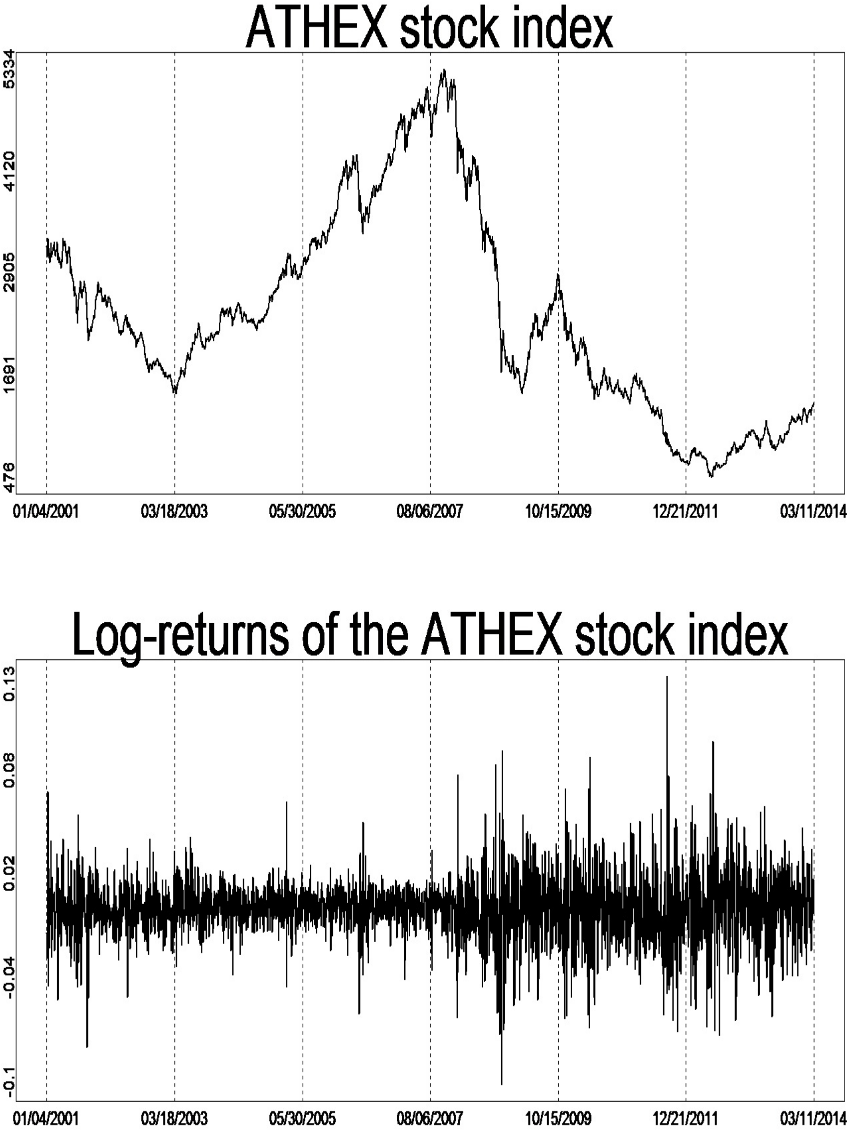


Figure 3.11: ***Greek stock index and its log-returns from 04/01/2001 to 11/03/2014***

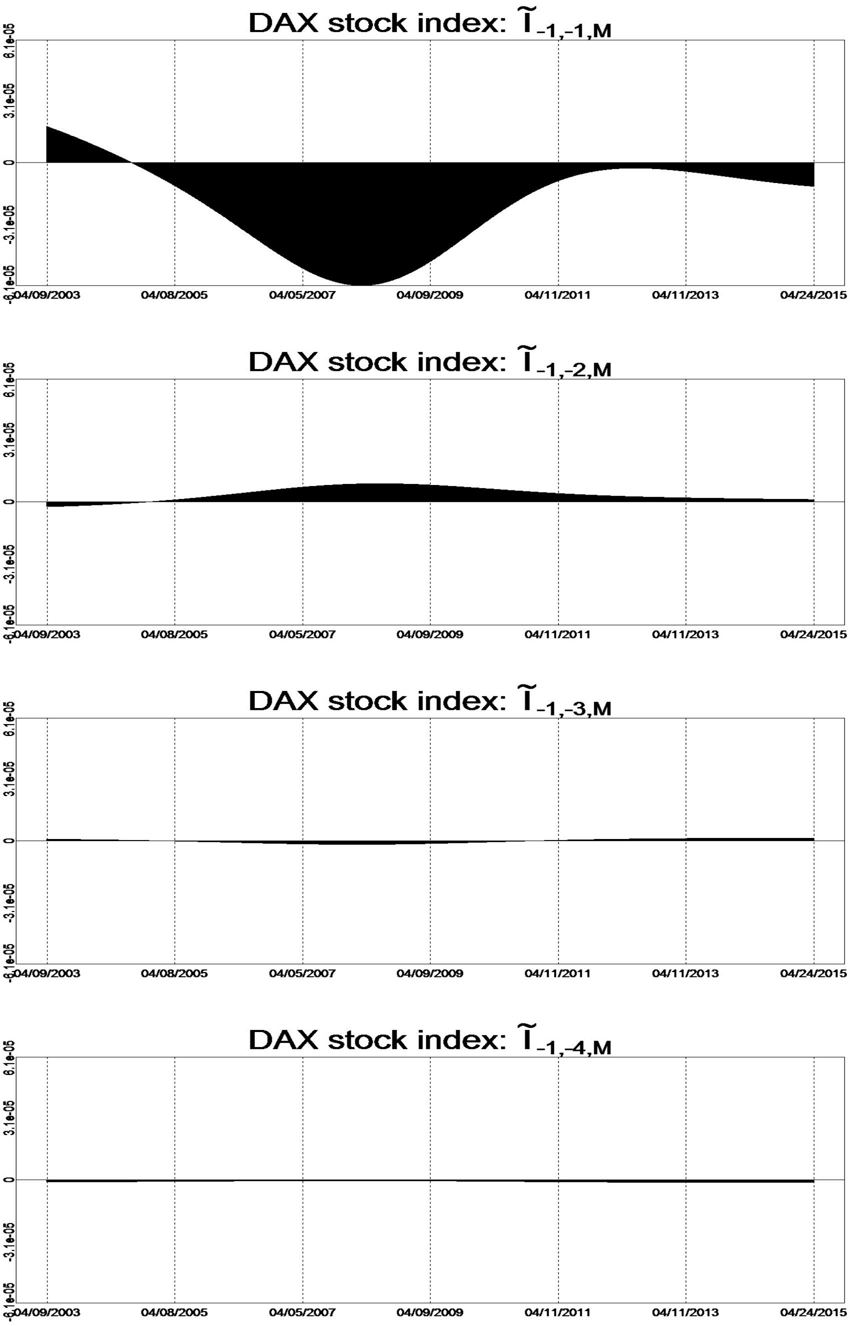


Figure 3.12: ***(German stock index) Estimated inter-scale EWS function from 11/04/2003 to 11/03/2014***

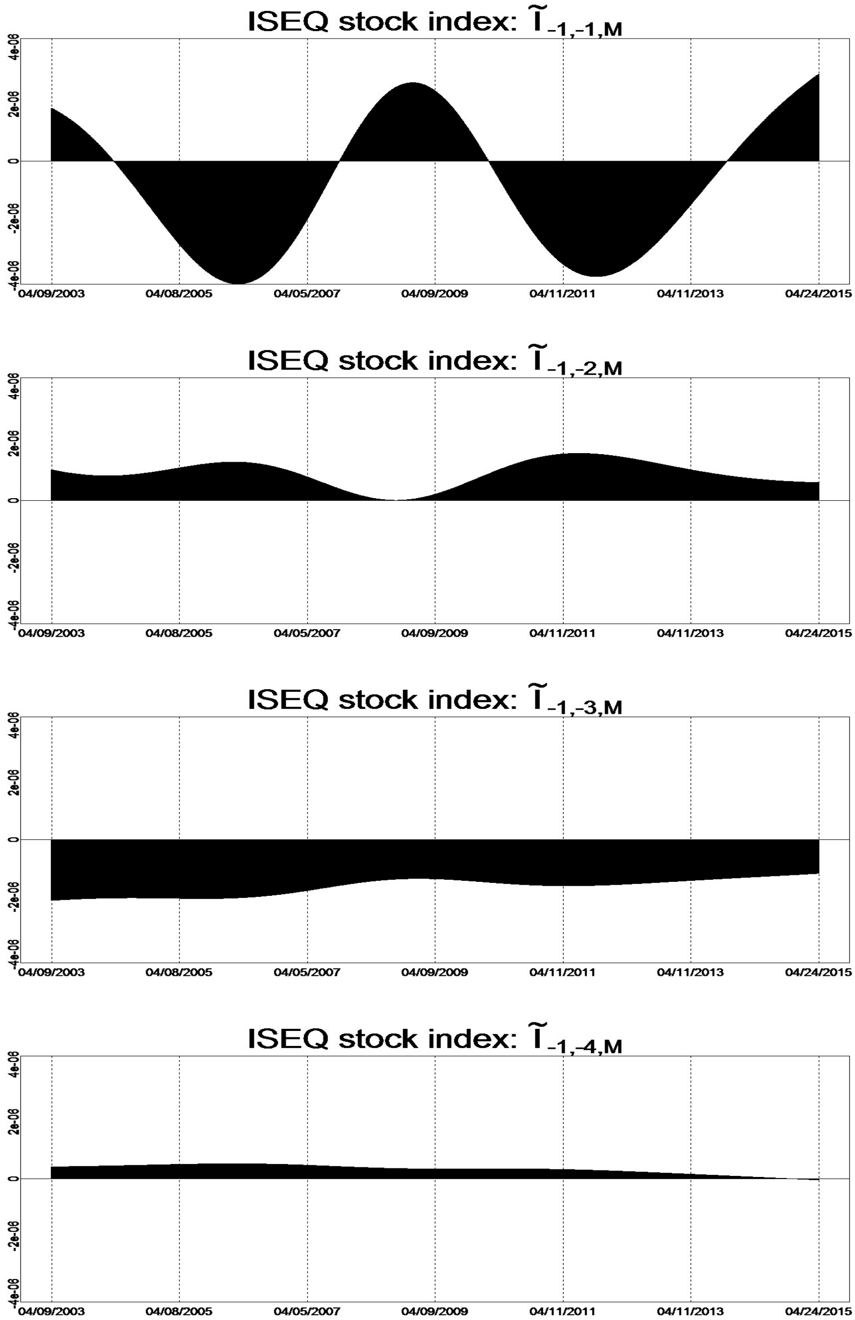


Figure 3.13: ***(Irish stock index) Estimated inter-scale EWS function from 11/04/2003 to 11/03/2014***

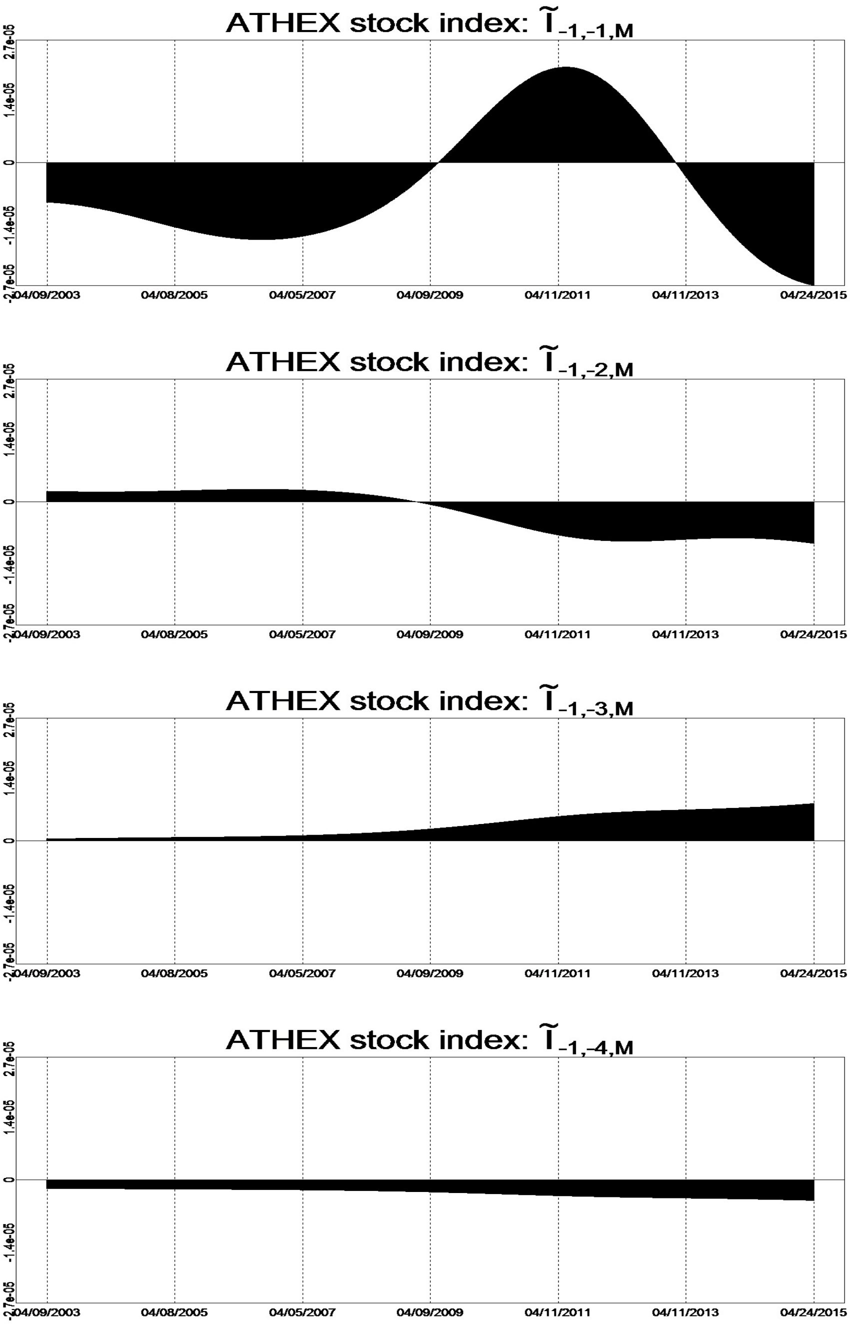


Figure 3.14: ***(Greek stock index) Estimated inter-scale EWS function from 11/04/2003 to 11/03/2014***

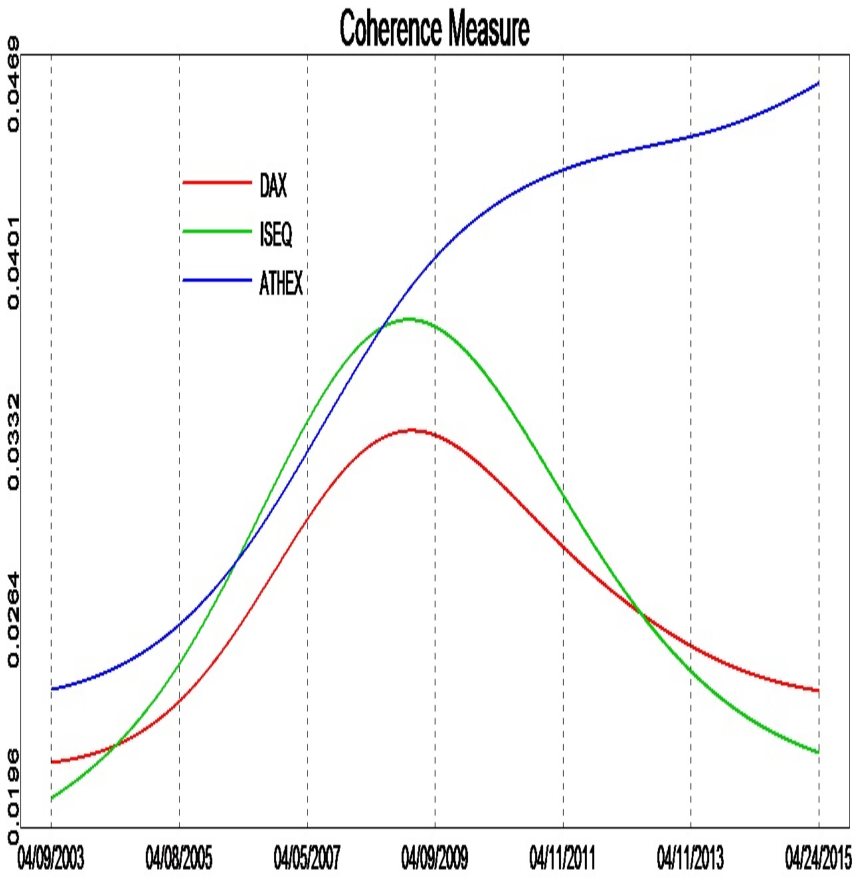


Figure 3.15: ***Coherence Measures from 11/04/2003 to 11/03/2014***

3.8 Proofs

3.8.1 Proof of Proposition 3.3

The covariance between $X(t)$ and $X(t+\tau)$ takes the form

$$\begin{aligned} C_n(\tau,z) &= \sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{0,j}\left(\tfrac{k}{n}\right)\psi_{j,k}(t)\psi_{j,k}(t+\tau) \\ &+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{i,j}\left(\tfrac{k}{n}\right)\psi_{j,k}(t)\psi_{j+i,k}(t+\tau) \\ &+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{-i,j+i}\left(\tfrac{k}{n}\right)\psi_{j+i,k}(t)\psi_{j,k}(t+\tau) \end{aligned}$$

where $S_{0,j}\left(\frac{k}{n}\right) = W_j^2\left(\frac{k}{n}\right)$and $S_{i,j}\left(\frac{k}{n}\right) = S_{-i,j+i}\left(\frac{k}{n}\right) = W_j\left(\frac{k}{n}\right)W_{j+i}\left(\frac{k}{n}\right)\Gamma_{i,j}\left(\frac{k}{n}\right)$. Given that $\psi_{j,t+k}(t+\tau) = \psi_{j,k}(\tau)$ and using a change of variables, we can write $\int_0^1 |C_n(\tau,z) - C(\tau,z)|$ as

$$\begin{aligned} \int_0^1 \Bigg| &\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty}\psi_{j,k}(0)\psi_{j,k}(\tau)\left(S_{0,j}\left(z+\tfrac{k}{n}\right)-S_{0,j}(z)\right) \\ &+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty}\left(\psi_{j,k}(0)\psi_{j+i,k}(\tau)+\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right)\left(S_{i,j}\left(z+\tfrac{k}{n}\right)-S_{i,j}(z)\right)\Bigg| \end{aligned}$$

where $z=\frac{t}{n}$. Given the assumption that the functions $S_{i,j}(z)$ are Lipschitz continuous and hence that $\int_0^1\left|S_{i,j}\left(z+\frac{k}{n}\right)-S_{i,j}(z)\right| \le \frac{C_{i,j}|k|}{n}$ we can bound $\int_0^1 |C_n(\tau,z)-C(\tau,z)|$ from above by

$$\begin{aligned} &\sum_{j=-1}^{-\infty}\frac{C_{0,j}}{n}\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \\ &\qquad+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\frac{C_{i,j}}{n}\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)+\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \end{aligned}$$

where

$$\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \le L_j\sum_{k=-\infty}^{+\infty}\left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right|$$

$$\begin{aligned} &\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)+\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \le \\ &\qquad L_{j+i}\sum_{k=-\infty}^{+\infty}\left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)+\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \end{aligned}$$

and hence we can bound $\sum\limits_{\tau=-\infty}^{+\infty} |C_n(\tau,z) - C(\tau,z)|$ from above by

$$\sum_{j=-1}^{-\infty} \frac{C_{0,j}L_j}{n} \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \\ + \sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} \frac{C_{i,j}L_{j+i}}{n} \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau) + \psi_{j+i,k}(0)\psi_{j,k}(\tau)\right|$$

Using that $\sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \leq 1$ and that $\sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau) + \psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \leq 1$ for $\tau \in \mathbb{Z}$ we have that

$$\sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \leq 2L_j \\ \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau) + \psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \leq 2L_{j+i}$$

Hence, $\sum_{\tau=-\infty}^{+\infty} \int_0^1 |C_n(\tau,z) - C(\tau,z)| dz \leq \frac{1}{n} \sum_{j=-1}^{-\infty} 2L_j^2 C_{0,j} + \frac{1}{n} \sum_{i=-1}^{-q} \sum_{j=-1}^{-\infty} 2L_{j+i}^2 C_{i,j}$ which leads to the result in 3.3 given the assumption that $\sum_{i=0}^{-q} \sum_{j=-1}^{-\infty} L_{j+i}^2 C_{i,j} < \infty$.

3.8.2 Proof of Proposition 3.4

To begin with, consider the two inverse representations

$$S_{q,i,j}(z) = \sum_{r=-q}^{0} \sum_{l=-\infty}^{-1} A_{i,r}(j,l)^{-1} \sum_\tau \Psi_{r,l}(\tau) C(\tau,z) \\ S_{n,q,i,j}(z) = \sum_{r=-q}^{0} \sum_{l=-\infty}^{-1} A_{i,r}(j,l)^{-1} \sum_\tau \Psi_{r,l}(\tau) C_n(\tau,z)$$

Using that $|\Psi_{r,l}(\tau)| \leq 1$ we can establish the upper bound

$$\int_0^1 \left|S_{n,q,i,j}(z) - S_{q,i,j}(z)\right| dz \leq \sum_{r=-q}^{0} \sum_{l=-\infty}^{-1} \left|A_{i,r}(j,l)^{-1}\right| \int_0^1 |C_n(\tau,z) - C(\tau,z)|.$$

Using the statement of Proposition 3.3 we have that $\int_0^1 |C_n(\tau,z) - C(\tau,z)| dz = O\left(\frac{1}{n}\right)$. Thus, given the assumption that $\sum_{i=-q}^{0} \sum_{j=-\infty}^{-1} |A_{i,r}(j,l)^{-1}| \leq C(q+1)2^{\frac{l}{2}}$ with $C > 0$ and if $q < \infty$ we obtain that

$$\int_0^1 \left|S_{n,q,i,j}(z) - S_{q,i,j}(z)\right| dz = O\left(\frac{1}{n}\right)$$

3.8.3 Proof of Theorem 3.1

Recall from Chapter 3 that for Haar wavelet functions the cross-correlation wavelets for $j=-1$ and $i=-J(n)+1$ are given by

$$\Psi_{i,j}(\tau)=\begin{cases} 2^{\frac{-J(n)-1}{2}}\left(2-2\left|2^{J(n)-1}-1-\tau\right|\right) & \text{if } \tau\in\left[2^{J(n)-1}-1,2^{J(n)-1}-1\right] \\ -2^{\frac{-J(n)-1}{2}}\left(1-\left|2^{J(n)}-1-\tau\right|\right) & \text{if } \tau\in\left[2^{J(n)}-1,2^{J(n)}-1\right] \\ -2^{\frac{J(n)-1}{2}}\left(1-\left|1-\tau\right|\right) & \text{if } \tau\in[1,1] \\ 0 & \text{otherwise} \end{cases}$$

Hence, we have that $A_{(-J(n)+1,-1),(-J(n)+1,-1)}=\sum_{\tau=-\infty}^{+\infty}\Psi_{-J(n)+1,-1}(\tau)^2=10\cdot 2^{-J(n)-1}$ and hence if $J(n)=\log_2(n)$ this becomes $A_{(-J(n)+1,-1),(-J(n)+1,-1)}=\frac{5}{n}$. Thus, using Horn's theorem we have that $\lambda_{\min}(A)\leq A_{(-J(n)+1,-1),(-J(n)+1,-1)}$ and hence

$$0\leq\lambda_{\min}(A)\leq\frac{5}{n}$$

3.8.4 Proof of Theorem 3.2

Consider the matrix $A_J=\left(A_{(i,j),(m,l)}\right)$, $j=-1,\ldots,-J$, $i=-1$, $l=-1,\ldots,-J$, $m=-1$ and define the matrix

$$A=\begin{pmatrix} A^{(1)}_{r,s}=\sum\limits_{\tau=-\infty}^{+\infty}\Psi_{0,r}\Psi_{0,s} & A^{(3)}_{r,s}=\sum\limits_{\tau=-\infty}^{+\infty}\Psi_{0,r}\Psi_{-1,s} \\ A^{(4)}_{r,s}=\sum\limits_{\tau=-\infty}^{+\infty}\Psi_{-1,r}(\tau)\Psi_{0,s}(\tau) & A^{(2)}_{r,s}=\sum\limits_{\tau=-\infty}^{+\infty}\Psi_{-1,r}(\tau)\Psi_{-1,s}(\tau) \end{pmatrix}$$

with $r=1,\ldots,J$ and $s=1,\ldots,J$ such that

$$A=\begin{pmatrix} A^{(1)} & A^{(3)} \\ A^{(4)} & A^{(2)} \end{pmatrix}$$

The components of the four submatrices are given by

$$A^{(1)}_{r,s} = \begin{cases} \frac{2^{2r-1}+1}{2^{s+1}} + \frac{1}{2} & \text{if } r < s \\ \frac{2^{2r}+5}{3 \cdot 2^{r+1}} + \frac{1}{2} & \text{if } r = s \\ \frac{2^{2s-1}+1}{2^{r+1}} + \frac{1}{2} & \text{if } r > s \end{cases}$$

$$A^{(2)}_{r,s} = \begin{cases} -\frac{1}{2^{3+s-2r}} & \text{if } r < s \\ \frac{2^r}{12} + \frac{1}{3} \cdot \frac{1}{2^{r+1}} & \text{if } r = s \\ -\frac{1}{2^{3+r-2s}} & \text{if } r > s \end{cases}$$

$$A^{(3)}_{r,s} = \begin{cases} \left(-\frac{\sqrt{2} \cdot 2^r}{48} - \frac{\sqrt{2}}{96 \cdot 2^{r-2}}\right) \cdot \frac{1}{2^{s-r-1}} & \text{if } r < s \\ -\frac{\sqrt{2} \cdot 2^{r-1}}{24} - \frac{\sqrt{2}}{12 \cdot 2^{r-1}} & \text{if } r = s \\ \frac{\sqrt{2}}{8} \cdot 2^{s-1} & \text{if } r = s+1 \\ \frac{3 \cdot \sqrt{2}}{8} \cdot 2^{s-1} \cdot \frac{1}{2^{r-s-2}} & \text{if } r > s+1 \end{cases}$$

$$A^{(4)}_{j,i} = \begin{cases} \frac{3 \cdot \sqrt{2}}{8} \cdot 2^{r-1} \cdot \frac{1}{2^{s-r-2}} & \text{if } r < s-1 \\ \frac{\sqrt{2}}{8} \cdot 2^{r-1} & \text{if } r = s-1 \\ -\frac{\sqrt{2} \cdot 2^{r-1}}{24} - \frac{\sqrt{2}}{12 \cdot 2^{r-1}} & \text{if } r = s \\ \left(-\frac{\sqrt{2} \cdot 2^s}{48} - \frac{\sqrt{2}}{96 \cdot 2^{s-2}}\right) \cdot \frac{1}{2^{r-s-1}} & \text{if } r > s \end{cases}$$

A_J can be written as as sum of three matrices

$$A_J = DTD + \Xi + \Theta$$

where

$$\Theta = \begin{pmatrix} \Theta^{(1)} & \Theta^{(3)} \\ \Theta^{(4)} & \Theta^{(2)} \end{pmatrix}$$

with $\Theta^{(k)},\ k=1,2,3,4$ being $J\times J-$dimensional matrices such that $\Theta^{(1)}_{r,s}=\frac{1}{2},\ \forall(r,s)$ and that $\Theta^{(2)}_{r,s}$, $\Theta^{(3)}_{r,s}$ and $\Theta^{(4)}_{r,s}$ are null matrices.

$$\Xi=\begin{pmatrix}\Xi^{(1)} & \Xi^{(3)}\\ \Xi^{(4)} & \Xi^{(2)}\end{pmatrix}$$

with $\Xi^{(k)},\ k=1,2,3,4$ being $J\times J-$dimensional matrices defined as

$$\Xi^{(1)}_{r,s}=\begin{cases}\frac{1}{2^{s+1}} & \text{if } r<s\\ \frac{5}{3\cdot 2^{r+1}} & \text{if } r=s\\ \frac{1}{2^{r+1}} & \text{if } r>s\end{cases}$$

$$\Xi^{(2)}_{r,s}=\begin{cases}\frac{1}{3\cdot 2^{r+1}} & \text{if } r=s\\ 0 & \text{elsewhere}\end{cases}$$

$$\Xi^{(3)}_{r,s}=\begin{cases}-\frac{\sqrt{2}}{96\cdot 2^{r-2}}\cdot\frac{1}{2^{s-r-1}} & \text{if } r<s\\ -\frac{\sqrt{2}}{12\cdot 2^{r-1}} & \text{if } r=s\\ 0 & \text{if } r>s\end{cases}$$

$$\Xi^{(4)}_{r,s}=\begin{cases}0 & \text{if } r<s\\ -\frac{\sqrt{2}}{12\cdot 2^{r-1}} & \text{if } r=s\\ -\frac{\sqrt{2}}{96\cdot 2^{s-2}}\cdot\frac{1}{2^{r-s-1}} & \text{if } r>s\end{cases}$$

T is a block-Toeplitz matrix such that

$$T=\begin{pmatrix}T^{(1)} & T^{(3)}\\ T^{(4)} & T^{(2)}\end{pmatrix}$$

where $T^{(k)}$ $k = 1,2,3,4$ are Toeplitz matrices given by

$$T^{(1)}_{r,s} = \begin{cases} \frac{2^{\frac{3}{2}r-1}}{2^{\frac{3}{2}s+1}} & \text{if } r < s \\ \frac{1}{6} & \text{if } r = s \\ \frac{2^{\frac{3}{2}s-1}}{2^{\frac{3}{2}r+1}} & \text{if } r > s \end{cases}$$

$$T^{(2)}_{r,s} = \begin{cases} -\frac{1}{2^{3+\frac{3}{2}r-\frac{3}{2}s}} & \text{if } r < s \\ \frac{1}{12} & \text{if } r = s \\ -\frac{1}{2^{3+\frac{3}{2}r-\frac{3}{2}s}} & \text{if } r > s \end{cases}$$

$$T^{(3)}_{r,s} = \begin{cases} -\frac{\sqrt{2}\cdot 2^{\frac{3}{2}r}}{48} \cdot \frac{1}{2^{\frac{3}{2}s-1}} & \text{if } r < s \\ -\frac{\sqrt{2}}{24} & \text{if } r = s \\ \frac{\sqrt{2}}{8} \cdot \frac{2^{\frac{1}{2}s-1}}{2^{\frac{1}{2}r}} & \text{if } r = s+1 \\ \frac{3\cdot\sqrt{2}}{8} \cdot \frac{2^{\frac{3}{2}s-1}}{2^{\frac{3}{2}r-2}} & \text{if } r > s+1 \end{cases}$$

$$T^{(4)}_{r,s} = \begin{cases} \frac{3\cdot\sqrt{2}}{8} \cdot \frac{2^{\frac{3}{2}r-1}}{2^{\frac{3}{2}s-2}} & \text{if } r < s-1 \\ \frac{\sqrt{2}}{8} \cdot \frac{2^{\frac{1}{2}r-1}}{2^{\frac{1}{2}s}} & \text{if } r = s-1 \\ -\frac{\sqrt{2}}{24} & \text{if } r = s \\ -\frac{\sqrt{2}\cdot 2^{\frac{3}{2}s}}{48} \cdot \frac{1}{2^{\frac{3}{2}r-1}} & \text{if } r > s \end{cases}$$

$$D = \operatorname{diag}\left(\sqrt{2}, \ldots, \sqrt{2^J}, \sqrt{2}, \ldots, \sqrt{2^J}\right)$$

We can decompose A_J further in the sense that Ξ can be written as $\Xi = \tilde{D}\tilde{T}\tilde{D}$.

$$\tilde{D} = \operatorname{diag}\left(\frac{1}{\sqrt{2}}, \ldots, \frac{1}{\sqrt{2^J}}, \frac{1}{\sqrt{2}}, \ldots, \frac{1}{\sqrt{2^J}}\right)$$

$\tilde{T}$ is again a blockwise Toeplitz matrix such that

$$\tilde{T} = \begin{pmatrix} \tilde{T}^{(1)} & \tilde{T}^{(3)} \\ \tilde{T}^{(4)} & \tilde{T}^{(2)} \end{pmatrix}$$

where $\tilde{T}^{(k)}$ $k = 1,2,3,4$ are Toeplitz matrices given by

$$\tilde{T}^{(1)}_{r,s} = \begin{cases} \frac{2^{\frac{1}{2}r}}{2^{\frac{1}{2}s+1}} & \text{if } r < s \\ \frac{5}{6} & \text{if } r = s \\ \frac{2^{\frac{1}{2}s}}{2^{\frac{1}{2}r+1}} & \text{if } r > s \end{cases}$$

$$\tilde{T}^{(2)}_{s,r} = \begin{cases} \frac{1}{6} & \text{if } r = s \\ 0 & \text{elsewhere} \end{cases}$$

$$\tilde{T}^{(3)}_{r,s} = \begin{cases} -\frac{\sqrt{2}\cdot 2^{\frac{1}{2}r+2}}{96\cdot 2^{\frac{1}{2}s-1}} & \text{if } r < s \\ -\frac{\sqrt{2}}{24} & \text{if } r = s \\ 0 & \text{if } r > s \end{cases}$$

$$\tilde{T}^{(4)}_{r,s} = \begin{cases} 0 & \text{if } r < s \\ -\frac{\sqrt{2}}{24} & \text{if } r = s \\ -\frac{\sqrt{2}\cdot 2^{\frac{1}{2}s+2}}{96\cdot 2^{\frac{1}{2}r-1}} & \text{if } r > s \end{cases}$$

Hence, A can be decomposed as

$$A = DTD + \tilde{D}\tilde{T}\tilde{D} + \Theta \tag{3.8.1}$$

Now denote as $\lambda_{\min}(A)$ the smallest eigenvalue of the matrix A. Using Weyl's inequalities we can write

$$\lambda_{\min}(A) \geq \lambda_{\min}(D)\cdot\lambda_{\min}(T)\cdot\lambda_{\min}(D)+\lambda_{\min}(\tilde{D})\cdot\lambda_{\min}(\tilde{T})\cdot\lambda_{\min}(\tilde{D})+\lambda_{\min}(\Theta)$$

with $\lambda_{\min}(D)=\sqrt{2}$, $\lambda_{\min}(\tilde{D})=\frac{1}{\sqrt{2^J}}$ and $\lambda_{\min}(\Theta)=0$. Thus, in order to prove that A has an absolute lower bound δ that does not depend on J we will show that the smallest eigenvalues of the two blockwise Toeplitz matrices T and $\tilde{T}$ have absolute lower bounds δ_T and $\delta_{\tilde{T}}$ respectively.

Define $T=\begin{pmatrix} B & C \\ C^t & D \end{pmatrix}$ as a symmetric blockwise Toeplitz matrix with four blocks and consider a quadratic form in the vector $\begin{pmatrix} x & z \end{pmatrix}^t$ where x and z are both vectors of dimension $J\times 1$ with $\begin{pmatrix} \bar{x} & \bar{z} \end{pmatrix}^t \begin{pmatrix} x \\ z \end{pmatrix}=1$ where $\bar{x}$ denotes the complex conjugate of x

$$\begin{aligned}\begin{pmatrix} \bar{x} & \bar{z} \end{pmatrix}^t T \begin{pmatrix} x \\ z \end{pmatrix} &= \sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} b_{j-k}\cdot\bar{x}_k\cdot x_j+\sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} c_{j-k}\cdot\bar{x}_k\cdot z_j \\ &+\sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} c_{k-j}\cdot\bar{z}_k\cdot x_j+\sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} d_{j-k}\cdot\bar{z}_k\cdot z_j \\ &= \sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} \begin{pmatrix} \bar{x}_k & \bar{z}_k \end{pmatrix}^t \begin{pmatrix} b_{j-k} & c_{j-k} \\ c_{k-j} & d_{j-k} \end{pmatrix}\begin{pmatrix} x_j \\ z_j \end{pmatrix}\end{aligned}$$

Using Parseval's identity for each one of the components of the matrix $\begin{pmatrix} b_{j-k} & c_{j-k} \\ c_{k-j} & d_{j-k} \end{pmatrix}$ we can write this as

$$\begin{aligned}&\begin{pmatrix} \bar{x} & \bar{z} \end{pmatrix}^t T \begin{pmatrix} x \\ z \end{pmatrix} \\ &= \sum_{k=0}^{\frac{J}{2}-1}\sum_{j=0}^{\frac{J}{2}-1} \begin{pmatrix} \bar{x}_k & \bar{z}_k \end{pmatrix}^t \frac{1}{2\pi}\int_0^{2\pi} \begin{pmatrix} \kappa_B(\omega) & \kappa_C(\omega) \\ \bar{\kappa}_C(\omega) & \kappa_D(\omega) \end{pmatrix} e^{i(k-j)\omega} d\omega \begin{pmatrix} x_j \\ z_j \end{pmatrix} \\ &= \frac{1}{2\pi}\int_0^{2\pi} \begin{pmatrix} \overline{\sum_{k=0}^{\frac{J}{2}-1} x_k e^{-i\omega k}} & \overline{\sum_{k=0}^{\frac{J}{2}-1} z_k e^{-i\omega k}} \end{pmatrix}^t \begin{pmatrix} \kappa_B(\omega) & \kappa_C(\omega) \\ \bar{\kappa}_C(\omega) & \kappa_D(\omega) \end{pmatrix} \begin{pmatrix} \sum_{j=0}^{\frac{J}{2}-1} x_k e^{-i\omega j} \\ \sum_{k=0}^{\frac{J}{2}-1} z_k e^{-i\omega j} \end{pmatrix} d\omega\end{aligned}$$

Now define $\lambda(\omega^*)$ as the smallest eigenvalue of the matrix

$$\Omega(\omega) = \begin{pmatrix} \kappa_B(\omega) & \kappa_C(\omega) \\ \bar{\kappa}_C(\omega) & \kappa_D(\omega) \end{pmatrix} = \begin{pmatrix} \sum_{k=-\infty}^{+\infty} b_k \cdot e^{i\omega k} & \sum_{k=-\infty}^{+\infty} c_k \cdot e^{i\omega k} \\ \sum_{k=-\infty}^{+\infty} c_j \cdot e^{-i\omega k} & \sum_{k=-\infty}^{+\infty} d_k \cdot e^{i\omega k} \end{pmatrix}$$

Using that

$$\begin{pmatrix} \overline{\sum_{k=0}^{\frac{J}{2}-1} x_k e^{-i\omega k}} & \overline{\sum_{k=0}^{\frac{J}{2}-1} z_k e^{-i\omega k}} \end{pmatrix}^t \begin{pmatrix} \kappa_B(\omega) & \kappa_C(\omega) \\ \bar{\kappa}_C(\omega) & \kappa_D(\omega) \end{pmatrix} \begin{pmatrix} \sum_{k=0}^{\frac{J}{2}-1} x_k e^{-i\omega k} \\ \sum_{k=0}^{\frac{J}{2}-1} z_k e^{-i\omega k} \end{pmatrix} \geq \lambda(\omega^*)$$

we have that

$$\begin{pmatrix} x & z \end{pmatrix}^t T \begin{pmatrix} x \\ z \end{pmatrix} \geq \lambda(\omega^*)$$

The closed form expressions of the geometric series contained in $\Omega(\omega)$ are given by

$$\kappa_B(\omega) = \sum_{k=-\infty}^{+\infty} b_k \cdot e^{i\omega k} = -\frac{1}{3} + \frac{1}{4} \cdot \frac{16-2^{\frac{5}{2}}\cos(\omega)}{9-2^{\frac{5}{2}}\cos(\omega)}$$
$$\kappa_C(\omega) = \sum_{k=-\infty}^{+\infty} c_k \cdot e^{i\omega k} = \frac{\sqrt{2}}{48} - \frac{e^{i\omega}}{8} - \frac{36-\sqrt{2}\sqrt{8}}{48\left(2^{\frac{3}{2}}-e^{i\omega}\right)}$$
$$\bar{\kappa}_C(\omega) = \sum_{k=-\infty}^{+\infty} c_k \cdot e^{-i\omega k} = \frac{\sqrt{2}}{48} - \frac{e^{-i\omega}}{8} - \frac{36-\sqrt{2}\sqrt{8}}{48\left(2^{\frac{3}{2}}-e^{-i\omega}\right)}$$
$$\kappa_D(\omega) = \sum_{k=-\infty}^{+\infty} d_k \cdot e^{i\omega k} = \frac{1}{3} + \frac{\cos(\omega)}{4\sqrt{8}} - \frac{1}{8} \cdot \frac{16-2^{\frac{5}{2}}\cos(\omega)}{9-2^{\frac{5}{2}}\cos(\omega)}$$

Denote as

$$\lambda^+(\omega) = \frac{\kappa_B(\omega)+\kappa_D(\omega)+\sqrt{(\kappa_B(\omega)-\kappa_D(\omega))^2+4\cdot\kappa_C^2(\omega)}}{2}$$

$$\lambda^-(\omega) = \frac{\kappa_B(\omega)+\kappa_D(\omega)-\sqrt{(\kappa_B(\omega)-\kappa_D(\omega))^2+4\cdot\kappa_C^2(\omega)}}{2}$$

with $\lambda^{+}(\omega) \geq 0.0783$ and $\lambda^{-}(\omega) \geq 0.0101$ and hence $\lambda(\omega^{*}) = \delta_T = 0.0101$.

As far as $\tilde{T}$ is concerned, the closed form expressions of the geometric series are given by

$$\tilde{\kappa}_B(\omega) = -\frac{1}{6} + \frac{1}{2} \cdot \frac{4-2^{\frac{3}{2}}\cos(\omega)}{3-2^{\frac{3}{2}}\cos(\omega)}$$

$$\tilde{\kappa}_C(\omega) = -\frac{1}{6\sqrt{2}} - \frac{1}{6\left(\sqrt{2}-e^{i\omega}\right)}$$

$$\tilde{\kappa}_D(\omega) = \frac{1}{6}$$

Denote as

$$\tilde{\lambda}^{+}(\omega) = \frac{\tilde{\kappa}_B(\omega)+\tilde{\kappa}_D(\omega)+\sqrt{(\tilde{\kappa}_B(\omega)-\tilde{\kappa}_D(\omega))^2+4\cdot\tilde{\kappa}_C^2(\omega)}}{2}$$

$$\tilde{\lambda}^{-}(\omega) = \frac{\tilde{\kappa}_B(\omega)+\tilde{\kappa}_D(\omega)-\sqrt{(\tilde{\kappa}_B(\omega)-\tilde{\kappa}_D(\omega))^2+4\cdot\tilde{\kappa}_C^2(\omega)}}{2}$$

with $\tilde{\lambda}^{+}(\omega) \geq 0.5184$ and $\tilde{\lambda}^{-}(\omega) \geq 0.0673$ and hence $\tilde{\lambda}(\omega^{*}) = \delta_{\tilde{T}} = 0.0673$. The absolute lower bound of A is obtained by plugging these values into Weyl's inequality

$$\lambda_{\min}(A) \geq \sqrt{2}\cdot 0.0101\cdot\sqrt{2} + \frac{1}{\sqrt{2}^{J}}\cdot 0.0673\cdot\frac{1}{\sqrt{2}^{J}} + 0 \geq 0.0202$$

3.8.5 Proof of Corollary 3.1

Define the vector $D = \left\{\left\{2^{-\frac{j}{2}}\right\}_{j=-J}^{-1}, \left\{2^{-\frac{j}{2}}\right\}_{j=-J}^{-1}\right\}$ and consider the matrix $\Omega = D^{-1}AD^{-1}$. Given the decomposition of A in (3.8.1), we have that $\Omega = T + D^{-2}\tilde{T}D^{-2} + D^{-1}\Xi D^{-1}$. Using Weyl's inequality we can establish the lower bound

$$\lambda_{\min}(\Omega) \geq \lambda_{\min}(T)$$

Using the results of the proof of Theorem 3.3 we have that

$$\lambda_{\min}(\Omega) \geq 0.0101$$

Thus, given that $|\Omega_{i,r}(j,l)^{-1}| \leq \frac{1}{0.0101}$ we can write $\sum_{r=-1}^{0}\sum_{l=-\infty}^{-1}|A_{i,r}(j,l)^{-1}|$ $\leq \frac{2^{\frac{j}{2}}}{0.0101}2\sum_{j=-\infty}^{-1}2^{\frac{j}{2}}$ with $r,l<0$ and where $\sum_{j=-\infty}^{-1}2^{\frac{j}{2}} < +\infty$. Hence, we have that

$$\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}|A_{i,r}(j,l)^{-1}| \leq C2^{\frac{j}{2}+1}$$

3.8.6 Proof of Theorem 3.3

$\kappa(\omega)$ is a 3×3 dimensional matrix

$$\kappa(\omega) = \begin{pmatrix} \kappa_B(\omega) & \kappa_C(\omega) & \kappa_E(\omega) \\ \bar{\kappa}_C(\omega) & \kappa_D(\omega) & \kappa_F(\omega) \\ \bar{\kappa}_E(\omega) & \bar{\kappa}_F(\omega) & \kappa_G(\omega) \end{pmatrix}$$

where

$$\kappa_E(\omega) = \tfrac{3}{4} - \tfrac{\sqrt{8}-24\cdot\sqrt{2}-3\cdot\sqrt{8}\cdot\sqrt{2}}{32\cdot\sqrt{8}\cdot\sqrt{2}}\cdot e^{i\omega} + \tfrac{3}{4}\cdot\tfrac{1}{1-\frac{e^{i\omega}}{\sqrt{2}^3}}$$
$$\kappa_F(\omega) = \tfrac{5\cdot\sqrt{2}}{48} + \tfrac{4\cdot\sqrt{8}+\sqrt{8}\cdot\sqrt{2}+16\cdot\sqrt{2}+2\cdot\sqrt{8}\cdot\sqrt{2}}{128\cdot\sqrt{8}}\cdot e^{i\omega} + \tfrac{\sqrt{2}}{8}\cdot\tfrac{1}{1-\frac{e^{i\omega}}{\sqrt{2}^3}}$$

$$\kappa_G(\omega) = \tfrac{-72-\sqrt{8}\cdot\sqrt{2}}{48\cdot\sqrt{8}} - \tfrac{1}{8}\cdot e^{i\omega} + \tfrac{36-\sqrt{8}\cdot\sqrt{2}}{24\cdot\sqrt{8}}\cdot\tfrac{1}{1-\frac{e^{i\omega}}{\sqrt{2}^3}}$$

with $\lambda^{(1)}(\omega) \geq 0.1103$, $\lambda^{(2)}(\omega) \geq 0.0211$ and $\lambda^{(3)}(\omega) \geq 0.002$ and hence $\lambda(\omega^*) = \delta_T = 0.002$. Thus, the absolute lower bound of A is given by

$$\lambda_J(A) \geq 0.004$$

Using the same argument as in the proof of Theorem 3.2 we get that

$$\lambda_{\min}(\Omega) \geq 2^{1-\alpha}\cdot 0.002$$

3.8.7 Proof of Corollary 3.2

Define the vector $D = \left\{\left\{2^{-\frac{j}{2}}\right\}_{j=-J}^{-1}, \left\{2^{-\frac{j}{2}}\right\}_{j=-J}^{-1}, \left\{2^{-\frac{j}{2}}\right\}_{j=-J}^{-1}\right\}$ and consider the matrix $\Omega = D^{-1}AD^{-1}$. Given the decomposition of A in (3.8.1), we have that $\Omega = T + D^{-2}\tilde{T}D^{-2} + D^{-1}\Xi D^{-1}$. Using Weyl's inequality we can establish the lower bound

$$\lambda_{\min}(\Omega) \geq \lambda_{\min}(T)$$

Using the results of the proof of Theorem 3.3 we have that

$$\lambda_{\min}(\Omega) \geq 0.004$$

Thus, given that $|\Omega_{i,r}(j,l)^{-1}| \leq \frac{1}{0.004}$ we can write $\sum_{r=-2}^{0}\sum_{l=-\infty}^{-1}|A_{i,r}(j,l)^{-1}|$ $\leq \frac{2^{\frac{j}{2}}}{0.004}3\sum_{j=-\infty}^{-1}2^{\frac{j}{2}}$ with $r,l<0$ and where $\sum_{j=-\infty}^{-1}2^{\frac{j}{2}} < +\infty$. Hence, we have that

$$\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}|A_{i,r}(j,l)^{-1}| \leq 3C2^{\frac{j}{2}}$$

3.8.8 Proof of Theorem 3.4

The expected value of $I_{i,j,k}$ takes the form

$$\begin{aligned}
&E\left(I_{i,j,k}\right)\\
&= 2E\left(\sum_t \psi_{j,k}(t)X(t)\sum_{t'}\psi_{j+i,k}(t')X(t')\right)\\
&= 2\sum_t \psi_{j,k}(t)\sum_{t'}\psi_{j+i,k}(t')\sum_l\sum_m S_{0,l}\left(\tfrac{m}{n}\right)\psi_{l,m}(t)\psi_{l,m}(t')\\
&+2\sum_t \psi_{j,k}(t)\sum_{t'}\psi_{j+i,k}(t')\sum_l\sum_r\sum_m S_{r,l}\left(\tfrac{m}{n}\right)\left(\psi_{l,m}(t)\psi_{l+r,m}(t')+\psi_{l,m}(t')\psi_{l+r,m}(t)\right)
\end{aligned}$$

If we put $z=\frac{k}{n}$ and set $t'=t+\tau$ this can be written as

$$\begin{aligned}
&2\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_l S_{0,l}\left(z\right)\Psi_{0,l}(\tau)\\
&+2\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}\left(z\right)\left(\Psi_{r,l}(\tau)+\Psi_{-r,l+r}(\tau)\right)\\
&+2\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)R(t,\tau)
\end{aligned}$$

where

$$\begin{aligned}
&R(t,\tau)\\
&= \sum_{l=-1}^{-\infty}\sum_{m=-k-t}^{n-1-k-t}\left(S_{0,l}\left(z+\tfrac{m}{n}\right)-S_{0,l}\left(z\right)\right)\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\\
&+\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}\sum_{m=-k-t}^{n-1-k-t}\left(S_{r,l}\left(z+\tfrac{m}{n}\right)-S_{r,l}\left(z\right)\right)\left(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)+\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)\right)
\end{aligned}$$

Hence, using the assumption that the functions $S_{0,j}(z)$ and $S_{i,j}(z)$ are Lipschitz continuous we have that $|R(t,\tau)|$ can be bounded from above by

$$\begin{aligned}&\leq \tfrac{1}{n}\sum_{l=-1}^{-\infty} C_{0,l} \sum_{m=-k-t}^{n-1-k-t} |m|\,|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)| \\ &+\tfrac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty} C_{r,l}\sum_{m=-k-t}^{n-1-k-t} |m|\,|(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)+\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau))|\end{aligned}$$

where

$$\sum_{m=-k-t}^{n-1-k-t} |m|\left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right| \leq L_l \sum_{m=-k-t}^{n-1-k-t}\left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right|$$

$$\begin{aligned}&\sum_{m=-k-t}^{n-1-k-t} |m|\left|\left(\psi_{l,k}(0)\psi_{l+r,k}(\tau)+\psi_{l+r,k}(0)\psi_{l,k}(\tau)\right)\right| \leq \\ &\qquad L_{r+l}\sum_{m=-k-t}^{n-1-k-t}\left|\left(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)+\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)\right)\right|\end{aligned}$$

Given that

$$\begin{aligned}&\sum_{m=-k-t}^{n-1-k-t}\left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right| \leq 1 \\ &\sum_{m=-k-t}^{n-1-k-t}\left|\left(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)+\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)\right)\right| \leq 1\end{aligned}$$

we can bound $2\left|\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)R(t,\tau)\right|$ from above by

$$2\left(\sum_{\tau=-L_{j+i}+1}^{L_{j+i}-1}\sum_t\left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right|\right)\left(\frac{1}{n}\sum_{l=-1}^{-\infty}C_{0,l}L_l+\frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}C_{r,l}L_{r+l}\right)$$

where $\sum_t\left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right| \leq 1$ and hence

$$\sum_{\tau=-L_{j+i}+1}^{L_{j+i}-1}\sum_t\left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right| \leq 2L_{j+i}.$$

Thus, we have that

$$2\left|\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)R(t,\tau)\right| \leq \frac{1}{n}\sum_{l=-1}^{-\infty} 4C_{0,l}L_l^2 + \frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty} 4C_{r,l}L_{r+l}^2$$

Using that $\sum\limits_{r=0}^{-q}\sum\limits_{l=-1}^{-\infty} 2C_{r,l}L_{r+l}^2 < \infty$ we thus have that

$$\left|\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)R(t,\tau)\right| = O\left(\frac{1}{n}\right).$$

Moreover, note that $\sum\limits_t \psi_{j,k}(t)\psi_{j+i,k}(t+\tau) = \Psi_{i,j}(\tau)$. Using that $\Psi_{0,l}(-\tau) = \Psi_{0,l}(\tau)$ and $\Psi_{r,l}(-\tau) + \Psi_{-r,l+r}(-\tau) = \Psi_{r,l}(\tau) + \Psi_{-r,l+r}(\tau)$ and that $\Psi_{i,j}(0) = \Psi_{i,j}(0) + \Psi_{-i,j+i}(0) = 0$ we thus have that

$$\begin{aligned}
&2\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_l S_{0,l}(z)\Psi_{0,l}(\tau) \\
&\qquad\qquad + 2\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}(z)\left(\Psi_{r,l}(\tau)+\Psi_{-r,l+r}(\tau)\right) \\
&= 2\sum_l S_{0,l}(z)\sum_\tau \Psi_{j+i}(\tau)\Psi_{0,l}(\tau) + 2\sum_r\sum_l S_{r,l}(z)\sum_\tau \Psi_{j+i}(\tau)\left(\Psi_{r,l}(\tau)+\Psi_{-r,l+r}(\tau)\right) \\
&= \sum_l S_{0,l}(z)\,2\sum_{\tau>0}\left(\Psi_{j+i}(\tau)+\Psi_{-i,j+i}(\tau)\right)\Psi_{0,l}(\tau) \\
&\qquad\qquad + \sum_r\sum_l S_{r,l}(z)\,2\sum_{\tau>0}\left(\Psi_{j+i}(\tau)+\Psi_{-i,j+i}(\tau)\right)\left(\Psi_{r,l}(\tau)+\Psi_{-r,l+r}(\tau)\right) \\
&= \sum_l S_{0,l}(z)\sum_\tau\left(\Psi_{j+i}(\tau)+\Psi_{-i,j+i}(\tau)\right)\Psi_{0,l}(\tau) \\
&\qquad\qquad + \sum_r\sum_l S_{r,l}(z)\sum_\tau\left(\Psi_{j+i}(\tau)+\Psi_{-i,j+i}(\tau)\right)\left(\Psi_{r,l}(\tau)+\Psi_{-r,l+r}(\tau)\right) \\
&= \sum_r\sum_l S_{r,l}(z)A_{i,r}(j,l)
\end{aligned}$$

3.8.9 Proof of Theorem 3.5

Write the variance of the raw wavelet periodogram as $\mathrm{Var}\left(I_{i,j,k}\right) = E\left(I_{i,j,k}^2\right) - \left(E\left(I_{i,j,k}\right)\right)^2$. Using Isserli's theorem for higher moments of the multivariate normal distribution, we have that

$$E\left(I_{i,j,k}^2\right) = 2\cdot\left(E\left(d_{j,k}d_{j+i,k}\right)\right)^2 + E\left(d_{j,k}d_{j,k}\right)E\left(d_{j+i,k}d_{j+i,k}\right)$$

and hence $\mathrm{Var}\left(I_{i,j,k}\right) = \left(E\left(d_{j,k}d_{j+i,k}\right)\right)^2 + E\left(d_{j,k}d_{j,k}\right)E\left(d_{j+i,k}d_{j+i,k}\right)$. Given the results of Theorem 3.4 we hence have that

$$\begin{aligned}\mathrm{Var}\left(I_{i,j,k}\right) &= \tfrac{1}{4}\left(\sum_r\sum_l S_{r,l}(z)A_{i,r}(j,l)\right)^2 \\ &+\sum_r\sum_l S_{r,l}(z)A_{0,r}(j,l)\sum_r\sum_l S_{r,l}(z)A_{0,r}(j+i,l)+O\left(\tfrac{1}{n}\right)\end{aligned}$$

3.8.10 Proof of Theorem 3.6

Bound the variance of the smoothed wavelet periodogram as

$$\mathrm{Var}\left(I_{i,j,k,M}\right) \leq \frac{1}{(2M+1)^2}\sum_{s=-M}^{M}\sum_{\kappa}\left|\mathrm{Cov}\left(I_{i,j,k+s}, I_{i,j,k+s+\kappa}\right)\right|$$

Using Isserli's theorem (Isserli (1918)) for higher moment of zero-mean Gaussian random variables we can write

$$\begin{aligned}\mathrm{Cov}\left(I_{i,j,k+s}, I_{i,j,k+s+\kappa}\right) &= E\left(I_{i,j,k+s}I_{i,j,k+s+\kappa}\right) - E\left(I_{i,j,k+s}\right)E\left(I_{i,j,k+s+\kappa}\right) \\ &= E\left(d_{j,k+s}d_{j,k+s+\kappa}\right)E\left(d_{j+i,k+s}d_{j+i,k+s+\kappa}\right) \\ &+E\left(d_{j,k+s}d_{j+i,k+s+\kappa}\right)E\left(d_{j+i,k+s}d_{j,k+s+\kappa}\right)\end{aligned}$$

Using similar techniques as in the proof of Theorem 3.4 we can show that for $z = \frac{k}{n}$

$$E\left(d_{j,k+s}d_{j+i,k+s+\kappa}\right) = \sum_{l=-\infty}^{-1}\sum_{r=-q}^{0} S_{r,l}(z)A_{i,r}^{(\kappa)}(j,l) + O\left(\frac{1}{n}\right)$$

where

$$A_{i,r}^{(\kappa)}(j,l) = \sum_{\tau}\left(\Psi_{i,j}(\tau)+\Psi_{-i,j+i}(\tau)\right)\left(\Psi_{r,l}(\tau+\kappa)+\Psi_{-r,l+r}(\tau+\kappa)\right)$$

Note that $|A_{i,r}^{(\kappa)}(j,l)| \leq 4\min(L_{j+i}, L_{l+r})\mathbb{I}\left[|\kappa| \leq \max(L_{j+i}, L_{l+r})\right]$, which means that we can write

$$\begin{aligned}\left|\sum_l\sum_r S_{r,l}(z)A_{i,r}^{(\kappa)}(j,l)\right| &\leq 4\mathbb{I}\left[|\kappa| \leq L_{j+i}\right]\sum_{l=j}^{-1}\sum_{r=-q}^{0} S_{r,l}(z)L_{l+r} \\ &+4L_{j+i}\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0} S_{r,l}(z)\mathbb{I}\left[|\kappa| \leq L_{l+r}\right]\end{aligned}$$

This means that $\left|E\left(d_{j,k+s}d_{j+i,k+s+\kappa}\right)E\left(d_{j+i,k+s}d_{j,k+s+\kappa}\right)\right|$ can be bounded by

$$\begin{aligned}
&4\mathbb{I}\left[|\kappa|\le L_{j+i}\right]\sum_{l=j}^{-1}\sum_{r=-q}^{0}\left|S_{r,l}\right|(z)\,L_{l+r}\sum_{l'=j}^{-1}\sum_{r'=-q}^{0}\left|S_{r',l'}\right|(z)\,L_{l'+r'}\\
&+4\mathbb{I}\left[|\kappa|\le L_{j+i}\right]L_{j+i}\sum_{l=j}^{-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|L_{l+r}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|\mathbb{I}\left[|\kappa|\le L_{l'+r'}\right]\\
&+4\mathbb{I}\left[|\kappa|\le L_{j+i}\right]L_{j+i}\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|\mathbb{I}\left[|\kappa|\le L_{l+r}\right]\sum_{l'=j}^{-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|L_{l'+r'}\\
&+4L_{j+i}^{2}\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|\mathbb{I}\left[|\kappa|\le L_{l+r}\right]\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|\mathbb{I}\left[|\kappa|\le L_{l'+r'}\right]
\end{aligned}$$

Summing this over κ then leads to

$$\begin{aligned}
B_{i,j}=\;&8L_{j+i}\sum_{l=j}^{-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|L_{l+r}\sum_{l'=j}^{-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|L_{l'+r'}\\
&+8L_{j+i}^{2}\sum_{l=j}^{-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|L_{l+r}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|\\
&+8L_{j+i}^{2}\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|\sum_{l'=j}^{-1}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|L_{l'+r'}\\
&+16L_{j+i}^{2}\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}\left|S_{r,l}(z)\right|\left|S_{r',l'}(z)\right|\min(L_{l+r},L_{l'+r'})
\end{aligned}$$

Given the assumption that for all l and r the EWS functions are such that $|S_{r,l}(z)|\le\frac{K}{L_{l+r}}$ with $L_{l-1+r}=2L_{l+r}$, we obtain that $\sum_{l=j}^{-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|\le Kj(q+1)$. Moreover, we can establish the upper bound $\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|\le\frac{2K(q+1)}{L_j}$. Next, we write the expression $\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}\left|S_{r,l}(z)\right|\left|S_{r',l'}(z)\right|\min(L_{l+r},L_{l'+r'})$ under the form

$$\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{0}S_{r,l}L_{l+r}\sum_{l'\ge l}\sum_{r'=-q}^{0}\left|S_{r',l'}(z)\right|+\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{0}S_{r',l'}L_{l'+r'}\sum_{l\ge l'}\sum_{r=-q}^{0}\left|S_{r,l}(z)\right|$$

which can be bounded from above by the sequence of upper bounds

$$K^{2}(q+1)^{2}\left(\sum_{l=-\infty}^{j-1}\frac{1}{L_{l}}+\sum_{l'=-\infty}^{j-1}\frac{1}{L_{l'}}\right)\le\frac{2K^{2}(q+1)^{2}}{L_{j}}$$

Thus, we have that $B_{i,j} = O\left(L_j q^2\right)$. Hence, using that $\mathrm{Var}\left(I_{i,j,k,M}\right) \leq \frac{2}{M} B_{i,j} + O\left(\frac{1}{n}\right)$ we get that $\mathrm{Var}\left(I_{i,j,k,M}\right) = O\left(\frac{L_j q^2 j^2}{M}\right)$.

3.8.11 Proof of Theorem 3.7

To begin with, write $E\left(I_{i,j,k,M}\right) = \frac{1}{2M+1}\sum_{s=-M}^{M} E\left(I_{i,j,k+s}\right)$. According to Theorem 3.4 we have that $E\left(I_{i,j,k+s}\right) = \sum_l \sum_r S_{r,l}\left(z+\frac{s}{n}\right) A_{i,r}(j,l) + O\left(\frac{1}{n}\right)$ with $z = \frac{k}{n}$. Thus, we can write

$$\begin{aligned} E\left(I_{i,j,k+s}\right) &= \sum_l \sum_r S_{r,l}(z) A_{i,r}(j,l) \\ &+ \sum_l \sum_r \left(S_{r,l}\left(z+\tfrac{s}{n}\right) - S_{r,l}(z)\right) A_{i,r}(j,l) + O\left(\tfrac{1}{n}\right) \end{aligned}$$

Using the Lipschitz continuity properties of the functions $S_{r,l}(z)$, we have that

$$\left|\sum_l \sum_r \left(S_{r,l}\left(z+\frac{s}{n}\right) - S_{r,l}(z)\right) A_{i,r}(j,l)\right| \leq \frac{|s|}{n} \sum_l \sum_r C_{r,l} |A_{i,r}(j,l)|$$

First of all note that $|A_{i,r}(j,l)| \leq 2L_{l+r}$. Thus, using that $\sum_l \sum_r C_{r,l} L_{l+r}^2 < \infty$ and that $\sum_{s=-M}^{M} |s| = (M+1)M$ we have that

$$\frac{1}{2M+1} \sum_{s=-M}^{M} \left|\sum_l \sum_r \left(S_{r,l}\left(z+\frac{s}{n}\right) - S_{r,l}(z)\right) A_{i,r}(j,l)\right| = O\left(\frac{M}{n}\right)$$

3.8.12 Proof of Theorem 3.8

Consider the sequence of second-order wavelet functions and the associated sequence of EWS coefficients

$$\Phi_r = \left\{\left\{\Phi_{i,j}(\tau)\right\}_{i=\max(-q,-\log_2(n)-j)}^{-1}\right\}_{j=-r\log_2(n)+q(r-1)}^{(r-1)(-\log_2(n)+q)-1}$$

$$S_r(z) = \left\{\left\{S_{i,j}(z)\right\}_{i=\max(-q,-\log_2(n)-j)}^{-1}\right\}_{j=-r\log_2(n)+q(r-1)}^{(r-1)(-\log_2(n)+q)-1}$$

with $r = 1, 2, \ldots$. Denote as $\left\{A_{1,r}\right\}_r$ the sequence of matrices containing the cross-inner products between the elements of Φ_1 and Φ_r. Denote as $I_{k,M} = \left\{\left\{I_{i,j,k,M}\right\}_{i=\max(-q,-\log_2(n)-j)}^{-1}\right\}_{j=-\log_2(n)}^{-1}$ the vector of smoothed periodogram estimators at time point k. Given the statement of Theorem 3.7 we can write the expected value of $I_{k,M}$ as

$$E\left(I_{k,M}\right) = \sum_{r=1}^{+\infty} A_{1,r} S_r(z) + O\left(\frac{M}{n}\right)$$

The expected value of the smoothed and corrected wavelet periodogram then takes the form

$$\begin{aligned} E\left(I_{k,M}\right) &= A_{1,1}^{-1}\left(\sum_{r=1}^{+\infty} A_{1,r} S_r + O\left(\tfrac{M}{n}\right)\right) \\ &= S_1 + A_{1,1}^{-1} \sum_{r=2}^{+\infty} A_{1,r} S_r + A_{1,1}^{-1} O\left(\tfrac{M}{n}\right) \end{aligned}$$

Thus, the L_2 norm of the vector $E\left(I_{k,M}\right) - S_1$ can be bounded from above by

$$\sum_{r=2}^{+\infty} ||A_{1,1}^{-1} A_{1,r}||_2 ||S_r||_2 + q||A_{1,1}^{-1}||_2 O\left(\frac{M \log_2(n)}{n}\right)$$

Consider a generic constant such that $||A_{1,1}^{-1} A_{1,r}||_2 \leq c$ for all r which allows to write the upper bound as

$$c \sum_{r=2}^{+\infty} ||S_r||_2 + \delta^{-1} O\left(\frac{M \log_2(n) q}{n}\right)$$

where δ denotes the absolute lower bound of the eigenvalues of $A_{1,1}$. Given the assumption that $|S_{i,j}(z)| \leq \frac{K}{L_{j+i}}$ with $L_{j-1+i} = 2L_{j+i}$ we have that $||S_r||_2 \leq q \log_2(n) \left(\frac{2^q}{n}\right)^{r-1}$ for $r = 2, 3, \ldots$ and hence $\sum_{r=2}^{+\infty} ||S_r||_2 = q \log_2(n) \frac{\frac{2^q}{n}}{1-\frac{2^q}{n}}$. Thus, we obtain that

$$||E\left(I_{k,M}\right) - S_1||_2 \leq \delta^{-1} O\left(\frac{M \log_2(n) q}{n}\right)$$

3.8.13 Proof of Theorem 3.9

Bound the variance of the smoothed and corrected wavelet periodogram as

$$\mathrm{Var}\left(\tilde{I}_{i,j,k,M}\right) \leq \sum_{r=-q}^{0}\sum_{l=-J}^{-1}\left|A_{i,r}^{-1}(j,l)\right|\sqrt{\mathrm{Var}\left(I_{r,l,k,M}\right)}\sum_{r'=-q}^{0}\sum_{l'=-J}^{-1}\left|A_{i,r'}^{-1}(j,l')\right|\sqrt{\mathrm{Var}\left(I_{r',l',k,M}\right)}$$

Using Theorem 3.6, this can be written as

$$\begin{aligned}\mathrm{Var}\left(\tilde{I}_{i,j,k,M}\right) \leq \quad & \tfrac{2}{M}\textstyle\sum_{r=0}^{-q}\sum_{l=-J}^{-1}\left|A_{i,r}^{-1}(j,l)\right|\sqrt{B_{r,l}}\sum_{r'=-q}^{0}\sum_{l'=-J}^{-1}\left|A_{i,r'}^{-1}(j,l')\right|\sqrt{B_{r',l'}}\\ & +O\left(\tfrac{1}{n}\right)\textstyle\sum_{r=0}^{-q}\sum_{l=-J}^{-1}\left|A_{i,r}^{-1}(j,l)\right|\sum_{r'=0}^{-q}\sum_{l'=-J}^{-1}\left|A_{i,r}^{-1}(j,l')\right|\end{aligned}$$

Given the statement of Theorem 3.6 we have that $\sqrt{B_{r,l}} = O\left(\frac{lq\sqrt{L_l}}{\sqrt{M}}\right)$. Moreover, recall that $L_l = \left(2^{-l}-1\right)\left(L[h]-1\right)$. Hence, if $\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}\left|A_{i,r}(j,l)^{-1}\right| \leq C2^{\frac{j}{2}}$ and if we set $J=\log_2(n)$, we have that $\mathrm{Var}\left(\tilde{I}_{i,j,k,M}\right) = O\left(\frac{2^j\log_2(n)q}{M}\right)$.

Chapter 4

Multivariate locally stationary wavelet processes

In the previous chapter we have focused on the modelling of the time-varying autocovariance function of univariate time series. In this section these univariate time series are considered to be components of a vector-valued process in the context of a multivariate time series model. Vector-valued stochastic processes are defined as a sequence of random vectors with an inherent order according to a time index, that is

$$X_{n,p} = \left\{X_p(t)\right\}_{t=0}^{n-1}$$

with $X_p(t) = (X^{(1)}(t), \ldots, X^{(p)}(t))$. In contrast to univariate time series modelling, multivariate time series analysis does not merely focus on the modelling of the autocovariance function of each of the constituent processes but also on the representation of the dynamic interdependence between them. The analysis of multivariate time series data is of crucial importance in many different fields in finance or economics. For example, understanding the co-movements of the returns of large numbers of assets is the key for portfolio allocation and risk management. In macroeconomics, panel time series models allow to study economic processes while accounting for dynamic effects that are not visible in cross sections. Moreover, understanding the pattern of dependence between the series of the vector-valued process also helps to better understand the behavior of each of the component series leading to more adequate forecasting results. Finally, multivariate stochastic processes are one of the main building blocks of dynamic factor modeling approaches which will be dealt with in Chapter 5.

As in the univariate case, we focus on the modelling of the time-varying aspects of the second-order structure of the vector-valued process. Priestley and Tong (1973) [105] propose an extension of the theory of evolutionary wavelet spectra to the case of bivariate non-stationary processes. Ombao et. al. (2005) [96] propose an approach to model multivariate non-stationary time series based on the smooth localized complex exponential (SLEX) transform, which forms a library of orthogonal complex-valued transforms that are simultaneously localized in time and frequency. Sanderson et al. (2010) [108] extend the univariate LSW model framework to bivariate processes by defining the notion of evolutionary wavelet cross-spectrum and proposing a locally stationary wavelet coherence measure. This bivariate LSW model framework has been extended to a multivariate setting by Park et al. (2014) [98] and Cho and Fryzlewicz (2015) [24].

Both, the bivariate approach defined in Sanderson et al. (2010) and the multivariate settings in Park et al. (2014) and Cho and Fryzlewicz (2015) [24], are based on the use of orthogonal vector-valued increment processes. As we are going to show in Section 4.1.4, the use of these orthogonal increment processes leads to a highly restrictive assumption on the form of the autocovariance function by implying that the autocovariance function of the time series at the macro-scale is such that

$$\mathrm{Cov}\left(X_n^u(t), X_n^v(t+\tau)\right) = \mathrm{Cov}\left(X_n^v(t+\tau), X_n^u(t)\right)$$

for $t = 0, \ldots, n-1$ $\tau \in \mathbb{Z}$, $u = 1, \ldots, p$ and $v = 1, \ldots, p$. However, these symmetry constraints might actually never be fulfilled in any multivariate time series relevant for practical purposes which puts a serious limitation on the use of the existing MLSW process models.

In this chapter we extend the LSW(q) process model, introduced in Section 3.4 to the multivariate setting. In the multivariate framework, each of the p constituent time series is associated with a specific increment process, denoted as $\{\varepsilon_{j,k}^{(u)}\}_{j,k}$ with $u = 1, \ldots, p$. In the existing multivariate LSW frameworks, the elements of two separate increment processes are independent from each other whenever they belong to different wavelet scales which means that $\mathrm{Cov}(\varepsilon_{j,k}^{(u)}, \varepsilon_{j+i,k}^{(v)}) = \Gamma_j^{(u,v)}(z)$ and $\mathrm{Cov}\left(\varepsilon_{j,k}^{(u)}, \varepsilon_{j+i,l}^{(v)}\right) = 0$ for each $i \leq -1$ and $k, l = 0, \ldots, n-1$. We allow for a more flexible correlation pattern between the constituent increment processes such that

$$\mathrm{Cov}\left(\varepsilon_{j,k}^{(u)},\varepsilon_{j+i,l}^{(u)}\right)=\begin{cases}\Gamma_{i,j}^{(u)}\left(\frac{k}{n}\right) & \text{if } -q\leq i\leq 0 \text{ and } k=l\\ 0 & \text{if } i<-q \text{ or } k\neq l\end{cases}$$

with $j=-1,-2,\ldots$, $i=-1,-2,\ldots$. We show that the insertion of a correlation pattern between the increment processes at the different wavelet scales allows to get rid of the symmetry constraints of the autocovariance function of the existing MLSW processes. The chapter is divided into four sections. In Section we introduce the notion of MLSW processes of order q. Section 4.2 proposes a consistent estimation procedure for the evolutionary wavelet spectrum of this new type of processes. In Section 4.3 we extend the MLSW(q) framework to high-dimensional settings in which $p=O(n^{\alpha})$. Section 4.4 contains an application of the MLWS(q) process model.

4.1 Multivariate locally stationary wavelet processes of order q

Definition 4.1 below extends Definition 3.2 into the multivariate setting.

Definition 4.1. *A sequence of doubly-indexed stochastic processes $\{(X_n^{(1)}(t),\ldots,X_n^{(p)}(t))^t\}_{t=0}^{n-1}$, $n>0$, with zero-mean elements is in the class of multivariate locally stationary wavelet processes of order q if there exists a representation*

$$X_n^{(u)}(t)=\sum_{j=-\infty}^{-1}\sum_{k=0}^{n-1}W_j^{(u)}(z)\,\varepsilon_{j,k}^{(u)}\,\psi_{j,k}(t). \tag{4.1.1}$$

The collection $\{\psi_{j,k}\}_{j,k}$, $j=-1,-2,\ldots$, $k=0,\ldots,n-1$ is a set of discrete non-decimated wavelet functions. The ratio $z=\frac{k}{n}$ denotes rescaled time. The increment terms $\varepsilon_{j,k}^{(u)}$ are Gaussian random variables such that $E(\varepsilon_{j,k}^{(u)})=0$, $Var(\varepsilon_{j,k}^{(u)})=1$ and

$$Cov\left(\varepsilon_{j,k}^{(u)},\varepsilon_{j+i,l}^{(v)}\right)=\begin{cases}\Gamma_{i,j}^{(u,v)}(z) & \text{if } -q\leq i\leq 0 \text{ and } k=l\\ 0 & \text{if } i<-q \text{ or } k\neq l\end{cases}$$

with $j = -1, -2, \ldots$, $i = 0, -1, \ldots$, $u = 1, \ldots, p$ and $v = 1, \ldots, p$. The EWS functions $S_{i,j}^{(u,v)}(z) = W_j^{(u)}(z) W_{j+i}^{(v)}(z) \Gamma_{i,j}^{(u,v)}$ are Lipschitz continuous on $(0,1)$ with Lipschitz constants $C_{i,j}^{(u,v)}$ satisfying

$$\sum_{i=-q}^{0} \sum_{j=-\infty}^{-1} L_{j+i}^2 C_{i,j}^{(u,v)} < \infty$$

where L_{j+i} denotes the support of the wavelet functions at level $-j-i$.

According to Definition 4.1, MLSW processes take the form of a collection of univariate LSW processes as defined in Definition 3.2. The correlations between the multiscale increment processes of the individual LSW models lead to the correlations between the elements of $\{X_{n,p}^{(u)}(t)\}_{u=1}^{p}$ and $\{X_{n,p}^{(v)}(t+\tau)\}_{u=1}^{p}$ for all $u, v = 1, \ldots, p$. Note that as in Sanderson et al. (2010) [108], Park et al. (2014) [98] and Cho and Fryzlewicz [24], we suppose that the collection of non-decimated wavelet functions is identical for all constituent univariate LSW processes.

4.1.1 Autocovariance representation

If we use the rescaled time notation we can write the covariance between $X_n^{(u)}(t)$ and $X_n^{(v)}(t+\tau)$ as $C_n^{(u)}(\tau, z) = \text{Cov}[X_n^{(u)}(nz), X_n^{(v)}(nz+\tau)]$ where $z = \frac{t}{n}$. The local autocovariance function of the MLSW(q) model is given by

$$\begin{aligned} C^{(u,v)}(\tau, z) &= \sum_{j=-\infty}^{-1} S_{0,j}^{(u,v)}(z) \Psi_{0,j}(\tau) + \sum_{i=-q}^{-1} \sum_{j=-\infty}^{-1} S_{i,j}^{(u,v)}(z) \Psi_{i,j}(\tau) \\ &\quad + \sum_{i=-q}^{-1} \sum_{j=-\infty}^{-1} S_{i,j}^{(v,u)}(z) \Psi_{-i,j+i}(\tau) \end{aligned} \tag{4.1.2}$$

$$z \in (0,1); \quad \tau \in \mathbb{Z}; \quad u = 1, \ldots, p; \quad v = 1, \ldots, p$$

where $\{\Psi_{0,j}, \{\Psi_{i,j}, \Psi_{-i,j+i}\}_{i=-q}^{-1}\}_{j=-1}^{-\infty}$ are families of cross-correlation wavelets (see Section 2.2). For the sake of notation simplicity, we denote $\{\{\Phi_{i,j}(\tau) = \Psi_{i,j}(\tau)\}_{i=-q}^{0}\}_{j=-\infty}^{-1}$ and $\{\{\Phi_{-i,j}(\tau) = \Psi_{-i,j+i}(\tau)\}_{i=-q}^{-1}\}_{j=-\infty}^{-1}$ and write the family of cross-correlation functions as

$$\Phi_q = \left\{ \left\{ \Phi_{i,j}(\tau) \right\}_{i=-q}^{q} \right\}_{j=-\infty}^{-1} \tag{4.1.3}$$

Similarly, we set $\{\{S^{(u,v)}_{i,j}(z) = S^{(u,v)}_{i,j}(z)\}^{0}_{i=-q}\}^{-1}_{j=-\infty}$ and $\{\{S^{(u,v)}_{-i,j}(z) = S^{(u,v)}_{-i,j+i}(z)\}^{-1}_{i=-q}\}^{-1}_{j=-\infty}$ which allows to write the LACF in (4.1.2) as

$$C^{(u,v)}(\tau,z) = \sum_{i=-q}^{q}\sum_{j=-\infty}^{-1} S^{(u,v)}_{i,j}(z)\Phi_{i,j}(\tau) \tag{4.1.4}$$

Given that $S^{(u,u)}_{i,j} = S^{(u,u)}_{-i,j}$, we obtain

$$C^{(u,u)}(\tau,z) = \sum_{j=-\infty}^{-1} S^{(u,u)}_{0,j}(z)\Phi_{0,j}(\tau) + \sum_{i=-q}^{-1}\sum_{j=-\infty}^{-1} S^{(u,u)}_{i,j}(z)\left(\Phi_{i,j}(\tau)+\Phi_{-i,j}(\tau)\right)$$

which corresponds to the local autocovariance function of univariate LSW(q) processes in (3.4.2).

Proposition 4.1. *Given the assumptions in Definition 4.1 and if $n \to \infty$ we have that $C^{(u,v)}_n(\tau,z) = Cov\left[X^{(u)}_n(nz), X^{(v)}_n(nz+\tau)\right]$ converges to the local autocovariance function, given that*

$$\sum_{\tau=-\infty}^{+\infty}\int_0^1 |C^{(u,v)}_n(\tau,z) - C^{(u,v)}(\tau,z)|dz = O\left(\frac{1}{n}\right). \tag{4.1.5}$$

4.1.2 Uniqueness of the autocovariance representation

Define the operator $A := (A_{i,r}(j,l))_{j,l<0}$ such that

$$A_{i,r}(j,l) = \sum_{\tau\in\mathbb{Z}} \Phi_{i,j}(\tau)\Phi_{r,l}(\tau)$$

and define $A_q := \{A_{i,r}\}_{-q\leq i,r\leq q}$ as the inner product operator of the family of functions in (4.1.3). Denote as $A_{q,J}$ the Gramian matrix of the sub-family of functions

$$\Phi^{(J)}_q = \left\{\left\{\Phi_{i,j}\right\}^{\min(J+j,q)}_{i=\max(-J-j,-q)}\right\}^{-1}_{j=-J}. \tag{4.1.6}$$

such that

$$A^{-1}_{q,J} = \left(\Phi^{(J)}_q\right)^t \Phi^{(J)}_q \tag{4.1.7}$$

According to Theorem 2.1, the family of CCWF Φ^J_q is linearly independent.

Thus, we have that the operator norm of $A_{q,J}^{-1}$ is bounded from above by some constant $\delta_{q,J} > 0$ for $J \geq 1$ and $q \geq 0$ which implies that A is an invertible operator. Hence, the EWS can be written as the inverse representation of the local autocovariance function that is

$$S_{q,i,j}^{(u,v)}(z) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{q} A_{i,r}^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Phi_{r,l}(\tau) C^{(u,v)}(\tau,z). \tag{4.1.8}$$

Denote as S_n the inverse representation of the autocovariance function $C_n(\tau,z)$

$$S_{q,n,i,j}^{(u,v)}(z) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{q} A_{i,r}^{-1}(j,l) \sum_{\tau \in \mathbb{Z}} \Phi_{r,l}(\tau) C_n^{(u,v)}(\tau,z). \tag{4.1.9}$$

Proposition 3.4 below states that for $S_{q,n,i,j}(z)$ to converge to the EWS, there needs to exists an absolute lower bound $\delta > 0$ such that $\delta \leq \lambda_{\min}(A_{q,J})$ where $\lambda_{\min}(A_{q,J})$ denotes the smallest eigenvalue of the Gramian matrix $A_{q,J}$.

Proposition 4.2. *Consider the Gramian matrix in 4.1.7 and suppose that $q < +\infty$. If there exists $\delta > 0$ such that for the smallest eigenvalue of $A_{q,J}$, denoted as $\lambda_{\min}(A_{q,J})$, we have $\lambda_{\min}(A_{q,J}) \geq \delta$ for $J \geq 1$ and if $\sum_{i=-q}^{q} \sum_{j=-\infty}^{-1} |A_{i,r}^{-1}(j,l)| < +\infty$, we obtain that*

$$\int_0^1 \left| S_{q,n,i,j}^{(u,v)}(z) - S_{q,i,j}^{(u,v)}(z) \right| dz = O\left(\frac{1}{n}\right)$$

As in Section 3.4, we are going to study some classes of MLSW(q) processes in more details.

4.1.3 MLSW(∞) processes

MLSW(∞) processes allow for correlations between the increments of all the wavelets scales and hence

$$\operatorname{Cov}\left(\varepsilon_{j,k}^{(u)}, \varepsilon_{j+i,l}^{(v)}\right) = \begin{cases} \Gamma_{i,j}^{(u,v)}(z) & \text{if } i \leq 0 \text{ and } k = l \\ 0 & \text{if } k \neq l \end{cases}$$

for $j = -1, -2, \ldots$, $i = -1, -2, \ldots$, $u = 1, \ldots, p$ and $v = 1, \ldots, p$.

Theorem 4.1. *Consider the Gramian matrix in* (4.1.7)*. If we let J depend on n as $J(n) = \log_2(n)$ and set $q(n) = \log_2(n) - 1$, we have that for Haar wavelet functions $\lambda_{\min}(A_{q(n),J(n)}) = O\left(\frac{1}{n}\right)$.*

Thus, using the statements of Proposition 4.2 and of Theorem 4.1, we can conclude that the autocovariance representation of the MLSW(0) process model is not unique. Thus, as for the univariate case, the degree of inter-scale correlation within the increment process cannot increase along with the sample size n.

4.1.4 MLSW(0) processes

MLSW(0) processes only allow for correlations between the increments within each wavelet scale which means that

$$\mathrm{Cov}\left(\varepsilon_{j,k}^{(u)}, \varepsilon_{j+i,l}^{(v)}\right) = \begin{cases} \Gamma_{i,j}^{(u,v)}(z) & \text{if } i = 0 \text{ and } k = l \\ 0 & \text{if } i < 0 \text{ or } k \neq l \end{cases}$$

for $j = -1, -2, \ldots$, $i = -1, -2, \ldots$, $u = 1, \ldots, p$ and $v = 1, \ldots, p$.

According to the representation in (4.1.2), the local autocovariance function of the MLSW(0) model is given by

$$C_n^{(u,v)}(\tau, z) = \sum_{j=-\infty}^{-1} S_{0,j}^{(u,v)}(z) \Psi_{0,j}(\tau).$$

If $S_{0,j}^{(u,v)}(z) = S_{0,j}^{(v,u)}(z)$ for all $u = 1, \ldots, p$, $v = 1, \ldots, p$, and if the collections of non-decimated wavelet functions are identical for each process $u = 1, \ldots, p$, we have that

$$C_n^{(u,v)}(\tau, z) = C_n^{(v,u)}(\tau, z). \tag{4.1.10}$$

4.1.5 Discussion

The symmetry constraint of the MLSW(0) model is a direct result of the orthogonality assumption on the vector-valued increment process combined with the use of the same set of non-decimated wavelet functions for each of the constituent time series. There are two possible ways in order to get rid of the symmetry assumption, we could use different sets of wavelet functions for each constituent time series or allow for a more flexible correlation structure among the increment terms. Indeed, recall that $S_{i,j}^{(u,v)}(z) = \mathrm{Cov}(\varepsilon_{j,k}^{(u)}, \varepsilon_{j+i,k}^{(u)})$ and $S_{i,j}^{(v,u)}(z) = \mathrm{Cov}(\varepsilon_{j,k}^{(v)}, \varepsilon_{j+i,k}^{(u)})$ and consider the representation of the local autocovariance function in (4.1.2). Given that $S_{i,j}^{(u,v)}(z) \neq S_{i,j}^{(v,u)}(z)$ the local autocovariance representation of the corresponding MLSW(q) process is not symmetric.

For the sake of argument consider the spectral representation of a second-order stationary multivariate time series (see Brockwell and Davis, 2009) [20]

$$X_n^{(u)}(t) = \int\limits_{\omega} e^{i\omega t} dZ^{(u)}(\omega) \tag{4.1.11}$$

where $\{\{dZ^{(u)}(\omega)\}_{u=1}^{p}\}_{\omega}$ denotes a complex-valued multiscale increment process of dimension p. At first glance the representation in (4.1.11) might seem similar to the MLSW(0) process model. However, given that the increment process in (4.1.11) is complex-valued and that $e^{i\omega t} = \cos(\omega t) + i\sin(\omega t)$, we can write (4.1.11) as

$$X_n^{(u)}(t) = \int\limits_{\omega} \cos(\omega t) dZ_1^{(u)}(\omega) + i \int\limits_{\omega} \sin(\omega t) dZ_2^{(u)}(\omega) \tag{4.1.12}$$

Thus, $X_n^{(u)}(t)$ is a sum of two multiscale processes with two separate collections of increments $\{dZ_1^{(u)}(\omega)\}_{\omega}$ and $\{dZ_2^{(u)}(\omega)\}_{\omega}$ respectively. Based on the representation in (4.1.12) we can write the autocovariance functions of the process at the macro-scale as

$$\begin{aligned} C_n^{(u,v)}(\tau) &= \int\limits_{\omega} \cos(\omega t)\cos(\omega(t+\tau)) E\left[dZ_1^{(u)}(\omega) dZ_1^{(v)}(\omega)\right] \\ &+ \int\limits_{\omega} \sin(\omega t)\sin(\omega(t+\tau)) E\left[dZ_2^{(u)}(\omega) dZ_2^{(v)}(\omega)\right] \\ &- i\int\limits_{\omega} \cos(\omega t)\sin(\omega(t+\tau)) E\left[dZ_1^{(u)}(\omega) dZ_2^{(v)}(\omega)\right] \\ &+ i\int\limits_{\omega} \cos(\omega(t+\tau))\sin(\omega t) E\left[dZ_2^{(u)}(\omega) dZ_1^{(v)}(\omega)\right] \end{aligned} \tag{4.1.13}$$

If we fix $E[dZ_1^{(u)}(\omega)dZ_1^{(v)}(\omega)] = E[dZ_2^{(u)}(\omega)dZ_2^{(v)}(\omega)]$ and $E[dZ_1^{(u)}(\omega)dZ_2^{(v)}(\omega)] = E[dZ_2^{(u)}(\omega)dZ_1^{(v)}(\omega)]$ and given that $\cos(\omega\tau) = \cos(\omega t)\cos(\omega(t+\tau)) + \sin(\omega t)\sin(\omega(t+\tau))$, $\sin(\omega\tau) = \sin(\omega t)\cos(\omega(t+$

$\tau)) + \cos(\omega t)\sin(\omega(t+\tau))$ we can write (4.1.13) as

$$C_n^{(u,v)}(\tau) = \int\limits_{\omega} E\left[dZ_1^{(u)}(\omega)dZ_1^{(v)}(\omega)\right]\cos(\omega\tau) - i\int\limits_{\omega} E\left[dZ_1^{(u)}(\omega)dZ_2^{(v)}(\omega)\right]\sin(\omega\tau) \tag{4.1.14}$$

Using the same type of reasoning we can show that

$$C_n^{(v,u)}(\tau) = \int\limits_{\omega} E\left[dZ_1^{(u)}(\omega)dZ_1^{(v)}(\omega)\right]\cos(\omega\tau) + i\int\limits_{\omega} E\left[dZ_1^{(u)}(\omega)dZ_2^{(v)}(\omega)\right]\sin(\omega\tau) \tag{4.1.15}$$

Thus, the covariance matrix of the stationary time series is a Hermitian matrix.

If there was no intra-frequency correlation between the two increment processes, that is $E\left[dZ_1^{(u)}(\omega)dZ_2^{(v)}(\omega)\right] = 0$ we obtain that

$$C_n^{(u,v)}(\tau) = C_n^{(v,u)}(\tau) = \int\limits_{\omega} E\left[dZ_1^{(u)}(\omega)dZ_1^{(v)}(\omega)\right]\cos(\omega(\tau))$$

and hence $C_n^{(u,v)}(\tau) = C_n^{(v,u)}(\tau)$. Hence, linking the two multiscale representation models in (4.1.12) is crucial to obtain a non-symmetric autocovariance function.

Finally, note that the order of the MLSW model has an impact on the eigenspectrum of the Gramian matrix of the collection of functions in (4.1.6). Table 4.1 contains the values of the smallest eigenvalues of A_q for MLSW models of order zero to five.

q	0	1	2	3	4	5
$\lambda_{(\min)}(A_q)$	0.6925	0.06779	0.0128	0.0027	0.0006	0.0001

Table 4.1: ***Smallest eigenvalues of*** A_q ***for MLSW***(q) ***models with*** $q = 0, \ldots, 5$

The order of the increment process increases the level of model complexity. Thus, an increase in q leads to a decrease in the lowest eigenvalue of the corresponding inner product operator A_q.

4.2 Estimation of the MLSW(q) model

The estimation procedure of the EWS of LSW(q) models in the univariate and the multivariate framework are similar. Thus, as in the univariate setting (see Section 3.5), we start off with defining sets of wavelet coefficients and use the latter to define the raw wavelet periodogram. We then generate a smoothed version of the raw wavelet periodogram by taking moving averages of the latter within each of its layers. Finally, we correct the smoothed wavelet periodogram to obtain a consistent estimator of the EWS. Most of the Theorems derived in this section are direct extensions of the corresponding Theorems in the univariate framework.

We first define the collection of wavelet coefficients

$$d^{(u)}_{j,k,n} = \sum_{t=0}^{n-1} \psi_{j,k}(t) X^{(u)}(t) \tag{4.2.1}$$

with $j = -1, \dots, -J$, $k = 0, \dots, n-1$ and $u = 1, \dots, p$.

4.2.1 Raw wavelet periodogram

The raw wavelet periodogram of the MLSW(q) model consists in a set of random variables denoted as $I_{i,j,k,n}$ with $j = -1, \dots, -J$, $i = -1, \dots, \max(-J - j, -q)$, $k = 0, \dots, n-1$ and $u, v = 1, \dots, p$, such that for $u \neq v$

$$I^{(u,v)}_{i,j,k,n} = d^{(u)}_{j,k,n} d^{(v)}_{j+i,k,n}. \tag{4.2.2}$$

Theorem 4.2 below states that the raw wavelet periodogram is a biased estimator of the EWS of the MLSW(q) process model.

Theorem 4.2. *Given the assumptions in Definition 4.1, the expectation of the raw wavelet periodogram is*

$$E\left(I^{(u,v)}_{i,j,k,n}\right) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{q} S^{(u,v)}_{r,l}(z) A_{i,r}(j,l) + O\left(\frac{1}{n}\right) \tag{4.2.3}$$

for $j = -1, \dots, -J$, $i = -1, \dots, \max(-J - j, -q)$, $k = 0, \dots, n-1$ *and* $u, v = 1, \dots, p$ *with* $u \neq v$, $J \geq 0$ *and* $q \geq 0$.

Theorem 4.3. *Given the assumptions in Definition 4.1, the variance of the raw wavelet periodogram is*

$$\begin{aligned} Var\left(I_{i,j,k}^{(u,v)}\right) &= \tfrac{1}{4}\left(\sum_r \sum_l S_{r,l}^{(u,v)}(z) A_{i,r}(j,l)\right)^2 \\ &+ \sum_r \sum_l S_{r,l}^{(u,v)}(z) A_{0,r}(j,l) \sum_r \sum_l S_{r,l}^{(u,v)}(z) A_{0,r}(j+i,l) + O\left(\tfrac{1}{n}\right) \end{aligned}$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$, $k = 0, \ldots, n-1$ *and* $u, v = 1, \ldots, p$ *with* $J \geq 0$ *and* $q \geq 0$.

4.2.2 Smoothed wavelet periodogram

As in the univariate setting, we define the smoothed wavelet periodogram estimators as moving averages of the elements of the raw wavelet periodogram within each layer, that is we consider the sequence of random variables

$$I_{i,j,k,n,M}^{(u,v)} = \tfrac{1}{2M+1} \textstyle\sum_{s=-M}^{M} I_{i,j,k+s,n}^{(u,v)} \quad \text{with } k = 0, \ldots, n-1$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$, $k = 0, \ldots, n-1$ and $u, v = 1, \ldots, p$ with $u \neq v$.

Theorem 4.4. *Given the assumptions in Definition 4.1 and if* $|S_{i,j}^{(u,v)}(z)| \leq \frac{K}{L_{j+i}}$ *for any* $i = 0, \ldots, -q$ *with* $q < +\infty$, $j = -1, -2, \ldots$ *and* $u, v = 1, \ldots, p$, *the variance of the smoothed wavelet periodogram is*

$$Var\left(I_{i,j,k,n,M}^{(u,v)}\right) = O\left(\frac{2^{-j} j^2}{M}\right)$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$, $k = 0, \ldots, n-1$ *and* $u, v = 1, \ldots, p$ *with* $J \geq 0$ *and* $q \geq 0$.

Theorem 4.5. *Given the assumptions in Definition 4.1, the expected value of the smoothed wavelet periodogram is*

$$E\left(I_{i,j,k,n,M}^{(u,v)}\right) = E\left(I_{i,j,k,n}^{(u,v)}\right) + O\left(\frac{M}{n}\right)$$

for $j = -1, \ldots, -J$, $i = -1, \ldots, \max(-J-j, -q)$, $k = 0, \ldots, n-1$ *and* $u, v = 1, \ldots, p$ *with* $u \neq v$, $J \geq 0$ *and* $0 \leq q < +\infty$.

4.2.3 Smoothed and corrected wavelet periodogram

The smoothed and corrected wavelet periodogram of the MLSW(q) model is defined as

$$\tilde{I}^{(u,v)}_{i,j,k,n,M} = \sum_{j=-J}^{-1} \sum_{i=-q}^{q} A^{-1}_{i,r}(j,l) I^{(u,v)}_{r,l,k,M}$$

with $j = -1, \dots, -J$, $i = 0, \dots, \max(-J-j, -q)$, $k = 0, \dots, n-1$ and $u, v = 1, \dots, p$.

Theorem 4.6. *Given the assumptions in Definition 4.1 and if* $|S^{(u,v)}_{i,j}(z)| \leq \frac{K}{L_{j+i}}$ *for any* $i = 0, \dots, -q$ *with* $q < +\infty$, $j = -1, -2, \dots$ *and* $u, v = 1, \dots, p$, *and provided that there exists* $\delta > 0$ *such that* $\lambda_{\min}(A_{q,J}) \geq \delta$ *and that* $\sum_{r=-q}^{0} \sum_{l=-\infty}^{-1} \left|A_{i,r}(j,l)^{-1}\right| \leq C2^{\frac{j}{2}}$, *we have that*

$$E\left(\tilde{I}^{(u,v)}_{i,j,k,n,M}\right) = S^{(u,v)}_{i,j}(z) + O\left(\frac{M \log_2(n)}{n}\right)$$

for $j = -1, \dots, -J$, $i = -1, \dots, \max(-J-j, -q)$, $k = 0, \dots, n-1$ *and* u, v $= 1, \dots, p$ *with* $J \geq 0$ *and* $q \geq 0$.

Theorem 4.7. *Given the assumptions in Definition 4.1 and if* $|S^{(u,v)}_{i,j}(z)| \leq \frac{K}{L_{j+i}}$ *for any* $i = 0, \dots, -q$ *with* $q < +\infty$, $j = -1, -2, \dots$ *and* $u, v = 1, \dots, p$, *and provided that there exists* $\delta > 0$ *such that* $\lambda_{\min}(A_{q,J}) \geq \delta$ *and that* $\sum_{r=-q}^{0} \sum_{l=-\infty}^{-1} \left|A_{i,r}(j,l)^{-1}\right| \leq C2^{\frac{j}{2}}$, *we have that*

$$Var\left(\tilde{I}^{(u,v)}_{i,j,k,n,M}\right) = O\left(\frac{2^j \log_2(n)}{M}\right)$$

for $j = -1, \dots, -J$, $i = -1, \dots, \max(-J-j, -q)$, $k = 0, \dots, n-1$ *and* u, v $= 1, \dots, p$ *with* $J \geq 0$ *and* $q \geq 0$.

As for the univariate LSW model we can derive the optimal convergence rate of the smoothed and corrected wavelet periodogram by first setting $M = O\left(n^{\beta}\right)$. Given the statements of Theorems 3.8 and 3.9, the mean squared error of the smoothed and corrected wavelet periodogram at each layer is given by

$$\text{MSE}\left(\tilde{I}^{(u,v)}_{i,j,k,n,M}\right) = O\left(\log_2^2(n)\left(\frac{1}{n^{2(1-\beta)}} + \frac{2^j}{n^{\beta}\log_2(n)}\right)\right). \qquad (4.2.4)$$

Hence, the optimal convergence rate of the MSE can be obtained by solving $\max(\min(2(1-\beta),\beta))$ with respect to β leading to $\beta = \frac{2}{3}$ in which case

$$\mathrm{MSE}\left(\tilde{I}^{(u,v)}_{i,j,k,n,M}\right) = O\left(\frac{\log_2^2(n)}{n^{\frac{2}{3}}}\right)$$

Using Markov's inequality, we have that for all $\varepsilon > 0$

$$P\left(|\tilde{I}^{(u,v)}_{i,j,k,n,M} - S^{(u,v)}_{i,j}(z)| > \varepsilon\right) \leq \frac{\mathrm{MSE}\left(\tilde{I}^{(u,v)}_{i,j,k,n,M}\right)}{\varepsilon}$$

which, given the result in , proves convergence in probability of the smoothed and corrected wavelet periodogram.

4.3 High-dimensional MLSW processes

In Definition 4.1 we have considered MLSW processes in which the dimension of the constituent random vectors is fixed. Note that this type of multivariate settings are based on the inherent assumption that the relevant variables for given analysis are known beforehand. However, in most cases the set of relevant variables needs to be derived a posteriori on the basis of some kind of data driven model selection procedure. In these settings the initial set of potentially relevant variables may be quite large. In this section we adapt the MLSW setting, introduced in Definition 4.1, to these high-dimensional cases by considering triply-indexed instead of doubly-indexed random sequences, and hence we consider stochastic processes such that $\{(X_n^{(1)}(t),\ldots,X_n^{(p)}(t))^t\}_{t=0}^{n-1}$, $n > 0$ and $p = O(n^{\alpha})$ with $\alpha > 0$. Note that given the fact that p grows along with the sample size, the estimation of triply-indexed MLSW processes requires a different type of asymptotic reasoning in which both, n and p, go to infinity. The objective of this section is to adapt the consistency result of the smoothed and corrected wavelet periodogram obtained in Section 4.2.3 to the multivariate modelling framework.

To start with, consider the random vectors $\varepsilon_{j,t} = (\varepsilon^{(1)}_{j,t},\ldots,\varepsilon^{(p)}_{j,t})^t$ with $j = -1,\ldots,-\log_2(n)$ and define the increment vector $\varepsilon_t = (\varepsilon_{-1,t},\ldots,\varepsilon_{-\log_2(n),t})^t$ and denote as $S(z) = E(\varepsilon_t\varepsilon_t^t)$ the EWS matrix. Define as $\tilde{I}_{n,M}(z)$ the estimator of $S(z)$ based on the smoothed and corrected wavelet periodogram described in Section 4.2.3 and define the random matrix

$$\Xi_{n,M}(z) = \tilde{I}_{n,M}(z) - S(z) \tag{4.3.1}$$

Theorem 4.8. *Denote as* $||\Xi_{n,M}(z)||_2$ *the spectral norm of the random matrix defined in* (4.3.1). *Given the statement of Theorem 4.7 and if we set* $p = O(n^{\alpha})$, *we then have that*

$$E\left(||\Xi_{n,M}(z)||_2^2\right) = O\left(p^2 \log_2^4(n)\left(\frac{M^2}{n^2} + \frac{1}{M\log_2(n)}\right)\right)$$

If the dimensionality of the process is fixed and under the assumption that $M = O(n^{\beta})$, we obtain that $E\left(||\Xi_{n,M}(z)||_2^2\right) = O\left(\log_2^4(n)\left(\frac{1}{n^{2(1-\beta)}} + \frac{1}{n^{\beta}\log_2(n)}\right)\right)$. If we set $\beta = \frac{2}{3}$, we get that

$$E\left(||\Xi_{n,M}(z)||_2^2\right) = O\left(\frac{\log_2^4(n)}{n^{\frac{2}{3}}}\right). \tag{4.3.2}$$

In the high-dimensional case, we assume that the dimensionality of the vector process is such that $p = O(n^{\alpha})$, which, given that $\beta = \frac{2}{3}$, leads to

$$E\left(||\Xi_{n,M}(z)||_2^2\right) = O\left(\frac{\log_2^4(n)}{n^{2\left(\frac{1}{3}-\alpha\right)}}\right). \tag{4.3.3}$$

Using Markov's inequality, we have that for each $\varepsilon > 0$

$$P\left(||\Xi_{n,M}(z)||_2 > \varepsilon\right) \leq \frac{E\left(||\Xi_{n,M}(z)||_2^2\right)}{\varepsilon}$$

Hence, given (4.3.3), we have that $||\Xi_{n,M}(z)||_2 \underset{p}{\rightarrow} 0$ given that $0 \leq \alpha < \frac{1}{3}$.

4.4 Real Data Analysis: Modelling the log-returns of the EURIBOR yield curve

In this section we use MLSW(q) processes to model the cross-covariance pattern between Euro Interbank Offered Rates (Euribors) with maturities 1 week, 1 month, 3 months, 6 months, 9 months, 12 months on the one hand and the aggregated nominal exchange rate between the Euro and the currencies of the nineteen leading trade partners of the Euro zone, forming the so-called EWK-19 group, on the other hand. We denote as $E^{(u)}(t)$ the value of the Euribor of a given maturity at time point t, with $u = 1,\dots,6$ and as $R(t)$ the value of the Euro exchange rate at time t. Figure 4.1 plots the Euribors between 21/07/2003 and 02/04/2015. The most striking feature is the steep decline of the Euribors from the third quarter 2009 on, which is mainly linked to the subsequent cuts in interest rates of the base rate by the European Central Bank (ECB). These subsequent cuts have been part of a major stimulus package designed to face the downturn of the economic activity within the Euro zone during and after the global financial crisis of 2008-2009. Another interesting observation is that before the financial crisis the yields belonging to different maturities are relatively close to each other, leading to an approximately flat shape of the Euribor yield curve. After the crisis however, the yield curve takes on a ”normal” shape, characterized by rising yields as maturity lengthens. Note however, that from the end of the fourth quarter 2013 the yield curve flattens again which may reflect market expectations of future declines of the Euribor rates. Figure 4.2 contains the aggregated exchange rate between the Euro and the respective currencies of the EWK-19 group. As for the Euribors there is a sharp decline in the exchange rate at the beginning of the financial crisis and more precisely between the third quarter 2007 until the start of the fourth quarter 2008. This sharp decrease is followed by a strong increase which lasts until the beginning of 2010. However, from the first quarter 2010 on we can observe a gradual decrease in the exchange rate up to December 2013. This period of gradual decrease is followed by a sharp decrease of the exchange rate between the end of 2013 and April 2014.

We denote the log-returns of the Euribors and the exchange rate as

$$\begin{aligned} X^{(u)}(t) &= \ln\left(E^{(u)}(t)\right) - \ln\left(E^{(u)}(t-1)\right) \quad \text{for } u = 1,\dots,6 \\ X^{(7)}(t) &= \ln\left(R^{(u)}(t)\right) - \ln\left(R^{(u)}(t-1)\right). \end{aligned}$$

We assume that the vector process of log-returns belongs t some class of MLSW(q) processes and write each element of the increment process as

$$\varepsilon_{j,k}^{(u)} = \sum_{m=1}^{J(n)p} \Delta_{m,j}^{(u)}(z) Z_{m,k}. \tag{4.4.1}$$

where $Z_{m,k}$ are independent Gaussian random variables with mean zero and unit variance. If we plug the representation in (4.4.1) into (4.1.1) we obtain

$$X_n^{(u)}(t) = \sum_{k=0}^{n-1} \sum_{m=1}^{J(n)p} Z_{m,k} \sum_{j=-\infty}^{-1} \Delta_{m,j}^{(u)}(z) \psi_{j,k}(t). \tag{4.4.2}$$

According to the representation in (4.4.2) the covariance between the random variables $X_n^{(u)}(t)$ and $Z_{m,k}$ is given by $\sum_{j=-\infty}^{-1} \Delta_{m,j}^{(u)}(z) \psi_{j,k}(t)$. Consider the random vectors $Y_{n,k}^{(u)} = \{\{X_n^{(u)}(t) Z_{m,k}\}_{m=1}^{J(n)p}\}_{t=0}^{n-1}$ with $u = 1, \ldots, 6$ and $Y_{n,k}^{(7)} = \{\{X_n^{(7)}(t) Z_{m,k}\}_{m=1}^{J(n)p}\}_{t=0}^{n-1}$ and define cosine similarity between the vectors $E(Y_{n,k}^{(u)})$ and $E(Y_{n,k}^{(7)})$ as

$$\begin{aligned} \mathscr{C}^{(u)}(z) &= \cos\left[\theta\left(E\left(Y_{n,k}^{(u)}\right), E\left(Y_{n,k}^{(7)}\right)\right)\right] \\ &= \frac{\sum_{t=0}^{n-1} \sum_{m=1}^{J(n)p} E\left(X_n^{(u)}(t) Z_{m,k}\right) E\left(X_n^{(7)}(t) Z_{m,k}\right)}{\left(\sum_{t=0}^{n-1} \sum_{m=1}^{J(n)p} E^2\left(X_n^{(u)}(t) Z_{m,k}\right)\right)^{\frac{1}{2}} \left(\sum_{t=0}^{n-1} \sum_{m=1}^{J(n)p} E^2\left(X_n^{(7)}(t) Z_{m,k}\right)\right)^{\frac{1}{2}}} \end{aligned} \tag{4.4.3}$$

with $u = 1, \ldots, 6$. Note that $\sum_{t=0}^{n-1} \sum_{m=1}^{J(n)p} E\left(X_n^{(u)}(t) Z_{m,k}\right) E\left(X_n^{(7)}(t) Z_{m,k}\right)$ can be written as

$$\sum_{m=1}^{J(n)p} \sum_{j=-J(n)}^{-1} \sum_{j'=-J(n)}^{-1} \Delta_{m,j}^{(u)}(z) \Delta_{m,j'}^{(7)}(z) \sum_{t=0}^{n-1} \psi_{j,k}(t) \psi_{j',k}(t).$$

Given the orthonormality of the elements of $\{\psi_{j,k}(t)\}_{j=-J(n)}^{-1}$ this becomes $\sum_{m=1}^{J(n)p} \sum_{j=-J(n)}^{-1} \Delta_{m,j}^{(u)}(z) \Delta_{m,j}^{(v)}(z)$. Using the same kind of reasoning for the terms in the denominator of the representation in (4.4.3) we can write

$$\mathscr{C}^{(u)}(z) = \frac{\sum_{m=1}^{J(n)p} \sum_{j=-J(n)}^{-1} \Delta_{m,j}^{(u)}(z) \Delta_{m,j}^{(7)}(z)}{\left(\sum_{m=1}^{J(n)p} \sum_{j=-J(n)}^{-1} \left(\Delta_{m,j}^{(u)}(z)\right)^2\right)^{\frac{1}{2}} \left(\sum_{m=1}^{J(n)p} \sum_{j=-J(n)}^{-1} \left(\Delta_{m,j}^{(7)}(z)\right)^2\right)^{\frac{1}{2}}} \tag{4.4.4}$$

with $u = 1, \ldots, 6$. Note that the vector $E(Y_{n,k}^{(u)})$ could be interpreted as a kind of impulse response function, given that the elements of the latter provide an indication of the output of a given time series at time point t when submitted to m random shocks at time k represented by the random variables $Z_{1,k},\dots,Z_{m,k}$. For a detailed discussion on impulse response functions see Lütkepohl (2008) [79] and Lütkepohl and Poskitt (1991) [80]. Consequently, the coefficient in represents the cosine of the angle between the impulse response functions of a given Euribor log-return time series and the Euro log-return time series.

Consider the corrected and smoothed periodogram vector $\tilde{I}_{n,M}(z)$, as defined in 4.3. Define as $\{\lambda_m\}_{m=1}^{J(n)p}$ the set of ordered eigenvalues and as $\{u_m\}_{m=1}^{J(n)p}$ the set of associated eigenvectors. We define the estimator of the consistency measure in (4.4) as

$$\mathscr{C}^{(u)}(z) = \frac{\sum_{m=1}^{J(n)p} \lambda_m\left(I_{n,M}(z)\right) \sum_{j=-J(n)}^{-1} u_{m,j,u}\left(I_{n,M}(z)\right) u_{m,j,7}\left(I_{n,M}(z)\right)}{\left(\sum_{m=1}^{J(n)p} \lambda_m\left(I_{n,M}(z)\right) \sum_{j=-J(n)}^{-1} \left(u_{m,j,u}^2\left(I_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}} \left(\sum_{m=1}^{J(n)p} \lambda_m\left(I_{n,M}(z)\right) \sum_{j=-J(n)}^{-1} \left(u_{m,j,7}^2\left(I_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}}} \tag{4.4.5}$$

For the real data analysis we use the time series of log-returns of the Euribors between from 21/07/2003 to 02/04/2015 and the log-returns of the Euro-EWK exchange rate for that same period. We assume that the vector process of log-returns is in the class of MLSW(1) processes with D8 wavelet basis. Figure 4.3 plots the estimate of the cosine similarity in (4.4.5) between the Euro-EWK exchange rate and each of the Euribors from 21/07/2003 to 02/04/2015.

Note first that at the start of the period of analysis, that is 2003 to 2004, the cosine similarities between the Euro exchange rate and the different Euribors are all relatively low and take on rather similar values. However, from the beginning of 2005 on, the cosine similarities start to increase apart from the one for the Euribor with maturity one month. Moreover, note the very strong increase of the cosine similarities for maturities 9 months and 12 months from the second quarter of 2007 on. If we compare two periods, July 2003 - December 2004 on the one hand and August 2012 - April 2015 on the other hand, we notice three key elements. First, the cosine similarities for maturities 3 months to 12 months take on much larger values in the second period than during the first one. Secondly, the cosine similarities for maturities 1 week and 1 month take approximately the same values within both periods. Thirdly, the differences between the cosine similarities of different maturities is much more important in the second period than in the first one.

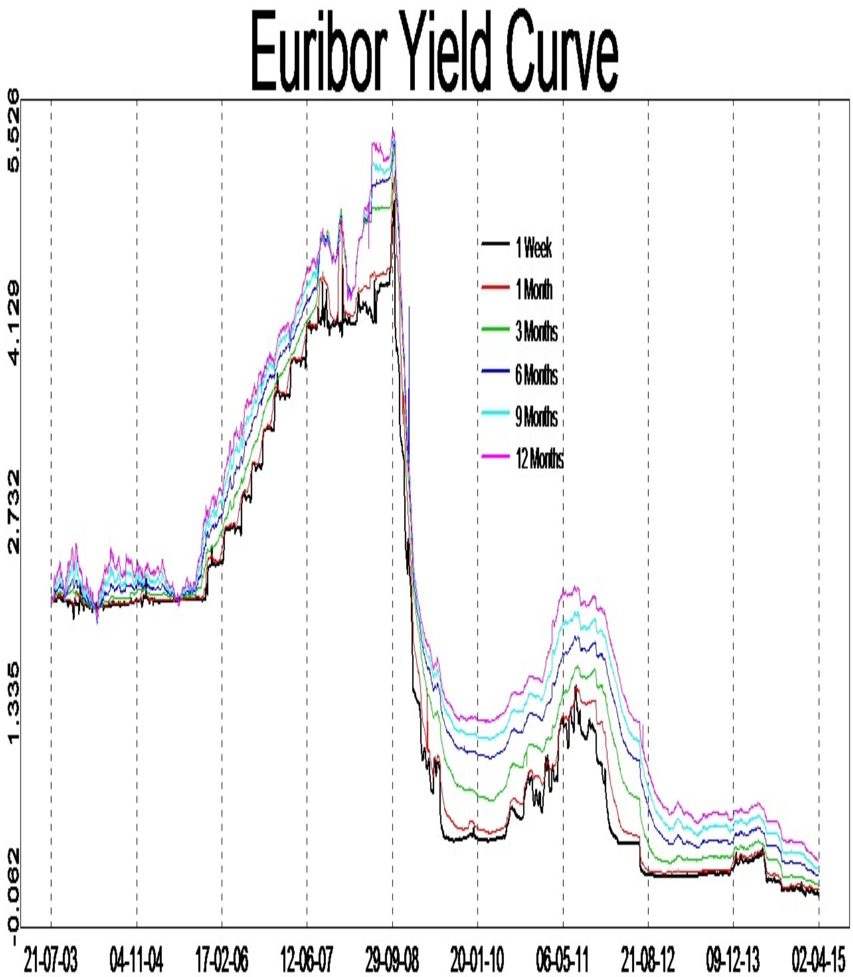


Figure 4.1: ***Euribors from 21/07/2003 to 02/04/2015***

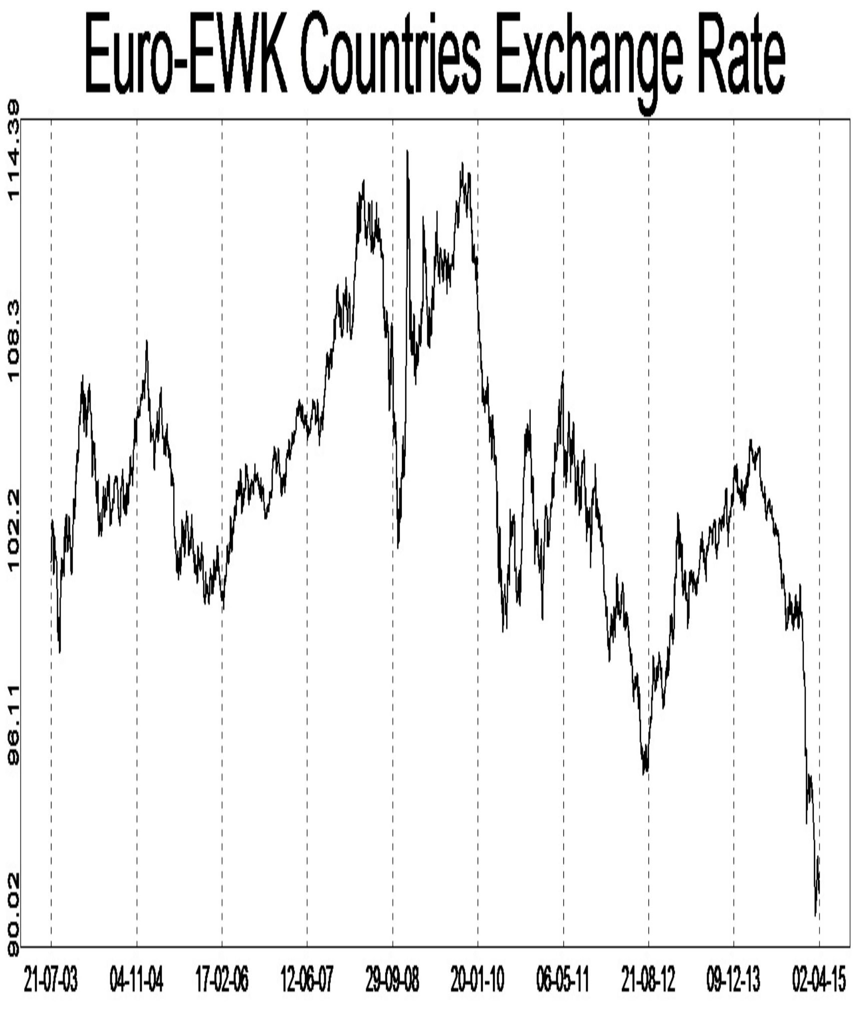


Figure 4.2: ***Euro-EWK exchange rate from 21/07/2003 to 02/04/2015***

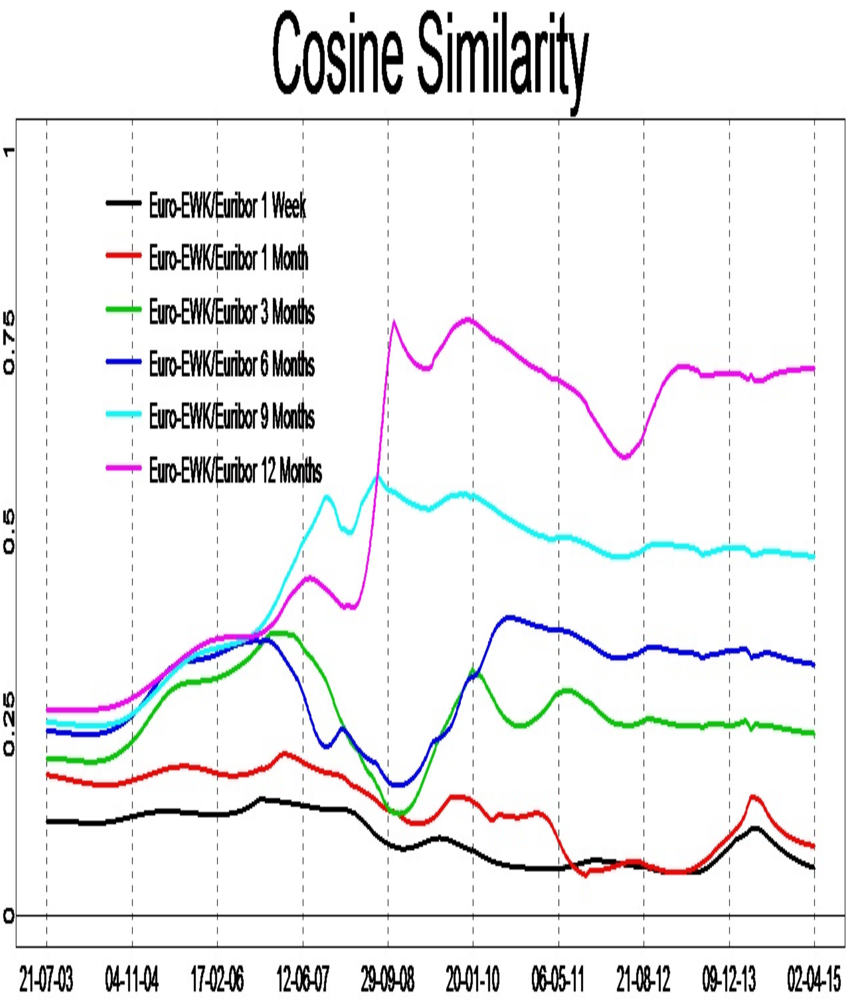


Figure 4.3: ***Cosine similarities from 21/07/2003 to 02/04/2015***

4.5 Proofs

4.5.1 Proof of Proposition 4.1

The covariance between $X^{(u)}(t)$ and $X^{(v)}(t+\tau)$ takes the form

$$\begin{aligned}C_n^{(u,v)}(\tau,z) &= \sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{0,j}^{(u,v)}\left(\tfrac{k}{n}\right)\psi_{j,k}(t)\psi_{j,k}(t+\tau)\\ &+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{i,j}^{(u,v)}\left(\tfrac{k}{n}\right)\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\\ &+\sum_{i=-1}^{-q}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty} S_{i,j}^{(v,u)}\left(\tfrac{k}{n}\right)\psi_{j+i,k}(t)\psi_{j,k}(t+\tau)\end{aligned}$$

where $S_{0,j}^{(u,v)}\left(\frac{k}{n}\right) = W_j^{(u)}\left(\frac{k}{n}\right)W_j^{(v)}\left(\frac{k}{n}\right)\Gamma_{0,j}^{(u,v)}\left(\frac{k}{n}\right)$, $S_{i,j}^{(u,v)}\left(\frac{k}{n}\right) = W_j^{(u)}\left(\frac{k}{n}\right) W_{j+i}^{(v)}\left(\frac{k}{n}\right)\Gamma_{i,j}^{(u,v)}\left(\frac{k}{n}\right)$ and $S_{i,j}^{(v,u)}\left(\frac{k}{n}\right) = W_j^{(v)}\left(\frac{k}{n}\right)W_{j+i}^{(u)}\left(\frac{k}{n}\right)\Gamma_{i,j}^{(v,u)}\left(\frac{k}{n}\right)$. Given that $\psi_{j,t+k}(t+\tau)=\psi_{j,k}(\tau)$ and using a change of variables, we can write $\int_0^1 |C_n(\tau,z)-C(\tau,z)|$ as

$$\begin{aligned}\int_0^1 \Bigg| &\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty}\psi_{j,k}(0)\psi_{j,k}(\tau)\left(S_{0,j}^{(u,v)}\left(z+\tfrac{k}{n}\right)-S_{0,j}^{(u,v)}(z)\right)\\ &+\sum_{i=-1}^{-\infty}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty}\psi_{j,k}(0)\psi_{j+i,k}(\tau)\left(S_{i,j}^{(u,v)}\left(z+\tfrac{k}{n}\right)-S_{i,j}^{(u,v)}(z)\right)\\ &+\sum_{i=-1}^{-\infty}\sum_{j=-1}^{-\infty}\sum_{k=-\infty}^{+\infty}\psi_{j+i,k}(0)\psi_{j,k}(\tau)\left(S_{i,j}^{(u,v)}\left(z+\tfrac{k}{n}\right)-S_{i,j}^{(u,v)}(z)\right)\Bigg|\end{aligned}$$

where $z=\frac{t}{n}$. Given the assumption that the functions $S_{i,j}^{(u,v)}(z)$ are Lipschitz continuous and hence that $\int_0^1\left|S_{i,j}^{(u,v)}\left(z+\frac{k}{n}\right)-S_{i,j}^{(u,v)}(z)\right| \le \frac{C_{i,j}^{(u,v)}|k|}{n}$ we can bound $\int_0^1 |C_n(\tau,z)-C(\tau,z)|$ from above by

$$\begin{aligned}&\sum_{j=-1}^{-\infty}\frac{C_{0,j}^{(u,v)}}{n}\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right|\\ &+\sum_{i=-1}^{-\infty}\sum_{j=-1}^{-\infty}\frac{C_{i,j}^{(u,v)}}{n}\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right|\\ &+\sum_{i=-1}^{-\infty}\sum_{j=-1}^{-\infty}\frac{C_{i,j}^{(v,u)}}{n}\sum_{k=-\infty}^{+\infty}|k|\left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right|\end{aligned}$$

where

$$\sum_{k=-\infty}^{+\infty} |k| \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \leq L_j \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right|$$

$$\sum_{k=-\infty}^{+\infty} |k| \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right| \leq L_{j+i} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right|$$

$$\sum_{k=-\infty}^{+\infty} |k| \left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \leq L_{j+i} \sum_{k=-\infty}^{+\infty} \left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right|$$

and hence we can bound $\sum_{\tau=-\infty}^{+\infty} |C_n^{(u,v)}(\tau,z) - C^{(v,u)}(\tau,z)|$ from above by

$$\begin{aligned}
&\sum_{j=-1}^{-\infty} \frac{C_{0,j}^{(v,v)} L_j}{n} \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \\
&+ \sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} \frac{C_{i,j}^{(u,v)} L_{j+i}}{n} \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right| \\
&+ \sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} \frac{C_{i,j}^{(v,u)} L_{j+i}}{n} \sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right|
\end{aligned}$$

Using that $\sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| \leq 1$, $\sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right| \leq 1$ and $\sum_{k=-\infty}^{+\infty} \left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| \leq 1$ for $\tau \in \mathbb{Z}$ we have that

$$\begin{aligned}
\sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j,k}(\tau)\right| &\leq 2L_j \\
\sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j,k}(0)\psi_{j+i,k}(\tau)\right| &\leq 2L_{j+i} \\
\sum_{\tau=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \left|\psi_{j+i,k}(0)\psi_{j,k}(\tau)\right| &\leq 2L_{j+i}
\end{aligned}$$

Hence, $\sum_{\tau=-\infty}^{+\infty} \int_0^1 |C_n^{(u,v)}(\tau,z) - C^{(u,v)}(\tau,z)|dz \leq \frac{1}{n} \sum_{j=-1}^{-\infty} 2C_{0,j}^{(u,v)} L_j^2$ $+\frac{1}{n} \sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} 2C_{i,j}^{(u,v)} L_{j+i}^2 + \frac{1}{n} \sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} 2C_{i,j}^{(v,u)} L_{j+i}^2$ which leads to the result in (1) given the assumptions that $\sum_{i=-1}^{-\infty} \sum_{j=-1}^{-\infty} C_{i,j}^{(u,v)} L_{j+i}^2$.

4.5.2 Proof of Theorem 4.1

Given the form of the Haar autocorrelation wavelets (see Section 3.3) we have that $A_{(-J(n)+1,-J(n)+1),(-1,-1)} = \sum_{\tau=-\infty}^{+\infty} \Psi_{-J(n)+1,-1} = 3 \cdot 2^{-J(n)-1}$ and hence if $J(n) = \log_2(n)$ this becomes $A_{(-J(n)+1,-J(n)+1),(-1,-1)} = \frac{3}{2n}$. Thus, using Horn's theorem we have that $\lambda_{\min}(A) \leq A_{(-J(n)+1,-1),(-J(n)+1,-1)}$ and hence

$$0 \leq \lambda_{\min}(A) \leq \frac{3}{2n}$$

4.5.3 Proof of Theorem 4.2

The expected value of $I_{i,j,k}^{(u,v)}$ takes the form

$$
\begin{aligned}
&E\left(I_{i,j,k}^{(u,v)}\right)\\
&= E\left(\sum_t \psi_{j,k}(t)X^{(u)}(t)\sum_{t'}\psi_{j+i,k}(t')X^{(v)}(t')\right)\\
&= \sum_t \psi_{j,k}(t)\sum_{t'}\psi_{j+i,k}(t')\sum_l\sum_m S_{0,l}^{(u,v)}\left(\tfrac{m}{n}\right)\psi_{l,m}(t)\psi_{l,m}(t')\\
&+\sum_t \psi_{j,k}(t)\sum_{t'}\psi_{j+i,k}(t')\sum_l\sum_r\sum_m S_{r,l}^{(u,v)}\left(\tfrac{m}{n}\right)\psi_{l,m}(t)\psi_{l+r,m}(t')\\
&+\sum_t \psi_{j,k}(t)\sum_{t'}\psi_{j+i,k}(t')\sum_l\sum_r\sum_m S_{r,l}^{(v,u)}\left(\tfrac{m}{n}\right)\psi_{l,m}(t')\psi_{l+r,m}(t)
\end{aligned}
$$

If we put $z=\frac{k}{n}$ and set $t'=t+\tau$ this can be written as

$$
\begin{aligned}
&\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_l S_{0,l}^{(u,v)}(z)\,\Psi_{0,l}(\tau)\\
&+\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}^{(u,v)}(z)\,\Psi_{r,l}(\tau)\\
&+\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}^{(v,u)}(z)\,\Psi_{-r,l+r}(\tau)\\
&+\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)R^{(u,v)}(t,\tau)
\end{aligned}
$$

where

$$
\begin{aligned}
&R^{(u,v)}(t,\tau)\\
&=\sum_{l=-1}^{-\infty}\sum_{m=-k-t}^{n-1-k-t}\left(S_{0,l}^{(u,v)}\left(z+\tfrac{m}{n}\right)-S_{0,l}^{(u,v)}(z)\right)\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\\
&+\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}\sum_{m=-k-t}^{n-1-k-t}\left(S_{r,l}^{(u,v)}\left(z+\tfrac{m}{n}\right)-S_{r,l}^{(u,v)}(z)\right)\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)\\
&+\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}\sum_{m=-k-t}^{n-1-k-t}\left(S_{r,l}^{(v,u)}\left(z+\tfrac{m}{n}\right)-S_{r,l}^{(v,u)}(z)\right)\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)
\end{aligned}
$$

Hence, using the assumption that the functions $S_{i,j}^{(u,v)}(z)$ are Lipschitz continuous we have that $|R^{(u,v)}(t,\tau)|$ can be bounded from above by

$$
\begin{aligned}
&\le \frac{1}{n}\sum_{l=-1}^{-\infty}C_{0,l}^{(u,v)}\sum_{m=-k-t}^{n-1-k-t}|m|\,|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)|\\
&+\frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}C_{r,l}^{(u,v)}\sum_{m=-k-t}^{n-1-k-t}|m|\,|\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau)|\\
&+\frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty}C_{r,l}^{(v,u)}\sum_{m=-k-t}^{n-1-k-t}|m|\,|\psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)|
\end{aligned}
$$

where

$$\sum_{m=-k-t}^{n-1-k-t} |m| \left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right| \leq L_l \sum_{m=-k-t}^{n-1-k-t} \left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right|$$

$$\sum_{m=-k-t}^{n-1-k-t} |m| \left|\left(\psi_{l,k}(0)\psi_{l+r,k}(\tau) + \psi_{l+r,k}(0)\psi_{l,k}(\tau)\right)\right| \leq$$
$$L_{r+l} \sum_{m=-k-t}^{n-1-k-t} \left|\left(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau) + \psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)\right)\right|$$

Given that

$$\sum_{m=-k-t}^{n-1-k-t} \left|\psi_{l,m+k}(0)\psi_{l,m+k}(\tau)\right| \leq 1$$
$$\sum_{m=-k-t}^{n-1-k-t} \left|\left(\psi_{l,m+k}(0)\psi_{l+r,m+k}(\tau) + \psi_{l+r,m+k}(0)\psi_{l,m+k}(\tau)\right)\right| \leq 1$$

we can bound $\left|\sum_t \psi_{j,k}(t) \sum_\tau \psi_{j+i,k}(t+\tau) R(t,\tau)\right|$ from above by

$$\left(\sum_{\tau=-L_{j+i}+1}^{L_{j+i}-1} \sum_t \left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right|\right) \left(\frac{1}{n}\sum_{l=-1}^{-\infty} C_{0,l}L_l + \frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty} L_{r+l}\left(C_{r,l}^{(u,v)} + C_{r,l}^{(v,u)}\right)\right)$$

where $\sum_t \left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right| \leq 1$ and hence

$$\sum_{\tau=-L_{j+i}+1}^{L_{j+i}-1} \sum_t \left|\psi_{j,k}(t)\psi_{j+i,k}(t+\tau)\right| \leq 2L_{j+i}.$$

Thus, we have that

$$\left|\sum_t \psi_{j,k}(t) \sum_\tau \psi_{j+i,k}(t+\tau) R(t,\tau)\right| \leq \frac{1}{n}\sum_{l=-1}^{-\infty} C_{0,l}L_l^2 + \frac{1}{n}\sum_{r=-1}^{-q}\sum_{l=-1}^{-\infty} L_{r+l}^2\left(C_{r,l}^{(u,v)} + C_{r,l}^{(v,u)}\right)$$

Using that $\sum_{r=0}^{-q}\sum_{l=-1}^{-\infty} L_{r+l}^2\left(C_{r,l}^{(u,v)} + C_{r,l}^{(v,u)}\right) < \infty$ we thus have that

$$\left|\sum_t \psi_{j,k}(t) \sum_\tau \psi_{j+i,k}(t+\tau) R(t,\tau)\right| = O\left(\frac{1}{n}\right).$$

Note that $\sum_t \psi_{j,k}(t)\psi_{j+i,k}(t+\tau) = \Psi_{i,j}(\tau)$, which allows to write

$$\begin{aligned}
&\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_l S_{0,l}^{(u,v)}(z)\Psi_{0,l}(\tau)\\
&+\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}^{(u,v)}(z)\Psi_{r,l}(\tau)\\
&+\sum_t \psi_{j,k}(t)\sum_\tau \psi_{j+i,k}(t+\tau)\sum_r\sum_l S_{r,l}^{(v,u)}(z)\Psi_{-r,l+r}(\tau)\\
&=\sum_r\sum_l S_{r,l}^{(u,v)}(z)A_{i,r}(j,l)
\end{aligned}$$

4.5.4 Proof of Theorem 4.3

Write the variance of the raw wavelet periodogram as $\mathrm{Var}\left(I_{i,j,k}^{(u,v)}\right) = E\left(\left(I_{i,j,k}^{(u,v)}\right)^2\right) - \left(E\left(I_{i,j,k}^{(u,v)}\right)\right)^2$. Using Isserli's theorem for higher moments of the multivariate normal distribution (Isserli (1918)), we have that

$$E\left(\left(I_{i,j,k}^{(u,v)}\right)^2\right) = 2\cdot\left(E\left(d_{j,k}^{(u)}d_{j+i,k}^{(v)}\right)\right)^2 + E\left(d_{j,k}^{(u)}d_{j,k}^{(u)}\right)E\left(d_{j+i,k}^{(v)}d_{j+i,k}^{(v)}\right)$$

and hence $\mathrm{Var}\left(I_{i,j,k}^{(u,v)}\right) = \left(E\left(d_{j,k}^{(u)}d_{j+i,k}^{(v)}\right)\right)^2 + E\left(d_{j,k}^{(u)}d_{j,k}^{(u)}\right)E\left(d_{j+i,k}^{(v)}d_{j+i,k}^{(v)}\right)$. Given the results of Theorem 3.4 we hence have that

$$\begin{aligned}
\mathrm{Var}\left(I_{i,j,k}^{(u,v)}\right) &= \tfrac{1}{4}\left(\sum_r\sum_l S_{r,l}^{(u,v)}(z)A_{i,r}(j,l)\right)^2\\
&+\sum_r\sum_l S_{r,l}^{(u,v)}(z)A_{0,r}(j,l)\sum_r\sum_l S_{r,l}^{(u,v)}(z)A_{0,r}(j+i,l)+O\left(\tfrac{1}{n}\right)
\end{aligned}$$

4.5.5 Proof of Theorem 4.4

Bound the variance of the smoothed wavelet periodogram as

$$\mathrm{Var}\left(I_{i,j,k,M}^{(u,v)}\right) \le \frac{1}{(2M+1)^2}\sum_{s=-M}^{M}\sum_{\kappa}\left|\mathrm{Cov}\left(I_{i,j,k+s}^{(u,v)}, I_{i,j,k+s+\kappa}^{(u,v)}\right)\right|$$

Using Isserli's theorem (Isserli (1918)) for higher moment of zero-mean Gaussian random variables we can write

$$\begin{aligned}\mathrm{Cov}\left(I^{(u,v)}_{i,j,k+s}, I^{(u,v)}_{i,j,k+s+\kappa}\right) &= E\left(I^{(u,v)}_{i,j,k+s} I^{(u,v)}_{i,j,k+s+\kappa}\right) - E\left(I^{(u,v)}_{i,j,k+s}\right) E\left(I^{(u,v)}_{i,j,k+s+\kappa}\right)\\ &= E\left(d^{(u)}_{j,k+s} d^{(u)}_{j,k+s+\kappa}\right) E\left(d^{(v)}_{j+i,k+s} d^{(v)}_{j+i,k+s+\kappa}\right)\\ &+ E\left(d^{(u)}_{j,k+s} d^{(v)}_{j+i,k+s+\kappa}\right) E\left(d^{(v)}_{j+i,k+s} d^{(u)}_{j,k+s+\kappa}\right)\end{aligned}$$

Using similar techniques as in the proof of Theorem 4.3 we can show that for $z = \frac{k}{n}$

$$E\left(d^{(u)}_{j,k+s} d^{(v)}_{j+i,k+s+\kappa}\right) = \sum_{l=-\infty}^{-1} \sum_{r=-q}^{q} S^{(u,v)}_{r,l}(z) A^{(\kappa)}_{i,r}(j,l) + O\left(\frac{1}{n}\right)$$

where

$$A^{(\kappa)}_{i,r}(j,l) = \sum_{\tau} \left(\Psi_{i,j}(\tau) + \Psi_{-i,j+i}(\tau)\right)\left(\Psi_{r,l}(\tau+\kappa) + \Psi_{-r,l+r}(\tau+\kappa)\right)$$

Note that $|A^{(\kappa)}_{i,r}(j,l)| \leq 4\min(L_{j+i}, L_{l+r}) \mathbb{I}\left[|\kappa| \leq \max(L_{j+i}, L_{l+r})\right]$, which means that we can write

$$\begin{aligned}\left|\sum_l \sum_r S^{(u,v)}_{r,l}(z) A^{(\kappa)}_{i,r}(j,l)\right| &\leq 4\mathbb{I}\left[|\kappa| \leq L_{j+i}\right] \sum_{l=j}^{-1} \sum_{r=-q}^{q} S^{(u,v)}_{r,l}(z) L_{l+r}\\ &+ 4L_{j+i} \sum_{l=-\infty}^{j-1} \sum_{r=-q}^{q} S^{(u,v)}_{r,l}(z) \mathbb{I}\left[|\kappa| \leq L_{l+r}\right]\end{aligned}$$

This means that $\left|E\left(d^{(u)}_{j,k+s} d^{(u)}_{j+i,k+s+\kappa}\right) E\left(d^{(v)}_{j+i,k+s} d^{(v)}_{j,k+s+\kappa}\right)\right|$ can be bounded by

$$\begin{aligned}&4\mathbb{I}\left[|\kappa| \leq L_{j+i}\right] \sum_{l=j}^{-1} \sum_{r=-q}^{q} |S^{(u,v)}_{r,l}|(z) L_{l+r} \sum_{l'=j}^{-1} \sum_{r'=-q}^{0} |S^{(u,v)}_{r',l'}|(z) L_{l'+r'}\\ &+4\mathbb{I}\left[|\kappa| \leq L_{j+i}\right] L_{j+i} \sum_{l=j}^{-1} \sum_{r=-q}^{q} |S^{(u,v)}_{r,l}(z)| L_{l+r} \sum_{l'=-\infty}^{j-1} \sum_{r'=-q}^{0} |S^{(u,v)}_{r',l'}(z)| \mathbb{I}\left[|\kappa| \leq L_{l'+r'}\right]\\ &+4\mathbb{I}\left[|\kappa| \leq L_{j+i}\right] L_{j+i} \sum_{l=-\infty}^{j-1} \sum_{r=-q}^{q} |S^{(u,v)}_{r,l}(z)| \mathbb{I}\left[|\kappa| \leq L_{l+r}\right] \sum_{l'=j}^{-1} \sum_{r'=-q}^{q} |S^{(u,v)}_{r',l'}(z)| L_{l'+r'}\\ &+4L^2_{j+i} \sum_{l=-\infty}^{j-1} \sum_{r=-q}^{q} |S^{(u,v)}_{r,l}(z)| \mathbb{I}\left[|\kappa| \leq L_{l+r}\right] \sum_{l'=-\infty}^{j-1} \sum_{r'=-q}^{q} |S^{(u,v)}_{r',l'}(z)| \mathbb{I}\left[|\kappa| \leq L_{l'+r'}\right]\end{aligned}$$

Summing this over κ then leads to

$$\begin{aligned}
B_{i,j} = \quad & 8L_{j+i}\sum_{l=j}^{-1}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)|L_{l+r}\sum_{l'=j}^{-1}\sum_{r'=-q}^{q}|S_{r',l'}^{(u,v)}(z)|L_{l'+r'} \\
& +8L_{j+i}^2\sum_{l=j}^{-1}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)|L_{l+r}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{q}|S_{r',l'}^{(u,v)}(z)| \\
& +8L_{j+i}^2\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)|\sum_{l'=j}^{-1}\sum_{r'=-q}^{q}|S_{r',l'}^{(u,v)}(z)|L_{l'+r'} \\
& +16L_{j+i}^2\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{q}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{q}|S_{r,l}^{(u,v)}(z)||S_{r',l'}^{(u,v)}(z)|\min(L_{l+r},L_{l'+r'})
\end{aligned}$$

Given the assumption that for all l and r the EWS functions are such that $|S_{r,l}^{(u,v)}(z)| \leq \frac{K}{L_{l+r}}$ with $L_{l-1+r} = 2L_{l+r}$, we obtain that $\sum_{l=j}^{-1}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)| \leq Kj(q+1)$. Moreover, we can establish the upper bound $\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)| \leq \frac{2K(q+1)}{L_j}$. Next, we write the expression $\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{q}\sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{q}|S_{r,l}^{(u,v)}(z)||S_{r',l'}^{(u,v)}(z)|\min(L_{l+r},L_{l'+r'})$ under the form

$$\sum_{l=-\infty}^{j-1}\sum_{r=-q}^{q}S_{r,l}^{(u,v)}L_{l+r}\sum_{l'\geq l}\sum_{r'=-q}^{q}|S_{r',l'}^{(u,v)}(z)| + \sum_{l'=-\infty}^{j-1}\sum_{r'=-q}^{q}S_{r',l'}^{(u,v)}L_{l'+r'}\sum_{l\geq l'}\sum_{r=-q}^{q}|S_{r,l}^{(u,v)}(z)|$$

which can be bounded from above by the sequence of upper bounds

$$K^2(q+1)^2\left(\sum_{l=-\infty}^{j-1}\frac{1}{L_l}+\sum_{l'=-\infty}^{j-1}\frac{1}{L_{l'}}\right) \leq \frac{2K^2(q+1)^2}{L_j}$$

Thus, we have that $B_{i,j} = O\left(L_j q^2\right)$. Hence, using that $\mathrm{Var}\left(I_{i,j,k,M}^{(u,v)}\right) \leq \frac{2}{M}B_{i,j} + O\left(\frac{1}{n}\right)$ we get that $\mathrm{Var}\left(I_{i,j,k,M}^{(u,v)}\right) = O\left(\frac{L_j q^2 j^2}{M}\right)$.

4.5.6 Proof of Theorem 4.5

To begin with, write $E\left(I_{i,j,k,M}^{(u,v)}\right) = \frac{1}{2M+1}\sum_{s=-M}^{M}E\left(I_{i,j,k+s}^{(u,v)}\right)$. According to Theorem 3.4 we have that $E\left(I_{i,j,k+s}^{(u,v)}\right) = \sum_{l}\sum_{r}S_{r,l}^{(u,v)}\left(z+\frac{s}{n}\right)A_{i,r}(j,l) + O\left(\frac{1}{n}\right)$

with $z = \frac{k}{n}$. Thus, we can write

$$\begin{aligned} E\left(I_{i,j,k+s}^{(u,v)}\right) &= \sum_l \sum_r S_{r,l}^{(u,v)}(z) A_{i,r}(j,l) \\ &+ \sum_l \sum_r \left(S_{r,l}^{(u,v)}\left(z+\tfrac{s}{n}\right) - S_{r,l}^{(u,v)}(z)\right) A_{i,r}(j,l) + O\left(\tfrac{1}{n}\right) \end{aligned}$$

Using the Lipschitz continuity properties of the functions $S_{r,l}^{(u,v)}(z)$, we have that

$$\left| \sum_l \sum_r \left(S_{r,l}^{(u,v)}\left(z+\frac{s}{n}\right) - S_{r,l}^{(u,v)}(z)\right) A_{i,r}(j,l) \right| \leq \frac{|s|}{n} \sum_l \sum_r C_{r,l}^{(u,v)} |A_{i,r}(j,l)|$$

First of all note that $|A_{i,r}(j,l)| \leq 2L_{l+r}$. Thus, using that $\sum_l \sum_r C_{r,l}^{(u,v)} L_{l+r}^2 < \infty$ and that $\sum_{s=-M}^{M} |s| = (M+1)M$ we have that

$$\frac{1}{2M+1} \sum_{s=-M}^{M} \left| \sum_l \sum_r \left(S_{r,l}^{(u,v)}\left(z+\frac{s}{n}\right) - S_{r,l}^{(u,v)}(z)\right) A_{i,r}(j,l) \right| = O\left(\frac{M}{n}\right)$$

4.5.7 Proof of Theorem 4.6

Consider the sequence of second-order wavelet functions and the associated sequence of EWS coefficients

$$\Phi_r = \left\{ \left\{\Phi_{i,j}(\tau)\right\}_{i=\max(-q,-\log_2(n)-j)}^{\min(q,\log_2(n)+j)} \right\}_{j=-r\log_2(n)+q(r-1)}^{(r-1)(-\log_2(n)+q)-1}$$

$$S_r^{(u,v)}(z) = \left\{ \left\{ S_{i,j}^{(u,v)}(z) \right\}_{i=\max(-q,-\log_2(n)-j)}^{\min(q,\log_2(n)+j)} \right\}_{j=-r\log_2(n)+q(r-1)}^{(r-1)(-\log_2(n)+q)-1}$$

with $r = 1, 2, \ldots$. Denote as $\left\{A_{1,r}\right\}_r$ the sequence of matrices containing the cross-inner products between the elements of Φ_1 and Φ_r. Denote as $I_{k,M}^{(u,v)} = \left\{ \left\{ I_{i,j,k,M}^{(u,v)} \right\}_{i=\max(-q,-\log_2(n)-j)}^{\min(q,\log_2(n)+j)} \right\}_{j=-\log_2(n)}^{-1}$ the vector of smoothed periodogram estimators at time point k. Given the statement of Theorem 4.5

we can write the expected value of $I_{k,M}$ as

$$E\left(I_{k,M}^{(u,v)}\right) = \sum_{r=1}^{+\infty} A_{1,r} S_r^{(u,v)}(z) + O\left(\frac{M}{n}\right)$$

The expected value of the smoothed and corrected wavelet periodogram then takes the form

$$\begin{aligned} E\left(I_{k,M}^{(u,v)}\right) &= A_{1,1}^{-1}\left(\sum_{r=1}^{+\infty} A_{1,r} S_r^{(u,v)} + O\left(\tfrac{M}{n}\right)\right) \\ &= S_1^{(u,v)} + A_{1,1}^{-1}\sum_{r=2}^{+\infty} A_{1,r} S_r^{(u,v)} + A_{1,1}^{-1} O\left(\tfrac{M}{n}\right) \end{aligned}$$

Thus, the L_2 norm of the vector $E\left(I_{k,M}^{(u,v)}\right) - S_1$ can be bounded from above by

$$\sum_{r=2}^{+\infty} ||A_{1,1}^{-1} A_{1,r}||_2 ||S_r^{(u,v)}||_2 + q||A_{1,1}^{-1}||_2 O\left(\frac{M\log_2(n)}{n}\right)$$

Consider a generic constant such that $||A_{1,1}^{-1}A_{1,r}||_2 \le c$ for all r which allows to write the upper bound as

$$c\sum_{r=2}^{+\infty} ||S_r^{(u,v)}||_2 + \delta^{-1} O\left(\frac{M\log_2(n)q}{n}\right)$$

where δ denotes the absolute lower bound of the eigenvalues of $A_{1,1}$. Given the assumption that $|S_{i,j}^{(u,v)}(z)| \le \frac{K}{L_{j+i}}$ with $L_{j-1+i} = 2L_{j+i}$ we have that $||S_r^{(u,v)}||_2 \le q\log_2(n)\left(\frac{2q}{n}\right)^{r-1}$ for $r = 2,3,\ldots$ and hence $\sum_{r=2}^{+\infty} ||S_r^{(u,v)}||_2 = q\log_2(n)\frac{\frac{2q}{n}}{1-\frac{2q}{n}}$. Thus, we obtain that

$$||E\left(I_{k,M}^{(u,v)}\right) - S_1^{(u,v)}||_2 \le \delta^{-1} O\left(\frac{M\log_2(n)q}{n}\right)$$

4.5.8 Proof of Theorem 4.7

Bound the variance of the smoothed and corrected wavelet periodogram as

$$\mathrm{Var}\left(\tilde{I}^{(u,v)}_{i,j,k,M}\right) \leq \sum_{r=-q}^{0}\sum_{l=-J}^{-1}\left|A^{-1}_{i,r}(j,l)\right|\sqrt{\mathrm{Var}\left(I^{(u,v)}_{r,l,k,M}\right)}\sum_{r'=-q}^{0}\sum_{l'=-J}^{-1}\left|A^{-1}_{i,r'}(j,l')\right|\sqrt{\mathrm{Var}\left(I^{(u,v)}_{r',l',k,M}\right)}$$

Using Theorem 4.4, this can be written as

$$\begin{aligned}\mathrm{Var}\left(\tilde{I}^{(u,v)}_{i,j,k,M}\right) \leq\ & \tfrac{2}{M}\textstyle\sum_{r=0}^{-q}\sum_{l=-J}^{-1}\left|A^{-1}_{i,r}(j,l)\right|\sqrt{B_{r,l}}\sum_{r'=-q}^{0}\sum_{l'=-J}^{-1}\left|A^{-1}_{i,r'}(j,l')\right|\sqrt{B_{r',l'}}\\ &+O\left(\tfrac{1}{n}\right)\textstyle\sum_{r=0}^{-q}\sum_{l=-J}^{-1}\left|A^{-1}_{i,r}(j,l)\right|\sum_{r'=0}^{-q}\sum_{l'=-J}^{-1}\left|A^{-1}_{i,r}(j,l')\right|\end{aligned}$$

Given the statement of Theorem 3.6 we have that $\sqrt{B_{r,l}} = O\left(\frac{lq\sqrt{L_l}}{\sqrt{M}}\right)$. Moreover, recall that $L_l = \left(2^{-l}-1\right)\left(L[h]-1\right)$. Hence, if $\sum_{r=-q}^{0}\sum_{l=-\infty}^{-1}\left|A_{i,r}(j,l)^{-1}\right| \leq C2^{\frac{j}{2}}$ and if we set $J = \log_2(n)$, we have that $\mathrm{Var}\left(\tilde{I}^{(u,v)}_{i,j,k,M}\right) = O\left(\frac{2^j\log_2(n)q}{M}\right)$.

4.5.9 Proof of Theorem 4.8

We denote as $||\Xi^2_{n,M}(z)||_{\max}$ the max norm of $\Xi^2_{n,M}(z) = \Xi^t_{n,M}(z)\Xi_{n,M}(z)$. Given that $\Xi^2_{n,M}(z)$ has $p\cdot\log_2(n)$ rows and columns we can bound the spectral norm as $||\Xi_{n,M}(z)||_2^2 \leq p\log_2(n)||\Xi^2_{n,M}(z)||_{\max}$. Using the Cauchy-Schwarz inequality we can bound each component of $E\left(\Xi^2_{n,M}(z)\right)$ as

$$\begin{aligned}&\sum_i\sum_v E\left(\tilde{I}^{(u,v)}_{n,M,i,j}-S^{(u,v)}_{i,j}\right)\left(\tilde{I}^{(u',v)}_{n,M,i,j'}-S^{(u',v)}_{i,j'}\right)\\ &\qquad\leq\sum_i\sum_v\left(E\left(\tilde{I}^{(u,v)}_{n,M,i,j}-S^{(u,v)}_{i,j}\right)^2E\left(\tilde{I}^{(u',v)}_{n,M,i,j'}-S^{(u',v)}_{i,j'}\right)^2\right)^{\frac{1}{2}}\end{aligned}$$

with $E\left(\tilde{I}^{(u,v)}_{n,M,i,j}-S^{(u,v)}_{i,j}\right)^2 = \left(E\left(\tilde{I}^{(u,v)}_{n,M,i,j}\right)-S^{(u,v)}_{i,j}\right)^2+\mathrm{Var}\left(\tilde{I}^{(u,v)}_{n,M,i,j}\right)$. Given the statements of Theorems 4.6 and 4.7 we have that $E\left(\tilde{I}^{(u,v)}_{n,M,i,j}-S^{(u,v)}_{i,j}\right)^2 = O\left(\frac{M^2\log_2^2(n)}{n^2}\right)+O\left(\frac{\log_2(n)}{M}\right)$ and hence that

$$E\left(||\Xi^2_{n,M}(z)||_{\max}\right) = O\left(p\log_2^3(n)\left(\frac{M^2}{n^2}+\frac{1}{M\log_2(n)}\right)\right)$$

which then leads to

$$E\left(||\Xi_{n,M}^2(z)||_2\right) = O\left(p^2 \log_4^3(n)\left(\frac{M^2}{n^2} + \frac{1}{M\log_2(n)}\right)\right)$$

Chapter 5

The locally stationary wavelet factor model

Over the past decade, factor models have become a standard econometric tool to model the co-movement of the elements of multivariate macroeconomic time series. Factor models present an interesting alternative to VAR or VARMA models (see Sims, 1980 [117], Lütkepohl and Poskitt, 1991 [80], Toda and Phillips, 1993, [127] and Lütkepohl and Saikkonen, 1997 [81]) in that they provide a parsimonious parametrization of a multivariate time series model in situations where the variables are driven by a small number of factors. The parsimony of the factor representation is especially important in high-dimensional settings in which the number of variables is large compared to the sample size given that it allows to reduce the size of the set of parameters to be estimated.

Take $\{X(t)\}_{t=0}^{+\infty}$ to be a vector-valued Gaussian process such that $X(t) \in \mathbb{R}^p$ and denote as $X^{(u)}(t)$ the $u-$th element of $X(t)$ such that

$$X^{(u)}(t) = \sum_{k=0}^{+\infty} \phi(k)\varepsilon_{t-k}^{(u)} \tag{5.1}$$

where ε_k are zero-mean Gaussian random variables such that $E(\varepsilon_k^{(u)}\varepsilon_{k'}^{(v)}) = 0$ for all $u,v = 1,\ldots,p$. The underlying idea of factor modelling is that the increment terms of the linear filter representation in (5.1) can be decomposed as

$$\varepsilon_k^{(u)} = \sum_{m=1}^{K} B_{m,p}^{(u)}(k) F_{m,k} + \xi_k^{(u)} \tag{5.2}$$

where $B_{m,p}^{(u)}(k) \in \mathbb{R}$ and with $F_{m,k}$ and $\xi_k^{(u)}$ being zero mean Gaussian random variables. The random variables $\{F_{m,k}\}_{m=1}^{K}$ are referred to as common factors since they impact several elements of the random vector $X(t)$. On the other hand, the elements of $\{\xi_k^{(u)}\}_{u=1}^{p}$ are called idiosyncratic components, since they are specific to only one of the elements of $X(t)$. The objective of factor modelling is not so much to obtain an estimator of the covariance pattern of the increment vectors, but rather to estimate the factor loadings $B_{m,p}^{(u)}(k)$ with $u = 1, \ldots, p$ and $m = 1, \ldots, K$. Note that if the focus lies on the estimation of the covariance matrix of the increment terms, the number of parameters that need to be estimated is p^2, whereas in factor analysis this number is reduced to pK with $K << p$.

We now provide an overview of the different subbranches of factor analysis, starting with the factor model defined in Ross (1976) [106]. One of the main assumptions in this class of factor models is that the coefficients in (5.1) are such that $\phi(k) = 1$ if $k = 0$ and $\phi(k) = 0$ otherwise. This assumption on the coefficients of the moving average representation leads to a *static* representation, in the sense that the elements of the vector process $\{X(t)\}_0^{+\infty}$ are independent. Moreover, the factor loadings are supposed to be time-invariant and hence we get that

$$X^{(u)}(t) = \sum_{m=1}^{K} B_{m,p}^{(u)} F_{m,t} + \xi_t^{(u)} \tag{5.3}$$

The common factors F_m are assumed to be uncorrelated with the idiosyncratic components $\xi_t^{(u)}$. A convenient normalization is to assume that the factors are uncorrelated to each other and that they have zero mean and unit variance. Similarly, the idiosyncratic components are assumed to be orthogonal to each other and they are supposed to be zero mean and have time-invariant variance. Factor models which include the assumption of independent idiosyncratic components, are commonly referred to as *strict* factor models. Given the fact that the common factors and the idiosyncratic components are uncorrelated, and that both, the factor loadings and the variances of the idiosyncratic components are time-invariant, the covariance matrix of the random vector $X(t) = (X_1(t), \ldots, X_p(t))^t$, denoted as S_p, can be decomposed as

$$S_p = B_p B_p^t + \Sigma_p \tag{5.4}$$

where B_p is the $p \times K$-dimensional matrix whose $(u,m)-$th element is the factor loading $B_{m,p}^{(u)}$ and Σ_p is the covariance matrix of the idiosyncratic components. Note that given the strict factor model specification, Σ_p is a diagonal matrix containing the variances of the idiosyncratic components. Thus, the covariances between the components of the random vector $X(t)$ are given by

$$\text{Cov}\left(X^{(u)}(t), X^{(u')}(t)\right) = \sum_{m=1}^{K} B_{m,p}^{(u)} B_{m,p}^{(u')}$$

with $u \neq u'$. Hence, given the independence of the idiosyncratic components, the covariance between all pairs of elements $X(t)$ is defined as the inner product of the vectors of factor loadings $\{B_{m,p}^{(u)}\}_{m=1}^{K}$ and $\{B_{m,p}^{(u')}\}_{m=1}^{K}$.

Given that $B_{m,p}^{(u')} = \text{Cov}(X^{(u)}(t), F_{m,t})$, the strict factor model specification leads to the assumption that the covariances of the elements of $X(t)$ are purely a by-product of the covariances of the latter and the common factors. However, as Chamberlain and Rothschild (1983) [23] point out, this assumption might be unrealistic in most cases relevant for practical purposes and propose an *approximative* factor representation instead. The latter allows for correlations between the idiosyncratic components. Thus, in this type of models the off-diagonal components of S_p are given by

$$\text{Cov}\left(X^{(u)}(t), X^{(u')}(t)\right) = \sum_{m=1}^{K} B_{m,p}^{(u)} B_{m,p}^{(u')} + \Sigma_{u,u'}. \tag{5.5}$$

where $\Sigma_{u,u'}$ denotes the covariance between $\xi_t^{(u)}$ and $\xi_t^{(u')}$. Note however that in the approximate factor model, the assumption of independence of the idiosyncratic components is replaced by the assumption that $\left\{\Sigma_p\right\}_{p=1}^{+\infty}$ is a sequence of positive-definite matrices with uniformly bounded eigenvalues. Thus, the diagonality assumption on Σ_p of the strict factor model is replaced by a less restrictive assumption on the behavior of the eigenvalues of Σ_p, allowing for a limited amount of correlation between the idiosyncratic components.

For most economic or financial time series, the assumption of serial independence of the elements of the $p-$dimensional vector process $\{X(t)\}_{t=0}^{n-1}$, underlying the static factor model, is not valid. Hence, for this type of data, a dynamic model specification seems to be more appropriate. The transition from a static to a dynamic factor model is obtained by dropping the assumption that the moving average coefficients in (5.1) satisfy $\phi_t(k) = 1$ if $t = k$ and $\phi_t(k) = 0$ otherwise. The dynamic factor models defined in Forni and Lippi (2001) [53] and Stock and Watson (2002) [121] use time-invariant factor loadings in (5.2) which leads to a representation of the elements of $X(t)$ as

$$X^{(u)}(t) = \sum_{m=1}^{K} B_{m,p}^{(u)} \left(\sum_{k=0}^{+\infty} \phi(k) F_{m,t-k} \right) + \sum_{k=0}^{+\infty} \phi(k) \xi_{t-k}^{(u)}$$

The authors adopt an approximative factor model specification and hence allow for time-invariant serial correlations between the idiosyncratic components.

As it has been mentioned in Chapters 3 and 4, the assumption of stationarity appears to be unrealistic for most economic and financial time series especially if the period of analysis is large. Indeed over large time intervals which may contain transitions between recessions and booms, economic and financial crises, changes in the monetary policy of the central bank, it seems unlikely that either the factor loadings or the covariance pattern of the idiosyncratic components remain constant in time. Motta et al. (2011) [93] propose an approximate factor model for large cross-sectional and time dimension based on a locally stationary model framework with smoothly changing factor loadings and covariance function of the idiosyncratic components. They define a static factor model specification such that

$$X^{(u)}(t) = \sum_{m=1}^{K} B_{m,p}^{(u)}(t) F_{m,t} + \xi_t^{(u)}.$$

which means that they do not allow for serial correlation in the multivariate time series. Eichler et al. (2011) [43] extend the idea of smoothly changing factor loading of Motta et al. (2011) [93] to the class of dynamic approximate factor models. Motivated from the stationary case (see Forni et al., 2000 [52]), the authors suggest an estimator of the factor loadings based on a time-varying principal components approach in the frequency domain using the eigenvectors of the estimator of the time-varying spectral density matrix.

In this chapter we propose a new kind of factor model representation based on a multiscale decomposition of vector processes using the high dimensional MLSW(q) model framework defined in Chapter 4.3. The chapter is divided into four sections. In Section 5.1 we introduce the LSW factor model. Section 5.2 deals with the link between the LSW factor model and principal components analysis of the evolutionary wavelet spectrum matrix. In Section 5.3 we present a consistent estimation procedure of both, the number of common factors and the standardized factor loading vectors, of the LSW factor model. Section 5.4 contains an application in which we use the LSW factor model to measure the degree of economic integration within four different groups of countries.

5.1 LSW Factor Model

Definition 5.1. *A sequence of triply-indexed stochastic processes* $\{(X_n^{(1)}(t),\dots,X_n^{(p)}(t))\}$, $t=0,\dots,n-1$, $n>0$, $p=O(n^{\alpha})$, *with mean zero elements is in the class of LSW factor models if there exists a representation*

$$X_n^{(u)}(t)=\sum_{j=-\log_2(n)}^{-1}\sum_{k=0}^{n-1}W_j^{(u)}(z)\,\varepsilon_{j,k}^{(u)}\psi_{j,k}(t).$$

The collection $\{\psi_{j,k}\}_{j,k}$, $j=-1,\dots,-\log_2(n)$, $k=0,\dots,n-1$ *is a set of non-decimated wavelet functions. The ratio* $z=\frac{k}{n}$ *denotes rescaled time. The increment terms* $\varepsilon_{j,k}^{(u)}$ *are Gaussian random variables such that* $E(\varepsilon_{j,k}^{(u)})=0$, $Var(\varepsilon_{j,k}^{(u)})=1$ *and*

$$Cov\left(\varepsilon_{j,k}^{(u)},\varepsilon_{j+i,l}^{(v)}\right)=\begin{cases}\Gamma_{i,j}^{(u,v)}(z) & \text{if } -q\le i\le 0 \text{ and } k=l\\ 0 & \text{if } i<-q \text{ or } k\neq l\end{cases}\qquad(5.1.1)$$

for $j=-1,\dots,-\log_2(n)$, $i=0,\dots,-\log_2(n)-j$, $u,v=1,\dots,p$ *and where*

$$\varepsilon_{j,k}^{(u)}=\sum_{m=1}^{K}B_{j,m}^{(u)}(z)\,F_{m,k}+\xi_{j,k}^{(u)}\qquad(5.1.2)$$

with $F_{m,k}$ and $\xi^{(u)}_{j,k}$ being zero-mean Gaussian random variables such that

$$Cov\left(F_{m,k}, F_{m',k}\right) = \begin{cases} 1 & \text{if } m = m' \\ 0 & \text{otherwise} \end{cases} . \tag{5.1.3}$$

$$Cov\left(\xi^{(u)}_{j,k}, \xi^{(v)}_{j',k}\right) = \sigma^{(u,v)}_{j,j'}(z) \quad \text{for } j, j' = -1, \ldots, -\log_2(n) \text{ and } u, v = 1, \ldots, p \tag{5.1.4}$$

$$Cov\left(F_{m,k}, \xi^{(u)}_{j,k}\right) = 0 \quad \text{for all } m,\ j,\ k \text{ and } u \tag{5.1.5}$$

The factor loadings and the covariance functions of the idiosyncratic components are Lipschitz continuous on $(0,1)$.

Consider the random vectors $\varepsilon^{(u)}_{n,k} = (\varepsilon^{(u)}_{-1,k}, \ldots, \varepsilon^{(u)}_{-\log_2(n),k})^t$ with $u = 1, \ldots, p$ and define $\varepsilon_{n,k}$ as a collection of the vectors $\varepsilon^{(u)}_{n,k}$, $u = 1, \ldots, p$, that is

$$\varepsilon_{n,k} = \left\{\varepsilon^{(u)}_{n,k}\right\}_{u=1}^{p} .$$

Moreover, define the vectors of common factors $F_k = (F_{1,k}, \ldots, F_{K,k})$ and idiosyncratic components $\xi_{n,k} = (\xi_{n,-1,k}, \ldots, \xi_{n,-\log_2(n),k})$ where $\xi_{n,j,k} = (\xi^{(1)}_{j,k}, \ldots, \xi^{(p)}_{j,k})$ such that

$$\varepsilon_{n,k} = B_n(z)F_k + \xi_{n,k} \tag{5.1.6}$$

where $B_n(z)$ is a $\log_2(n)p \times K$ matrix such that $B_n(z) = (B_{n,1}(z), \ldots, B_{n,K}(z))$ where $B_{n,m}(z) = (\{B^{(1)}_{m,j}(z)\}^{-1}_{j=-\log_2(n)}, \ldots, \{B^{(p)}_{m,j}(z)\}^{-1}_{j=-\log_2(n)})$. The squared Euclidean norms of the vectors $B_{n,m}(z)$, $m = 1, \ldots, K$, are such that

$$||B_{n,m}(z)|| = O\left(p^{\kappa}\right). \tag{5.1.7}$$

$0 < \kappa \leq 1$. *Moreover, we have that for all $m = 1, \ldots, K$ and $m' > m$*

$$B^t_{n,m}(z)B^t_{n,m'}(z) = 0. \tag{5.1.8}$$

The covariance matrix of ε_k, denoted as $S_n(z) = E\left(\varepsilon_{n,k}\varepsilon^t_{n,k}\right)$, is given by

$$S_n(z) = B_n(z)B^t_n(z) + \Sigma_n(z) \tag{5.1.9}$$

where $\Sigma_n(z)$ denotes the covariance matrix of the idiosyncratic components. The largest eigenvalue of $\Sigma_n(z)$ is such that for all n

$$\lambda_m(\Sigma_n(z)) = O(1) \tag{5.1.10}$$

Moreover, we have that

$$\sum_{m=1}^{\log_2(n)p} \lambda_m(\Sigma_n(z)) = O(p^{\eta}) \tag{5.1.11}$$

with $0 \leq \eta \leq 1$

Note that the representation in (5.1.6) is in itself meaningless. Indeed, we could set $B_n(z)$ to be the null matrix of dimension $\log_2(n)p \times K$ in which case $\varepsilon_{n,k} = \xi_{n,k}$ and hence the vectors of increment processes would merely be collections of idiosyncratic components. This reveals the importance of the assumption in (5.1.7). Indeed if $\kappa > 0$, we assume that the common factors are persistent in the sense that each factor has a non-negligible effect on the elements of a non-vanishing sub-collection of the set $\{\varepsilon_{n,k}^{(u)}\}_{u=1}^{p}$ as p increases. Thus, the representation in (5.1.6) has an underlying factor structure only if the degree of persistence is such that $\kappa > 0$.

The assumption in (5.1.10) states that a given idiosyncratic component can only be correlated to a finite set of other idiosyncratic components. To illustrate this a bit further, suppose that there exists a collection of subsets of the random variables $\xi_{j,k}^{(u)}$ and that within these subsets the constituent random variables are correlated to each other. The assumption in (5.1.10) states that the cardinalities of these sets cannot increase along with p. Hence, it is only due to this assumption that the random variables $\xi_{j,k}^{(u)}$ represent circumstances which are unique to a given element or a small group of $\varepsilon_{n,k}$. Thus, it is only due to the in (5.1.10) that we can refer to the random variables $\xi_{j,k}^{(u)}$ as idiosyncratic components.

5.2 LSW factor model and principal components analysis

5.2.1 Eigendecomposition of $S_n(z)$

Denote as $\lambda\left(B_n(z)\right) = \{\lambda_m\left(B_n(z)\right)\}_{m=1}^{K}$ the set of ordered singular values of the $\log_2(n)p \times K$ matrix $B_n(z)$ in (5.1.9) such that $\lambda_m(B_n(z)) \geq \lambda_{m+1}(B_n(z))$. Note that, given the orthogonality assumption in (5.1.8), the set of ordered singular values of $\lambda(B_n(z))$ is identical to the set of Euclidean norms $\{||B_{n,m}(z)||^2\}_{m=1}^{K}$. Moreover, note that the set of associated eigenvectors is given by $\{\frac{B_{n,m}(z)}{||B_{n,m}(z)||^2}\}_{m=1}^{K}$.

Theorem 5.1 states that if both, the degree of high-dimensionality (α) and the degree of persistence (κ), are larger than zero, the number of common factors K is identical to the number of eigenvalues of $S_n(z)$ that diverge as p increases.

Theorem 5.1. *Consider the EWS matrix $S_n(z)$ in* (5.1.9) *and denote as $\lambda_m\left(S_n(z)\right)$ its eigenvalues with $\lambda_m\left(S_n(z)\right) \geq \lambda_{m+1}\left(S_n(z)\right)$ for $m = 1,\ldots,\log_2(n) \times p$. Given the assumptions of the factor model representation in Definition 5.1, we have that*

$$\begin{aligned} &\lambda_m\left(S_n(z)\right) = O\left(n^{\alpha\kappa}\right) \quad \textit{for } m = 1,\ldots,K \\ &\lambda_m\left(S_n(z)\right) = O(1) \qquad\quad \textit{for } m > K \end{aligned}$$

Theorem 5.2 states that if $\alpha > 0$ and $\kappa > 0$, the m-th eigenvector of $S_n(z)$ converges in Euclidean norm towards the standardized vector of factor loadings $\frac{B_{n,m}(z)}{||B_{n,m}(z)||}$.

Theorem 5.2. *Consider the EWS matrix $S_n(z)$ in* (5.1.6) *and denote as $\lambda_m(S_n(z))$ its eigenvalues with $\lambda_m(S_n(z)) \geq \lambda_{m+1}\left(S_n(z)\right)$ for $m = 1,\ldots,\log_2(n) \times p$. Moreover, denote as $u_m(S_n(z))$ the eigenvector associated with the $m-$th eigenvalue. Given the assumptions of the factor model representation in Definition 5.1, we have that*

$$\left|\left|u_m\left(S_n(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| = O\left(\frac{1}{n^{\alpha\kappa}}\right).$$

5.2.2 Local factor inertia

We define the local inertia function as $I(z) = \sum_{u=1}^{p} C^{(u,u)}(0,z)$, which, according to the factor representation in (5.1.2), is given by

$$I(z) = \sum_{m=1}^{K} ||B_{n,m}||^2 + \sum_{u=1}^{p} \sum_{j=-\log_2(n)}^{-1} \left(\sigma_{j,j}^{(u,u)}(z)\right)^2. \tag{5.2.1}$$

The local factor inertia, denoted as $I_F(z)$, represents the part of $I(z)$ which is accounted for by the common factors and hence

$$I_F(z) = \frac{\sum_{m=1}^{K} ||B_{n,m}(z)||^2}{I(z)}. \tag{5.2.2}$$

Theorem 5.3. *Consider the EWS matrix $S_n(z)$ in* (5.1.6) *and denote as $\lambda_m(S_n(z))$ its eigenvalues with $\lambda_m(S_n(z)) \geq \lambda_{m+1}(S_n(z))$ for $m = 1,\ldots,\log_2(n)\times p$. Given the assumptions of the LSW factor model in Definition 5.1 we have that*

$$\left| \frac{\sum_{m=1}^{K} \lambda_m(S_n)}{\sum_{m=1}^{\log_2(n)p} \lambda_m(S_n)} - I_F(z) \right| = O\left(\frac{1}{n^{\alpha \max(\kappa,\eta)}}\right).$$

5.2.3 Local distance measure

According to the factor model representation in Definition 5.1, each element of an LSW process can be represented as

$$X_n^{(u)}(t) = \sum_{k=0}^{n-1} \sum_{m=1}^{K} \left(\sum_{j=-\log_2(n)}^{-1} \psi_{j,k}(t) B_{j,m}^{(u)}(z) \right) F_{m,k} + \sum_{k=0}^{n-1} \sum_{j=-\log_2(n)}^{-1} \psi_{j,k}(t)\xi_{j,k}. \tag{5.2.3}$$

To start with, consider the random vectors $X_n = \{X_n^{(u)}\}_{u=1}^{p}$ where $X_n^{(u)} = \{X_n^{(u)}(t)\}_{t=0}^{n-1}$ and $F_k = \{F_{m,k}\}_{m=1}^{K}$. Define the vector of covariance coefficients $\mathrm{Cov}(X_n, F_{m,k})$ and denote as $||\mathrm{Cov}(X_n, F_{m,k})||_2$ its Euclidean norm, that is

$$||\mathrm{Cov}\left(X_n, F_{m,k}\right)||_2$$
$$= \left(\sum_{t=0}^{n\hat{-}1}\sum_{u=1}^{p}\left(\mathrm{Cov}\left(X_n^{(u)}(t), F_{m,k}\right)\right)^2\right)^{\frac{1}{2}}$$
$$= \left(\sum_u\sum_j\sum_{j'}\left(B_{j,m}^{(u)}(z) - B_{j,m}^{(u)}(z)\right)\left(B_{j',m}^{(u)}(z) - B_{j',m}^{(u)}(z)\right)\sum_t \psi_{j,k}(t)\psi_{j',k}(t)\right)^{\frac{1}{2}}.$$

Using that $\sum_t \psi_{j,k}(t)\psi_{j',k}(t)$ is equal to 1 if $j = j'$ and equal to 0 if $j \neq j'$ this reduces to

$$\begin{aligned}||\mathrm{Cov}\left(X_n, F_{m,k}\right)||_2 &= \left(\sum_u\sum_j\left(B_{j,m}^{(u)}(z) - B_{j,m}^{(u)}(z)\right)^2\right)^{\frac{1}{2}}\\ &= ||B_{n,m}(z)||_2\end{aligned}$$

Define the matrix $\Xi_n^{(u)}(z) = \left(\frac{\mathrm{Cov}\left(X_n^{(u)}, F_{1,k}\right)}{||\mathrm{Cov}\left(X_n, F_{1,k}\right)||_2}, \ldots, \frac{\mathrm{Cov}\left(X_n^{(u)}, F_{K,k}\right)}{||\mathrm{Cov}\left(X_n, F_{K,k}\right)||_2}\right)$ and consider the Frobenius norm

$$||\Xi_n^{(u)}(z) - \Xi_n^{(v)}(z)||_F$$
$$= \left(\sum_{m=1}^{K}\frac{1}{||\mathrm{Cov}\left(X_n, F_{m,k}\right)||_2^2}\sum_t\left(\mathrm{Cov}\left(X_n^{(u)}(t), F_{m,k}\right) - \mathrm{Cov}\left(X_n^{(v)}(t), F_{m,k}\right)\right)^2\right)^{\frac{1}{2}}$$
$$= \left(\sum_{m=1}^{K}\frac{1}{||B_{n,m}(z)||_2^2}\sum_j\left(B_{j,m}^{(u)}(z) - B_{j,m}^{(v)}(z)\right)^2\right)^{\frac{1}{2}}.$$

Thus, if we define the vectors $B_{n,m}^{(u)}(z) = \{B_{j,m}^{(u)}(z)\}_{j=-\log_2(n)}^{-1}$ and consider the matrix $\Delta_n^{(u)}(z) = \left(\frac{B_{n,1}^{(u)}(z)}{||B_{n,1}(z)||_2}, \ldots, \frac{B_{n,K}^{(u)}(z)}{||B_{n,K}(z)||_2}\right)$, we obtain that

$$||\Xi_n^{(u)}(z) - \Xi_n^{(v)}(z)||_F = ||\Delta_n^{(u)} - \Delta_n^{(v)}||_F$$

Theorem 5.4. *Define the matrix* $\Delta_n^{(u)}(z) = \left(\frac{B_{n,1}^{(u)}(z)}{||B_{n,1}(z)||_2}, \ldots, \frac{B_{n,K}^{(u)}(z)}{||B_{n,K}(z)||_2}\right)$. *Denote as* $\{\lambda_m(S_n(z)\}_{m=1}^{\log_2(n)p}$ *the set of ordered eigenvalues of the EWS matrix such that* $\lambda_m(S_n(z)) > \lambda_{m+1}(S_n(z))$ *for* $m = 1, \ldots, \log_2(n)p$, *and let* $\{u_m(S_n(z))\}_{m=1}^{\log_2(n)p}$ *be the set of associated eigenvectors. Define the* *vectors* $U^{(u)}_{n,m} = \{(u_m(S_n(z)))_{-j+(u-1)p}\}^{-1}_{j=-\log_2(n)}$ *and consider the matrix* $\overline{\Delta}^{(u)}_n(z) = (U^{(u)}_{n,1}(z)(z), \dots, U^{(u)}_{n,K}(z))$. *Given the assumptions of the factor model representation in Definition 5.1 and if we define the Frobenius norms* $||\Delta^{(u)}_n(z) - \Delta^{(v)}_n(z)||_F$ *and* $||\overline{\Delta}^{(u)}_n(z) - \overline{\Delta}^{(v)}_n(z)||_F$, *we have that*

$$\left| ||\overline{\Delta}^{(u)}_n(z) - \overline{\Delta}^{(v)}_n(z)||_F - ||\Delta^{(u)}_n(z) - \Delta^{(v)}_n(z)||_F \right| = O\left(\frac{1}{n^{\alpha\kappa}}\right)$$

5.3 Estimation of the LSW factor model

Denote as $\tilde{I}_{n,M}(z)$ the corrected and smoothed periodogram estimator of the EWS matrix $S_n(z)$ as defined in Section 4.3. In this section we illustrate how the eigenvalues and the eigenvectors of $\tilde{I}_{n,M}(z)$ can be used to design a consistent estimation procedure of the LSW factor model.

To start with, recall from Section 5.2.1 that the number of common factors in the LSW factor model is identical to the number of diverging eigenvalues of the EWS matrix. Thus, a natural way of estimating the number of common factors is to count the number of eigenvalues of $\tilde{I}_{n,M}(z)$ being larger than a certain threshold, with the latter growing along with p. Hence, if we denote as $\mathbb{I}[x]$ the indicator function, the estimator of K can be defined as

$$\hat{K} = \sum_{m=1}^{\log_2(n)p} \mathbb{I}\left[\lambda_m\left(\tilde{I}_{n,M}\right) > n^{\gamma}\right] \tag{5.3.1}$$

with $\gamma > 0$. Theorem 5.5 states that the estimator $\hat{K}$ converges in probability to the number of common factors as long as the degree of high-dimensionality fulfills $\alpha < \frac{1}{3}$.

Theorem 5.5. *Denote as* $\lambda_m(\tilde{I}_{n,M}(z))$, *with* $m = 1, \dots, \log_2(n)p$, *the decreasingly ordered eigenvalues of* $\tilde{I}_{n,M}(z)$ *and consider the random variable* $\hat{K} = \sum_{m=1}^{\log_2(n)p} \mathbb{I}[\lambda_m\left(\tilde{I}_{n,M}(z)\right) > n^{\gamma}]$. *If we denote as K the number of common factors and if* $\gamma < \kappa$, *with the assumptions of the LSW factor model in Definition 5.1 being verified, we have that*

$$\hat{K} \underset{p}{\rightarrow} K$$

given $0 \leq \alpha < \frac{1}{3}$.

Theorem 5.5 tells us that the convergence of $\hat{K}$ depends on the appropriate choice of γ which depends on an a priori knowledge of the degree of persistence of the common factors. However, note that γ could also be obtained based on a data-driven selection procedure using cross-validation or bootstrap approaches. We will not discuss these issues in this work and leave this question open for further research.

Theorem 5.6 states that if $\alpha < \frac{1}{3}$, the m-th eigenvector of $\tilde{I}_{n,M}$ converges in Euclidean norm towards the standardized vector of factor loadings $\frac{B_{n,m}(z)}{||B_{n,m}(z)||}$.

Theorem 5.6. *If we denote as* $\lambda_m(\tilde{I}_{n,M}(z))$*, with* $m = 1, \ldots, \log_2(n)p$*, the decreasingly ordered eigenvalues of* $\tilde{I}_{n,M}(z)$ *and as* $u_m\left(\tilde{I}_{n,M}(z)\right)$ *the associated eigenvectors. Given the assumptions of the factor model representation in Definition 5.1, we have that*

$$\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| \underset{p}{\rightarrow} 0$$

given $0 \leq \alpha < \frac{1}{3}$.

Theorems 5.7 and 5.8 show convergence in probability of the estimators of the local factor inertia and the local distance measure respectively.

Theorem 5.7. *If we denote as* $\lambda_m(\tilde{I}_{n,M}(z))$*, with* $m = 1, \ldots, \log_2(n)p$*, the decreasingly ordered eigenvalues of* $\tilde{I}_{n,M}(z)$ *and given the assumptions of the factor model representation in Definition 5.1, we have that*

$$\left|\frac{\sum\limits_{m=1}^{K} \lambda_m\left(\tilde{I}_{n,M}(z)\right)}{\sum\limits_{m=1}^{\log_2(n)p} \hat{\lambda}_m\left(\tilde{I}_{n,M}(z)\right)} - I_F(z)\right| \underset{p}{\rightarrow} 0$$

given $0 \leq \alpha < \frac{1}{3}$.

Theorem 5.8. *Define the matrix* $\Delta_n^{(u)}(z) = \left(\frac{B_{n,1}^{(u)}(z)}{||B_{n,1}(z)||_2}, \ldots, \frac{B_{n,K}^{(u)}(z)}{||B_{n,K}(z)||_2}\right)$*. Denote as* $\{\lambda_m(I_{n,M}(z))\}_{m=1}^{\log_2(n)p}$ *the set of decreasingly ordered eigenvalues of* $I_{n,M}(z)$ *and let* $\{u_m(I_{n,M}(z))\}_{m=1}^{\log_2(n)p}$ *be the set of associated eigenvectors.*

Define the vectors $\hat{U}_{n,m}^{(u)} = \{(u_m(I_{n,M}(z)))_{-j+(u-1)p}\}_{j=-\log_2(n)}^{-1}$ *and consider the matrix* $\hat{\Delta}_n^{(u)}(z) = (\hat{U}_{n,1}^{(u)}(z)(z), \ldots, \hat{U}_{n,K}^{(u)}(z))$*. If we define the Frobenius norms* $||\Delta_n^{(u)}(z) - \Delta_n^{(v)}(z)||_F$ *and* $||\hat{\Delta}_n^{(u)}(z) - \hat{\Delta}_n^{(v)}(z)||_F$ *given the assumptions of the factor model representation in Definition 5.1, we have that*

$$\left| ||\hat{\Delta}_n^{(u)}(z) - \hat{\Delta}_n^{(v)}(z)||_F - ||\Delta_n^{(u)}(z) - \Delta_n^{(v)}(z)||_F \right| \underset{p}{\to} 0$$

given $0 \leq \alpha < \frac{1}{3}$.

5.4 Real Data Analysis: Modelling the degree of economic integration of the European currency area

Economic integration is commonly described as a process whereby countries coordinate and link their economic policies. For a more detailed discussion on the notion of economic integration see for example Lipsey (1957) [77], Johnson (1965) [71] and Balassa (1967) [13]. The stronger the harmonization of fiscal, economic and monetary policies between countries the higher the degree of economic integration. The degree of economic integration can be roughly categorized into seven stages (see for example Balassa, 2013 [14]), that is

Preferential Trade Area: Preferential trade areas consist in several countries within a geographical region agreeing to reduce or eliminate tariff barriers on selected goods imported from other members of the area.

Free Trade Area: Free trade areas exist when several countries within a given region agree to reduce or eliminate barriers to trade on all goods coming from other members (see for example the North Atlantic Free Trade Agreement (NAFTA) including USA, Canada, and Mexico).

Customs Union: In contrast to the free trade areas, customs union do not only remove tariff barriers for its members but also accept a unified external tariff against non-members. This also implies that the members if a customs union negotiate as a single bloc with third parties, such as the WTO.

Common Market: A common market occurs when member countries trade freely in all economic resources such as goods, services, capital, and labour. As well as removing tariffs, non-tariff barriers are also reduced and eliminated. Common markets are also characterized by a significant level of harmonisation of micro-economic policies, and common rules regarding monopoly power and anti-competitive practices and common policies affecting key industries, such as the Common Agricultural Policy (CAP) and Common Fisheries Policy (CFP) of the European Single Market (ESM).

Economic Union: Economic Union is a term applied to a trading bloc that has both a common market between members, and a common trade policy towards non-members but where members are free to pursue independent macro-economic policies.

Monetary Union: Monetary union involves adopting a single, shared currency, such as the Euro for the Euro-16 countries, and the East Caribbean Dollar for eleven islands in the East Caribbean. This means that there is a common exchange rate, a common monetary policy set up by a single central bank, such as the European Central Bank or the East Caribbean Central Bank.

Fiscal Union: A fiscal union is an agreement to harmonize tax rates, to establish common levels of public sector spending and jointly agree on national budget deficits or surpluses. Note however that control over fiscal policy and public spending is considered central to national sovereignty, and hence in the world today there is no substantial fiscal union between independent nations.

Economic and Monetary Union: Economic and Monetary Union involves a single economic market, a common trade policy, a single currency and a common monetary policy.

Complete Economic Integration: Complete economic integration involves a single economic market, a common trade policy, a single currency, a common monetary policy (EMU) together with a single fiscal policy, tax and benefit rates and hence a complete harmonization of all policies, rates, and economic trade rules.

In this section we use both, the local factor inertia (see Section 5.2.2) and the local distance measure (see Section 5.2.3), to model the degree of economic integration of four groups of countries

Group 1: Belgium, France, Germany, Netherlands and Spain

Group 2: Brazil, Canada, Chile, Mexico and United States

Group 3: Australia, China, India, Japan and Singapore

Group 4: Eleven members of the European currency area

The dataset contains the log-returns of the stock indexes of these countries on a period ranging from 2003, 2004 or 2005, depending on the respective group of countries, to April 2015.

The local factor inertia provides an indication of the potential presence of common factors in the sense that the larger the local factor inertia at a given point in time and for a given sample, the higher the likelihood that there are some common factors in the underlying increment process of the MLSW model. The presence of common factors among the members of a group of countries is a necessary condition for a high degree of economic integration within a group of countries. Indeed, the elimination of tariff barriers, a common market or a single monetary policy all have the objective to enhance the economic activity among the member states. Given these strong economic bounds between the member states this should reflect in the exposition of the latter to the same random economic shocks leading to the factor model representation in Definition 5.1 and hence eventually to a non-vanishing local factor inertia as the number of variables p increases. Note however that the presence of common factors is not a sufficient indicator for the presence of a strong economic integration among countries. Indeed, note that a group of countries could be affected by the same economic shocks even if they economies are not integrated at all. It may for example be argued that the state of the economy of nearly every country in the world is influenced by major factors such as global financial crises and the current state of the world economy.

Thus, we define an additional criterion based on the local distance measures and refer to the lattera as average local distance measure. To do this we first compute the Frobenius norms $||\hat{\Delta}_n^{(u)}(z) - \hat{\Delta}_n^{(v)}(z)||_F$ defined in Section 5.2.3 for $u, v = 1, \ldots, p$. After this we compute the average distance measure by taking the mean over $v = 1, \ldots, p$ for each u which leads to the vector

$$\left\{ \frac{1}{p} \sum_{v=1}^{p} ||\hat{\Xi}^{(u)}(z) - \hat{\Xi}^{(v)}(z)||_F \right\}_{u=1}^{p} . \tag{5.4.1}$$

The average local distance is a measure of homogeneity of a group of countries with small values of the latter indicating a high level of homogeneity. Note that a small value of the local distance measure is a necessary but not a sufficient condition for the presence of economic integration. However, a joint use of both indicators, the local factor inertia and the local distance measure, can provide a good description of the state of the process of economic integration within a given group of countries.

5.4.1 Belgium, France, Germany, Netherlands and Spain

We consider the stock indices of five Central and Western European countries.

Index	Country
AEX	Netherlands
BEL20	Belgium
CAC40	France
DAX	Germany
IBEX	Spain

Table 5.1: ***Stock indices of Belgium, France, Germany, Netherlands and Spain***

Figure 5.1 plots the local factor inertia between June 2005 and April 2015. To start with, note that the local factor inertia reaches its peak during the third quarter of 2008. Recall that September 2008 has been marked by two key events, which have unfolded the late-2000s global financial crisis, that is the bankruptcies of Lehman brothers and American International Group (AIG). Lehman Brothers constituted the fourth largest investment bank in the US at that time. AIG was the world's largest insurance company during the late 2000s and was a major issuer of credit default swaps (CDS) on collateralized debt obligations. From the third quarter of 2008 on the lhe local factor inertia decreases from a peak value of 0.43 to an absolute minimum of 0.27 attained during the start of the third quarter of 2010. After this, the local factor inertia increases again to reach a value of 0.43 in April 2015 which is identical to the value it had in September 2008.

Figure 5.2 plots the average local distance measures for these five countries between June 2005 and April 2015. Note that during the entire period of analysis the distance measures remain below 0.1.

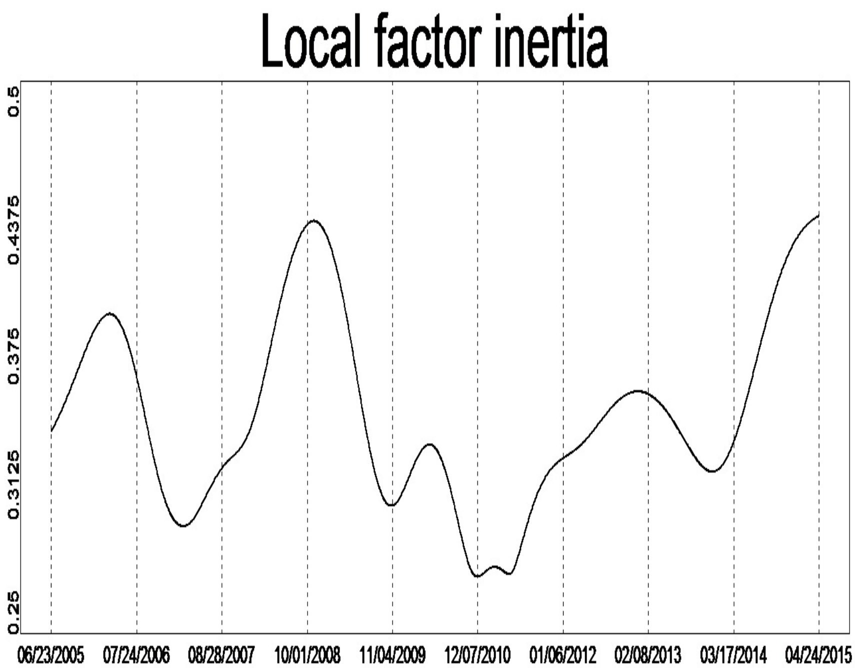


Figure 5.1: ***(Local factor inertia) Belgium, France, Germany, Netherlands and Spain***

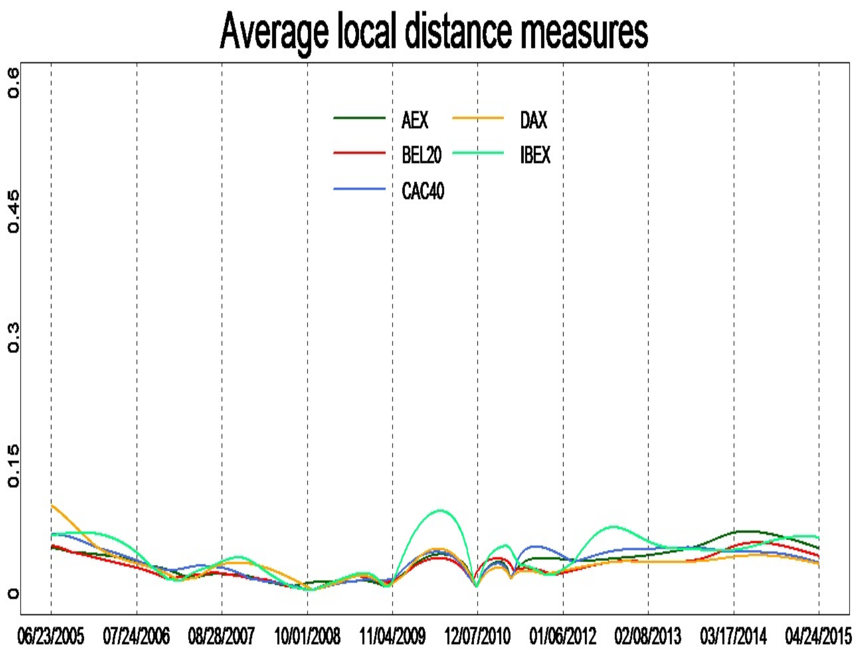


Figure 5.2: ***(Local distance measure) Belgium, France, Germany, Netherlands and Spain***

5.4.2 Brazil, Canada, Chile, Mexico and United States

We consider the stock indices of five different countries from the Americas.

Index	Country
BONESPA	Brazil
DJIA	United States
IPC	Mexico
IPSA	Chile
TSX	Canada

Table 5.2: ***Stock indices of Brazil, Canada, Chile, Mexico and United States***

Figure 5.3 plots the local factor inertia between January 2004 and April 2015. Note that as for the countries analyzed in Section 5.4.1 the local factor inertia reaches its peak in the third quarter of 2008 and is characterized by a sharp decline which lasts until the start of the third quarter in 2010. After this the local factor inertia recovers and rises again. Note however that, in contrast to the countries in 5.4.1, the local factor inertia does not attain its pre-crisis values during summer 2008.

Figure 5.4 plots the average local distance measures for these five countries between January 2004 and April 2015. First of all note that during most of the period of analysis, the average local distance measures are above 0.1 which marks the main difference between the two groups of countries studies in 5.4.1 and 5.4.2 respectively. The smaller values of the average local distance measures might be indicate a higher level of economic integration of the countries in 5.4.1. Another striking feature is that, apart from the period 2007-2008 marking the peak of the financial crisis in the US, the factor loading pattern of the Brazilian stock index reveals to be very different from the factor loading patterns of the remaining countries.

5.4.3 Australia, China, India, Japan and Singapore

We consider the stock indices of five different countries from South-East Asia and Australia.

Index	Country
ASX	Australia
BSE	India
NIKKEI	Japan
SHCOMP	China
STRAITS TIMES	Singapore

Table 5.3: ***Stock indices of Australia, China, India, Japan and Singapore***

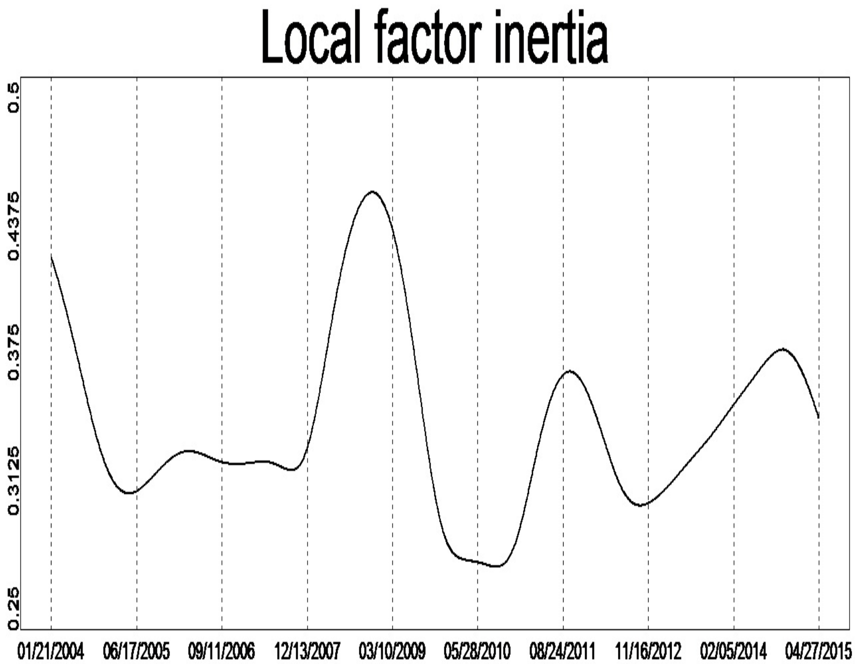


Figure 5.3: ***(Local factor inertia) Brazil, Canada, Chile, Mexico and United States***

Figure 5.5 contains the local factor inertia between April 2003 and April 2015. Note that as for the countries analyzed in Sections 5.4.1 and 5.4.2 the local factor inertia reaches its highest value during the third quarter of 2008. However, in contrast to the two other groups of countries, the decline after September 2008 is not as sharp. Moreover, note that throughout the entire period of analysis the local factor inertia is never as strong as for the two other groups of countries.

Figure 5.6 plots the average local distance measures between April 2003 and April 2015.

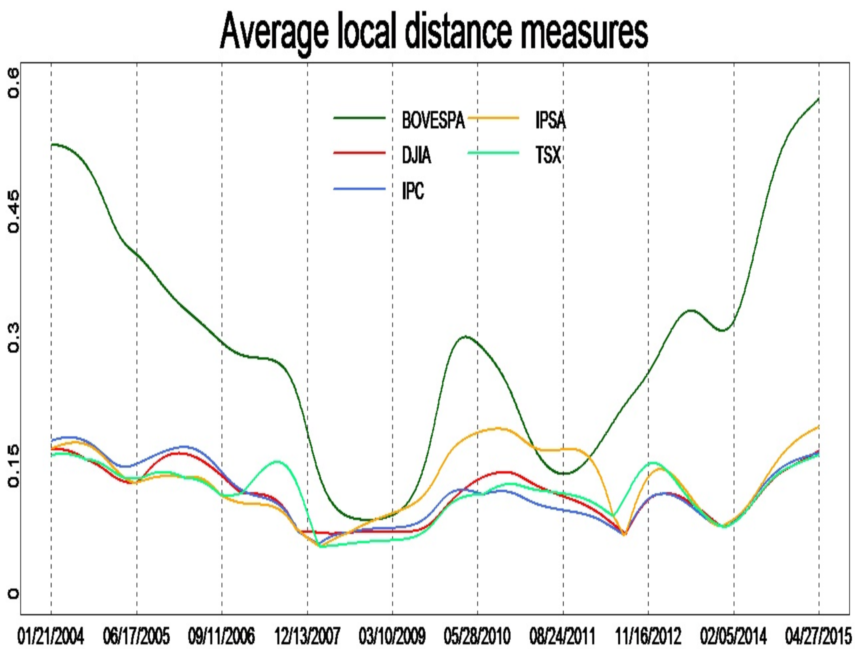


Figure 5.4: ***(Local distance measure) Brazil, Canada, Chile, Mexico and United States***

5.4.4 European currency area

We consider the stock indices of ten members of the European currency area.

Figure 5.7 contains the local factor inertia between July 2002 and April 2015 and Figure 5.8 contains the average local distance measure on that same time span. First, note that the average distance measures of all the countries are characterized by a more of less steady decrease up to the start of the financial crisis in the third quarter 2008. Second, there is a very strong increase in the average distance measure of the Greek stock index from the aftermath of the financial crisis in 2009 on until April 2015.

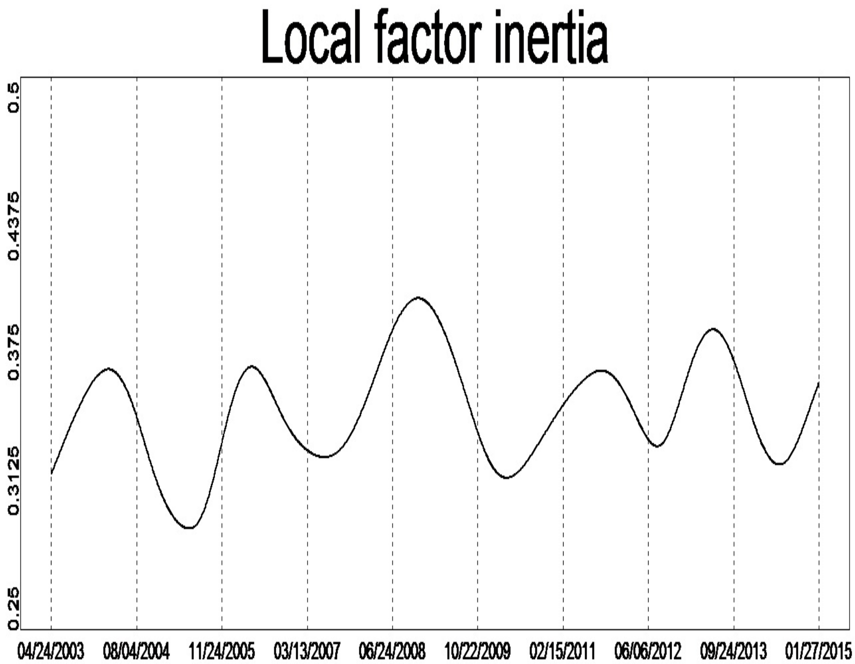


Figure 5.5: ***(Local factor inertia) Australia, China, India, Japan and Singapur***

We now briefly illustrate how the local distance measure can be used in a multivariate time series clustering setting. The objective of time series clustering is to classify the members of the Euro area into homogeneous groups using the local distance measure as a measure of similarity. More precisely if we adopt a k-means clustering approach (see Lloyd, 1982 [78]), we aim at partitioning the vectors of factor loadings at any point z,

$$B^{(u)}(z) = \left\{ B^{(u)}_{m,j}(z) :\ m = 1, \dots, K;\ j = -1, \dots, -\log_2(n) \right\}$$

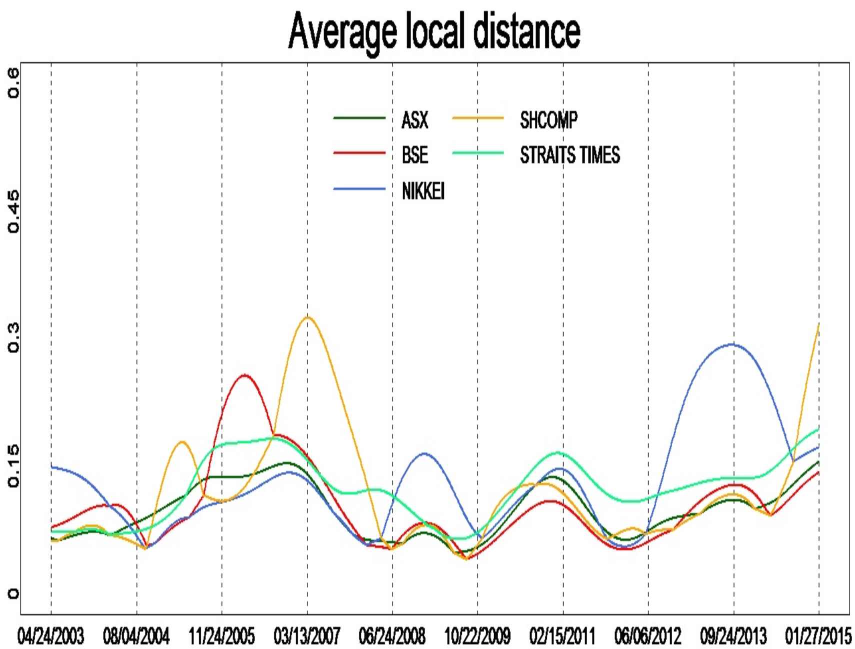


Figure 5.6: ***(Local distance measure) Australia, China, India, Japan and Singapur***

with $u = 1, \dots, p$, into L sets denoted as $M(z) = (M_1(z), \dots, M_L(z))$ so as to minimize the within-cluster sum of squares

$$\underset{M(z)}{\operatorname{argmin}} \sum_{i=1}^{L} \sum_{B^{(u)}(z) \in M_i(z)} \left| \left| B^{(u)}(z) - \bar{B}^{(i)}(z) \right| \right|_2$$

where $\bar{B}^{(i)}(z)$ is the mean factor loadings vector of the elements in $M_i(z)$.

We use this adaptive clustering approach to provide a time-aggregated measure of the degree of similarity of the factor loading structure of pairs of countries by counting the number of time points for which two given coun-

Index	Country
AEX	Netherlands
ATHEX	Greece
ATX	Austria
BEL20	Belgium
CAC40	France
DAX	Germany
MIB	Italy
IBEX	Spain
ISEQ	Ireland
OMX	Finland
PSI	Portugal

Table 5.4: ***Stock indices of ten of the founding members of the European currency area***

	AEX	ATHEX	ATX	BEL20	CAC40	DAX	MIB	IBEX	ISEQ	OMX	PSI
AEX	1	0.01	0.15	0.68	0.89	0.64	0.74	0.76	0.22	0.38	0.32
ATHEX	0.01	1	0.38	0.04	0.01	0.04	0.02	0.01	0.30	0.22	0.34
ATX	0.15	0.38	1	0.28	0.16	0.26	0.26	0.28	0.52	0.45	0.34
BEL20	0.68	0.04	0.28	1	0.77	0.66	0.68	0.71	0.27	0.32	0.22
CAC40	0.89	0.01	0.16	0.77	1	0.72	0.75	0.79	0.25	0.32	0.22
DAX	0.64	0.04	0.26	0.66	0.72	1	0.65	0.65	0.40	0.29	0.15
MIB	0.74	0.02	0.26	0.68	0.75	0.65	1	0.81	0.36	0.26	0.27
IBEX	0.76	0.01	0.28	0.71	0.79	0.65	0.81	1	0.32	0.37	0.22
ISEQ	0.22	0.30	0.52	0.27	0.25	0.40	0.36	0.32	1	0.47	0.27
OMX	0.38	0.22	0.45	0.32	0.32	0.29	0.26	0.37	0.47	1	0.45
PSI	0.32	0.34	0.34	0.22	0.22	0.15	0.27	0.22	0.27	0.45	1

Table 5.5: ***Adaptive k-means clustering of Euro member states***

tries are located within the same cluster. Thus, we compute the proportion

$$P^{(u,v)} = \frac{\sum_{z} \sum_{i=1}^{l} \mathbb{I}\left(B^{(u)}(z) \in M_i(z)\right) \mathbb{I}\left(B^{(v)}(z) \in M_i(z)\right)}{n}$$

where $\mathbb{I}(x)$ denotes the indicator function. The larger the proportion $P^{(u,v)}$ for a given pair of countries, the more similar their factor loading patterns over a given time horizon. Table 5.5 contains the computed values of $P^{(u,v)}$ for all pairs of countries using the log-return data from July 2002 until April 2015. Figure 5.9 contains the representation of the information contained in Table 5.5 under the form of a pixel image.

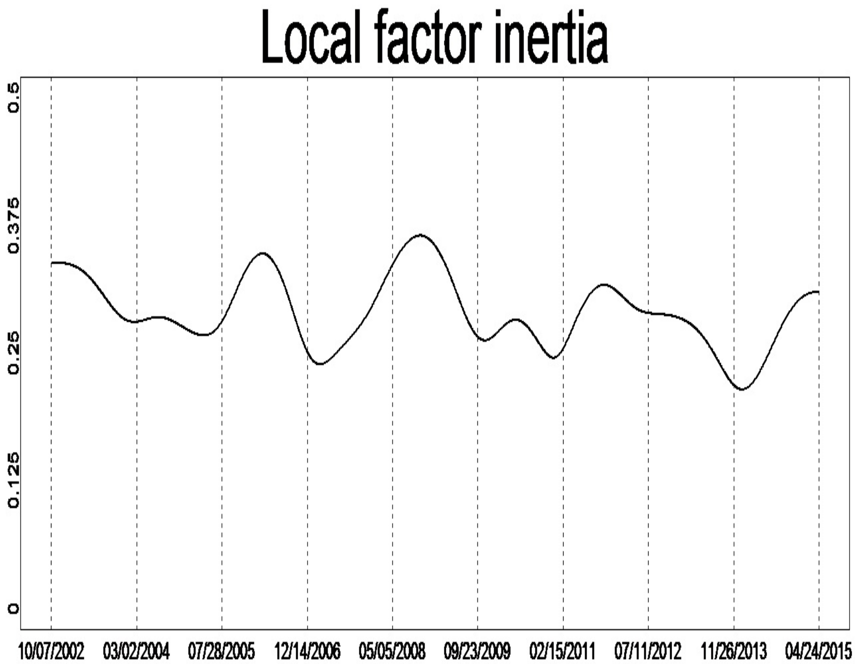


Figure 5.7: ***(Local factor inertia) European currency area***

Note for example that within the group of countries, Netherlands. Belgium, France, Germany, Italy and Spain, the value of $P^{(u,v)}$ never drops below 0.65, which may allow to conclude that the economic integration among those countries is quite strong compared to the group of remaining countries for which $P^{(u,v)}$ does not exceed 0.47.

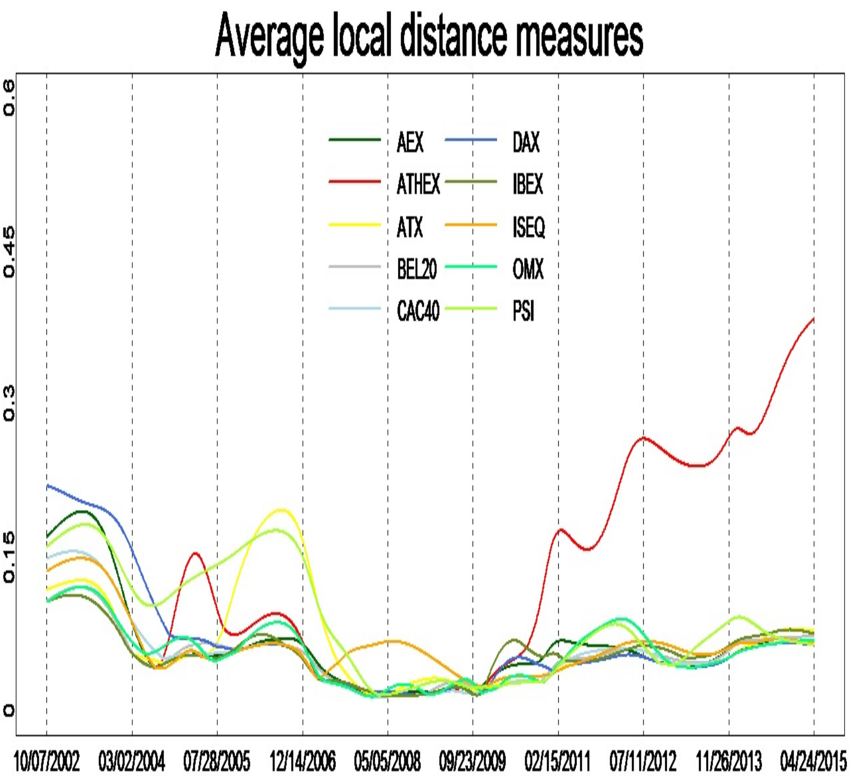


Figure 5.8: ***(Local distance measure) European currency area***

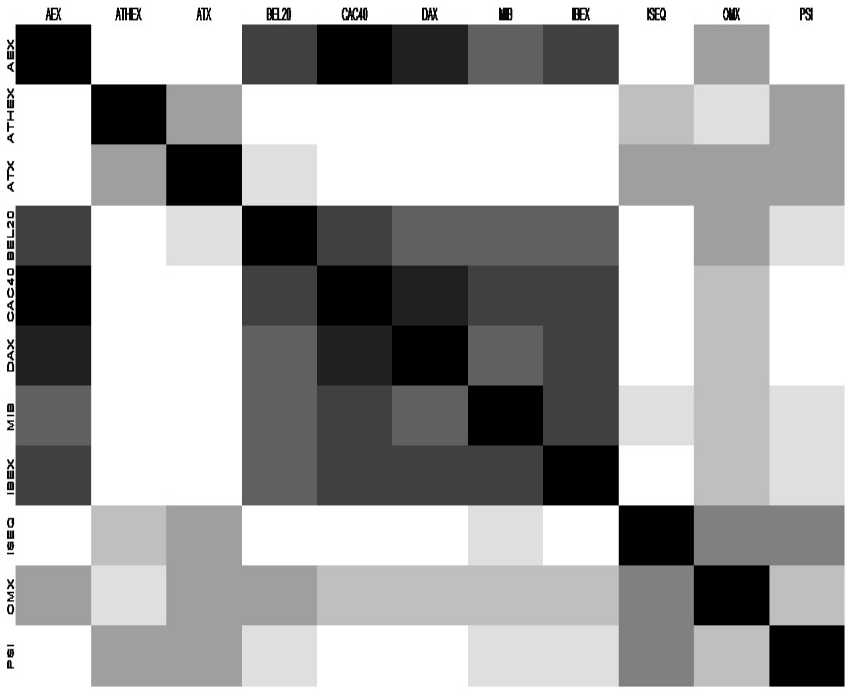


Figure 5.9: ***Adaptive k-means clustering of Euro member states***

5.5 Proofs

5.5.1 Proof of Theorem 5.1

Given the factor model representation of the increment process we can write $S_n(z) = B_n(z)B_n^t(z) + \Sigma_n(z)$. Given the assumption that $S_n(z)$, $B_n(z)B_n^t(z)$ and $\Sigma_n(z)$ are definite positive matrices we have that for any $\log_2(n)p \times 1-$vector x the inequality $x^t S_n(z)x \geq x^t B_n(z)B_n^t(z)x$ holds. Hence, we have that $\lambda_m(S_n(z)) \geq ||B_{n,m}(z)||^2$. Thus, given the persistence assumption on the common factors in Definition 5.1 and using that $p = O(n^{\alpha})$ we obtain that $\lambda_m(S_n(z)) = O(n^{\alpha\kappa})$ for $m = 1, \ldots, K$.

Moreover, given that only the first K eigenvalues of $B_n(z)B_n^t(z)$ are non-zero, we have that for $m > K$ the equality $\lambda_m(S_n(z)) = \lambda_m(\Sigma_n(z))$ holds. Hence, given the non-persistence assumption on the idiosyncratic factors in Definition 5.1, we have that $\lambda_m(S_n(z)) = O(1)$.

5.5.2 Proof of Theorem 5.2

The Davis-Kahan $\sin(\theta)$ theorem for Hermitian matrices (see Davis and Kahan (1970)) allows to write

$$\left|\left|u_m(S_n(z)) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| \leq \frac{\sqrt{2}||\Sigma_n(z)||}{\min\left(\left|\lambda_{m-1}(S_n(z)) - ||B_{n,m}(z)||^2\right|, \left|||B_{n,m}(z)||^2 - \lambda_{m+1}(S_n(z))\right|\right)}$$

Use the triangle inequality to establish the lower bounds as follows

$$\left|\lambda_{m-1}(S_n(z)) - ||B_{n,m}(z)||^2\right| \geq \left|||B_{n,m-1}(z)||^2 - ||B_{n,m}(z)||^2\right| - \left|||B_{n,m-1}(z)||^2 - \lambda_{m-1}(S_n(z))\right|$$

$$\left|||B_{n,m}(z)||^2 - \lambda_{m+1}(S_n(z))\right| \geq \left|||B_{n,m}(z)||^2 - ||B_{n,m+1}(z)||^2\right| - \left|\lambda_{m+1}(S_n(z)) - ||B_{n,m+1}(z)||^2\right|.$$

Using Weyl's inequalities, we have that $\left|||B_{n,m-1}(z)||^2 - \lambda_{m-1}(S_n(z))\right|$ and $\left|\lambda_{m+1}(S_n(z)) - ||B_{n,m+1}(z)||^2\right|$ can be bounded from above by $\lambda_m(\Sigma_n(z))$. Given the assumption that $\lambda_m(\Sigma_n(z)) = O(1)$ for all m we thus have that

$$\left| ||B_{n,m-1}(z)||^2 - \lambda_{m-1}\left(S_n(z)\right) \right| = O(1)$$

$$\left| \lambda_{m+1}\left(S_n(z)\right) - ||B_{n,m+1}(z)||^2 \right| = O(1).$$

Hence, under the assumption that $\left| ||B_{n,m-1}(z)||^2 - ||B_{n,m}(z)||^2 \right| = O\left(p^{\kappa}\right)$ and
$\left| ||B_{n,m}(z)||^2 - ||B_{n,m+1}(z)||^2 \right| = O\left(p^{\kappa}\right)$ for $m = 1, \ldots, K$ and given that $p = O\left(n^{\alpha}\right)$, we obtain that

$$\left|\left| u_m\left(S_n(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||} \right|\right| = O\left(\frac{1}{n^{\alpha\kappa}}\right).$$

5.5.3 Proof of Theorem 5.3

To start with, we bound $\left| \frac{\sum_{m=1}^{K} \lambda_m(S_n)}{\sum_{m=1}^{\log_2(n)p} \lambda_m(S_n)} - I_F(z) \right|$ from above by

$$\frac{\sum_{m=1}^{K} \left| \lambda_m(S_n(z)) - ||B_{n,m}(z)||^2 \right|}{\sum_{m=1}^{\log_2(n)p} \lambda_m(S_n(z))} +$$

$$\frac{\sum_{m=1}^{K} ||B_{n,m}(z)||^2 \sum_{m=1}^{K} \left| ||B_{n,m}(z)||^2 - \lambda_m(S_n(z)) \right|}{\left(\sum_{m=1}^{K} ||B_{n,m}(z)||^2 + \sum_{u=1}^{p} \sum_{j=-\log_2(n)}^{-1} \left(\sigma_{j,j}^{(u,u)}\right)^2 \right) \left(\sum_{m=1}^{\log_2(n)p} \lambda_m(S_n) \right)}$$

Based on the decomposition of the EWS matrix in (5.1.9) and using Weyl's inequalities, we obtain that $\left| \lambda_m\left(S_n\right) - ||B_{n,m}(z)||^2 \right| \leq \lambda_1\left(\Sigma_n(z)\right)$. Given the assumption in (5.1.10) we have that $\lambda_1\left(\Sigma_n(z)\right) = O(1)$. Using the persistence assumption in (5.1.7) we have that $\sum_{m=1}^{K} ||B_{n,m}(z)||^2 = O\left(p^{\kappa}\right)$. Thus, given that $\sum_{m=1}^{\log_2(n)p} \lambda_m\left(S_n\right) = O(p^{\max(\kappa,\eta)})$ and using that $p = O\left(n^{\alpha}\right)$, we have that

$$\left| \frac{\sum_{m=1}^{K} \lambda_m\left(S_n\right)}{\sum_{m=1}^{\log_2(n)p} \lambda_m\left(S_n\right)} - I_F(z) \right| = O\left(\frac{1}{n^{\alpha \max(\kappa,\eta)}}\right)$$

5.5.4 Proof of Theorem 5.4

The inverse triangle inequality of the Frobenius norm allows to establish the upper bound

$$\Big| ||\overline{\Delta}_n^{(u)}(z) - \overline{\Delta}_n^{(v)}(z)||_F - ||\Delta_n^{(u)}(z) - \Delta_n^{(u)}(z)||_F \Big| \leq \\ ||\overline{\Delta}_n^{(u)}(z) - \Delta_n^{(u)}(z) + \Delta_n^{(v)}(z) - \overline{\Delta}_n^{(v)}(z)||_F.$$

Using the triangle inequality of the Frobenius norm we can write the upper bound as $||\overline{\Delta}_n^{(u)}(z) - \Delta_n^{(u)}(z)||_F + ||\Delta_n^{(v)}(z) - \overline{\Delta}_n^{(v)}(z)||_F$. If we denote $\Delta_{n,m}^{(u)}(z) = \frac{B_{n,m}^{(u)}(z)}{||B_{n,m}(z)||_2}$ and $\overline{\Delta}_{n,m}^{(u)}(z) = U_{n,m}^{(u)}(z)$, we can write $||\overline{\Delta}_n^{(u)}(z) - \Delta_n^{(u)}(z)||_F = \sqrt{\sum_{m=1}^K ||\overline{\Delta}_{n,m}^{(u)}(z) - \Delta_{n,m}^{(u)}(z)||_2^2}$. Hence, we have that

$$\begin{aligned} ||\overline{\Delta}_n^{(u)}(z) - \Delta_n^{(u)}(z)||_F &\leq \sum_{m=1}^K ||\overline{\Delta}_{n,m}^{(u)}(z) - \Delta_{n,m}^{(u)}(z)||_2 \\ ||\overline{\Delta}_n^{(v)}(z) - \Delta_n^{(v)}(z)||_F &\leq \sum_{m=1}^K ||\overline{\Delta}_{n,m}^{(v)}(z) - \Delta_{n,m}^{(v)}(z)||_2 \end{aligned}$$

Given Theorem 5.2 we have that $||\overline{\Delta}_{n,m}^{(u)}(z) - \Delta_{n,m}^{(u)}(z)||_2 = O\left(\frac{1}{n^{\alpha\kappa}}\right)$ and $||\overline{\Delta}_{n,m}^{(u)}(z) - \Delta_{n,m}^{(u)}(z)||_2 = O\left(\frac{1}{n^{\alpha\kappa}}\right)$ and hence

$$\Big| ||\overline{\Delta}_n^{(u)}(z) - \overline{\Delta}_n^{(v)}(z)||_F - ||\Delta_n^{(u)}(z) - \Delta_n^{(u)}(z)||_F \Big| = O\left(\frac{1}{n^{\alpha\kappa}}\right).$$

5.5.5 Proof of Theorem 5.5

Take $\left\{\lambda_m\left(\tilde{I}_{n,M}(z)\right)\right\}_{m=1}^{\log_2(n)p}$ and $\left\{\lambda_m\left(S_n(z)\right)\right\}_{m=1}^{\log_2(n)p}$ to be the sets of decreasingly ordered eigenvalues of $\tilde{I}_{n,M}(z)$ and $S_n(z)$ respectively. First, note that

$$\left\{\hat{K} \neq K\right\} \Leftrightarrow \left\{\lambda_K\left(\tilde{I}_{n,M}(z)\right) \leq n^\gamma\right\} \cup \left\{\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) > n^\gamma\right\}$$

and hence that $P\left(\left\{\hat{K} \neq K\right\}\right) \leq P\left(\lambda_K\left(\tilde{I}_{n,M}(z)\right) \leq n^\gamma\right) + P\left(\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) > n^\gamma\right)$.

Consider the random variable $Y_{n,M}(z) = \lambda_K\left(\tilde{I}_{n,M}(z)\right) - \lambda_K\left(S_n(z)\right)$ with

$E\left(Y_{n,M}(z)\right) = O\left(\frac{\log_2^2(n)}{n^{\frac{1}{3}-\alpha}}\right)$ and $\mathrm{Var}\left(Y_{n,M}(z)\right) = O\left(\frac{\log_4^2(n)}{n^{2\left(\frac{1}{3}-\alpha\right)}}\right)$ and write

$$\begin{aligned} &P\left(\lambda_K\left(\tilde{I}_{n,M}(z)\right) \leq n^{\gamma}\right) \\ &\quad = P\left(\frac{Y_{n,M}(z)-E\left(Y_{n,M}(z)\right)}{\sqrt{\mathrm{Var}\left(Y_{n,M}(z)\right)}} \frac{n^{\gamma}-\lambda_K\left(S_n(z)\right)-E\left(Y_{n,M}(z)\right)}{\sqrt{\mathrm{Var}\left(Y_{n,M}(z)\right)}}\right) \end{aligned}$$

Given the statement of Theorem 5.1 we have that $\lambda_K\left(S_n(z)\right) = O\left(n^{\kappa}\right)$ and hence if $\kappa > \gamma$ we have that $\frac{n^{\gamma}-\lambda_K\left(S_n(z)\right)-E\left(Y_{n,M}(z)\right)}{\sqrt{\mathrm{Var}\left(Y_{n,M}(z)\right)}} \to -\infty$ and thus $P\left(\lambda_K\left(\tilde{I}_{n,M}(z)\right) \leq n^{\gamma}\right) \to 0$.

Next, use the triangle inequality of the absolute value and Markov's inequality to establish the subsequent upper bounds

$$\begin{aligned} P\left(\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) > n^{\gamma}\right) &\leq P\left(\left|\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) - \lambda_{K+1}\left(S_n(z)\right)\right| + \lambda_{K+1}\left(S_n(z)\right) > n^{\gamma}\right) \\ &\leq \frac{\sqrt{E\left(\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_{K+1}\left(S_n(z)\right)\right)^2}+\lambda_{K+1}\left(S_n(z)\right)}{n^{\gamma}}. \end{aligned}$$

Using the statement of Theorem 5.8 we have that $\lambda_{K+1}\left(S_n(z)\right) = O(1)$. Moreover, $\sqrt{E\left(\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) - \lambda_{K+1}\left(S_n(z)\right)\right)^2} = O\left(\frac{\log_2^2(n)}{n^{\frac{1}{3}-\alpha}}\right)$ and hence if $\gamma > 0$ we obtain that

$$P\left(\lambda_{K+1}\left(\tilde{I}_{n,M}(z)\right) > n^{\gamma}\right) \to 0.$$

5.5.6 Proof of Theorem 5.6

Use the triangle inequality of the Euclidean norm to write

$$\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| \leq \left|\left|u_m\left(\tilde{I}_{n,M}(z)\right) - u_m\left(S_n(z)\right)\right|\right| + \left|\left|u_m\left(S_n(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right|.$$

Thus, we can bound $P\left(\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| > \varepsilon\right)$ from above by

$$P\left(\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right) - u_m\left(S_n(z)\right)\right|\right| + \left|\left|u_m\left(S_n(z)\right) - \frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right| > \varepsilon\right)$$

Using the Davis-Kahan $\sin(\theta)$ theorem and the Cauchy-Schwarz inequality we have that

$$E\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right)-u_m\left(S_n(z)\right)\right|\right| \leq \left(\frac{E\left(\lambda_1\left(\tilde{I}_{n,M}(z)-S_n(z)\right)^2\right)}{E\min\left(\left|\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_m\left(S_n(z)\right)\right|,\left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)\right|\right)^2}\right)^{\frac{1}{2}}$$

Use the triangle inequality to establish the lower bounds as follows

$$\begin{aligned}
&\left|\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_m\left(S_n(z)\right)\right| \geq \\
&\qquad \left|\lambda_{m-1}\left(S_n(z)\right)-\lambda_m\left(S_n(z)\right)\right|-\left|\lambda_{m-1}\left(S_n(z)\right)-\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)\right| \\
&\left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)\right| \geq \\
&\qquad \left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(S_n(z)\right)\right|-\left|\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_{m+1}\left(S_n(z)\right)\right|
\end{aligned}$$

Using Weyl's inequality, we have that both, $\left|\lambda_{m-1}\left(S_n(z)\right)-\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)\right|$ and $\left|\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_{m+1}\left(S_n(z)\right)\right|$, can be bounded from above by $\lambda_1\left(\tilde{I}_{n,M}(z)-S_n(z)\right)$. This allows to bound

$$E\min\left(\left|\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_m\left(S_n(z)\right)\right|,\left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)\right|\right)^2$$

from below by

$$E\left(\min\left(\left|\lambda_{m-1}\left(S_n(z)\right)-\lambda_m\left(S_n(z)\right)\right|,\left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(S_n(z)\right)\right|\right)-\lambda_1\left(\tilde{I}_{n,M}(z)-S_n(z)\right)\right)^2.$$

Given the assumption in (5.1.7) we have that $\left|\lambda_{m-1}\left(S_n(z)\right)-\lambda_m\left(S_n(z)\right)\right| = O\left(n^{\alpha\kappa}\right)$ and $\left|\lambda_m\left(S_n(z)\right)-\lambda_{m+1}\left(S_n(z)\right)\right| = O\left(n^{\alpha\kappa}\right)$. Thus, the lower bound can be rewritten as $E\left(O\left(n^{\alpha\kappa}\right)-\lambda_1\left(\tilde{I}_{n,M}(z)-S_{J(n)}(z)\right)\right)^2 = O\left(n^{2\kappa\alpha}\right) - 2O\left(n^{\kappa\alpha}\right)E\left(\lambda_1\left(\tilde{I}_{n,M}(z)-S_n(z)\right)\right)+E\left(\lambda_1\left(\tilde{I}_{n,M}(z)-S_n(z)\right)\right)^2$. Using the statement of Theorem 4.8 we have that $E\left(\lambda_1\left(\tilde{I}_{n,M}(z)-S_{J(n)}(z)\right)\right)=O\left(\frac{\log_2(n)}{n^{\frac{1}{3}-\alpha}}\right)$ and $E\left(\lambda_1\left(\tilde{I}_{n,M}(z)-S_{J(n)}(z)\right)\right)^2 = O\left(\frac{\log_2^4(n)}{n^{2\left(\frac{1}{3}-\alpha\right)}}\right)$. Thus, we obtain that

$$E\min\left(\left|\lambda_{m-1}\left(\tilde{I}_{n,M}(z)\right)-\lambda_m\left(S_{J(n)}(z)\right)\right|^2,\left|\lambda_m\left(S_{J(n)}(z)\right)-\lambda_{m+1}\left(\tilde{I}_{n,M}(z)\right)\right|\right)^2 = O\left(n^{\alpha\kappa}\right).$$

Using Chebyshev's inequality we hence have that

$$P\left(\left|\left|u_m\left(\tilde{I}_{n,M}(z)\right)-\frac{B_{n,m}(z)}{||B_{n,m}(z)||}\right|\right|>\varepsilon\right)=O\left(\frac{\log_2^2(n)}{n^{\frac{1}{3}+\alpha(\frac{\kappa}{2}-1)}}\right)$$

5.5.7 Proof of Theorem 5.7

Use the triangle inequality of the absolute value to bound $\left|\frac{\sum_{m=1}^{K}\lambda_m\left(\tilde{I}_{n,M}\right)}{\sum_{m=1}^{J(n)p}\hat{\lambda}_m\left(S_{J(n)}\right)}-I_F(z)\right|$ from above by

$$\left|\frac{\sum_{m=1}^{K}\left(\lambda_m\left(\tilde{I}_{n,M}(z)\right)-\lambda_m(S_n(z))\right)}{\sum_{m=1}^{\log_2(n)p}\lambda_m\left(\tilde{I}_{n,M}(z)\right)}\right|+\left|\frac{\sum_{m=1}^{K}\lambda_m(S_n(z))\sum_{m=1}^{\log_2(n)p}\lambda_m\left(\tilde{I}_{n,M}(z)\right)-\lambda_m(S_n(z))}{\sum_{m=1}^{\log_2(n)p}\lambda_m\left(\tilde{I}_{n,M}(z)\right)\sum_{m=1}^{\log_2(n)p}\lambda_m(S_n(z))}\right|$$
$$+\left|\frac{\sum_{m=1}^{K}\lambda_m(S_n(z))}{\sum_{m=1}^{\log_2(n)p}\lambda_m(S_n(z))}-I_F(z)\right|$$

Given the Cauchy-Schwarz inequality we can write the expected value of the random variable above as

$$\left|\frac{\sum_{m=1}^{K}\sum_{m'=1}^{K}\left(E\left(\lambda_m\left(\tilde{I}_{n,M}(z)\right)-\lambda_m(S_n(z))\right)^2\right)^{\frac{1}{2}}\left(E\left(\lambda_{m'}\left(\tilde{I}_{n,M}(z)\right)-\lambda_{m'}(S_n(z))\right)^2\right)^{\frac{1}{2}}}{\sum_{m=1}^{\log_2(n)p}\sum_{m'=1}^{\log_2(n)p}\left(E\left(\lambda_m\left(\tilde{I}_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}}E\left(\left(\lambda_{m'}\left(\tilde{I}_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}}}\right|+$$
$$\left|\frac{\sum_{m=1}^{K}\sum_{m=1}^{K}\lambda_m(S_n(z))\lambda_{m'}(S_n(z))\sum_{m=1}^{\log_2(n)p}\sum_{m'=1}^{\log_2(n)p}\left(E\left(\lambda_m\left(\tilde{I}_{n,M}(z)\right)-\lambda_m(S_n(z))\right)^2\right)^{\frac{1}{2}}\left(E\left(\lambda_{m'}\left(\tilde{I}_{n,M}(z)\right)-\lambda_{m'}(S_n(z))\right)^2\right)^{\frac{1}{2}}}{\sum_{m=1}^{\log_2(n)p}\sum_{m'=1}^{\log_2(n)p}\left(E\left(\lambda_m\left(\tilde{I}_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}}\left(E\left(\lambda_{m'}\left(\tilde{I}_{n,M}(z)\right)\right)^2\right)^{\frac{1}{2}}\sum_{m=1}^{\log_2(n)p}\sum_{m'=1}^{\log_2(n)p}\lambda_m(S_n(z))\lambda_{m'}(S_n(z))}\right|$$
$$+\left|\frac{\sum_{m=1}^{K}\lambda_m(S_n(z))}{\sum_{m=1}^{\log_2(n)p}\lambda_m(S_n(z))}-I_F(z)\right|$$

5.5.8 Proof of Theorem 5.8

Use the triangle inequality of the absolute value to establish the upper bound

$$\left|||\hat{\Delta}^{(u)}(z)-\hat{\Delta}^{(v)}(z)||_F-||\Delta^{(u)}(z)-\Delta^{(v)}(z)||_F\right|\leq$$
$$||\hat{\Delta}^{(u)}(z)-\Delta^{(u)}(z)||_F+||\Delta^{(v)}(z)-\hat{\Delta}^{(v)}(z)||_F$$

with $||\hat{\Delta}^{(u)}(z)-\Delta^{(u)}(z)||_F \leq ||\hat{\Delta}^{(u)}(z)-\bar{\Delta}^{(u)}(z)||_F+||\bar{\Delta}^{(u)}(z)-\Delta^{(u)}(z)||_F$ and $||\hat{\Delta}^{(v)}(z)-\Delta^{(v)}(z)||_F \leq ||\hat{\Delta}^{(v)}(z)-\bar{\Delta}^{(v)}(z)||_F+||\bar{\Delta}^{(v)}(z)-\Delta^{(v)}(z)||_F$. Note that $||\hat{\Delta}^{(u)}(z)-\bar{\Delta}^{(u)}(z)||_F+||\bar{\Delta}^{(u)}(z)-\Delta^{(u)}(z)||_F \leq$

$\sum_{m=1}^{K}\left(||\hat{\Delta}_m^{(u)}(z)-\bar{\Delta}_m^{(u)}(z)||_2+||\bar{\Delta}_m^{(u)}(z)-\Delta_m^{(u)}(z)||_2\right)$ and that

$$\begin{aligned}
P&\left(\sum_{m=1}^{K}\left(||\hat{\Delta}_m^{(u)}(z)-\bar{\Delta}_m^{(u)}(z)||_2+||\bar{\Delta}_m^{(u)}(z)-\Delta_m^{(u)}(z)||_2\right)\geq\varepsilon\right)\\
&\qquad\leq\sum_{m=1}^{K}P\left(\left(||\hat{\Delta}_m^{(u)}(z)-\bar{\Delta}_m^{(u)}(z)||_2+||\bar{\Delta}_m^{(u)}(z)-\Delta_m^{(u)}(z)||_2\right)\geq\varepsilon\right).\\
&\qquad\qquad\leq\sum_{m=1}^{K}\frac{E\left(||\hat{\Delta}_m^{(u)}(z)\right)-\bar{\Delta}_m^{(u)}(z)||_2+||\bar{\Delta}_m^{(u)}(z)-\Delta_m^{(u)}(z)||_2}{\varepsilon}.
\end{aligned}$$

Chapter 6

Wavelet-based estimator of high-dimensional covariance matrices

Classic multivariate analysis is based on the assumption that, given some a priori knowledge, the set of relevant variables for a given statistical analysis can be derived in advance. However, in most statistical models the subset of relevant variables is not known beforehand. In these cases, we face a large set of possibly relevant variables instead of a small set of effectively relevant variables. This potentially leads to a large number of variables within the model framework and hence to a high-dimensional dataset. In Section 4.3 we have introduced high-dimensional MLSW processes. We have proven that in order for the smoothed and corrected periodogram estimator of the EWS matrix to be spectral norm consistent in high-dimensional settings, we need to have that $0 \leq \alpha \leq \frac{1}{3}$ with $p = O(n^{\alpha})$ where p denotes the number of constituent time series of the MLSW framework and n denotes the sample size. Hence, even though the number of constituent time series in the MLSW model can be potentially quite large, the ratio $\frac{p}{n}$ still needs to take on low values and tend to zero as both, n and p go to infinity. The small upper bound for α has negative effects on the performance of the LSW factor modelling approach introduced in Chapter 5. Indeed, if we have a look at the statements of the theorems in Section 5.3 , we see that a necessary condition for the convergence in probabilities of the respective estimators is given by $0 \leq \alpha < \frac{1}{3}$. In order to obtain a larger upper bound for α we have to move away from standard multivariate modelling techniques and to adopt a high-dimensional viewpoint. The latter consists in introducing some kind of model selection procedure with the aim of reducing the number of parameters to be estimated.

This chapter deals with the general topic of the use of model selection procedures in the estimation of covariance matrices in high-dimensional settings. The concepts developed in this chapter differ from the concepts of the preceding chapters in the sense that we leave the LSW model framework and introduce the assumption that the underlying multivariate process is white noise and hence has no serial correlation.

As it has been mentioned above, the main difference between classic multivariate and high-dimensional time series analysis is that the cardinality of the set of variables that play a role in the analysis is not limited per se. One classic problem in which the set of variables could be very large, is the issue of optimal asset allocation in finance. Indeed, apart from settings in which the analyst discards some financial assets or even whole asset classes in advance, the set of variables, which might have a relevance for the the construction of optimal portfolios, potentially contains all existing assets. In general, asset allocation and portfolio optimization are one of the fundamental problems of modern finance. Market participants base their investment decisions on a balance between risk and return in the attempt to maximize portfolio expected return for a given amount of portfolio risk, or equivalently minimize risk for a given level of expected return, by carefully choosing the proportions of various assets. The ground-breaking work on mean-variance analysis by Markowitz (1952) [89] proposes a mathematical formulation of this concept of diversification by translating the trade-off between risk and return into a quadratic programming problem. Since then, one of the main theoretical and practical criticism that has been leveled against the mean-variance approach is the ill-conditionedness of the optimization framework which manifests itself in a very high sensitivity towards estimation errors in the input parameters. Due to this large degree of sensitivity even small estimation errors in the input parameters can lead to a very large error in the solution vector of the optimization framework. However, most portfolio optimization problems involve a very large set of assets compared to a relatively small number of data points which leads to important estimation errors and hence to wrong investment decisions. Hence, reducing these estimation errors in the input parameters, especially in high-dimensional settings is one of the fundamental issues in modern portfolio optimization.

Several methods have been suggested to reduce the estimation error of the input parameters. Ledoit and Wolf (2004) [74] proposes a shrinkage-type estimator of the covariance matrix using a factor model as a shrinkage target. El Karoui (2008) [44] and Bickel and Levina (2008) assume that the population covariance matrix has a sparse structure and propose a thresholding approach for the sample covariance matrix. Cai and Liu (2011) [17] propose a thresholding procedure which is adaptive to the variability of individual entries. Brodie et al. (2009) [20] reformulate the portfolio optimization problem as a constrained least-squares regression problem under an L_1-penalty. They illustrate that the $L1$-penalty stabilizes the optimization framework and leads to sparse asset allocation schemes. Fan et. al (2012) [49] uses an approximate factor model combined with a sparsity assumption on the covariance matrix of the idiosyncratic components.

6.1 Maximum likelihood estimator in high-dimensional settings

We denote as $\{X(t)\}_{t=0}^{n-1}$ a $p-$dimensional random process with independent and identically distributed random vectors $X(t) \in \mathbb{R}^p$ such that $E(X(t)) = \mu_p$ and $\text{Var}(X(t)) = \Sigma_p$. Denote as $T(\hat{\Sigma}_p)$ some estimator of the covariance matrix Σ_p. Define the matrix $\Xi_p = T(\hat{\Sigma}_p) - \Sigma_p$ and denote as $||\Xi_p|| = \lambda_{\max}(\Xi_p)$ its spectral norm with $\lambda_{\max}(\Xi_p)$ being the largest eigenvalue of Ξ_p.

Definition 6.1. *We say that $T\left(\hat{\Sigma}_p\right)$ is a spectral norm consistent estimator of Σ_p if for any $\varepsilon > 0$ we have that*

$$\lim_{\substack{n \to +\infty \\ p \to +\infty}} P\left(||\Xi_p|| > \varepsilon\right) = 0.$$

Denote as $\lambda_i\left(T\left(\hat{\Sigma}_p\right)\right)$ and as $\lambda_i\left(\Sigma_p\right)$ the decreasing eigenvalues of the estimated covariance matrix and the population covariance matrix respectively with $i = 1, \ldots, p$. Using Weyl's inequality for the sums of Hermitian matrices (see Weyl, 1912) [130] we have that

$$\max_i \left|\lambda_i\left(T\left(\hat{\Sigma}_p\right)\right) - \lambda_i\left(\Sigma_p\right)\right| \leq \left|\left|\Xi_p\right|\right|.$$

Hence given Definition 6.1, this means that if $T\left(\hat{\Sigma}_p\right)$ is a spectral norm consistent estimator of Σ_p we have that

$$\lim_{\substack{n\to+\infty \\ p\to+\infty}} P\left(\max_i \left|\lambda_i\left(T\left(\hat{\Sigma}_p\right)\right) - \lambda_i\left(\Sigma_p\right)\right| > \varepsilon\right) = 0$$

as $n \to +\infty$ and $p \to +\infty$. Denote as $u_i\left(T\left(\hat{\Sigma}_p\right)\right)$ and $u_i\left(\Sigma_p\right)$ the eigenvectors of $T\left(\hat{\Sigma}_p\right)$ and Σ_p respectively with $i = 1,\ldots,p$. Using the Davis and Kahan $\sin(\theta)$ theorem (see Davis and Kahan, 1970 [29]) we have that the Euclidean norm of the vector $u_i\left(T\left(\hat{\Sigma}_p\right)\right) - u_i\left(\Sigma_p\right)$ is bounded from above as

$$\left|\left|u_i\left(T\left(\hat{\Sigma}_p\right)\right) - u_i\left(\Sigma_p\right)\right|\right| \leq \frac{\sqrt{2}||\Xi_p||}{\min\left(\left|\lambda_{i-1}\left(T\left(\hat{\Sigma}_p\right)\right) - \lambda_i\left(\Sigma_p\right)\right|, \left|\lambda_i\left(\Sigma_p\right)\right) - \lambda_{i+1}\left(T\left(\hat{\Sigma}_p\right)\right|\right)}$$

where

$$\left|\lambda_{i-1}\left(T\left(\hat{\Sigma}_p\right)\right) - \lambda_i\left(\Sigma_p\right)\right| \geq \left|\lambda_{i-1}\left(\Sigma_p\right) - \lambda_i\left(\Sigma_p\right)\right| - \left|\lambda_{i-1}\left(\Sigma_p\right) - \lambda_{i-1}\left(T\left(\hat{\Sigma}_p\right)\right)\right|$$

$$\left|\lambda_i\left(\Sigma_p\right) - \lambda_{i+1}\left(T\left(\hat{\Sigma}_p\right)\right)\right| \geq \left|\lambda_i\left(\Sigma_p\right) - \lambda_{i+1}\left(\Sigma_p\right)\right| - \left|\lambda_{i+1}\left(T\left(\hat{\Sigma}_p\right)\right) - \lambda_{i+1}\left(\Sigma_p\right)\right|.$$

Thus, if $T\left(\hat{\Sigma}_p\right)$ is a spectral norm consistent estimator of Σ_p, we obtain that for $i = 1,\ldots,p$

$$\lim_{\substack{n\to+\infty \\ p\to+\infty}} P\left(\left|u_i\left(T\left(\hat{\Sigma}_p\right)\right) - u_i\left(\Sigma_p\right)\right| > \varepsilon\right) = 0.$$

The maximum likelihood estimator (MLE) of Σ is given by

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{t=0}^{n-1}\left(X(t) - \hat{\mu}\right)\left(X(t) - \hat{\mu}\right)^t \tag{6.1.1}$$

where $\hat{\mu} = \frac{1}{n}\sum_{t=0}^{n-1} X(t)$ denotes the sample mean. The MLE estimator is not consistent in spectral norm. To illustrate this consider for example the case of a stationary $p-$dimensional time series with $X \sim N(0, I_p)$ where I_p denotes the identity matrix of dimension $p \times p$ with eigenvalues given by $\lambda_i = 1$ for $i = 1, \ldots, p$. Let $\hat{\lambda}_i$ denote the decreasing eigenvalues of the MLE estimator $\hat{\Sigma}$. We denote as $\rho_n = \frac{p}{n}$ the ratio of number of assets to the sample size with the assumption that $\rho_n \to \rho \in (0,1]$ as $n \to +\infty$ and $p \to +\infty$. Consider the random measure $F_p(x) = \frac{1}{p}\sum_{i=1}^{p} \mathbb{I}(\hat{\lambda}_i \leq x)$. Marchenko and Pastur (1967) [88] show that as $n \to +\infty$ and $p \to +\infty$

$$F_p \to F^{(\rho)} \text{ in probability}$$

where $F^{(\rho)}$ has known density with support $\left[\left(1-\rho^{\frac{1}{2}}\right)^2, \left(1+\rho^{\frac{1}{2}}\right)^2\right]$.

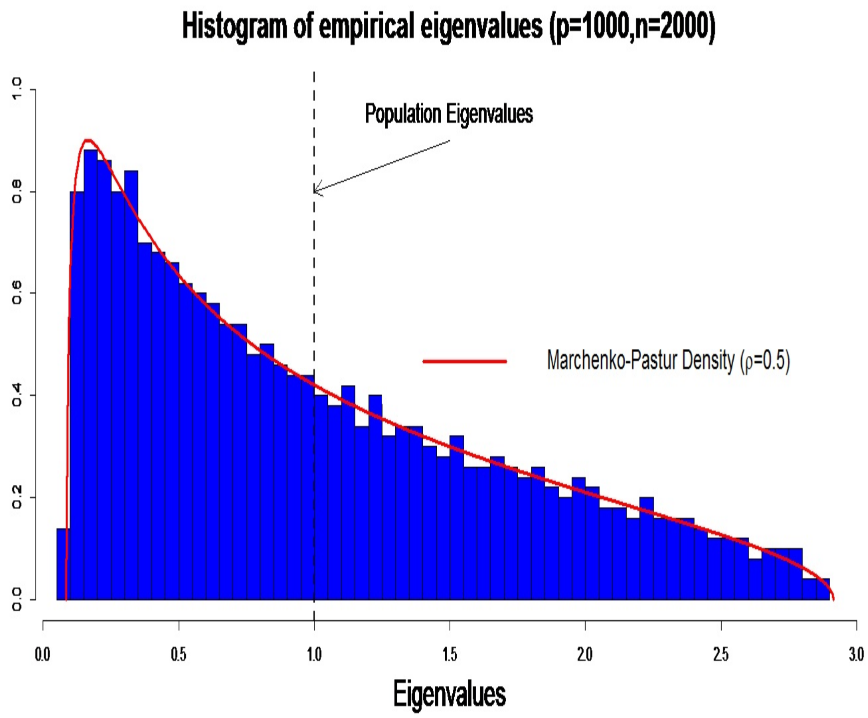


Figure 6.1: ***Illustration of the Marchenko-Pastur law***

6.2 Sparse estimation of covariance matrices in high-dimensional settings

Bickel and Levina (2008) [17] propose an estimator of the population covariance matrix which consists in putting to zero the components of the sample estimator which are far away from the diagonal by defining a banding operator, taking the form

$$\tau_B(\hat{\Sigma}_{i,j}) = \begin{cases} 0 & \text{if } |i-j| > k \\ \hat{\Sigma}_{i,j} & \text{if } |i-j| \leq k \end{cases} .$$

The authors show that if Σ belongs to the class of approximately bandable matrices that is

$$\mathscr{C} = \left\{ \Sigma : \max_j \left\{ \sum_{|i-j|>m} \Sigma_{i,j} \right\} < Cm^{-\alpha}, \ \forall m \right\}$$

with $m \asymp \left(n^{-\frac{1}{2}} \log(p)\right)^{\frac{-1}{\alpha+1}}$ and given certain tail decay conditions for the elements of $X(t)$ then

$$\left|\left|\tau_B(\hat{\Sigma}_p) - \Sigma_p\right|\right| \underset{\mathrm{p}}{\rightarrow} 0$$

as $n \to +\infty$ and $p \to +\infty$.

Note that the form of the sparsity described in Bickel and Levina (2008) [17] depends on the order of the variables in the random vector $X(t)$. El Karoui (2008) [44] defines a notion of sparsity that is order-independent based on the concept of $\beta-$sparsity. This more flexible type of sparsity allows to define a componentwise thresholding operator

$$\tau_E(\hat{\Sigma}_{i,j}) = \begin{cases} 0 & \text{if } |\hat{\Sigma}_{i,j}| \leq \theta \\ \hat{\sigma}_{i,j} & \text{if } |\hat{\Sigma}_{i,j}| > \theta \end{cases} .$$

The author shows that under the assumption that Σ is β-sparse with $\beta < \frac{1}{2} - \eta$ and $\eta > 0$. If $\theta = Cn^{-\alpha}$ where $\alpha = \beta + \varepsilon, \ \varepsilon < \frac{\eta}{2}$ then

$$\left|\left|\tau_E(\hat{\Sigma}_p) - \Sigma_p\right|\right| \underset{\mathrm{p}}{\rightarrow} 0$$

as $n \to +\infty$ and $p \to +\infty$.

Note that the estimators introduced in Bickel and Levina (2008) [17] and El Karoui (2008) [44] are based on the inherent assumption that the population covariance matrix Σ_p is characterized by some form of sparsity. However, when it comes to asset return series, the assumption of zero-values components within Σ_p may be unrealistic. Indeed in a global economy, there exists a small set of common factors affecting all the market participants. To explain this a bit further, take the factor model representation introduced in Section 5.1. The exposure of all market participants towards the same common factors generates dependencies among the latter even if the market participants are not directly linked to each other through correlations of their corresponding idiosyncratic components. For a more detailed discussion on the inadequacy of the sparsity assumption in the case of financial time series see Fan et al. (2012) [49].

Thus, in order to account for the potential inadequacy of the sparsity assumption when dealing with covariance matrices of series of asset returns, we are going to develop an estimator of covariance matrices in high-dimensional settings which does not rely on an underlying sparsity of the population covariance matrix. Instead we assume that the population covariance matrix is blockwise constant which, given an appropriate choice of the wavelet basis, leads to a sparse representation of the latter in the transform domain.

6.3 Multiscale estimation of covariance matrices in high-dimensional settings

As mentioned above, the estimator relies on the presence of a blockwise constant pattern within the population covariance matrix. Section 6.3.1 introduces the notion of blockwise constant covariance matrices.

6.3.1 Blockwise constant matrices

Definition 6.2. *We say that the covariance matrix of the random vector $X(t)$ has a blockwise constant form if the elements of $X(t)$ can be partitioned into K random vectors $X(t) = (X^{(1)}, \dots, X^{(K)})^t$ where $X^{(i)} = (X_1^{(i)}, \dots, X_{N_i}^{(i)})^t$*

with $\sum_{i=1}^{K} N_i = p$ *such that*

$$Cov(X_r^{(i)}(t), X_s^{(j)}(t)) = \rho_{i,j} \quad \textit{for } r,s = 1,\ldots,N_i \textit{ and } i,j = 1,\ldots,K \textit{ with } i \neq j$$
$$Cov(X_r^{(i)}(t), X_s^{(i)}(t)) = \rho_{i,i} \quad \textit{for } i = 1,\ldots,K \textit{ and } r,s = 1,\ldots,N_i \textit{ with } r \neq s$$

To begin with, note that there is no natural order to the elements of the vector of asset returns. Hence, a given matrix might not have a blockwise constant form under some given permutation of the assets but, after a well-chosen rearrangement of its rows and columns, reveal an underlying blockwise constant form. Thus, we say that a covariance matrix Σ_p has an underlying blockwise constant form if there exists some permutation of $X^{(1)}(t), \ldots, X^{(p)}(t)$, denoted as π, with permutation matrix P

$$\begin{array}{ll} P(i,j) = 1 & \text{if} \quad j = \pi(i) \\ P(i,j) = 0 & \text{if} \quad j \neq \pi(i). \end{array}$$

such that $P\Sigma_p P^t$ is blockwise constant as described in Definition 6.2.

In order to find π in the presence of an underlying blockwise constant form we proceed as follows. To begin with, denote as D the $p \times p$ matrix containing the pairwise Euclidean distances between the columns of Σ.

$$\begin{array}{c} D_{i,j} = \frac{1}{p}\left(\Sigma_i - \Sigma_j\right)^t \left(\Sigma_i - \Sigma_j\right) \\ i = 1,\ldots,p \\ j = 1,\ldots,p. \end{array}$$

The permutation π is then defined as the shortest path between the columns of Σ by associating each column of Σ to its nearest neighbor

$$\begin{array}{l} \pi(1) = l \\ \\ \pi(2) = \operatorname{argmin}_{j \notin \{\pi(1)\}} \left(D_{\pi(1),j}\right) \\ \ldots \\ \pi(k) = \operatorname{argmin}_{j \notin \{\pi(1),\ldots,\pi(k-1)\}} \left(D_{\pi(k-1),j}\right) \\ \ldots \end{array}$$

Denote as $\hat{\pi}$ the permutation and as $\hat{P}$ the corresponding permutation matrix obtained when applying the algorithm described above to the MLE of Σ_p

instead of the population covariance matrix itself. Theorem 6.1 below states that $\hat{P}$ converges in probability towards P as both, n and p go to infinity given some moment constraints on the random variables in $X(t)$.

Theorem 6.1. *Suppose that the random variables $X_1, \ldots, X_p$ have finite moments up to order 8q. We then have that*

$$\hat{P} \underset{p}{\rightarrow} P$$

as $n \rightarrow +\infty$ and $p \rightarrow +\infty$, given that $P(\hat{\pi} \neq \pi.) = O\left(\frac{1}{n^q}\right)$.

6.3.2 Unbalanced Haar wavelet functions

Girardi and Swelden (1997) [63] introduce the unbalanced Haar wavelet basis. In contrast to the standard Haar basis, the breakpoints of the unbalanced Haar wavelet vectors do not necessarily occur in the middle point of the support

$$\psi_{j,k_j}(t) = \begin{cases} \frac{1}{\sqrt{E_{(j,k_j)} - S_{(j,k_j)}}} & \text{if } t \in \left[S_{(j,k_j)}, B_{(j,k_j)}\right] \\ -\frac{1}{\sqrt{E_{(j,k_j)} - S_{(j,k_j)}}} & \text{if } t \in \left]B_{(j,k_j)}, E_{(j,k_j)}\right] \\ 0 & \text{otherwise} \end{cases}$$

where $B_{(j,k_j)}$ is the breakpoint of the wavelet vector and where $S_{(j,k_j)}$ and $E_{(j,k_j)}$ are the lower bound and the upper bound of its support. The lower and upper bounds are defined according to a recursive scheme across the different wavelet scales. To begin with, denote as $\mathscr{G}_{j-1}^{(1)}$ the subset of wavelet functions at scale $j-1$ such that $B_{(j,k_j)} - S_{(j,k_j)} > 0$ and $E_{(j,k_j)} - B_{(j,k_j)} - 1 > 0$ with and as $\mathscr{G}_{j-1}^{(2)}$ the subset of wavelet functions at scale $j-1$ such that such that $B_{(j,k_j)} - S_{(j,k_j)} = 0$ or $E_{(j,k_j)} - B_{(j,k_j)} - 1 = 0$ with $K_{j-1} = \text{Card}(\mathscr{G}_{j-1}^{(1)}) + \text{Card}(\mathscr{G}_{j-1}^{(2)})$ where K_{j-1} is the total number of wavelet functions at scale $j-1$.

Each wavelet function $\psi_{j-1,k_{j-1}} \in \mathscr{G}_{j-1}^{(1)}$ leads to two new wavelet functions at the adjacent scale j denoted as ψ_{j,k_j} and $\psi_{j,k_j}+1$ with

$$\begin{aligned}
S_{(j,k_j)} &= S_{(j-1,2k_{j-1}-1)}\\
E_{(j,k_j)} &= B_{(j-1,2k_{j-1}-1)}\\
S_{(j,k_j+1)} &= B_{(j-1,2k_{j-1})}+1\\
E_{(j,k_j)} &= E_{(j-1,2k_{j-1})}.
\end{aligned}$$

On the other hand, each wavelet vector $\psi_{j-1,k_{j-1}} \in \mathscr{G}^{(1)}_{j-1}$ leads to one new wavelet vector at the adjacent scale j denoted as ψ_{j,k_j} with

$$\begin{aligned}
S_{(j,k_j)} &= S_{(j-1,2k_{j-1}-1)}\\
E_{(j,k_j)} &= B_{(j-1,2k_{j-1}-1)}.
\end{aligned}$$

Thus, the number of wavelet vectors at scale j is given by $K_j = 2\mathrm{Card}(\mathscr{G}^{(1)}_{j-1})+\mathrm{Card}(\mathscr{G}^{(2)}_{j-1})$.

Hence, in contrast to the standard Haar wavelet basis (see Haar, 1910 [65]) the breakpoints of the unbalanced Haar wavelet vectors are not fixed in advance. We do this on the basis of a segmentation procedure which aims at clustering the neighboring elements of $X(t)$ into smaller and smaller classes. To do this we first, denote as d the vector of Euclidean distances of all pairs of adjacent columns of a given matrix. Based on the length of the support of $\psi_{j,k}$ there are three possible scenarios

- If $B_{j,k_j}-S_{j,k_j}>1$ and $E_{j,k_j}-B_{j,k_j}>1$ then ψ_{j,k_j} leads to two vectors at level $j+1$ with $S_{j+1,k_j}=S_{j,k_j}$, $E_{j+1,k_j}=B_{j,k_j}$, $S_{j+1,k_j+1}=B_{j,k_j}+1$ and $E_{j+1,k_j+1}=E_{j,k_j}$ and

$$\begin{aligned}
B_{j+1,k_j} &= \min_i (d_i) \qquad i = S_{j,k_j},\dots,B_{j,k_j}\\
B_{j+1,k_j+1} &= \min_i (d_i) \quad i = B_{j,k_j}+1,\dots,E_{j,k_j}.
\end{aligned}$$

- If $B_{j,k_j}-S_{j,k_j}>1$ and $E_{j,k_j}-B_{j,k_j}=1$ then ψ_{j,k_j} leads to one single vector at level $j+1$ with $S_{j+1,k_j}=S_{j,k_j}$, $E_{j+1,k_j}=B_{j,k_j}$, $S_{j+1,k_j+1}=B_{j,k_j}+1$ and $E_{j+1,k_j+1}=E_{j,k_j}$ and

$$B_{j+1,k_j} = \min_i (d_i) \quad i = S_{j,k_j},\dots,B_{j,k_j}.$$

- If $B_{j,k_j}-S_{j,k_j}=1$ and $E_{j,k_j}-B_{j,k_j}>1$ then ψ_{j,k_j} leads to one single vector at level $j+1$ with $S_{j+1,k_j} = S_{j,k_j}$, $E_{j+1,k_j} = B_{j,k_j}$, $S_{j+1,k_j+1} = B_{j,k_j}+1$ and $E_{j+1,k_j+1} = E_{j,k_j}$ and

$$B_{j+1,k_j} = \min_i (d_i) \quad i = B_{j,k_j}+1,\ldots,E_{j,k_j}.$$

- If $B_{j,k_j} - S_{j,k_j} = 1$ and $E_{j,k_j} - B_{j,k_j} = 1$ then ψ_{j,k_j} leads to no vector at level $j+1$

Denote as $\hat{B}$ the vector of breakpoints and as $\hat{\psi}$ the corresponding matrix of unbalanced Haar wavelet vectors obtained when applying the algorithm described above to the MLE estimator of Σ_p instead of the population covariance matrix itself. Theorem 6.2 states that $\hat{\psi}$ converges to ψ as both n and p go to infinity with $p = O(n^{\alpha})$.

Theorem 6.2. *Suppose that the random variables $X_1,\ldots,X_p$ have finite moments up to order 8q. We then have that*

$$\hat{\psi} \underset{p}{\rightarrow} \psi$$

as $n \rightarrow +\infty$ and $p \rightarrow +\infty$, given that $P\left(\hat{B} \neq B.\right) = O\left(\frac{1}{n^q}\right)$.

6.3.3 Thresholding operator

Suppose that the covariance matrix of $X(t)$ is blockwise constant and consider the sequence of unbalanced Haar wavelet functions

$$\psi_{j,k_j}, \psi_{j,2k_j-1}, \psi_{j,2k_j}, \psi_{j,4k_j-3}, \psi_{j,4k_j-2}, \psi_{j,4k_j-1}, \psi_{j,4k_j}, \cdots$$

with supports given in Figure 6.2.

Define the groups of wavelets coefficients $G_{j,k_j} = \left\{\psi_k^t \Sigma \psi_l\right\}_{k=1}^{p}$ for all ψ_l such that its support is contained in $\left[S_{j,k_j}, E_{j,k_j}\right]$. Now suppose that the lower bound and the upper bound of the support of ψ_{j,k_j} are given by $S_{j,k_j} = \sum_{i=1}^{l-1} N_i$ and $E_{j,k_j} = \sum_{i=1}^{l} N_i$. We then have that all the elements of G_{j,k_j} are equal to 0.

Denote as $\hat{G}_{j,k_j}$ as the empirical counterparts based on a projection of the sample covariance matrix on the same set of functions as used to construct

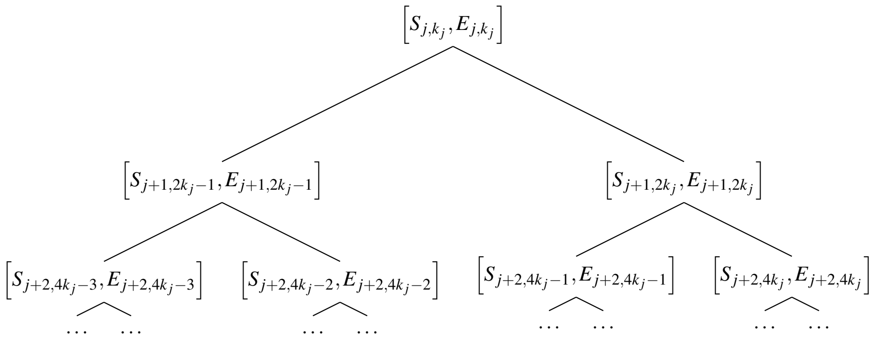


Figure 6.2: ***Tree-structure of supports of wavelet functions***

the group G_{j,k_j}. The group-structured thresholding operator then takes the form

$$\tau_\theta\left(\hat{G}_{j,k_j}\right) = \begin{cases} 0 & \text{if } \max\left(|\hat{G}_{j,k_j}|\right) \leq \theta \\ \hat{G}_{j,k_j} & \text{if } \max\left(|\hat{G}_{j,k_j}|\right) > \theta \end{cases}$$

with the threshold θ being defined as

$$\theta = C \cdot n^{\frac{-1+\eta}{2}}.$$

Theorem 6.3. *Denote as $S = \psi^t P \Sigma P^t \psi$ and $\hat{S} = \psi^t P \hat{\Sigma} P^t \psi$. Suppose that $p = O(n^\alpha)$ with $\alpha \in \mathbb{R}^+$ and that the random variables $X_1, \ldots, X_p$ have finite moments up to order $4q$ with $q > 2 \cdot \alpha$. If we set $\theta = C \cdot n^{\frac{-1+\eta}{2}}$ with $C > 0$ and where $\frac{\alpha}{q} < \eta < \frac{1}{2}$, we have that*

$$\left|\left|\tau_\theta\left(\hat{S}\right) - \tau_O\left(\hat{S}\right)\right|\right| \underset{p}{\rightarrow} 0$$

where τ_O is an oracle operator defined as

$$\tau_0\left(\hat{S}_{i,j}\right) = \begin{cases} 0 & \textit{if } S_{i,j} = 0 \\ \hat{S}_{i,j} & \textit{if } S_{i,j} \neq 0 \end{cases}$$

6.3.4 Spectral norm consistency

In this section we discuss the convergence in spectral norm of the wavelet estimator $\hat{\psi}^t \hat{P} \hat{\Sigma} \hat{P}^t \hat{\psi}^t$ towards $\psi^t P \Sigma P^t \psi$.

To begin with, we discuss the link between the blockwise constant structure of Σ_p and the sparsity of the latter in the wavelet domain when projected on a collection of unbalanced Haar wavelets. Denote as $\psi^t \Sigma_p \psi$ the projection of Σ_p onto the matrix of unbalanced Haar wavelet vectors ψ with decreasingle ordered with respect to the lengths of the support of the corresponding wavelet vector. Given that Σ_p is blockwise constant with $K \times K$ blocks and if the breakpoints of the first K wavelet vectors of ψ are equal to $B_k = \sum_{i=1}^{k} N_i$, then $\psi^t \Sigma_p \psi$ takes the form

$$\psi^t \Sigma_p \psi = \left(\begin{array}{c|c} \begin{matrix} \Gamma_p \\ [K \times K] \end{matrix} & \begin{matrix} 0_p \\ [K \times (p-K)] \end{matrix} \\ \hline \begin{matrix} 0_p \\ [(p-K) \times K] \end{matrix} & \begin{matrix} D_p \\ [(p-K) \times (p-K)] \end{matrix} \end{array} \right) \tag{6.3.1}$$

where Γ_p is a full matrix of dimension $K \times K$, 0_p denotes the null matrix of dimension $(K) \times (p-K)$ and $(p-K) \times K$ respectively, and D_p is a diagonal matrix of dimension $(p-K) \times (p-K)$. Define $\psi^t \hat{\Sigma}_p \psi$ as the projection of the sample estimator and consider the oracle thresholding operator

$$\tau_O\left(\psi^t \hat{\Sigma} \psi\right) = \begin{cases} \psi^t P \hat{\Sigma}_p P^t \psi & \text{if } \psi^t \Sigma_p \psi \neq 0 \\ 0 & \text{if } \psi^t \Sigma_p \psi = 0 \end{cases}$$

and denote $\Xi_p = \tau_O\left(\psi^t \hat{\Sigma}_p \psi\right) - \psi^t \Sigma_p \psi$.

Theorem 6.4. *Suppose that $p = O(n^{\alpha})$ with $\alpha \in \mathbb{R}^+$ and that the elements of $X_p(t)$ have finite moments up to order $4q$. If Σ is a blockwise constant matrix with the number of blocks given by K^2 where $K = O(p^{\beta})$ and if $\beta < \frac{1}{2 \cdot \alpha}$, we have that*

$$||\Xi_p|| \underset{p}{\rightarrow} 0$$

as $n \rightarrow +\infty$ and $p \rightarrow +\infty$ given that $E||\Xi_p||^2 = O\left(\frac{1}{n^{q(1-2\alpha\beta)}}\right)$.

Proposition 6.1. *The spectral norm consistency of $\hat{\psi}^t \hat{P} \hat{\Sigma}_p \hat{P}^t \hat{\psi}^t$ can be* *proven by showing that*

1. $\lim_{\substack{n \to \infty \\ p \to \infty}} P\left(||\tau_O\left(\psi^t P\hat{\Sigma}_p P^t \psi\right) - \psi^t P\Sigma_p P^t \psi|| > \varepsilon\right) = 0$
2. $\lim_{\substack{n \to \infty \\ p \to \infty}} P\left(\hat{P}\hat{\psi} \neq P\psi\right) = 0$
3. $\lim_{\substack{n \to \infty \\ p \to \infty}} P\left(\tau\left(\psi^t P\hat{\Sigma}_p P^t \psi\right) \neq \tau_O\left(\psi^t P\hat{\Sigma}_p P^t \psi\right)\right) = 0$

Hence, combining the statements of Theorems 6.1, 6.2, 6.3 and 6.4 allows to conclude that the wavelet estimator $\hat{\psi}^t \hat{P}\hat{\Sigma}_p \hat{P}^t \hat{\psi}^t$ converges in spectral norm towards $\psi^t P\Sigma_p P^t \psi^t$ as $n \to +\infty$ and $p \to +\infty$.

6.4 Portfolio optimization in the wavelet domain

Markowitz (1952) [89] defines the optimal portfolio as the vector w^* which solves the quadratic program

$$\begin{aligned} \min_{w \in \mathbb{R}^p} \quad & [w^t \Sigma_p w] \\ \text{s.t.} \quad & w^t \mu_p = r \\ & w^t e = 1 \end{aligned} \tag{6.4.1}$$

where $e = (1, \ldots, 1)$.

We transfer the Markowitz optimization algorithm into the wavelet domain by decomposing the solution vector as

$$w = \psi v \tag{6.4.2}$$

with v being a vector of dimension p. The optimal portfolio in the standard domain is then given by $w = \psi v^*$ where v^* solves the quadratic program in the wavelet domain

$$\begin{aligned} &\min \quad [v^t\left((\psi^t \Sigma_p \psi)\right) v] \\ &v \in \mathbb{R}^p \\ &\;\text{s.t.} \quad v^t(\psi^t \mu_p) = r \\ &\qquad\;\; v^t(\psi^t e) = r. \end{aligned}$$

6.5 Simulation Studies

In this section we compare the performance of the wavelet estimator to the performance of shrinkage type estimators using simulation studies. The basic idea of shrinkage estimation is to combine the sample estimator of the covariance matrix with some a priori information about the stock returns process. For a detailed overview of shrinkage estimators of covariance matrices we refer to Ledoit and Wolf (2004) [74]. The general form of the shrinkage estimator is given by

$$\Sigma(\alpha) = (1-\alpha)\cdot\hat{\Sigma} + \alpha\cdot\Sigma_\Gamma$$

where $\hat{\Sigma}$ is the sample estimator and Σ_Γ denotes the so-called target estimator with $0 \leq \alpha \leq 1$. We consider two different target estimators, truncated SVD estimators and the identity matrix. The truncated SVD estimators are given by $\Sigma_\Gamma = \sum_{k=1}^{\delta} \lambda_k u_k u_k^t$ where λ_k denotes the k-th eigenvalue and u_k is the associated eigenvector of the sample covariance matrix with $\delta = 1, 5, 10$.

We use two different portfolio optimization schemes, smallest variance portfolios and mean-variance portfolios. For minimum variance portfolios the optimal asset allocation scheme is defined as the vector w for which leads the random variable has minimum variance. Hence, the optimization framework is defined as

$$\begin{aligned} &\min \quad [w^t \Sigma_X(\theta) w] \\ &w \in \mathbb{R}^p \\ &\;\text{s.t.} \quad w^t e = 1. \end{aligned}$$

We focus on two different forms of population covariance matrices, the first one being the blockwise constant matrices introduced in Section 6.3.1. We refer to the second type of matrices as blurred blockwise constant matrices in the sense that these matrices are obtained by adding noise to the blockwise constant matrices. The latter are obtained according to a two-step procedure.

For a given iteration l of the simulation studies, the training set $\{\bar{X}_l^{(1)},\ldots,\bar{X}_l^{(n)}\}$ consists of n independent realizations of a $p-$dimensional normal distribution $N(\mu_p,\Sigma_p)$. The training set is used as an input of the mean variance or the minimum variance optimization algorithm which leads to the solution vector $\bar{w}_l$. The evaluation set contains one single realization $\tilde{X}_l$ of the same distribution. We use the evaluation set to construct the out-of-sample portfolio return $\bar{w}_l^t\tilde{X}_l$. The out-of-sample mean and variance are given by

$$\bar{\mu}_w = \frac{1}{L}\sum_{l=1}^{L}\bar{w}_l^t\bar{X}_l$$
$$\bar{\sigma}_w^2 = \frac{1}{L-1}\sum_{l=1}^{L}\left(\bar{w}_l^t\bar{X}_l - \bar{x}_w\right)^2.$$

Tables 6.1 and 6.2 contain the out-of-sample standard deviation for the minimum variance optimization setting with $p = 20$ and $n = 20$ for both, blockwise constant matrices and blurred blockwise constant matrices.

Estimator	SD
Wavelet	0.7875
Sample	-
Shrinkage: Single-Factor	0.8730
Shrinkage: Five-Factor	0.9216
Shrinkage: Ten-Factor	1.0641
Shrinkage: Identity	0.9782
Population	0.7373

Table 6.1: ***Minimum variance portfolio***
Blockwise constant matrix
p=20
n=20

In both cases the wavelet estimator outperforms the shrinkage-type estimators by quite a large margin. Note that the good performance of the wavelet estimator for blurred blockwise constant matrices illustrates the robustness

Estimator	SD
Wavelet	0.8304
Sample	-
Shrinkage: Single-Factor	0.9867
Shrinkage: Five-Factor	1.1095
Shrinkage: Ten-Factor	1.3935
Shrinkage: Identity	1.2185
Population	0.7604

Table 6.2: ***Minimum variance portfolio***
Blurred blockwise constant matrix
p=20
n=20

of the latter towards deviations from the blockwise constant matrix assumption.

Tables 6.3 to 6.6 contain the same type of results but for the mean variance optimization framework with known and unknown mean return vector. Note that in all scenarios the wavelet estimator outperforms the shrinkage-type estimators.

Estimator	Mean	SD
Wavelet	0.9492	0.8592
Sample	-	-
Shrinkage: Single-Factor	0.9574	0.8935
Shrinkage: Five-Factor	0.9530	0.9674
Shrinkage: Ten-Factor	0.9543	1.1165
Shrinkage: Identity	0.9528	1.0276
Population	0.9954	0.7669

Table 6.3: ***Mean-variance portfolio with known mean vector***
Blockwise constant matrix
p=20
n=20

Estimator	Mean	SD
Wavelet	0.9481	0.7972
Sample	-	-
Shrinkage: Single-Factor	0.9392	0.8827
Shrinkage: Five-Factor	0.9442	0.9423
Shrinkage: Ten-Factor	0.9544	1.0985
Shrinkage: Identity	0.9518	0.9971
Population	1.0005	0.7293

Table 6.4: ***Mean-variance portfolio with known mean vector***
Blurred blockwise constant matrix
p=20
n=20

Estimator	Mean	SD
Wavelet	0.9749	0.8434
Sample	-	-
Shrinkage: Single-Factor	0.9739	0.9076
Shrinkage: Five-Factor	0.9744	0.9856
Shrinkage: Ten-Factor	0.9754	1.1329
Shrinkage: Identity	0.9753	1.0437
Population	1.0067	0.7793

Table 6.5: ***Mean-variance portfolio with unknown mean vector***
Blockwise constant matrix
p=20
n=20

Estimator	Mean	SD
Wavelet	0.9481	0.7972
Sample	-	-
Shrinkage: Single-Factor	0.9392	0.8827
Shrinkage: Five-Factor	0.9442	0.9423
Shrinkage: Ten-Factor	0.9544	1.0985
Shrinkage: Identity	0.9518	0.9971
Population	1.0005	0.7293

Table 6.6: ***Mean-variance portfolio with unknown mean vector***
Blurred blockwise constant matrix
p=20
n=20

6.6 Adaptive asset allocation

One of the underlying assumptions of the mean-variance optimization framework is that the mean return vector μ_p and the cross=-covariance matrix Σ_p do not change in time. However, as it has been discussed in Sections 3, 4, and 4 the covariance stationarity assumption may not be a valid assumption especially when dealing with financial or economic data. Consider the random vector $X(t) = \{X_i(t)\}_{i=1}^{p}$ and denote as $\mu_p(t)$ and $\Sigma_p(t)$ the time-varying mean return vector and covariance matrix of $X(t)$. Write as $w^*(t)$ the solution vector of the mean-variance framework

$$\begin{array}{ll} \min & [w^t\Sigma(t)w] \\ w \in \mathbb{R}^p & \\ \text{s.t.} & w^t\mu(t) = r \\ & w^te = 1. \end{array} \tag{6.6.1}$$

Recall that the random variable $w^t(t)X(t)$ is such that it has the smallest variance among all possible vectors w for a given value of mean return r. The optimal asset allocation scheme for period $t+l$ is defined as the solution vector of (6.6.1) with input given by $\mu_p(t+l)$ and $\Sigma_p(t+l)$ respectively. Given the non-stationary behavior of the time series $\{X(t)\}_{t=0}^{n-1}$ we have that $w(t) \neq w(t+l)$. Thus, in a non-stationary setting the investment portfolio has to be rebalanced on a constant basis in order to take into account the current values of the time-varying behavior of the mean return vector and the covariance matrix.

In this section we build an adaptive portfolio optimization scheme using the log-return data of the companies in Table 6.7 from 1990 to 2014. As for the simulation studies in Section 6.5 we do a comparative study of the wavelet estimator on the one hand and shrinkage type estimators on the other hand. To evaluate the performance we again use the out of sample mean and variance, defined as

$$\hat{\mu}(t) = \tfrac{1}{n}\sum_{k=-m1}^{n-1} \hat{w}^t(t)X(t+1)$$
$$\hat{\sigma}(t) = \tfrac{1}{n-1}\sum_{t=0}^{n-1} \left(\hat{w}^t(t)X(t+1) - \hat{\mu}(t)\right)^2.$$

Tables 6.8 to 6.13 contain the values of the out-of-sample mean and variance for both, minimum variance and mean-variance portfolios, and show the good performance of the wavelet estimator compared to the shrinkage-type estimators for different sample sizes.

3M CO.	American Express CO.
AT&T	Boeing CO.
Caterpillar Inc.	Chevron
Cisco Systems Inc.	Coca-Cola Co.
E.I. DuPont de Nemours & CO.	Exxon Mobil
General Electric Co.	Home Depot Inc.
Intel Corp.	IBM Corp.
Johnson&Johnson	JPMorgan Chase
McDonald's Corp.	Merck & Co. Inc.
Microsoft Corp.	Pfizer Inc.
Procter & Gamble CO.	Travelers Cos.
Verizon Communications	United Technologies Corp.
UnitedHealth Group Inc.	Visa Inc.
Wal-Mart Stores Inc.	

Table 6.7: ***Components of the Dow Jones Industrial Average***

Estimator	SD
Wavelet	0.0127
Sample	0.0452
Shrinkage: Identity	0.0163
Shrinkage: Single-Factor	0.0139
Shrinkage: Five-Factor	0.0150
Shrinkage: Ten-Factor	0.0159

Table 6.8: ***Minimum variance portfolio: n=30***

Estimator	SD
Wavelet	0.0123
Sample	-
Shrinkage: Identity	0.0174
Shrinkage: Single-Factor	0.0137
Shrinkage: Five-Factor	0.0151
Shrinkage: Ten-Factor	0.0164

Table 6.9: ***Minimum variance portfolio: n=15***

Estimator	SD
Wavelet	0.0120
Sample	-
Shrinkage: Identity	0.0152
Shrinkage: Single-Factor	0.0135
Shrinkage: Five-Factor	0.0148
Shrinkage: Ten-Factor	-

Table 6.10: ***tab:Minimum variance portfolio: n=8***

Estimator	Mean	SD
Wavelet	0.00032	0.0135
Uniform	0.00028	0.0120
Sample	0.00015	0.0397
Shrinkage: Identity	0.00025	0.0149
Shrinkage: Single-Factor	0.00024	0.0144
Shrinkage: Five-Factor	0.00025	0.0154
Shrinkage: Ten-Factor	0.00022	0.0163

Table 6.11: ***Mean variance portfolio: n=30***

Estimator	Mean	SD
Wavelet	0.00046	0.0131
Uniform	0.0028	0.0120
Sample	-	-
Shrinkage: Identity	0.00021	0.0149
Shrinkage: Single-Factor	0.00027	0.0141
Shrinkage: Five-Factor	0.00021	0.0155
Shrinkage: Ten-Factor	0.00014	0.0170

Table 6.12: ***Mean variance portfolio: n=15***

Estimator	Mean	SD
Wavelet	0.00045	0.0132
Uniform	0.00028	0.0120
Sample	-	
Shrinkage: Identity	0.000028	0.0147
Shrinkage: Single-Factor	0.00025	0.0139
Shrinkage: Five-Factor	0.00026	0.0154
Shrinkage: Ten-Factor	-	

Table 6.13: ***Mean variance portfolio: n=8***

6.7 Proofs

6.7.1 Lemma 6.1

Lemma 6.1. *Define $\{X(t) : t \in T\}$ as a $p-$dimensional stationary stochastic process with iid random variables. Furthermore, consider* *the matrix of unbalanced Haar wavelet vectors ψ defined as*

$$\begin{gathered}\psi_{r,l} = \left(\frac{1}{B_r - S_r + 1}\right) I\left[S_r \le l \le B_r\right] - \left(\frac{1}{E_r - B_r}\right) I\left[B_r + 1 \le l \le E_r\right] \\ l = 1, \ldots, p \\ r = 1, \ldots, p-1\end{gathered}$$

where S_i, B_i and E_i represent the starting point, the break point and the end point respectively.

Define the random variables

$$\begin{aligned}Y_i(t) &= \psi_i^t \left[X(t) - E(X)\right] \\ Y_j(t) &= \left[X(t) - E(X)\right]^t \psi_j.\end{aligned}$$

The wavelet coefficient $M_{r,s}$ is given by

$$E\left(Y_i(t) Y_j(t)\right) = \psi_i^t\ E\left\{\left[X(t) - E(X)\right]\left[X(t) - E(X)\right]^t\right\}\ \psi_j = \psi_i^t\ \Sigma\ \psi_j.$$

Consider the random matrix

$$\begin{aligned}\left[X(t) - \bar{X}\right]_{t=1,\ldots,n} &= \left[(X(t) - E(X)) - (\bar{X} - E(X))\right]_{t=1,\ldots,n} \\ &= \left[(X(t) - E(X)) - \left(\frac{1}{n}\sum_{t=1}^{n} X(t) - E(X)\right)\right]_{t=1,\ldots,n}\end{aligned}$$

and define the random vectors

$$\begin{aligned}\left[Y_i(t) - \frac{1}{n}\sum_{t=1}^{n} Y_i(t)\right]^t_{t=1,\ldots,n} &= \psi_i^t \left[(X(t) - E(X)) - \left(\frac{1}{n}\sum_{t=1}^{n} X(t) - E(X)\right)\right]_{t=1,\ldots,n} \\ \left[Y_j(t) - \frac{1}{n}\sum_{t=1}^{n} Y_j(t)\right]_{t=1,\ldots,n} &= \left[(X(t) - E(X)) - \left(\frac{1}{n}\sum_{t=1}^{n} X(t) - E(X)\right)\right]^t_{t=1,\ldots,n} \psi_j.\end{aligned}$$

If we define

$$\begin{aligned}\hat{M}_{i,j} &= \frac{1}{n-1}\ \left[Y_i(t) - \frac{1}{n}\sum_{t=1}^{n} Y_i(t)\right]^t_{t=1,\ldots,n} \left[Y_j(t) - \frac{1}{n}\sum_{t=1}^{n} Y_j(t)\right]_{t=1,\ldots,n} \\ &= \frac{1}{n-1}\sum_{t=1}^{n} Y_i(t) Y_j(t) - \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{t'=1}^{n} Y_i(t) Y_j(t')\end{aligned}$$

we have that $E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q} = \frac{c_{i,j}}{n^q}$.

If $E\left(\hat{M}_{i,j}\right) = 0$, we can write

$$\begin{aligned}
&E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q} \\
&= E\left(\hat{M}_{i,j}\right)^{2q} \\
&= E\left(\frac{1}{n-1}\sum_{t=1}^{n} Y_i(t)Y_j(t) - \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{t'=1}^{n} Y_i(t)Y_j(t')\right)^{2q}
\end{aligned}$$

Define the set $T_k = \left\{\left(t_1, \ldots, t_{2q+k}\right)\right\} \in [1, \ldots, n]^{2q+k}$ and write

$$E\left(\hat{M}_{i,j}\right)^{2q} = \sum_{k=0}^{2q} \frac{1}{n^{2q+k}} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{\left(t_1,\ldots,t_{2q+k}\right)\in T_k} E\left(\prod_{r\in[1,2q-k]} Y_i(t_r) Y_j(t_r) \prod_{r\in[2q-k+1,2q]} Y_i(t_r) Y_j(t_{r+k})\right)$$

Moreover, consider the sets

$$S^{(k)} = \{1, \ldots, 2q+k\}$$

and

$$S_1^{(k,l)}, \ldots, S_{h_{k,l}}^{(k,l)} \quad \text{with} \quad \begin{aligned} &S_j^{(k,l)} \subseteq S \\ &\bigcap_{j=1}^{h_l} S_j^{(k,l)} = \emptyset \\ &Card\left(S_j^{(k,l)}\right) \geq 2 \end{aligned}$$

Denote $V_{k,l} = \bigcup_{j=1}^{h_{k,l}} S_j^{(k,l)}$ with $Card\left(V_{k,l}\right) = \sum_{j=1}^{h_{k,l}} Card\left(S_j^{(k,l)}\right) \leq 2q+k$

On the basis of $V_{k,l}$ we then define $T_{k,l}$ as

$$\begin{aligned}
T_{k,l} = \Big\{ &\left(t_1, \ldots, t_{2q+k}\right) \in [1, \ldots, n]^{2q+k} \\
&\text{where } t_a = t_b \text{ if } \exists S_j^{(k,l)} \mid \text{a} \in S_j^{(k,l)} \text{ and b} \in S_j^{(k,l)} \\
&t_a \neq t_b \text{ if } \nexists S_j^{(k,l)} \mid \text{a} \in S_j^{(k,l)} \text{ and b} \in S_j^{(k,l)} \Big\}
\end{aligned}$$

with $\bigcup_{l=1} T_{k,l} = T$, $\bigcap_{l=1} T_{k,l} = \emptyset$ and

$$\text{Card}\left(T_{k,l}\right) = \frac{n!}{\left(n-2q+k+x_{k,l}\right)!} = c_{k,l} n^{2q+k-x_{k,l}}$$

where $x_{k,l} = \sum_{j=1}^{h_{k,l}} \text{Card}\left(S_j^{(k,l)}\right) - h_{k,l}$

Thus, we have that

$$
\begin{aligned}
&\sum_{(t_1,\ldots,t_{2q+k})\in T_k} E\left(\prod_{r\in[1,2q-k]} Y_i(t_r)Y_j(t_r) \prod_{r\in[2q-k+1,2q]} Y_i(t_r)Y_j(t_{r+k})\right) \\
&= \sum_l \sum_{(t_1,\ldots,t_{2q+k})\in T_{k,l}} E\left(\prod_{r\in[1,2q-k]} Y_i(t_r)Y_j(t_r) \prod_{r\in[2q-k+1,2q]} Y_i(t_r)Y_j(t_{r+k})\right) \\
&= \sum_l \sum_{(t_1,\ldots,t_{2q+k})\in T_{k,l}} \prod_{\substack{r\in V_{k,l}^C \\ r\in[1,2q-k]}} E\left(Y_i(t_r)Y_j(t_r)\right) \\
&\prod_{\substack{r\in V_{k,l}^C \\ r\in[2q-k+1,2q]}} E\left(Y_i(t_r)\right) \prod_{\substack{r+k\in V_{k,l}^C \\ r\in[2q-k+1,2q]}} E\left(Y_j(t_{r+k})\right) \\
&\qquad E\left(\prod_{\substack{r\in V_{k,l} \\ r\in[1,2q-k]}} Y_i(t_r)Y_j(t_r) \prod_{\substack{r\in V_{k,l} \\ r\in[2q-k+1,2q]}} Y_i(t_r) \prod_{\substack{r+k\in V_{k,l} \\ r\in[2q-k+1,2q]}} Y_j(t_{r+k})\right)
\end{aligned}
$$

If $E\left(\hat{M}_{i,j}\right) \neq 0$

$$
\begin{aligned}
&E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q} \\
&= E\left[\frac{1}{n-1}\sum_{t=1}^{n} Y_i(t)Y_j(t) - \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{t'=1}^{n} Y_i(t)Y_j(t')\right. \\
&\left. - E\left(\frac{1}{n-1}\sum_{t=1}^{n} Y_i(t)Y_j(t) - \frac{1}{n(n-1)}\sum_{t=1}^{n}\sum_{t'=1}^{n} Y_i(t)Y_j(t')\right)\right]^{2q}
\end{aligned}
$$

Denote as $P(\{1,\ldots,2q\})$ the power set of $\{1,\ldots,2q\}$ and as $P(\{1,\ldots,2q\})_{u,u'}$ its $u'-$th subset with $\text{Card}\left(P(\{1,\ldots,2q\})_{u,u'}\right) = u$. If we define the set $T_k = \left\{\left(t_1,\ldots,t_{2q+k}\right)\right\} \in [1,\ldots,n]^{2q+k}$ we can write $E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q}$ as

$$
\begin{aligned}
&\sum_{k=0}^{2q} \frac{1}{n^{2q+k}} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{(t_1,\ldots,t_{2q+k})\in T_k} \sum_{u=0}^{2q}\sum_{u'} (-1)^u E\left(\prod_{\substack{r\in P(\{1,\ldots,2q\})_{u,u'} \\ r\in[1,2q-k]}} Y_i(t_r)Y_j(t_r) \prod_{\substack{r\in P(\{1,\ldots,2q\})_{u,u'} \\ r\in[2q-k+1,2q]}} Y_i(t_r)Y_j(t_{r+k})\right) \\
&\qquad \prod_{\substack{r\in P(\{1,\ldots,2q\})_{u,u'}^C \\ r\in[1,2q-k]}} E\left(Y_i(t_r)Y_j(t_r)\right) \prod_{\substack{r\in P(\{1,\ldots,2q\})_{u,u'}^C \\ r\in[2q-k+1,2q]}} E\left(Y_i(t_r)Y_j(t_{r+k})\right)
\end{aligned}
$$

If we denote as $P\left(V_{k,l}\right)$ the power set of $V_{k,l}$ and as $P\left(V_{k,l}\right)_{v,v'}$ its $v'-$th subset with Card $\left(P\left(V_{k,l}\right)_{v,v'}\right) = v$, this can be written as

$$\begin{aligned}
&\sum_{l} \sum_{\left(t_1,\ldots,t_{2q+k}\right)\in T_{k,l}} \sum_{v=0}^{2q} \sum_{v'} (-1)^v E\left(\prod_{\substack{r\in P\left(V_{k,l}\right)_{v,v'} \\ r\in[1,2q-k]}} Y_i(t_r)Y_j(t_r) \prod_{\substack{r\in P\left(V_{k,l}\right)_{v,v'} \\ r\in[2q-k+1,2q]}} Y_i(t_r)Y_j(t_{r+k}) \right) \\
&\qquad \prod_{\substack{r\in P\left(V_{k,l}\right)^C_{v,v'} \\ r\in[1,2q-k]}} E\left(Y_i(t_r)Y_j(t_r)\right) \prod_{\substack{r\in P\left(V_{k,l}\right)^C_{v,v'} \\ r\in[2q-k+1,2q]}} E\left(Y_i(t_r)Y_j(t_{r+k})\right) \\
&\qquad \sum_{\substack{u,u':\ P\left(V_{k,l}\right)_{v,v'} \subseteq P(\{1,\ldots,2q\})_{u,u'} \\ \text{and} \\ P\left(V_{k,l}\right)^C_{v,v'} \subseteq P(\{1,\ldots,2q\})^C_{u,u'}}} \\
&= \sum_{l} \sum_{\left(t_1,\ldots,t_{2q+k}\right)\in T_{k,l}} \sum_{m=0}^{\mathrm{Card}\left(V_{k,l}\right)} (-1)^m \frac{2q!}{m!(2q-m)!} \sum_{v=0}^{2q} \sum_{v'} (-1)^v E\left(\prod_{\substack{r\in P\left(V_{k,l}\right)_{v,v'} \\ r\in[1,2q-k]}} Y_i(t_r)Y_j(t_r) \prod_{\substack{r\in P\left(V_{k,l}\right)_{v,v'} \\ r\in[2q-k+1,2q]}} Y_i(t_r)Y_j(t_{r+k}) \right) \\
&\qquad \prod_{\substack{r\in P\left(V_{k,l}\right)^C_{v,v'} \\ r\in[1,2q-k]}} E\left(Y_i(t_r)Y_j(t_r)\right) \prod_{\substack{r\in P\left(V_{k,l}\right)^C_{v,v'} \\ r\in[2q-k+1,2q]}} E\left(Y_i(t_r)Y_j(t_{r+k})\right)
\end{aligned}$$

Now write $E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q}$ as

$$\begin{aligned}
&E\left[\hat{M}_{i,j} - E\left(\hat{M}_{i,j}\right)\right]^{2q} \\
&= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \frac{1}{n^{2q+k}} \sum_{l} \sum_{\left(t_1,\ldots,t_{2q+k}\right)\in T_{k,l}} P_{k,l}\left(t_1,\ldots,t_{2q+k}\right) \\
&\text{(using the iid assumption)} \\
&= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{l} \frac{\mathrm{Card}\left(T_{k,l}\right)}{n^{2q+k}} P_{k,l}
\end{aligned}$$

Note that Card $\left(V_{k,l}\right) < 2q+k$ leads to $P_{k,l} = 0$ and hence we can write

$$\begin{aligned}
E\left(\hat{M}_{i,j}\right)^{2q} &= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{l} \frac{\mathrm{Card}\left(T_{k,l}\right)}{n^{2q+k}} P_{k,l} \\
&= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{l|\mathrm{Card}\left(V_{k,l}\right)=2q+k} \frac{\mathrm{Card}\left(T_{k,l}\right)}{n^{2q+k}} P_{k,l} \\
&= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{l|\mathrm{Card}\left(V_{k,l}\right)=2q+k} \frac{c_{k,l} n^{2q+k-x_{k,l}}}{n^{2q+k}} P_{k,l} \\
&= \sum_{k=0}^{2q} (-1)^k \frac{2q!}{(2q-k)!k!} \sum_{l|\mathrm{Card}\left(V_{k,l}\right)=2q+k} \frac{c_{k,l}}{n^{x_{k,l}}} P_{k,l}
\end{aligned}$$

As it has been mentioned above, we have that

$$\sum_{j=1}^{h_{k,l}} \operatorname{Card}\left(V_{k,l}\right) - h_{k,l} = x_{k,l}$$

If $\operatorname{Card}\left(V_{k,l}\right) = 2q+k$ this becomes

$$2q+k-h_{k,l} = x_{k,l}$$

If $2q+k$ is an even number and if we set $\operatorname{Card}\left(S_j^{(k,l)}\right) = 2,\ j = 1\ldots,\frac{2q+k}{2}$, we have that

$$x_{k,l} = q+\frac{k}{2}$$

Note that if there exists $S_j^{(k,l)}$ such that $\operatorname{Card}\left(S_j^{(k,l)}\right) > 2$, we have that $h_{k,l} < \frac{2q+k}{2}$ and hence

$$q+\frac{k}{2} < x_{k,l}$$

Thus, we can establish the lower bound

$$q+\frac{k}{2} \leq x_{k,l}$$

Given that $x_{k,l} \leq 2q+k-1$ we can write

$$q+\frac{k}{2} \leq x_{k,l} \leq 2q+k-1$$

If $2q+k$ is an odd number and if we set $\operatorname{Card}\left(S_i^{(k,l)}\right) = 3$ and $\operatorname{Card}\left(S_j^{(k,l)}\right) = 2,\ j = 1\ldots,\frac{2q+k}{2}-1$, we have that

$$x_{k,l} = q+\frac{k+1}{2}$$

Note that if $\operatorname{Card}\left(S_i^{(k,l)}\right) > 3$ or if there exists $S_j^{(k,l)}$ such that $\operatorname{Card}\left(S_j^{(k,l)}\right) > 2$, we have that $h_l < q+\frac{k-1}{2}$ and hence

$$q+\frac{k+1}{2} < x_{k,l}$$

Thus, we can establish the lower bound

$$q+\frac{k+1}{2}\leq x_{k,l}$$

Given that $x_l \leq 2q+k-1$ we can write

$$q+\frac{k+1}{2}\leq x_{k,l}\leq 2q+k-1$$

Thus, we can write $E\left[\hat{M}_{i,j}-E\left(\hat{M}_{i,j}\right)\right]^{2q}$ as

$$\begin{aligned}
&\sum_{k=0}^{2q}(-1)^k\frac{2q!}{(2q-k)!k!}\sum_{l\mid\mathrm{Card}\left(v_{k,l}\right)=2q+k}\frac{c_{k,l}}{n^{x_{k,l}}}P_{k,l}\\
&=\sum_{k=0}^{2q}(-1)^k\frac{2q!}{(2q-k)!k!}\sum_{\substack{2q+k\text{ even}\\ q+\frac{k}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}}}P_{k,l}+\sum_{\substack{2q+k\text{ odd}\\ q+\frac{k+1}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}}}P_{k,l}\\
&=\sum_{k=0}^{2q}(-1)^k\frac{2q!}{(2q-k)!k!}\frac{1}{n^{q+\frac{k}{2}}}\sum_{\substack{2q+k\text{ even}\\ q+\frac{k}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}-q-\frac{k}{2}}}P_{k,l}+\sum_{\substack{2q+k\text{ odd}\\ q+\frac{k+1}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}-q-\frac{k}{2}}}P_{k,l}\\
&=\frac{1}{n^q}\sum_{k=0}^{2q}(-1)^k\frac{2q!}{(2q-k)!k!}\frac{1}{n^{\frac{k}{2}}}\sum_{\substack{2q+k\text{ even}\\ q+\frac{k}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}-q-\frac{k}{2}}}P_{k,l}+\sum_{\substack{2q+k\text{ odd}\\ q+\frac{k+1}{2}\leq x_{k,l}\leq 2q+k-1}}\frac{c_{k,l}}{n^{x_{k,l}-q-\frac{k}{2}}}P_{k,l}\\
&=\frac{c_{i,j}}{n^q}
\end{aligned}$$

6.7.2 Lemma 6.2

Lemma 6.2. *If we define as*

$$\begin{aligned}
\hat{D}_{i,i'} &= \frac{1}{p}\sum_{j=1}^{p}\left(\frac{1}{n}\sum_{t=1}^{n}\left[X_i(t)-X_i(t)\right]X_j(t)-\frac{1}{n^2}\sum_{t=1}^{n}\left[X_i(t)-X_i(t)\right]X_j(t')\right)^2\\
&= \frac{1}{p}\sum_{j=1}^{p}\left(\frac{1}{n}\sum_{t=1}^{n}Y_i(t)X_j(t)-\frac{1}{n^2}\sum_{t=1}^{n}Y_i(t)X_j(t')\right)^2\\
&= \frac{1}{p}\sum_{j=1}^{p}Q_j
\end{aligned}$$

the distance measure between the i-th and the i'-th column of the sample covariance matrix we have that

$$E\left[\hat{D}_{i,i'}-E\left(\hat{D}_{i,i'}\right)\right]^{2q}=\frac{c_{i,i'}}{n^{2q}}.$$

Write $E\left[\hat{D}_{i,i'}-E\left(\hat{D}_{i,i'}\right)\right]^{2q}$ as $E\left[\frac{1}{p}\sum_{j=1}^{p}Q_j-E\left(Q_j\right)\right]^{2q}$ and use Hoelder's inequality to establish the upper bound

$$E\left[\frac{1}{p}\sum_{j=1}^{p} Q_j - E\left(Q_j\right)\right]^{2q} \leq \prod_{j=1}^{2q}\left(E\left[Q_j - E\left(Q_j\right)\right]^{2q}\right)^{\frac{1}{2q}}$$

Note that $E\left[Q_j - E\left(Q_j\right)\right]^{2q}$ can be written as

$$\begin{aligned}
E\Bigg[&\tfrac{1}{n}\sum_{t1=1}^{n}\sum_{t2=1}^{n} Y_i(t_1)X_j(t_1)Y_i(t_2)X_j(t_2) \\
&-\tfrac{2}{n^3}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n} Y_i(t_1)Y_j(t_1)Y_i(t_2)Y_j(t_3) \\
&+\tfrac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n}\sum_{t_3=1}^{n} Y_i(t_1)Y_j(t_2)Y_i(t_3)Y_j(t_4) \\
&-E\left(\tfrac{1}{n}\sum_{t1=1}^{n}\sum_{t2=1}^{n} Y_i(t_1)X_j(t_1)Y_i(t_2)X_j(t_2)\right) \\
&+E\left(\tfrac{2}{n^3}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n} Y_i(t_1)Y_j(t_1)Y_i(t_2)Y_j(t_3)\right) \\
&-E\left(\tfrac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{t_3=1}^{n}\sum_{t_3=1}^{n} Y_i(t_1)Y_j(t_2)Y_i(t_3)Y_j(t_4)\right)\Bigg]^{2q}
\end{aligned}$$

Thus, using Lemma 1, we have that $E\left[Q_j - E\left(Q_j\right)\right]^{2q} = \frac{c_j}{n^{2q}}$. Hence,

$$E\left[\frac{1}{p}\sum_{j=1}^{p} Q_j - E\left(Q_j\right)\right]^{2q} \leq \prod_{j=1}^{2q}\left(\frac{c_j}{n^{2q}}\right)^{\frac{1}{2q}} = \frac{c_{i,i'}}{n^{2q}}$$

6.7.3 Proof of Proposition 6.1

Write $\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}^t - P\Sigma P^t|| > \varepsilon\right\}$ as

$$\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}^t - P\Sigma P^t|| > \varepsilon\right\} \cap \left[\left\{\hat{P}\hat{\psi} = P\psi\right\} \cup \left\{\hat{P}\hat{\psi} \neq P\psi\right\}\right]$$

or

$$\begin{aligned}
&\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}^t - P\Sigma P^t|| > \varepsilon\right\} \cap \left\{\hat{P}\hat{\psi} = P\psi\right\} \cup \\
&\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}^t - P\Sigma P^t|| > \varepsilon\right\} \cap \left\{\hat{P}\hat{\psi} \neq P\psi\right\}
\end{aligned}$$

Using that

$$\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}-P\Sigma P^t||>\varepsilon\right\}\\ \cap\left\{\hat{P}\hat{\psi}\neq P\psi\right\}\subseteq\left\{\hat{P}\hat{\psi}\neq P\psi\right\}$$

we have that

$$\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}-P\Sigma P^t||>\varepsilon\right\}\\ \subseteq\left\{||\psi\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)\psi-P\Sigma P^t||>\varepsilon\right\}\cup\left\{\hat{P}\hat{\psi}\neq P\psi\right\}$$

Note that

$$\left\{||\psi\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\psi^t-\psi\psi^tP\Sigma P^t\psi\psi^t||>\varepsilon\right\}\\ \subseteq\left\{||\psi||\,||\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\psi^t-\psi\psi^tP\Sigma P^t\psi||\,||\psi^t||>\varepsilon\right\}$$

with $||\psi||=1$ and hence

$$\left\{||\hat{\psi}\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\hat{\psi}-P\Sigma P^t||>\varepsilon\right\}\\ \subseteq\left\{||\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)-\psi P\Sigma P^t\psi^t||>\varepsilon\right\}\cup\left\{\hat{P}\hat{\psi}\neq P\psi\right\}$$

Decompose the event $\left\{||\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)-\psi^tP\Sigma P^t\psi||>\varepsilon\right\}$ as

$$\left\{\left[||\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)-\psi^tP\Sigma P^t\psi||>\varepsilon\right\}\cap\left\{\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)=\tau_O\left(\psi^tP\hat{\Sigma}P^t\psi\right)\right]\right\}\\ \cup\\ \left\{\left[||\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)-\psi^tP\Sigma P^t\psi||>\varepsilon\right\}\cap\left\{\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)\neq\tau_O\left(\psi^tP\hat{\Sigma}P^t\psi\right)\right]\right\}$$

which means that

$$\left\{||\psi\tau\left(\hat{\psi}^t\hat{P}\hat{\Sigma}\hat{P}^t\hat{\psi}\right)\psi^t-P\Sigma P^t||>\varepsilon\right\}\\ \subseteq\left\{||\tau_O\left(\psi^tP\hat{\Sigma}P^t\psi\right)-\psi^tP\Sigma P^t\psi||>\varepsilon\right\}\\ \cup\left\{\hat{P}\hat{\psi}\neq P\psi\right\}\\ \cup\left\{\tau\left(\psi^tP\hat{\Sigma}P^t\psi\right)\neq\tau_O\left(\psi^tP\hat{\Sigma}P^t\psi\right)\right\}$$

6.7.4 Proof of Theorem 6.1 and Theorem 6.2

By construction the event $\left\{\hat{P}\hat{\Psi}=P\psi\right\}$ is equivalent to the event $\left\{\exists\lambda:\mathrm{Rk}(\hat{D}_\lambda)\neq\mathrm{Rk}(D_\lambda)\right\}$. Thus, we have that

$$\begin{aligned}P\left(\left\{\hat{P}\hat{\Psi}=P\psi\right\}\right)&=P\left(\left\{\exists\lambda\neq\lambda':\hat{D}_\lambda<\hat{D}_{\lambda'}\text{ and }D_\lambda>D_{\lambda'}\right\}\right)\\&\leq\sum_\lambda\sum_{\lambda'<\lambda}P\left(\left\{\hat{D}_\lambda<\hat{D}_{\lambda'}\text{ and }D_\lambda>D_{\lambda'}\right\}\right)\end{aligned}$$

Denote $\delta_{\lambda\lambda'} = E\left(\hat{D}_\lambda - \hat{D}_{\lambda'}\right)$ and write

$$\begin{aligned}
&\sum_\lambda \sum_{\lambda'<\lambda} P\left(\delta_{\lambda\lambda'} + \hat{D}_\lambda - E\hat{D}_\lambda < \hat{D}_{\lambda'} - E\hat{D}_{\lambda'} \text{ and } D_\lambda > D_{\lambda'}\right) \\
&\leq \sum_\lambda \sum_{\lambda'<\lambda} P\left(\hat{D}_\lambda - E\hat{D}_\lambda < -\tfrac{\delta_{\lambda\lambda'}}{2} \text{ and } D_\lambda > D_{\lambda'}\right) \\
&\quad + \sum_\lambda \sum_{\lambda'<\lambda} P\left(\hat{D}_\lambda - E\hat{D}_\lambda > \tfrac{\delta_{\lambda\lambda'}}{2} \text{ and } D_\lambda > D_{\lambda'}\right)
\end{aligned}$$

Note that the cardinality of the double sum is $\frac{p'(p'-1)}{2}$ with $p' = \frac{p(p-1)}{2}$ and hence

$$\sum_\lambda \sum_{\lambda'<\lambda} P\left(\hat{D}_\lambda - E\hat{D}_\lambda < -\frac{\delta_{\lambda\lambda'}}{2} \text{ and } D_\lambda > D_{\lambda'}\right) \leq \frac{p'(p'-1)}{2} \frac{2^{2q}}{\delta_{\lambda\lambda'}^{2q}} \frac{c}{n^q}$$

6.7.5 Proof of Theorem 6.3

Consider the projected matrices $\hat{S} = \psi^t P \hat{\Sigma} P^t \psi$ and $S = \psi^t P \Sigma P^t \psi$ and denote as $\left\{\hat{C}_k\right\}_{k=1}^p$ a collection of subsets of the coefficients $\hat{S}$ with $\hat{S} = \bigcup_{k=1}^p \hat{C}_k$. Moreover define $\hat{M}_k = \max\left(\hat{C}_k\right)$ and denote as θ_k the threshold for $\hat{M}_k$.

<u>Type *I* Error</u>

The event of a type I error on the individual level is defined as $\left\{|\hat{M}_k| > \theta_k \text{ and } M_k = 0\right\}$. Write

$$\begin{aligned}
P\left(\left\{|\hat{M}_k| > \theta_k \text{ and } M_k = 0\right\}\right) &= 2P\left(\left\{\hat{M}_k^{2q} > \theta_k^{2q} \text{ and } M_k = 0\right\}\right) \\
&\leq \frac{E\left(\hat{M}_k^{2q}\right)}{\theta_k^{2q}} \\
&= \frac{c_k}{n^q \cdot \theta_k^{2q}}
\end{aligned}$$

If we denote $c = \max_k(c_k)$ and $\theta = \min_k(\theta_k)$ then the probability of at least one type I error occurring can be bounded as

$$P\left(\exists k : \left\{|\hat{M}_k| > \theta_k \text{ and } M_k = 0\right\}\right) \leq p \cdot \frac{c}{n^q \cdot \theta^{2q}}$$

Using that $p = F \cdot n^\alpha$ and if we set $\theta = Kn^{\frac{-1+\eta}{2}}$ the upper bound becomes

$$F \cdot n^\alpha \cdot \frac{c}{n^q \cdot K^{2q} n^{-\frac{1}{2}\cdot 2q} \cdot n^{\frac{\eta}{2}\cdot 2q}} = \frac{F \cdot c}{K^{2q}} \cdot n^{\alpha - \eta \cdot q}$$

Thus, for $\eta > \frac{\alpha}{q}$ we have that

$$\lim_{\substack{n \to \infty \\ p \to \infty}} P\left(\exists k : \left\{|\hat{M}_k| > \theta_k \text{ and } M_k = 0\right\}\right) = 0$$

Type *II* Error

The event of a type II error is defined as $\left\{|\hat{M}_k| \leq \theta_k | M_k \neq 0\right\}$ with $P\left(|\hat{M}_k| \leq \theta_k\right) = 2P\left(\frac{\hat{M}_k^{2q} - E\left(\hat{M}_k^{2q}\right)}{\sqrt{E\left(\hat{M}_k^{2q} - E\left(\hat{M}_k^{2q}\right)\right)}} \leq \frac{\theta_k^{2q} - E\left(\hat{M}_k^{2q}\right)}{\sqrt{E\left(\hat{M}_k^{2q} - E\left(\hat{M}_k^{2q}\right)\right)}}\right)$. Using Lemma 6.1 we have that $\sqrt{E\left(\hat{M}_k^{2q} - E\left(\hat{M}_k^{2q}\right)\right)} = \frac{c_k}{\sqrt{n}}$ and put $\theta_k = Kn^{-\frac{1}{2}+\eta}$ which allows to write $P\left(|\hat{M}_k| \leq \theta_k\right) = P\left(|\hat{M}_k| \leq c_k^{-1}\left(Kn^{2q\eta} - E\left(\hat{M}_k^{2q}\right)n^q\right)\right)$. Thus, given that if $M_k \neq 0$ we have that $E\left(\hat{M}_k^{2q}\right) > 0$ and if $\eta < \frac{1}{2}$ we obtain

$$\lim_{n \to +\infty} P\left(|\hat{M}_k| \leq c_k^{-1}\left(Kn^{2q\eta} - E\left(\hat{M}_k^{2q}\right)n^q\right)|M_k \neq 0\right) = 0.$$

6.7.6 Proof of Theorem 6.4

To start with, we bound $P\left(||\Xi|| > \varepsilon\right)$ as

$$P\left(||\Xi|| > \varepsilon\right) = P\left(||\Xi||^{2q} > \varepsilon^{2q}\right) \leq P\left(\text{trace}(\Xi^{2q}) > \varepsilon^{2q}\right) \leq \frac{E\left[\text{trace}(\Xi^{2q})\right]}{\varepsilon^{2q}}$$

Given the oracle property of the operator $\tau_O(,)$, the matrixΞ_p has the same sparsity structure than the projected population matrix in (6.3.1). Moreover, for any power $||\Xi||^{2q}$ with $q \in \mathbb{N}^+$ the sparsity pattern is also given by the form in (6.3.1). Thus, for $d = 1, \ldots, K$ we have that

$$\Xi_{d,d}^{2q} = \sum_{i_1=1}^{K} \cdots \sum_{i_{2q-1}=1}^{K} \prod_{l=1}^{2q} \Xi_{i_1,\ldots,i_{2q-1},l}$$

Taking the expectation leads to

$$\begin{aligned}
E\left(\Xi_{d,d}^{2q}\right) &= E\left(\sum_{i_1=1}^{K}\cdots\sum_{i_{2q-1}=1}^{K}\prod_{l=1}^{2q}\Xi_{i_1^k,\ldots,i_{2q-1},l}\right)\\
&= \sum_{i_1=1}^{K}\cdots\sum_{i_{2q-1}=1}^{K}E\left(\prod_{l=1}^{2q}\Xi_{i_1^k,\ldots,i_q^k,l}\right)\\
&\le \sum_{i_1=1}^{K}\cdots\sum_{i_{2q-1}=1}^{K}\prod_{l=1}^{2q}E\left(\Xi_{i_1^k,\ldots,i_{2q-1},l}^{2q}\right)^{\frac{1}{2q}}\\
&= \sum_{i_1=1}^{K}\cdots\sum_{i_{2q-1}=1}^{K}\prod_{l=1}^{2q}\left(\frac{c_{i_1^k,\ldots,i_{2q-1},l}}{n^q}\right)^{\frac{1}{2q}}\\
&\le \sum_{i_1=1}^{K}\cdots\sum_{i_{2q-1}=1}^{K}\frac{c_{i_1^k,\ldots,i_{2q-1}}}{n^q}\\
&\le K^{2q-1}\frac{c_d}{n^q}
\end{aligned}$$

For $d = K+1,\ldots,p$, we have that

$$\begin{aligned}
E\left(\Xi_{d,d}^{2q}\right) &= E\left(\prod_{l=1}^{2q}\Xi_{i_1,\ldots,i_{2q-1},l}\right)\\
&\le \prod_{l=1}^{2q}E\left(\Xi_{l}^{2q}\right)^{\frac{1}{2q}}\\
&\le \frac{c_d}{n^q}
\end{aligned}$$

Hence,

$$\begin{aligned}
E\left[\mathrm{trace}(\Xi^{2q})\right] &= \sum_{d=1}^{K}K^{2q-1}\frac{c_d}{n^q}+\sum_{d=K+1}^{p}\frac{c_d}{n^q}\\
&\le K^{2q}\frac{c}{n^q}+K\frac{c'}{n^q}
\end{aligned}$$

Thus, if we set $K = Mp^{\beta}$ we have that

$$E\left[\mathrm{trace}(\Xi^{2q})\right] \le \frac{c\cdot M^{2q}p^{2q\cdot\beta}}{n^q}+\frac{c'\cdot Mp^{\beta}}{n^q}$$

Given that $p = Fn^{\alpha}$ this becomes

$$E\left[\mathrm{trace}(\Xi^{2q})\right] \le \frac{c\cdot M^{2q}F^{2q\cdot\beta}n^{2q\cdot\beta\cdot\alpha}}{n^q}+\frac{c'\cdot F^{2q\cdot\beta}n^{\beta\cdot\alpha}}{n^q}$$

If $\beta < \frac{1}{2\cdot\alpha}$ then $2q\cdot\beta\cdot\alpha - q < 0$ and hence

$$\lim_{\substack{n\to\infty \\ p\to\infty}} P\left(||\tau_O\left(\psi^t P\hat{\Sigma}P^t\psi\right) - \psi^t P\Sigma P^t\psi|| > \varepsilon\right) = 0$$

Chapter 7

Conclusion

In this thesis we introduce several new concepts in different fields of statistical analysis.

Wavelet Analysis: We generalize the notion of autocorrelation wavelet functions of discrete wavelet vectors by introducing the concept of cross-correlation functions.

Time Series Modelling: All existing LSW model frameworks in both, the univariate and the multivariate setting, are based on the use of an orthogonal increment process. As we have seen in Chapters 3 and 4, the orthogonality assumption leads to restrictive assumptions on the form of the local autocovariance function, especially in the multivariate case. We show that the local autocovariance function of LSW processes with non-orthogonal increment processes remains asymptotically unique. Thus, we introduce an enlarged class of locally stationary processes which contains the standard LSW process model as a sub-class. In Chapter 5 we illustrate how introducing a factor model specification of the increment process in the MLSW framework leads to a dynamic factor model specification with time-varying factor loadings. Based on the LSW factor model we introduce new concept such as the local factor inertia.

Estimation: In Chapters 3 and 4 we provide a periodogram-based estimator of the evolutionary wavelet spectrum for the generalized LSW process model. We show that the smoothed and corrected version of the estimator is consistent. Chapter 6 defines an estimator of the projections of covariance matrices projected on an unbalanced Haar wavelet basis in the case of white noise processes. We show that our estimator is consistent in operator norm as both, the sample size and the number of variables goes to infinity, given that the population covariance matrix has a sparse representation in the wavelet domain. Simulation studies and real data analyses show the good performance of the thresholding estimator compared to shrinkage-type estimators.

Application: This thesis contains two main applications. In Chapter 5 we use the LSW factor model to define the notion of local factor inertia and local distance measure between pairs of time-varying factor loadings. These two concepts are then used to model the degree of economic integration of the European currency area. Moreover, we propose an adaptive clustering approach of the Euro member states based on a combined use of the local distance measure and the k-means clustering algorithm. In Chapter 6 we use the thresholding estimator to build an adaptive asset allocation scheme for the components of the Dow Jones Industrial Average.

We conclude this thesis by discussing some possible future research directions. In terms of LSW modelling a very important extension would be to consider LSW processes with serially correlated increments. A short inspection of the asymptotic autocovariance representation of these kind of processes reveals that the local autocovariance function would involve shifted versions of the autocorrelation wavelets of the standard LSW model framework. Hence, showing asymptotic uniqueness of the covariance representation of these kind of processes would first of all involve the deep study of the dependence structure of autocorrelation wavelets at both, different scales and different shifts. A study of the literature on time series models with correlated increment terms, especially on fractional Brownian motion, shows that including serial correlation among the increments could be important when trying to model processes with long-range dependence. Another key issue would be to combine the sparse modelling approaches described in Chapter 6 and some aspects of LSW modelling. For example, in Chapter 5 we have seen that the inverse representation of the autocovariance function of multivariate LSW processes converges in operator norm to the EWS matrix only if the number of variables in the MLSW model remains constant as the length of the time series goes to infinity or at least grows at a lower rate. If we include sparsity assumptions on the EWS matrix, we could define a MLSW model framework suitable for high-dimensional settings where the number of variables grows at a larger rate than the length of the time series.

Index

Bibliography

[1] Ahn, S. C., Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors. *Econometrica, Vol. 81, No. 3 (1203-1227)*

[2] Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics, Vol. 34, No. 1 (122-148)*

[3] Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I. (1992). Image Coding Using Wavelet Transform. *IEEE Transactions on image processing, Vol. 1, No. 2 (205-220)*

[4] Baraniuk, R. (1999). Optimal Tree Approximation with Wavelets. *Wavelet Applications in Signal and Image Processing VII, Vol. 3813 (196-207)*

[5] Bai, J. (2003). Inferential theory for factor models of large dimension. *Econometrica, Vol. 71, No. 1 (135-171)*

[6] Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica, Vol. 70, No. 1 (191-221)*

[7] Bai, J., Ng, S. (2002). Determining the number of primitive shocks in factor models. *Journal of Business and Economic Statistics, Vol. 25, No. 1 (52-60)*

[8] Bai, Z. D., Silverstein, J.W. (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional random matrices. *Annals of Probability, Vol. 26, No. 1 (316-345)*

[9] Bai, Z. D., Silverstein, J. W. (1999). Exact separation of eigenvalues of large-dimensional sample covariance matrices. *Annals of Probability, Vol. 27, No. 3 (1536-1555)*

[10] Bai, Z. D., Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability, Vol. 32, No. 1A (553-605)*

[11] Bai, Z. D., Silverstein, J. W. (2010). Spectral Analysis of Large-Dimensional Random Matrices. *Springer, New York, Second Edition*

[12] Bai, Z. D., Yin, Y.Q. (1988). A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis, Vol. 26, No. 2 (166-168)*

[13] Balassa, ?. (1967). Trade Creation and Trade Diversion in the European Common Market. *The Economic Journal, Vol. 77 (121)*

[14] Balassa, B. (2013) The Theory of Economic Integration. *Routledge Revivals (Routledge)*

[15] Berkner, K., Wells, R. O. (2002). Smoothness Estimates for Soft-Threshold Denoising Via Translation Invariant Wavelet Transforms. *Applied and Computational Harmonic Analysis, Vol. 12, No. 1 (1-24)*

[16] Beylkin, G., Coifman, R., Rokhlin, V. (1991) Fast wavelet transforms and numerical algorithms. *Communications of Pure and Applied Mathematics, Vol. 44, No. 2 (141-183)*

[17] Bickel, P. J., Levina, E. (2008). Regularized estimation of large covariance matrices. *The Annals of Statistics, Vol. 36, No. 1 (199-227)*

[18] Box, G., Jenkins, G., Reinsel, G.C. (1994). Time Series Analysis: Forecasting and Control (Third edition.). *Prentice-Hall*

[19] Brockwell, P., Davis, R. (2009). Time series: Theory and Methods. *Springer*

[20] Brodie, J., Daubechies, I., De Mol, C., Giannone, D. and Loris, I. (2009). Sparse and stable Markowitz portfolios. *Proceedings of the National Academy of Sciences of the USA, Vol. 106 (12267-12272)*

[21] Breitung, J., Pigorsch, J. (2013). A canonical correlation approach for selecting the number of dynamic factors. *Oxford Bulletin of Economics and Statistics, Vol. 75, No. 1 (23-36)*

[22] Cardinali, A., Nason, G. (2010). Costationarity of locally stationary time series. *Journal of Time Series Econometrics, Vol. 2, No. 2 (1-35)*

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

[24] Cho, H., Fryzlewicz, P. (2015). Multiple change-point detection for high-dimensional time series via Sparsified Binary Segmentation. *Journal of the Royal Statistical Society Series B, Vol. 77 (475-507)*

[25] Cooley, J. W., Tukey, J. W. (1965). An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation, Vol. 19, No. 90 (297-301)*

[26] Cont, C. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance, Vol. 1, No. 2 (223-236)*

[27] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics, Vol. 25, No. 1 (1-37)*

[28] Dahlhaus, R. (2012). Locally stationary processes. *Handbook of Statistics (351-413)*

[29] Davis, C., Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. *SIAM Journal of Numerical Analysis, Vol. 7, No. 1 (1-46)*

[30] Daubechies, I. (1988). Orthonormal bases of compactly supported wavelets. *Communications on Pure and Applied Mathematics, Vol. 41 No. 7 (909-996)*

[31] Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory, Vol. 36, No. 5 (961-1005)*

[32] Daubechies, I. (1992). Ten Lectures on Wavelets. *SIAM*

[33] Daubechies, I., Sweldens, W. (1998). Factoring wavelet transforms into lifting steps. *Journal of Fourier Analysis and Applications, Vol. 4, No. 3 (247-269)*

[34] Deslauriers, G., Dubuc, S. (1989). Symmetric iterative interpolation processes. *Constructive Approximation, Vol. 5 (49-68)*

[35] DeVore, R. A., Jawerth, B., Lucier, B. J. (1992). Image compression through wavelet transform coding. *IEEE Transactions on Information Theory, Vol. 38, No. 2 (719-746)*

[36] Donoho, D. L. (1995). De-Noising by Soft-Thresholding. *IEEE Transactions on Information Theory, Vol. 41, No. 3 (613-627)*

[37] Donoho, D. L., Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika, Vol. 81, No. 3 (425-455)*

[38] Donoho, D. L., Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association, Vol. 90, No. 432 (1200-1224)*

[39] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage: asymptopia. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 57, No. 2 (301-369)*

[40] Donoho, D. L., Johnstone, I. M. (1994). Minimax estimation via wavelet shrinkage. *The Annals of Statistics, Vol. 26, No. 3 (879-921)*

[41] Dubuc, S. (1986). Interpolation through an iterative scheme. *Journal of Mathematical Analysis, Vol. 114, No. 1 (185-204)*

[42] Eckley, I., Nason, G. (2005). Effcient computation of the discrete autocorrelation wavelet inner product matrix. *Statistics and Computing, Vol. 15, No. 2 (83-92)*

[43] Eichler, M., Motta, G., von Sachs, R. (2011). Fitting dynamic factor model to nonstationary time series. *Journal of Econometrics, Vol. 163, No. 1 (51-70)*

[44] El Karoui, N. (2008). Operator norm consistent estimation of large-dimensional sparse covariance matrices. *The Annals of Statistics, Vol. 36, No. 6 (2717-2756)*

[45] Engle, R. F., Watson, M. W. (1981). A one-factor multivariate time series model of metropolitan wage rate. *Journal of the American Statistical Association, Vol. 76, No. 376 (774-781)*

[46] Fama, E. F. (1965). The behavior of stock market prices. *The Journal of Business, Vol. 38, No. 1 (34-67)*

[47] Fama, E. F., French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics, Vol. 33, No. 1 (3-56)*

[48] Fan, J., Fan, Y. and Lv, J. (2012). Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association, Vol. 107, No. 498 (592-606)*

[49] Fan, J., Zhang, J., and Yu, K. (2012). High-dimensional covariance matrix estimation in approximate factor models. *The Annals of Statistics, Vol. 39, No. 6 (3320-3356)*

[50] Fan, J., Liao, Y. and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Vol. 75 No. 4 (603-680)*

[51] Fan, J., Mincheva, M. (2011). High-dimensional covariance matrix estimation using a factor model. *Journal of Econometrics, Vol. 147 (186-197)*

[52] Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000). The Generalized Dynamic-Factor Model: Identification and Estimation. *The Review of Economics and Statistics, Vol. 82, No. 4 (540-554)*

[53] Forni, M., Lippi, M. (2001). The Generalized Dynamic Factor Model: Representation Theory *Econometric Theory, Vol. 17, No. 6 (1113-1141)*

[54] Forni, M., Reichlin, L. (1998). Dynamic common factors in large cross-sections. *Empirical Economics, Vol. 21 (27-42)*

[55] Forni, M., Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies, Vol. 65 (453-474)*

[56] Fourier, J. B. (1822). Théorie Analytique de la Chaleur. *Paris: Firmin Didot*

[57] Fryzlewicz, P. (2005). Modelling and forecasting financial log-returns as locally stationary wavelet processes. *Journal of Applied Statistics, Vol. 32 (503-528)*

[58] Fryzlewicz, P. (2008). A waveletFisz approach to spectrum estimation. *Journal of Time Series Analysis, Vol. 29 No. 5 (868-880)*

[59] Fryzlewicz, P., Nason, G. (2006). HaarFisz estimation of evolutionary wavelet spectra. *Journal of the Royal Statistical Society. Series B, Vol. 68 No. 4 (611-634)*

[60] Fryzlewicz, P., van Bellegem, S., and von Sachs, R. (2003). Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics, Vol. 55 No. 4 (737-764)*

[61] Geman, S. (1980). A limit theorem for the norm of random matrices. *Annals of Probability, Vol. 8, No. 2, (252-261)*

[62] Geweke, J. (1977). The dynamic factor analysis of economic time series. *Dennis J. Aigner and Arthur S. Goldberger (Eds.), Latent Variables in Socio-Economic Models*

[63] Girardi, M., Sweldens, W. (1997). A new class of unbalanced Haar wavelets that form an unconditional basis for L_p on general measure spaces. *Journal of Fourier Analysis and Applications, Vol. 3, No. 4 (457-474)*

[64] Grossmann A., Morlet J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM Journal of Mathematical Analysis, Vol. 15, No.4 (723-736)*

[65] Haar, A. (1910). Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen, Vol. 71, No. 1 (38-53)*

[66] Hamilton, J. D. (1994). Time Series Analysis. *Princeton University Press*

[67] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, Vol. 24 (417-441 and 498-520)*

[68] Horn, R. A., Johnson, C. R. (1990). Matrix Analysis. *Cambridge University Press*

[69] Huang, J., Yang, L. (2010). Correlation matrix with block structure and efficient sampling methods. *The Journal of Computational Finance, Vol. 14, No. 1 (81-94)*

[70] Isserlis, L. (1918). On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika, Vol. 12 (134-139)*

[71] Johnson, H. (1965). An Economic Theory of Protection, Tariff Bargaining and the Formation of Customs Unions. *Journal of Political Economy Vol. 73 (256283)*

[72] Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics, Vol. 29, No. 2 (295-327)*

[73] Jarque, C. M., Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters, Vol. 6, No. 3 (255-259)*

[74] Ledoit, O., Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management, Vol. 30, No. 4 (110-119)*

[75] Ledoit, O., Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis, Vol. 88, No. 2 (365-411)*

[76] Ledoit, O., Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics, Vol. 40, No. 2 (1024-1060)*

[77] Lipsey, R. G. (1957). The Theory of Customs Union: Trade Diversion and Welfare. *Economica, Vol. 24 (40-46)*

[78] Lloyd., S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory, Vol. 28, No. 2 (129137)*

[79] Lütkepohl, H., (2008). Impulse response function. *The New Palgrave Dictionary of Economics (2nd ed.)*

[80] Lütkepohl, H., Poskitt, D. S. (1991). Estimating orthogonal impulse responses via vector autoregressive models. *Econometric Theory, Vol. 7, No. 4 (487-496)*

[81] Lütkepohl, H., Saikkonen, P. (1997). Impulse response analysis in infinite order cointegrated vector autoregressive processes. *Journal of Econometrics, Vol. 81 (127157)*

[82] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, Vol. 1 (281-297)*

[83] Mallat, S. G. (1989a). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 11 No.7 (674-693)*

[84] Mallat, S. G. (1989b). Multiresolution approximations and wavelet orthonormal bases of L2. *Transactions of the American mathematical society, Vol. 315 No. 1 (69-87)*

[85] Mallat, S. G. (1989b). Multifrequency channel decompositions of images and wavelet models. *IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. 37 No. 12 (2091-2110)*

[86] Mallat, S. G. (2008). A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way. *Academic Press; 3rd edition*

[87] Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business, Vol. 36, No. 4 (394-419)*

[88] Marcenko, V. A., Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sbornik: Mathematics, Vol. 1, No. 4 (457-483)*

[89] Markowitz, H. (1952). Portfolio selection. *Journal of Finance, Vol. 7, No. 1 (77-91)*

[90] Meinshausen, N., Buehlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *The Annals of Statistics, Vol. 34, No. 3 (1436-1462)*

[91] Meyer, Y. (1986). Principe d'incertitude, bases Hilbertiennes et algebres d'operateurs. *Séminaire N. Bourbaki, Asterisque No. 145-146 (209-223)*

[92] Mikosch, T., Stărikă, C. (1999). Change of structure in financial time series, long range dependence and the GARCH model. *Revstat Statistical Journal, Vol. 2 (41-73)*

[93] Motta, G, Hafner. C., von Sachs, R. (2011). Locally stationary factor models: Identification and nonparametric estimation. *Econometric Theory, Vol. 27, No. 6 (1279-1319)*

[94] Nason, G. P., von Sachs, R. and Kroisandt, G. (2000). Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), Vol. 62, No.2 (271-292)*

[95] Ombao, H., Raz, J., von Sachs, R. and Guo, W. (2002). The SLEX Model of Non-Stationary Random Process. *Annals of the Institute of Statistical Mathematics, Vol. 54 (171-200)*

[96] Ombao, H., von Sachs, R and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *Journal of the American Statistical Association, Vol. 100, No. 470 (520-531)*

[97] Onatski, A. (2010. Testing hypotheses about the number of factors in large factor models. *Econometrica, Vol. 77, No. 5 (1447-1479)*

[98] Park, T.A., Eckley, I., Ombao, H. (2014). Estimating time-evolving coherence between signals via multivariate locally stationary wavelet processes *IEEE Transactions on Signal Processing, Vol. 62, No. 20 (5240-5250)*

[99] Parzen, E. (1959). Statistical inference on time series by Hilbert space methods, I. *Technical Report for the Office of Naval Research, No. 23 (252-382)*

[100] Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, Vol. 2 (559-572)*

[101] Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society. Series B (Methodological), Vol. 27 No. 2 (204-237)*

[102] Priestley, M. B. (1981a). Spectral Analysis and Time Series, Volume 1. *Academic Press*

[103] Priestley, M. B. (1981b). Spectral Analysis and Time Series, Volume 2. *Academic Press*

[104] Priestley, M. (1988). Non-linear and non-stationary time series analysis. *Academic Press*

[105] Priestley, M. B., Tong, H. (1973). On the analysis of bivariate non-stationary processes (with discussion) *Journal of the Royal Statistical Society, Series B, (Methodological), Vol. 35 (153-166)*

[106] Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, Vol. 13 (341-360)*

[107] Saito, N., Beylkin, G. (1992). Multiresolution representations using the auto-correlation functions of compactly supported wavelet. *IEEE Trans. Signal Processing, Vol. 41, No. 12 (3584-3590)*

[108] Sanderson, J., Fryzlewicz, P., Jones, M.W. (2010). Estimating linear dependence between nonstationary time series using the locally stationary wavelet model. *Biometrika, Vol. 82, No. 4 (435-446)*

[109] Sargent, T., Sims, C. (1992). Business cycle modeling without pretending to have too much a priori economic theory supported wavelet. *In Christopher A. Sims (Ed.), New Methods in Business Research (45-109)*

[110] Serfling, R. J. (2002). Approximation theorems of mathematical statistics. *Wiley Series in Probability and Statistics*

[111] Shapiro, S. S., Wilk, M. B. (1965) An analysis of variance test for normality (complete samples). *Biometrika, Vol. 52, No. 3-4 (591-611)*

[112] Sharpe, W. F. (1964) Capital Asset Prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, Vol. 19, No. 3 (425-442)*

[113] Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large-dimensional random matrices. *Journal of Multivariate Analysis, Vol. 55 (331-339)*

[114] Silverstein, J. W., Bai, Z. D. (1995). On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *Journal of Multivariate Analysis, Vol. 54 (175-192)*

[115] Silverstein, J. W., Choi, S. I. (1995). Analysis of the limiting spectral distribution of large dimensional random matrices *Journal of Multivariate Analysis, Vol. 54 (295-309)*

[116] Simoncelli, E., Freeman, W. T., Adelson, E. H., Heeger, D. J. (1992). Shiftable multi-scale transforms. *IEEE Transactions on Information Theory, Vol. 38, No. 2 (587-607)*

[117] Sims, C. (1980). Macroeconomics and Reality. *Econometrica Vol. 48, No. 1 (148)*

[118] Stein, C. (1975). Estimation of a covariance matrix. *Rietz lecture, 39th Annual Meeting IMS. Atlanta, Georgia.*

[119] Stein, C. (1986). Lectures on the theory of estimation of many parameters. *Journal of Mathematical Sciences, Vol. 34, No. 1 (1373-1403)*

[120] Stock, J. H., Watson, M. W. (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomics Annual, Vol. 4 (352-394)*

[121] Stock, J. H., Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association, Vol. 97, No. 460 (1167-1179)*

[122] Stock, J. H., Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics, Vol. 20, No. 2 (147-162)*

[123] Sweldens, W. (1996). The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets. *Applied and Computational Harmonic Analysis, Vol. 3, No. 2 (186-200)*

[124] Sweldens, W. (1996). The Lifting Scheme: A Construction Of Second Generation Wavelets. *SIAM Journal of Mathematical Analysis Vol. 29, No. 2 (511-546)*

[125] Szegö, G. (1915). Ein Grenzwertsatz ber die Toeplitzschen Determinanten einer reellen positiven Funktion. *Mathematische Annalen Vol. 76, No. 4 (490503)*

[126] Tibshirani, R. (1994). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society. Series B, Vol. 58, No. 1 (267-288)*

[127] Toda, H. Y., Phillips, P. C. B. (1993). Vector autoregressions and causality. *Econometrica Vol. 61 (13671393)*

[128] Van Bellegem, S. von Sachs, R. (2008). Locally adaptive estimation of evolutionary wavelet spectra. *Annals of Statistics, Vol. 36, No.4 (1879-1924)*

[129] Vetterli, M., Herley, C. (1992). Wavelets and filter banks: Theory and design. *IEEE Transactions on Signal Processing, Vol. 40, No.9 (2207-2232)*

[130] Weyl, H. (1912). Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen. *Mathematische Annalen, Vol. 71, No. 4 (441-479)*

[131] Wu, W. B., Xiao, H. (2012). Covariance matrix estimation in time series. *Handbook of Statistics (187-209)*

[132] Yuan, M., Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B, Vol. 68, No. 1 (49-67)*

[133] Zheng, Z., Fan, Y., Lv, J. (2014). High dimensional thresholded regression and shrinkage effect. *Journal of the Royal Statistical Society. Series B, Vol. 76, No. 3 (627-649)*