Analysis of Financial Time-Series using Fourier and Wavelet Methods

Philippe Masset[*]

Current Version: October 2008

Abstract

This paper presents a set of tools, which allow gathering information about the frequency components of a time-series. We focus on the concepts rather than giving too much weight to mathematical technicalities.

In a first step, we discuss spectral analysis and filtering methods. Spectral analysis can be used to identify and to quantify the different frequency components of a data series. Filters permit to capture specific components (e.g. trends, cycles, seasonalities) of the original time-series. Both spectral analysis and standard filtering methods have two main drawbacks: (i) they impose strong restrictions regarding the possible processes underlying the dynamics of the series (e.g. stationarity), and, (ii) they lead to a pure frequency-domain representation of the data, i.e. all information from the time-domain representation is lost in the operation.

In a second step, we introduce wavelets, which are relatively new tools in economics and finance. They take their roots from filtering methods and Fourier analysis. But they overcome most of the limitations of these two methods. Indeed their principal advantages are the following: (1) they combine information from both time-domain and frequency-domain and, (2) they are also very flexible and do not make strong assumptions concerning the data generating process for the series under investigation.

JEL classification: C14; G10; C60
Keywords: spectral analysis, wavelets

---

[*]University of Fribourg, Department of Finance, Bd. Pérolles 90, CH - 1700 Fribourg, philippe.masset@unifr.ch (corresponding author).

1   Introduction

The purpose of this paper is to present a set of methods and tools belonging to so-called *frequency domain* analysis and to explain why and how they can be used to enhance the more conventional *time domain* analysis. In essence, time domain analysis studies the evolution of an economic variable with respect to time, whereas frequency domain analysis shows at which frequencies the variable is active. We focus on concepts rather than technicalities and illustrate each method with fictitious examples and applications using real datasets.

The usual *time-domain* approach aims at studying the temporal properties of a financial or economic variable, whose realizations are recorded at a predetermined frequency. This approach does not convey any information regarding the frequency components of a variable. Thus it makes the implicit assumption that the relevant frequency to study the behavior of the variable matches with its sampling frequency. An issue arises, however, if the variable realizations depend (in a possibly complicate manner) on several frequency components rather than just one. In such a case, the time-domain approach will not be able to efficiently process the information contained into the original data series.

In this paper, we start by discussing methods belonging to the *frequency-domain* analysis. These tools are very appealing to study economic variables that exhibit a cyclical behavior and/or are affected by seasonal effects (e.g., GDP, unemployment). Spectral analysis and Fourier transforms can be used to quantify the importance of the various frequency components of the variable under investigation. In particular, they permit to infer information about the length of a cycle (e.g. business cycle) or a phase (e.g. expansion or recession). Presence of such patterns also imposes the use of appropriate methods when it comes to model the dynamics of the variable. Filtering methods have proven useful in this context. Notably, filters may serve to remove specific frequency components from the original data series.

In a second step, we introduce wavelets. During the last two decades, wavelets have become increasingly popular in scientific applications such as signal processing and functional analysis. More recently, these methods have also begun to be applied to financial datasets. They are indeed very attractive as they possess the unique ability to provide a complete representation of a data series from both the time and frequency perspectives simultaneously. Hence, they permit to break down the activity on the market into different frequency components and to study the dynamics of each of these components separately. They do not suffer from some of the limitations of standard frequency-domain methods (see section 3) and can be employed to study a financial variable, whose evolution through time is dictated by the interaction of a variety of different frequency components. These components may also behave according to non-trivial (non-cyclical) dynamics - e.g., regime shifts, jumps, long-term trends.

For instance, the presence of heterogeneous agents with different trading horizons may generate very complex patterns in the time-series of stock prices (see, e.g., Muller et al. (1995) and Lynch/Zumbach (2003)). This heterogeneity may in particular induce long-memory in stock returns volatility. In such a case, studying the properties of a time-series and trying to model it from the perspective of a single frequency can be misleading. Much information will be lost because of the naive and implicit aggregation of the different frequency components into

a single component. Furthermore, as these components may interact in a complicated manner and may be time-varying or even non-stationary, standard methods like Fourier analysis are not appropriate. Therefore, one has to resort to more flexible filtering methods like wavelets.

The remaining of this paper is structured as follows. In section 2, we discuss spectral analysis (see section 2.1) and filtering methods (see section 2.2). Section 3 is devoted to the presentation of wavelets: section 3.1 explains the relevant theoretical background, section 3.2 discusses the implementation of these methods and section 3.3 presents a complete case study.

2   Frequency-domain analysis

2.1   Spectral analysis: some basics and an example

Studying the properties of an economic variable in the *time-domain* is done using so-called time-series analysis. Similarly, the purpose of spectral analysis is to study the properties of an economic variable over the frequency spectrum, i.e. in the *frequency-domain.* In particular, the estimation of the population spectrum or the so-called power spectrum (also known as the energy-density spectrum) aims at describing how the variance of the variable under investigation can be split into a variety of frequency components. Many economics and econometrics books and articles have been published on the subject during the last 40 years (see Iacobucci (2003) for a short literature review). Our discussion is based primarily on Hamilton (1994) and Gençay et al. (2002).

FOURIER TRANSFORM

The basic idea of spectral analysis is to reexpress the original time-series[1] $x(t)$ as a new sequence $X(f)$, which determines the importance of each frequency component f in the dynamics of the original series. This is achieved using the discrete version of the Fourier transform,[2]

$$X(f) = \sum_{t=-\infty}^{\infty} x(t)e^{-i2\pi ft},\tag{1}$$

where f denotes the frequency at which $X(f)$ is evaluated. In order to get more insight into this decomposition, one may think about the De Moivre's (Euler's) theorem, which allows to write $e^{-i2\pi ft}$ as

$$e^{-i2\pi ft} = \cos(2\pi ft) - i\sin(2\pi ft).$$

Hence application of formula (1) tantamounts to project the original signal $x(t)$ onto a set of sinusoidal functions, each corresponding to a particular frequency component. Furthermore, one

---

[1]   In the following, we use the terms "time-series" and "signal" interchangeably.

[2]   The discrete version of the Fourier transform is used because the time-series is recorded at discrete time intervals.

can use the inverse Fourier transform to recover the original signal $x(t)$ from $X(f)$:

$$x(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(f)e^{-i2\pi ft}df. \tag{2}$$

Formula (2) shows that $X(f)$ determines how much of each frequency component is needed to synthetize the original signal $x(t)$.

POPULATION SPECTRUM AND SAMPLE PERIODOGRAM

Following Hamilton (1994), we define the *population* spectrum of a covariance-stationary process $y(t)$ as

$$s_y(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma_j e^{-i\omega j}, \tag{3}$$

where γ_j is the jth autocovariance of $y(t)$,[3] $\omega = 2\pi f$ is a real scalar, which is related to the frequency $f = 1/\tau$ at which the spectrum is evaluated and τ is the period length of one cycle at frequency f.[4] One may notice that the right part of equation (3) is indeed the discrete-time Fourier transform of the autocovariance series. There is also a close link between this expression and the autocovariance generating function, which is defined by

$$g_Y(z) = \sum_{j=-\infty}^{\infty} \gamma_j z^j,$$

where z denotes a complex scalar. This implies that one can easily recover the autocovariance generating function from the spectrum. In the same spirit, an application of the inverse discrete-time Fourier transform allows a direct estimation of the autocovariances from the population spectrum.

In practice, there are essentially two approaches that can be used to estimate the sample *periodogram* $\widehat{s}_y(\omega)$. The first is non-parametric because it infers the spectrum from a sample of realizations of the variable y, without trying to assign an explicit structure to the data generating process underlying its evolution. The estimation of the sample periodogram is straightforward as it is directly related to the squared magnitude of the discrete-time Fourier transform $|Y(f)|$ of the time-series $y(t)$,

$$\widehat{s}_y(\omega) = \frac{1}{2\pi} \frac{1}{T} |Y(f)|^2, \tag{4}$$

where T is the length of the time-series $y(t)$. $|Y(f)|^2$ is also known as the power spectrum of $y(t)$. This approach is usually called the "periodogram" method. As noted in Hamilton

---

[3]   The jth autocovariance of $y(t)$ is given by $\gamma_j = E[(y(t) - \mu)(y(t-j) - \mu)]$, where μ denotes the expected value of $y(t)$.

[4]   As an example, let's consider an economic variable, whose evolution is fully determined by the state of the economy. A complete business cycle lasts on average 30 months and therefore $f = 1/30$ months.

(1994), its accuracy seems questionable as the confidence interval for the estimated spectrum is typically very broad. Furthermore, the variance of the periodogram does not tend to zero as the length of the data series tends to infinity. This means that the periodogram is not a consistent estimator of the population spectrum. Therefore, modified versions of the periodogram have been put forward. For instance, smoothed periodogram estimates have been suggested as a way to reduce the noise of the original estimator and to improve its accuracy. The idea underlying this approach is that $s_y(\omega)$ will be close to $s_y(\lambda)$ when ω is close to λ. This suggests that $s_y(\omega)$ might be estimated with a weighted average of the values of $s_y(\lambda)$ for values of λ in a neighbourhood around ω, where the weights depend on the distance between ω and λ (Hamilton (1994)). The weights are typically determined by a kernel weighting function. Welch's method (Welch (1967) and Childers (1978)) is another alternative, which is based on a rather simple idea: instead of estimating a single periodogram for the complete sample, one divides the original sample into subsambles, estimates the periodogram for each subsample, and computes the average periodogram over all subsamples.

The second approach is based on some parameterization of the data generating process of $y(t)$. Methods belonging to this category are close in spirit to the population spectrum, i.e. to a direct application of equation (3). Typically some specification based on an autoregressive (ARMA) representation is chosen to represent the temporal dynamics of the variable. The model is then calibrated, i.e. the ARMA coefficients are estimated from the realizations of the process $y(t)$. These estimated coefficients are employed to calculate the spectrum. As long as the autocovariances are reasonably well estimated, the results would also be reasonably close to the true values. A detailed discussion of the various parametric methods (e.g., the covariance, Yull-Walker and Burg methods) is beyond the scope of this introduction but it is interesting to note that parametric methods are particularly effective when the length of the observed sample is short. This is because of their ability to distinguish the noise from the information contained in the data.

EXAMPLE

We now turn to the discussion of a simple example. We consider a time-series, which has the following dynamic

$$x_t = a \cdot \cos(\frac{2\pi t}{5}) + b \cdot \sin(\frac{2\pi t}{21}) + \varepsilon_t,$$

where ε_t is random term that follows a normal distribution with mean zero and unit variance. One may observe that the process is driven by two cyclical components, which repeat themselves respectively each 5 and 21 units of time. The full line in figure 1 shows the first 100 (simulated) realizations of x_t; the dotted lines are for the cos and sin functions. At first glance, it seems difficult to distinguish the realizations of x_t from a purely random process. Figure 2 reports the autocorrelations (left panel) and partial autocorrelations (right panel) of x_t (upper panel) and of the cos and sin components (bottom panel). Again, it remains difficult, when looking at this figure, to gather conclusive evidence concerning the appropriate model specification for x_t.

On the other hand, results from the Fourier analysis, reported in figure 3, clearly show that there are two cyclical components which drive the evolution of x_t and which repeat themselves
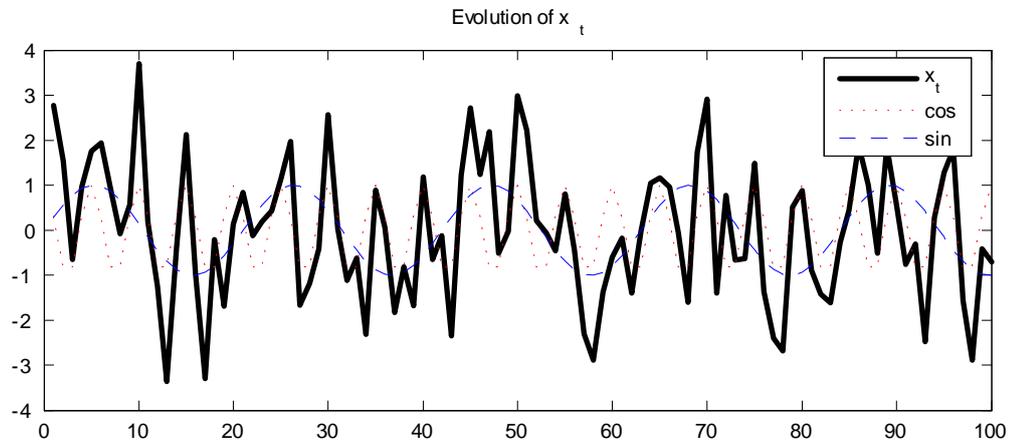
Figure 1: Sample path of x_t.
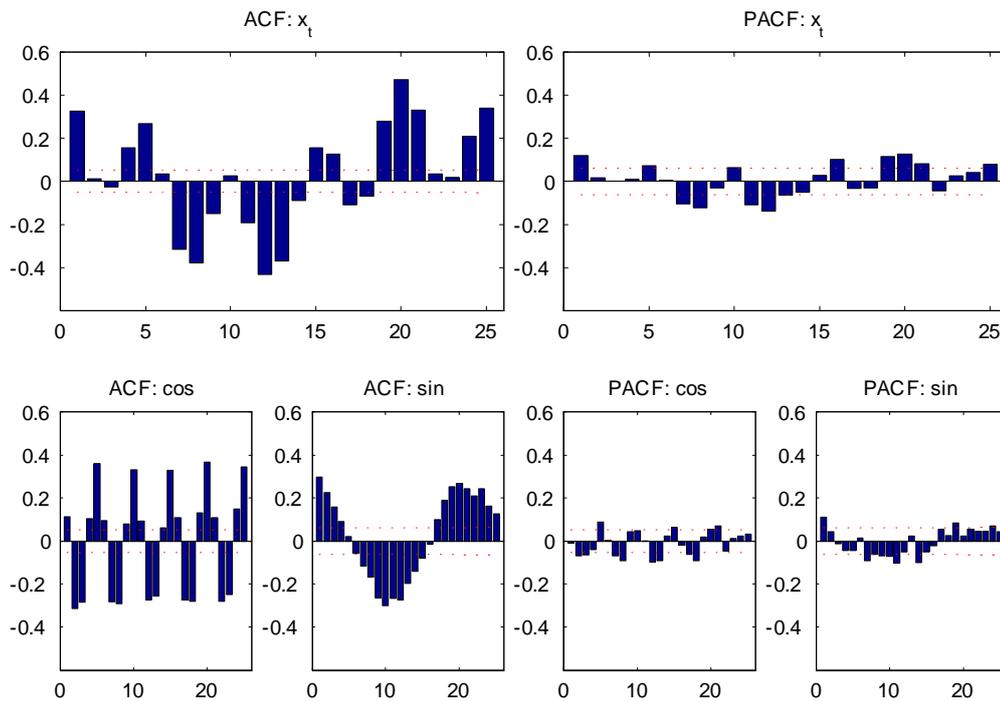
Figure 2: Autocorrelations and partial autocorrelations. The upper panel reports the autocorrelations (left panel) and partial autocorrelations (right panel) of x_t. The lower panel shows similar statistics for the cosinus and sinus functions.
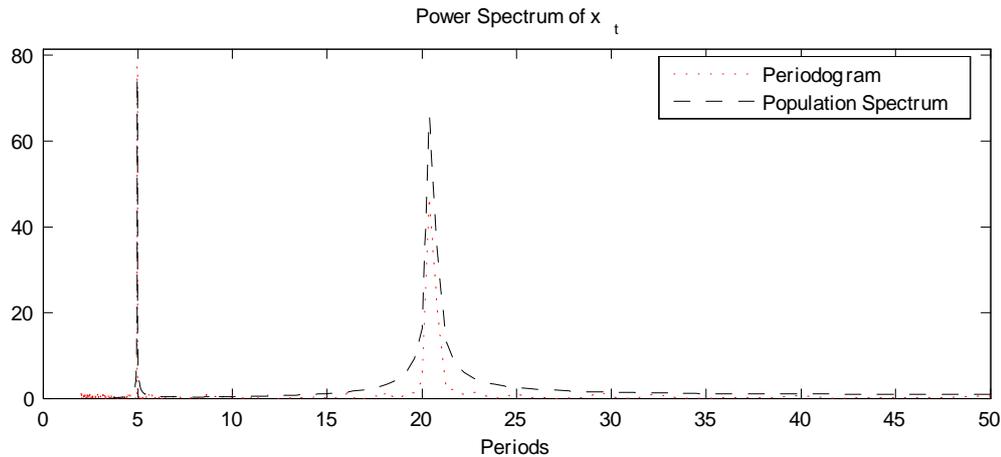
Figure 3: Spectral analysis. The periodogram has been estimated directly from the realizations of x_t. The population spectrum has been estimated on the basis of the theoretical autocovariances of x_t.

around each 5 and 21 units of time. This demonstrates the effectiveness of Fourier methods for the study of processes featuring cyclical components.

ILLUSTRATION: HOME PRICES IN NEW-YORK CITY

We now illustrate how spectral methods can be applied to real economic data series. We consider the Case-Shiller home price index for the city of New-York. The dataset covers the period January 1987 to May 2008 and the index is reported on a monthly basis. The upper panel of figure 4 shows the evolution of the index level over this time period, while the lower panel reports the time-series of index returns. Results from the Dickey-Fuller test cannot reject the null hypothesis that the index levels series is non-stationary. Application of the Fourier transform requires the series under study to be stationary. We therefore study the spectral properties of the index using the time-series of returns rather than the levels themselves.

We also estimate the autocorrelations and partial autocorrelations of the index returns up to 48 lags (i.e. 4 years of observations). These are reported in figure 5. The structure of both the autocorrelations and the partial aurtocorrelations indicates that the index returns are significantly autocorrelated and it also suggests a cyclical (or seasonal) behavior of the returns series. This observation is in line with previous results from the literature (see, e.g., Kuo (1996) and Gu (2002)). In order to gain more insight into the presence of such patterns, we compute the power spectrum of the series using parametric and non-parametric methods, see figure 6. The estimated power spectra returned by the two non-parametric methods (periodogram and Welch) are much noisier than the spectra obtained from the parametric methods (Yule-Walker and Burg). The Welch method also seems to result in an oversmoothed estimate of the power spectrum as compared to the other estimates (e.g. periodogram estimates). On the other hand, the difference between the Yule-Walker and the Burg methods is minimal. Nevertheless, the
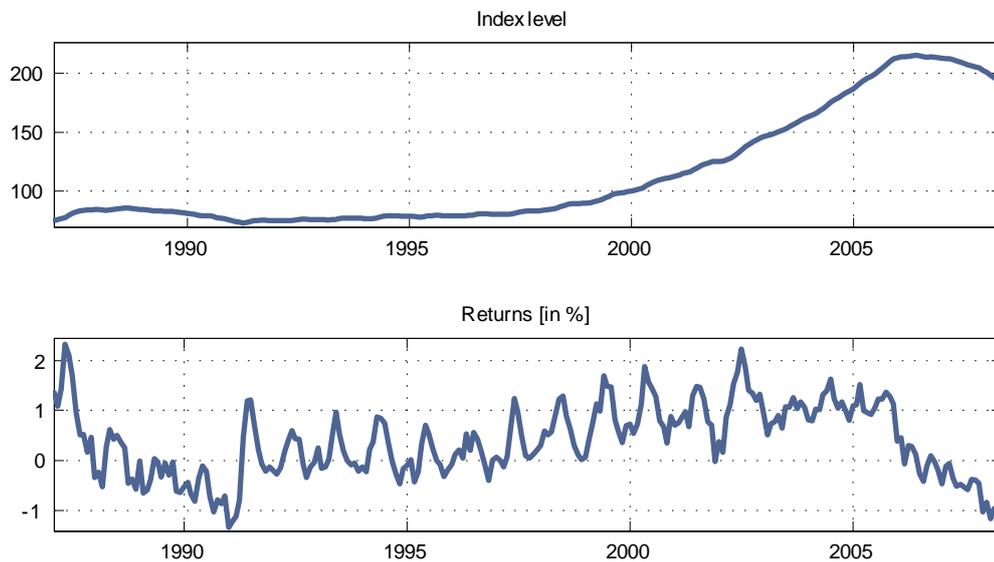
Figure 4: Case-Shiller home price index for the city of New-York. Price levels and
returns are reported in the upper and lower panel respectively.
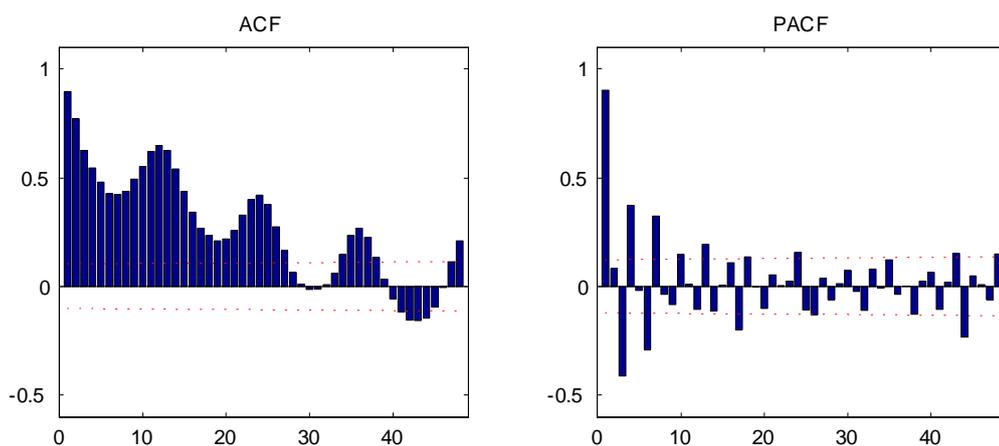
**Figure 5: Autocorrelations and partial autocorrelation of the returns on the New-
York home price index.**
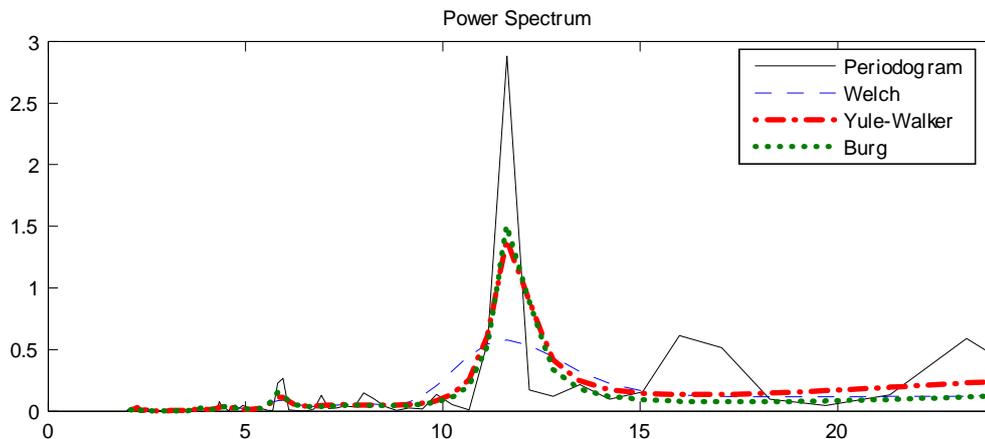
Figure 6: Spectral analysis of the return series. Comparison between parametric (Burg and Yule-Walker) and non-parametric (Periodogram and Welch) methods.

key message remains remarkably similar: strong seasonalities affect home prices and they have a frequency of recurrence of 12 months.

2.2   Filtering methods

A filter is a mathematical operator that serves to convert an original time-series x_t into another time-series y_t. The filter is applied by convolution of the original series x_t with a coefficient vector w,

$$y_t = (w * x)(t) = \sum_{k=-\infty}^{\infty} w_k x_{t-k}. \tag{5}$$

The purpose of this operation is to identify explicitly and to extract certain components from x_t. In the present context, one may want to remove from the original time-series some particular features (e.g. trends, business cycle, seasonalities and noise) that are associated with specific frequency components.

FREQUENCY RESPONSE FUNCTION

Filters in the time-domain can be characterized on the basis of their *impulse response function*, which traces the impact of a one-time unit impulse in x_t on subsequent values of y_t. Similarly, in the frequency domain, the analysis of the *frequency response function* (or transfer function) of a filter tells us which frequency components the filter captures from the original series. The

frequency response function is defined as the Fourier transform of the filter coefficients:[5]

$$H(f) = \sum_{k=-\infty}^{\infty} w_k e^{-i2\pi f k}. \tag{6}$$

The frequency response, $H(f)$, can be further split into two parts,

$$H(f) = G(f)e^{i\theta(f)}, \tag{7}$$

where $G(f)$ is the *gain* function, $e^{i\theta(f)}$ is the *phase* function and θ is called the phase angle (or equivalently the argument of $H(f)$). The gain function is the magnitude of the frequency response, i.e. $G(f) = |H(f)|$. If the application of the filter on x_t results in a phase shift, i.e. if the peaks and lows of x_t and y_t have a different timing, the phase angle θ will be different from zero. The use of uncentered moving average filters leads to this (often) undesirable feature because turning points will be recorded earlier in the original series than in the filtered series. On the other hand, centred (symmetric) moving averages have $\theta(f) = 0$; hence there is no phase shift for this class of filters. For instance, the frequency response of a two-period uncentered moving average filter with coefficients $w_k = \frac{1}{2}$ for $k = 0, 1$ is[6]

$$
\begin{aligned}
H(f) &= \sum_{k=0}^{1} 0.5 e^{-i2\pi f k} \\
&= 0.5 + 0.5 e^{-i2\pi f} \\
&= 0.5(e^{i\pi f} + e^{-i\pi f})e^{-i\pi f} \\
&= \cos(\pi f)e^{-i\pi f}.
\end{aligned}
$$

This result shows that there is indeed a phase shift as the phase angle is $\theta(f) = -\pi f$.[7] Hence the turning points from the original series will be shifted to the right in the filtered series.

Based on their gain functions, filters can be categorized as follows:

- High-pass filters should be able to capture the high-frequency components of a signal; i.e. the value of their gain function $G(f)$ should equal one for frequencies f close or equal to $1/2$.

- Low-pass filters should be able to capture the low-frequency components of a signal; i.e. $G(f) = 1$ for f close or equal to 0.

- Band-pass filters should be able to capture a range of frequency components of a signal; i.e. $G(f) = 1$ for $f_{lo} < f < f_{hi}$.

---

[5]   More generally, the transfer function is obtained as the ratio of the Laplace transform of the filtered signal y_t to the Laplace transform of the original signal x_t. The Fourier transform is indeed a special case of the bilateral Laplace transform.

[6]   See also Gençay et al. (2002).

[7]   In full generality, the phase angle can be computed as $\theta(f) = \arctan(\mathrm{Im}[H(f)]/\mathrm{Re}[H(f)])$, where $\mathrm{Im}[H(f)]$ and $\mathrm{Re}[H(f)]$ are respectively the imaginary part and the real part of $H(f)$.

- All-pass filters capture all the frequency components of a signal; i.e. $G(f) = 1$ for $\forall f$. Such filters leave the frequency components of the original signal unaltered.

Filters are commonly used in economics. The Hodrick/Prescott (1997) filter is probably the most well-known. It has a structure and a gain function which enable it to capture business cycle components.

EXAMPLE

A basic example of a high-pass filter is a filter that simply takes the difference between two adjacent values from the original series; its coefficients are $w_{hi} = [0.5, -0.5]$. Similarly, the most simple low-pass filter is a 2-period moving-average; in this case $w_{lo} = [0.5, 0.5]$. The gain functions for these two filters are reported in figure 7. In wavelet theory, $w_{hi}/\sqrt{2}$ and $w_{lo}/\sqrt{2}$ form the Haar wavelet family. In this case, the low-pass filter w_{lo} is basically an averaging filter, while the high-pass filter w_{hi} is a differencing filter. We will come back to this point in section 3. The gain functions for these two filters are reported in figure 7.
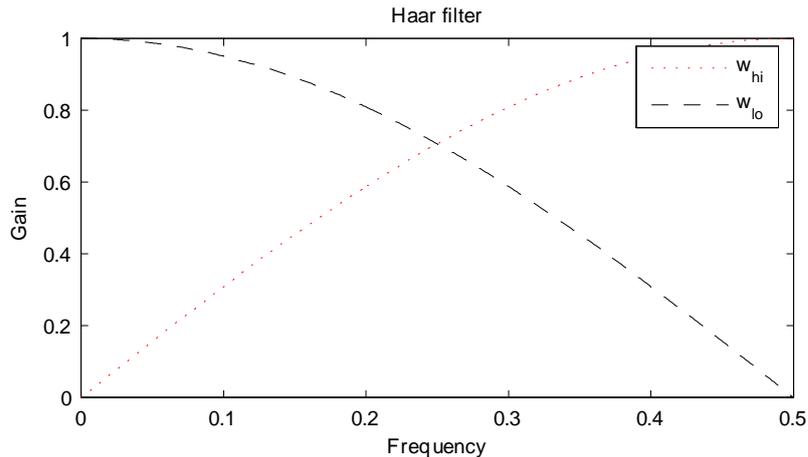
Figure 7: Gain function of the high-pass (differencing) and low-pass (averaging) Haar filters.

ILLUSTRATION

The full line in the left panels of figure 8 shows the (unadjusted) year-to-year changes in the US consumer price index (CPI) from January 2004 to December 2007. We add to the top panel the output series resulting from the application of both a centred and an uncentered (causal) 3-period moving average on the original data. The three filters coefficients have a value of $1/3$. This implies that the output series of the centred moving average at time t is basically the average change in CPI from months $t-1$ to $t+1$. Similarly, the uncentered moving returns the average change in CPI from $t-2$ to t. It is apparent from the figure that the uncentered moving average leads to a phase shift of 1 month. The bottom panel show the outputs of a 7-period moving average. The filter coefficients are equal to $1/7$. Again, we consider both centred and
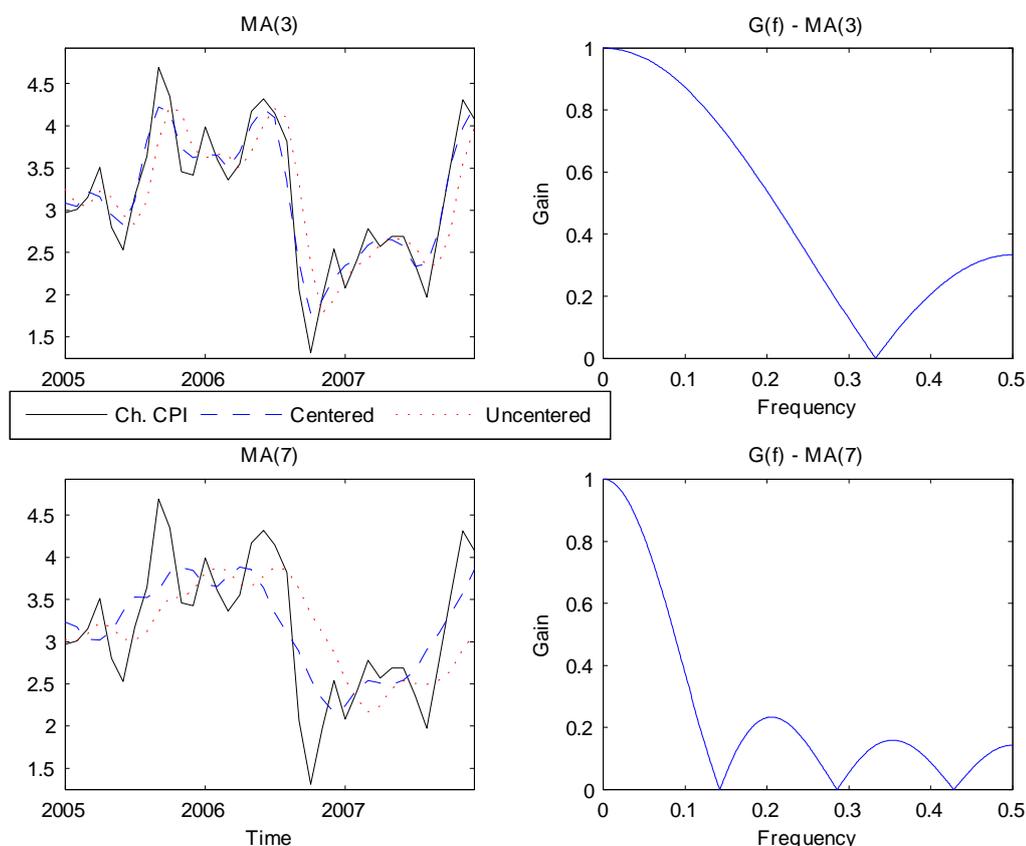
Figure 8: Two moving averagew and their frequency responses. Left panels show the outputs from a 3−period (upper panel) and a 7−period (lower panel) moving average filters applied on the change in US CPI. Right panels report the corresponding gain functions.

uncentered filters. The use of an uncentered moving average leads to a phase shift of 3 months as compared to the centred moving average.

The right part of the figure reports the gain functions for the 3-period and the 7-period moving averages. The gain function is similar for both the uncentered and the centred moving average. The only element that distinguishes these two filters is indeed their phase function. One may notice that (i) both filters are low-pass filters, and (ii) the longer filter captures more efficiently the low frequency components of the original signal than the short filter.

3 Scale-by-scale decomposition with wavelets

To a large extent, wavelets can be seen as a natural extension to spectral and Fourier analysis. This is for two reasons: (i) wavelets do not suffer the weaknesses of Fourier analysis, and (ii) wavelets provide a more complete decomposition of the original time-series than Fourier analysis does.

There are some problems with spectral methods and Fourier transforms. Notably, these methods require the data under investigation to be stationary. This is often not the case in economics and finance. In particular, volatility is known to exhibit complicated patterns like jumps, clustering and long memory. Furthermore, the frequency decomposition delivered by Fourier analysis only makes sense if the importance of the various frequency components remains stable over the sample period. Ex-ante, there is good reason to expect this assumption not to hold for a variety of economic and financial variables. For instance, volatility changes are likely to exhibit a different frequency spectrum when trading activity is intense than when the market is quiet. The short-time Fourier transform (which is also known as the Gabor or windowed Fourier transform) has been suggested to overcome these difficulties. The idea is basically to split the sample into subsamples and to compute the Fourier transform on these subsamples. Hence this extension achieves a better tradeoff between the time and the frequency representation of the original data. Nevertheless, this provides at best a partial solution to the aforementioned issues because it still makes strong restrictions regarding the possible data generating process over each subsample.

Wavelets do not make any of these assumptions. Furthermore, wavelets provide a complete decomposition of the original series, which is located both in time and in frequency. From a mathematical viewpoint, a wavelet is a function, which enables to split a given signal into several components, each reflecting the evolution trough time of the signal at a particular frequency. Wavelet analysis has originally been used in signal processing (e.g. image processing and data compression). Its applications to economics are relatively recent and are mainly due to econophysists. Nevertheless, the range of application of wavelets in finance is potentially wide: denoising and seasonality filtering, decorrelation and estimation of fractionally integrated models, identification of regime shift and jumps, robust estimation of the covariance and correlation between two variables at different time scales, etc.

From a physicist perspective, but with application to time-series analysis, Percival/Walden (2000) provide a mathematically rigorous and exhaustive introduction to wavelets. Struzik (2001) is another sophisticated introduction to wavelets. Struzik (2001) particularly emphasizes the unique ability of non-parametric methods (like wavelets) to let the data speak by themselves. Thus, such methods avoid making misleading interpretations of the coefficients obtained from the calibration of misspecified models. Gençay et al. (2002) discuss the use of wavelets for specific purposes in economics and finance and adopt a more intuitive approach (with many illustrations and examples). Ramsey (2002) surveys the most important properties of wavelets and discusses their fields of application in both economics and finance. Crowley (2007) proposes a genuine guide to wavelets for economists. His paper can be considered as a complete and easily understandable toolkit as he explains precisely in which circumstances to use wavelets and shows how to proceed. Schleicher (2002) is a complementary reference to those already

named, Schleicher focuses on some mathematical concepts underlying the use of wavelets and discuss them in detail using several examples.[8]

3.1 Theoretical background

WHAT IS A WAVELET?

As its name suggests, a wavelet is a small wave. In the present context, the term "small" essentially means that the wave grows and decays in a limited time frame. Figure 9 illustrates this notion by contrasting the values taken by a simple wavelet function (the Morlet function) and the values of the sin function, which can be considered as a sort of "big" wave. In order to clarify the notion of small wave, we start by introducing a function, which is called the *mother wavelet* and is denoted by $\psi(t)$ (see the next paragraph for more details). This function is defined on the real axis and must satisfy two conditions,

$$\int_{-\infty}^{\infty} \psi(t)dt = 0. \tag{8}$$

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt = 1. \tag{9}$$

Taken together, these conditions imply (i) that at least some coefficients of the wavelet function must be different from zero, and (ii) that these departures from zero must cancel out. Clearly the sin function does not meet these two requirements. A vast variety of functions meets conditions 8 and 9. Nevertheless, these conditions are very general and not sufficient for many pratical purposes. Therefore one has to impose additional conditions if one wants to run a specific analysis with wavelets. One of this condition is the so-called *admissibility condition*, which states that a wavelet function is admissible if its Fourier transform,

$$\Psi(f) = \int_{-\infty}^{\infty} \psi(t)e^{-i2\pi ft}dt, \tag{10}$$

is such that

$$C_\Psi = \int_0^\infty \frac{|\Psi(f)|^2}{f}df \quad \text{satisfies} \quad 0 < C_\Psi < \infty. \tag{11}$$

This conditions allows reconstructing a function from its continuous wavelet transform (see Percival/Walden (2000) for more details).

---

[8] This short literature review focuses only on textbook-style references. There are however quite a large amount of economic and financial papers that have employed wavelets for empirical purposes. In addition to the references cited in Ramsey (2002) and Crowley (2007), see, e.g., Gençay et al. (2003), Gençay et al. (2005), Vuorenmaa (2005), Nielsen/Frederiksen (2005), Oswiecimka et al. (2006), Helder/Jin (2007), Fan et al. (2007), Fernandez/Lucey (2007) and Subbotin (2008).
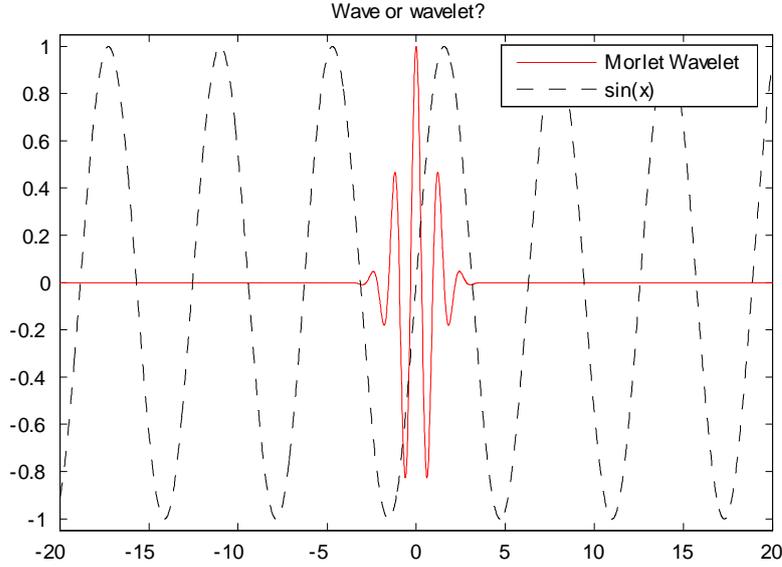
Figure 9: The Morlet wavelet and the sin function.

THE CONTINUOUS WAVELET TRANSFORM (CWT)

As a starting point, we discuss the CWT. In essence, the CWT aims at quantifying the change in a function at a particular frequency and at a particular point in time. In order to be able to achieve this, the mother wavelet $\psi(t)$ is dilated and translated,

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi(\frac{t-u}{s}), \tag{12}$$

where u and s are the location and scale parameters. The term $\frac{1}{\sqrt{s}}$ ensures that the norm of $\psi_{u,s}(t)$ is equal to one. The CWT, $W(u,s)$, which is a function of the two parameters u and s, is then obtained by projecting the original function $x(t)$ onto the mother wavelet $\psi_{u,s}(t)$,

$$W(u,s) = \int_{-\infty}^{\infty} x(t)\psi_{u,s}(t)dt. \tag{13}$$

If one wants to assess the variations of the function on a large scale (i.e. at a low-frequency), one will choose a large value for s, and vice-versa. By applying the CWT for a continuum of location and scale parameters to a function, one is able to decompose the function under study into elementary components. This is particularly interesting for studying a function with a complicated structure, because this procedure allows extracting a set of "basic" components that have a simpler structure than the original function. By "synthesizing" $W(u,s)$, it is also possible to reconstruct the original function $x(t)$ (see Gençay et al. (2002) for more details).

In empirical applications, there are several difficulties with the CWT. First, it is computationally impossible to analyze a signal using *all* wavelet coefficients. CWT is thus more suitable for studying functions than signals or (economics) time-series. Second, as noted by Gençay et al.

(2002), $W(u, s)$ is a function of two parameters and as such it contains a high amount of redundant information. We therefore turn to the discussion of the discrete wavelet transform (DWT). The DWT is grounded on the same concepts as the CWT but is more parsimonious (Gençay et al. (2003)).

THE DISCRETE WAVELET TRANSFORM (DWT)

The key difference between the CWT and the DWT lies in the fact that the DWT uses only a limited number of translated and dilated versions of the mother wavelet to decompose the original signal. The idea is to select u and s so that the information contained in the signal can be summarized in a minimum of wavelet coefficients. This objective is achieved by setting

$$s = 2^{-j} \quad \text{and} \quad u = k2^{-j},$$

where j and k are integers representing the set of discrete translations and discrete dilatations. Gençay et al. (2002) refer to this procedure as the *critical sampling* of the CWT. This implies that the wavelet transform of the original function or signal is calculated only at *dyadic* scales, i.e. at scales 2^j. A further implication is that for a time-series with T observations, the largest number of scales for the DWT is equal to the integer J such that $J = \lfloor log_2(T) \rfloor = \lfloor \log(T)/\log(2) \rfloor$. It is not possible to directly apply the DWT if the length of the original series is not dyadic (i.e. if $J < log_2(T) < J + 1$). In such case, one has either to remove some observations or to "complete" the original series in order to have a series of dyadic length. There exist several methods to deal with this kind of boundary problems (see section 3.2).

The DWT is based on two discrete wavelet filters, which are called the *mother wavelet* $h_l = (h_0, ..., h_{L-1})$ and the *father wavelet* $g_l = (g_0, ..., g_{L-1})$. The mother wavelet is characterized by three basic properties,

$$\sum_{l=0}^{L-1} h_l = 0, \quad \sum_{l=0}^{L-1} h_l^2 = 1, \quad \text{and} \quad \sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \text{ for all integers } n \neq 0. \tag{14}$$

These three properties ensure that (i) the mother wavelet is associated with a difference operator, (ii) the wavelet transform preserve the variance of the original data, and (iii) a multiresolution analysis can be performed on a finite variance data series. The first property implies that the mother wavelet (also called "differencing function") is a high-pass filter as it measures the deviations from the smooth components. On the other hand, the father wavelet ("scaling function") aims at capturing long scale (i.e. low frequency) components of the series and generates so-called scaling coefficients.[9] The mother and father wavelets must respect the following conditions:

$$\sum_{l=0}^{L-1} h_l = 0 \tag{15}$$

---

[9]   The low-pass filter can be directly obtained from the high-pass filter using the quadrature mirror relationship, see Percival and Walden (2000, p. 75).

$$\sum_{l=0}^{L-1} g_l = 1 \tag{16}$$

The application of both the mother and the father wavelets allow separating the low-frequency components of a time-series from its high-frequency components. Furthermore, a band-pass filter can be constructed by recursively applying a succession of low-pass and high-pass filters.

Let's assume that we have observed a sample of size T of some random variable $x(t)$:

$$\{x(1), x(2), ..., x(T)\}.$$

The *wavelet* and *scaling coefficients* at the first level of decomposition are obtained by convolution of the data series with the mother and the father wavelets,

$$w_1(t) = \sum_{l=0}^{L-1} h_l x(t') \quad \text{and} \quad v_1(t) = \sum_{l=0}^{L-1} g_l x(t'), \tag{17}$$

where $t = 0, 1, ..., T/2 - 1$ and t', the time subscript of x, is defined as $t' = 2t + 1 - l \bmod T$. The modulus operator is employed to deal with boundary conditions.[10] It ensures that the time subscript of x stays always positive. If, for some particular values of t and l, the expression $2t + 1 - l$ becomes negative, the application of the modulus operator returns $t' = 2t + 1 - l + T$. Thus, we are implicitly assuming that x can be regarded as periodic. Alternative methods to deal with boundary conditions are discussed thereafter. $w_1(t)$ and $v_1(t)$ are respectively the wavelet and the scaling coefficients at the first scale. Hence, $w_1(t)$ corresponds to the vector containing the components of x recorded at the highest-frequency. One may notice that the operation returns two series of coefficients that have length $T/2$. To continue the frequency-by-frequency decomposition of the original signal, one typically resorts to what is known as the pyramid algorithm.

PYRAMID ALGORITHM

After having applied the mother and father wavelets on the original data series, one has a series of high-frequency components and a series of lower-frequency components. The idea of the pyramid algorithm is to further decompose the (low-frequency) scaling coefficients $v_1(t)$ into high and low frequency components:

$$w_2(t) = \sum_{l=0}^{L-1} h_l v_1(t') \quad \text{and} \quad v_2(t) = \sum_{l=0}^{L-1} g_l v_1(t'), \tag{18}$$

where $t = 0, 1, ..., T/4 - 1$ and $t' = 2t + 1 - l \bmod T$. After two steps, the decomposition looks like $w = [w_1 \ w_2 \ v_2]$. One can then apply the pyramid algorithm again and again up to scale $J = \lfloor log_2(T) \rfloor$ to finally obtain $w = [w_1 \ w_2 \ ... \ w_J \ v_J]$. Figure 10 summarizes these steps. One may also apply the algorithm up to scale $J_p < J$ only. This is known as the *partial* DWT.

---

[10] For two integer a and b, a modulus b is basically the remainder after dividing a by b; i.e. $a \bmod b = a - c \cdot b$ with $c = \lfloor a/b \rfloor$.
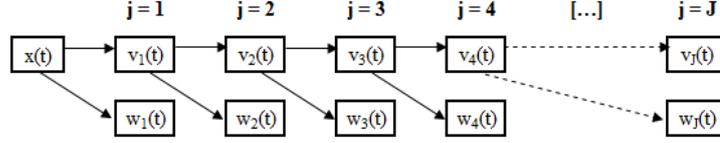
Figure 10: Flowchart of the pyramid algorithm

THE MAXIMAL OVERLAP DISCRETE WAVELET TRANSFORM (MODWT)

The standard DWT suffers from three drawbacks. First, it requires a series with a dyadic length (i.e. $T = 2^J$). Second, DWT is not shift invariant, i.e. if one shifts the series one period to the right, the multiresolution coefficients will be different. Third, it may introduce phase shifts in the wavelet coefficients: peaks or troughs in the original series may not be correctly aligned with similar events in the multiresolution analysis. To overcome these problems, the MODWT has been proposed. This wavelet transform can handle any sample size, it has an increased resolution at coarser scales (as compared to the DWT) and it is invariant to translation. It also delivers a more asymptotically efficient wavelet variance than the DWT.[11]

Actually, the main difference between the DWT and the MODWT lies in the fact that the MODWT considers all integer translations, i.e. $u = k$. This means that the MODWT keeps at each frequency a complete resolution of the series. Whatever the scale considered, the length of the wavelet and scaling coefficients vectors will be equal to the length of the original series. The wavelet and scaling coefficients at the first level of decomposition are obtained as follows

$$\widetilde{w}_1(t) = \sum_{l=0}^{L-1} h_l x(t') \quad \text{and} \quad \widetilde{v}_1(t) = \sum_{l=0}^{L-1} g_l x(t'), \tag{19}$$

where $t = 0, 1, ..., T$ and $t' = t - l \bmod T$. As for the DWT, the MODWT coefficients for scales $j > 1$ can be obtained using the pyramid algorithm. For instance, $\widetilde{w}_j$ and $\widetilde{v}_j$ are calculated as

$$\widetilde{w}_j(t) = \sum_{l=0}^{L-1} h_l \widetilde{v}_{j-1}(t') \quad \text{and} \quad \widetilde{v}_j(t) = \sum_{l=0}^{L-1} g_l \widetilde{v}_{j-1}(t'), \tag{20}$$

where $t' = t - 2^{j-1} l \bmod T$. In section 3.2, we present an example of the implementation (in MatLab) of the pyramid algorithm in the context of the MODWT.

MULTIRESOLUTION ANALYSIS (MRA)

Multiresolution analysis can be used to reconstruct the original time-series x from the wavelet and scaling coefficients, $\widetilde{w}$ and $\widetilde{v}$. In order to achieve this, one has to apply the inverse MODWT on $\widetilde{v}_J$ and $\widetilde{w}_j$, $j = 1, ..., J$.[12] As for the MODWT, the implementation of the inverse MODWT is done using a pyramid algorithm.

---

[11] See Crowley (2005) for more details about the properties of MODWT.

[12] Our presentation of the multiresolution analysis is restricted to the case of the MODWT. Nevertheless, a very similar procedure exists for the DWT, see Percival/Walden (2000).

In the first step, one applies the inverse MODWT to the scaling and wavelet coefficients at scale J in order to recover the scaling coefficients at scale $J-1$,

$$\widetilde{v}_{J-1}(t) = \sum_{l=0}^{L-1} h_l \widetilde{w}_J(t') + \sum_{l=0}^{L-1} g_l \widetilde{v}_J(t'), \tag{21}$$

$t = 0, 1, ..., N$ and $t' = t + 2^{J-1} l \bmod T$. The right hand-side of equation (21) can be expressed in matrix notation as

$$\widetilde{v}_{J-1}(t) = \widetilde{B}_J \widetilde{w}_J + \widetilde{A}_J \widetilde{v}_J.$$

We can extend this result to any scale j. For instance, at scale $j = 1$, we have

$$
\begin{aligned}
\widetilde{v}_1(t) &= \widetilde{B}_2 \widetilde{w}_2 + \widetilde{A}_2 \widetilde{v}_2 \\
&= \widetilde{B}_2 \widetilde{w}_2 + \widetilde{A}_2 (\widetilde{B}_3 \widetilde{w}_3 + \widetilde{A}_3 \widetilde{v}_3) \\
&= ... \\
&= \widetilde{B}_2 \widetilde{w}_2 + \widetilde{A}_2 \widetilde{B}_3 \widetilde{w}_3 + \widetilde{A}_2 \widetilde{A}_3 \widetilde{B}_4 \widetilde{w}_4 + ... + \widetilde{A}_2 ... \widetilde{A}_{J-1} \widetilde{B}_J \widetilde{w}_J + \widetilde{A}_2 ... \widetilde{A}_{J-1} \widetilde{A}_J \widetilde{v}_J.
\end{aligned}
$$

We can combine this result with $x = \widetilde{B}_1 \widetilde{w}_1 + \widetilde{A}_1 \widetilde{v}_1$ to finally get

$$x = \widetilde{B}_1 \widetilde{w}_1 + \widetilde{A}_1 \widetilde{B}_2 \widetilde{w}_2 + \widetilde{A}_1 \widetilde{A}_2 \widetilde{B}_3 \widetilde{w}_3 + ... + \widetilde{A}_1 ... \widetilde{A}_{J-1} \widetilde{B}_J \widetilde{w}_J + \widetilde{A}_1 ... \widetilde{A}_{J-1} \widetilde{A}_J \widetilde{v}_J. \tag{22}$$

Setting $D_j(t) = \widetilde{A}_1 ... \widetilde{A}_{J-1} \widetilde{B}_J \widetilde{w}_J$ and $S_J = \widetilde{A}_1 ... \widetilde{A}_{J-1} \widetilde{A}_J \widetilde{v}_J$, we can reconstruct the original time-series as

$$x = S_J + D_J + ... + D_1. \tag{23}$$

This "reconstruction" is known as multiresolution analysis (MRA). The elements of S_J are related to the scaling coefficients at the maximal scale and therefore represent the *smooth components* of x. The elements of D_j are the *detail* (or *rough*) *coefficients* of x at scale j.

On the basis of formula (23), one may also think of a way to compute an approximation or a smooth representation of the original data. This can be achieved by considering the scaling coefficients and the wavelet coefficients from scale J_s $(1 < J_s < J)$ to J only, i.e.

$$x_s = S_J + D_J + ... + D_s. \tag{24}$$

Formula (24) can be used, for instance, to filter out noise or seasonalities from a time-series. In image processing, formula (24) serves for data compression. Formula (22) has been specifically derived for the MODWT but similar results are available for the DWT (see Percival/Walden (2000)). Hence, formula (23) holds for both the DWT and the MODWT.

ANALYSIS OF VARIANCE

On the basis of the wavelet and scaling coefficients, it is also possible to decompose the variance into different frequency components. There are some slight differences between the variance

decomposition for the DWT and for the MODWT. We will therefore first present the main results for the DWT and then discuss their extension to the MODWT.

Using the wavelet and scaling coefficients of the discrete wavelet transform, it is possible to decompose the energy of the original series on a scale-by-scale basis:

$$\|x\|^2 = \sum_{t=0}^{T-1} x(t)^2 = \sum_{j=1}^{J} \sum_{t=0}^{T/2^J-1} w_j(t)^2 + v_J^2, \qquad (25)$$

where $\|x\|^2$ denotes the energy of x. The wavelet coefficients capture the deviations of x from its long-run mean at the different frequency resolutions. Therefore, at scale $j = J = log_2(T)$, the last remaining scaling coefficient is equal to the sample mean of x,

$$E(x) = v_J. \qquad (26)$$

On this basis, we can express the variance of x, $V(x)$, as

$$V(x) = E(x^2) - E(x)^2 = \sum_{j=1}^{J} E(w_j^2) = \sum_{j=1}^{J} V(w_j), \qquad (27)$$

where $V(w_j)$ denotes the variance of the wavelet coefficients at scale j, which is computed as

$$V(w_j) = \frac{1}{T} \sum_{t=0}^{T-1} w_j(t)^2. \qquad (28)$$

If we consider the wavelet and scaling coefficients obtained from a partial DWT, the variance of x can be expressed as

$$V(x) = \sum_{j=1}^{J_p} V(w_j) + V(v_{J_p}). \qquad (29)$$

The variance of the scaling coefficients has to be taken into account because v_{J_p} incorporates deviations of x from its mean at scales $J_p < j \leq J$.[13] $V(v_{J_p})$ is computed as

$$V(v_{J_p}) = \frac{1}{T} \sum_{t=0}^{T-1} v_J^2(t). \qquad (30)$$

An alternative way to decompose the energy of x is based on the smooth and detail coefficients of the MRA. As above, $\|x\|^2$ can be computed as the sum of the energy of the smooth and detail coefficients. This approach is, however, valid only for the DWT (see Gençay et al. (2002)).

---

[13] One may notice that the variance of the scaling coefficient at scale J is 0 as v_J is a scalar (the sample mean of x).

It is important to note that some of the wavelet coefficients involved in equation (27) are affected by boundary conditions. One should remove the corresponding wavelet coefficients in order to get an unbiased estimator of the wavelet variance,

$$\widehat{V}(x) = \sum_{j=1}^{J} \widehat{\sigma}_x(\lambda_j) = \sum_{j=1}^{J} \left[\frac{1}{2\lambda_j \widehat{T}_j} \sum_{t=L'_j}^{T/2^j - 1} w_j^2(t) \right], \tag{31}$$

where λ_j is the scale that is associated to the frequency interval $[1/2^{j+1} \quad 1/2^j]$. $L'_j = \left\lceil (L-2)(1-2^{-j}) \right\rceil$ is the number of DWT coefficients computed using the boundaries. Hence, $\widehat{T}_j = T/2^j - L'_j$ is the number of coefficients unaffected by the boundary.

We now turn to the analysis of variance in the context of the MODWT. Formula (27) remains perfectly valid. The MODWT keeps the same number of coefficients at each stage of the wavelet transform. The way one deals with boundary conditions must therefore be adapted. From the detail coefficients of a partial MODWT of order $J_p < \log_2(T)$, the wavelet variance can be estimated as follow

$$\widetilde{V}(x) = \sum_{j=1}^{J_p} \widetilde{\sigma}_x(\lambda_j) = \sum_{j=1}^{J_p} \left[\frac{1}{\widetilde{T}_j} \sum_{t=L_j-1}^{T-1} \widetilde{w}_j^2(t) \right], \tag{32}$$

where $L_j = (2^j - 1)(L-1) + 1$ is the number of scale λ_j wavelet coefficients, which are affected by boundary conditions. This number also corresponds to the length of the wavelet filter at scale λ_j. $\widetilde{T}_j = T - L_j + 1$ is thus the number of wavelet coefficients unaffected by the boundary.

3.2 Implementation and practical issues

CHOICE OF A WAVELET FILTER

There exist many different wavelet filters, each of them being particularly suitable for specific purposes of analysis. Wavelet filters differ in their properties and in their ability to match with the features of the time-series under study. Furthermore, when it comes to implement a *discrete* wavelet transform, one also has to decide about the filter length. Because of boundary conditions, longer filters are well adapted for long time-series. The simplest filter is the Haar wavelet, which is basically a difference and average filter of length two. In finance, most researchers have worked either with Daubechies (denoted as "D") or with Least-Asymetric ("LA") filters of length 4 to 8. Helder/Jin (2007) and Nielsen/Frederiksen (2005) employ D(4) filters. Gençay, Selçuk and Whitcher (2003, 2005) suggest that the LA(8) wavelet (i.e. a Least-Asymetric filter of length 8) is a good choice for analyzing financial time series, while Subbotin (2008) uses a LA(4) wavelet. Crowley (2005) argues that the impact of choosing another wavelet filter has a rather limited impact on the distribution of the variance of the time-series across the scales.

Depending on the purpose of analysis, it might be appealing to select a wavelet filter which satisfies one or more of the following properties:

- Symmetry: symmetric filters are appealing because they ensure that there will be no phase shift in the output series. Unfortunately, most wavelets are not symmetric. An exception

is the Haar wavelet. The requirement of a symmetric wavelet is, however, less essential if one uses a MODWT rather than a DWT because the MODWT ensures that the original series and its filter coefficients will be aligned.

- Orthogonality: this property refers to the fact that the wavelet and the scaling coefficients contain different information. This is an important feature as it ensures that the wavelet decomposition will preserve the energy (variance) of the original series (Crowley (2005)). Daubechies and Least-Asymmetric wavelets meet this requirement; that is, their scaling and wavelet coefficients are orthogonal by construction.

- Smoothness: The degree of smoothness is measured by the number of continuous derivatives of the basis function. As such, the Haar wavelet is the least smooth wavelet. The choice of a more or less smooth filter depends essentially on the data series to be represented. If the original time-series is very smooth, then one will opt for a smooth wavelet. For instance, the Haar wavelet is appropriate for the analysis of a pure jump process.

- Number of vanishing moments: The number of vanishing moments of the wavelet function has a direct implication on the ability of the wavelet to account for the behavior of the signal. That is, if a signal has a polynomial structure or if it can be approximated by a polynomial of order q, then the wavelet transform will be able to properly capture this polynomial structure only if it has q vanishing moments. For instance, Daubechies wavelets have a number of vanishing moments which is half the length of the filter. Thus the Haar and D(8) have respectively 1 and 4 vanishing moments.

The last two properties depend not only on the wavelet filter but also on its length. In fact, the most crucial point is probably not to choose the "right" filter but to choose a filter with an appropriate length. Increasing the filter length permits to better fit the data. Unfortunately, this also renders the influence of boundary conditions more severe. Hence a tradeoff has to be found.

Problems due to boundary conditions arise in two situations. The first case concerns the DWT. To use the DWT, one requires a time-series with a dyadic length. If the series does not meet this requirement, i.e. if its length N is such that $2^j < N < 2^{(j+1)}$, one has the choice between removing observations until $N = 2^j$ or completing the series so that $N = 2^{(j+1)}$. Removing data might be the best alternative but it leads to a loss of information.

The second case concerns both the DWT and the MODWT. The wavelet filter has to be applied on all observations, including observations recorded at the beginning ($t = 1$). A problem arises because the convolution operator requires that $L - 1$ observations are available before t. In this case, removing data is useless. Therefore one has to complete the data series. One may pad each end of the series with zeros; this technique is known as "zero padding". An alternative is to use the fit from a polynomial model to replace non-existing data at each end of the series ("polynomial approximation"). One may also complete each end of the series either by mirroring the last observations ("mirror" or "reflection") or by taking the values observed at the beginning of the other end of the series ("circular"). The choice depends on the data considered. For instance, if one works on stock returns, the use of a "mirror" seems to be the most suitable approach as it accounts for the presence of volatility clustering. Moreover, after the multiresolution decompo-

sition of the original signal, one may obviously discard the coefficients that are affected due to their proximity to the boundaries of the series.

Instead of selecting a priori a specific wavelet function, one may also use so-called optimal transforms. The idea is to choose the wavelet function which minimizes a loss function (e.g. the entropy cost function, see Crowley (2005)).

EXAMPLES OF WAVELET FILTERS AND THEIR GAIN FUNCTIONS

Figure 11 shows the coefficients of the Haar, D(4), D(8) and LA(8) wavelets for level $j = 4$. One may observe the very simple structure of the Haar wavelet. When comparing the latter with the D(4) and D(8) filters, it becomes clear that the longer the filter, the smoother it is. The LA(8) looks less asymmetric than the D(8). Nevertheless it is still far from being symmetric.
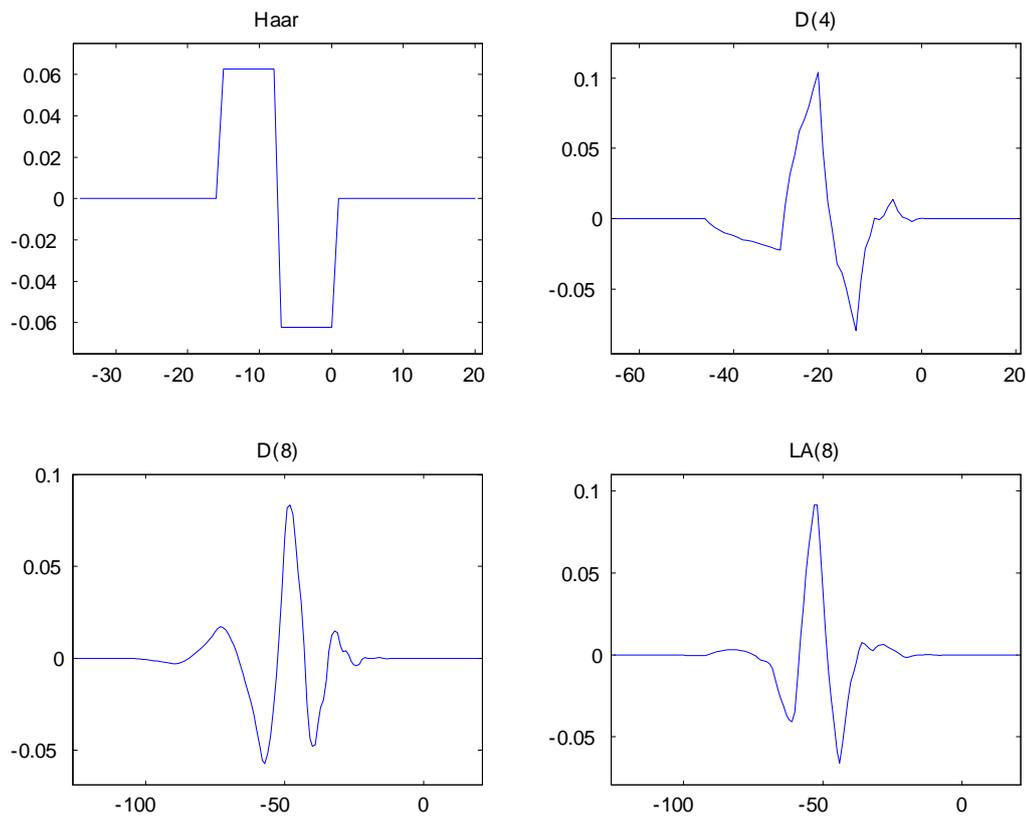
Figure 11: Mother wavelet filters. The Haar, D(4), D(8) and LA(8) filters at scale $j = 4$.

Studying the frequency response of these filters permit to assess their ability to capture the different frequency components. Figure 12 displays the gain function for each wavelet filter at scales 1 to 4. It is evident from the figure that the longer filters (D(8) and LA(8)) have better frequency localization. The gain functions of the D(8) and LA(8) are similar. In order to make
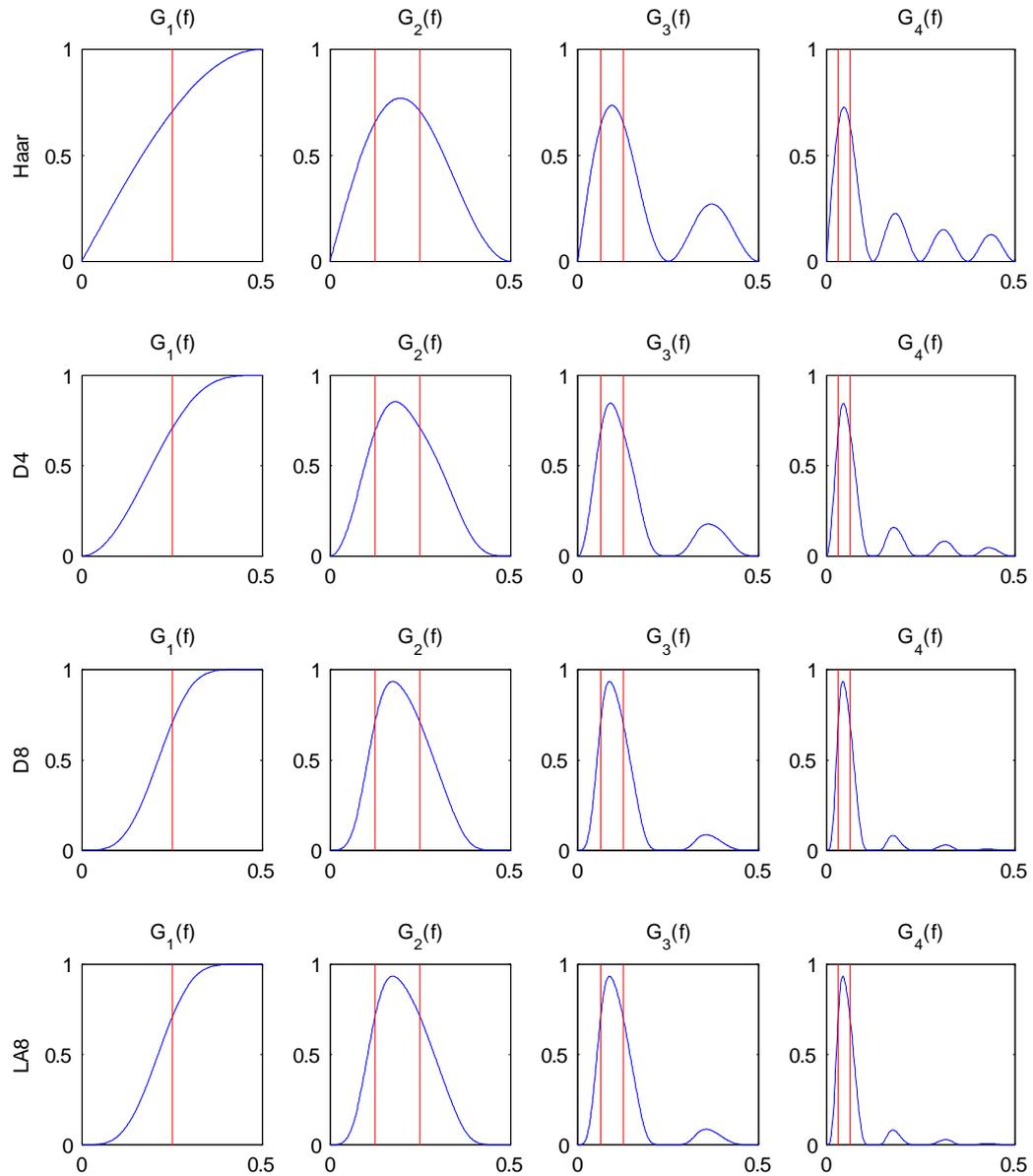
Figure 12: Gain functions for the Haar, D(4), D(8) and LA(8) wavelets.

this statement clearer, we contrast in figure 13 the gain functions of the Haar, D(4) and D(8) wavelets at scale $j = 5$. The D(8) captures much better the components corresponding to frequencies between $\frac{1}{2^j}$ and $\frac{1}{2^{j+1}}$.
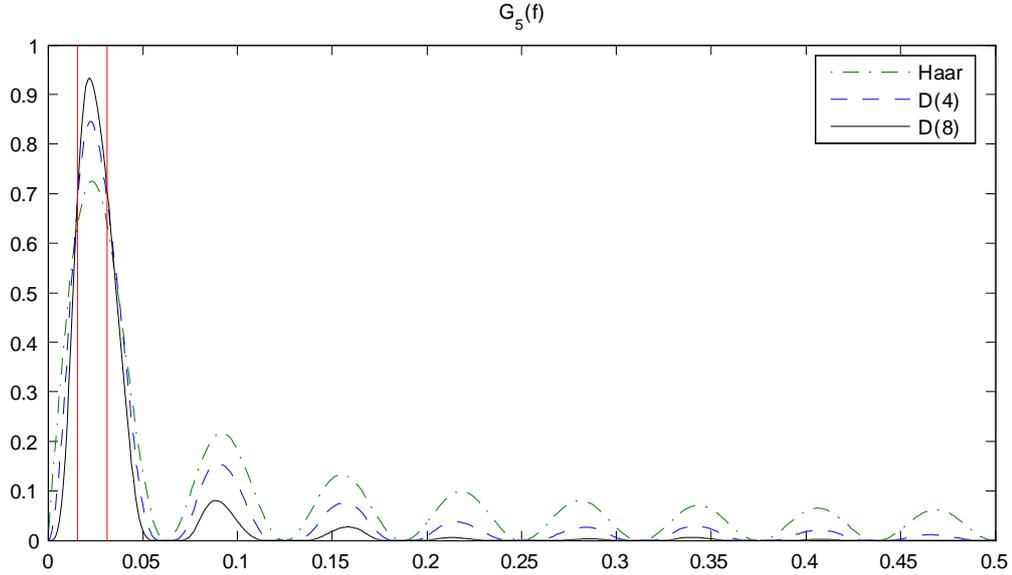
Figure 13: Gain function for the Haar, D(4) and D(8) wavelets at scale $j = 5$.

IMPLEMENTATION IN MATLAB

Hereafter, we show the main steps for an effective implementation of the MODWT in MatLab. Our code considers four wavelet filters: the Haar, D(4), D(8) and LA(8). The code deals with boundary conditions using either the reflection (also called mirror) or the circular approach.

1. Definition of the function and its inputs and outputs.

   Our function is called "waveletFilter" and it can be invoked through the instruction: "[W,V,boundCond]=waveletFilter(X,nameFilt,nLev,boundCondMethod)".

   It accepts four inputs: a time-series ("X"), the filter name ("nameFilt"), the number of levels for the wavelet analysis ("nLev") and the method to deal with boundary conditions ("boundCondMethod"), which can be either the reflection (boundCondMethod='reflection') or the circular method (boundCondMethod='circular'). The function returns three outputs: a matrix containing the wavelet coefficients ("W"), a vector of scaling coefficients ("V") and a matrix of dummies, which tells which wavelet and scaling coefficients are not affected by boundary effects.

2. Checking the validity of the inputs and definition of the mother and father wavelet filters.

   The first part of the function is devoted to checking that the arguments received by the function are correct. We also load the filter coefficients for the mother wavelet and com-

       pute the corresponding coefficients for the father wavelet using the quadrature mirror relationship.

3. Core of the code.

       The core part of the code is arranged around a main loop, which serves to compute the wavelet and scaling coefficients at each scale j for $j = 1$ to J ("nLev"). If the method to deal with boundary conditions is the reflection, then the original time-series is first mirrored. Otherwise, we go on with the original vector of observations. The code looks as follows:

```
T=size(X,1);

W=zeros(T,nLev);

boundCond=zeros(T,nLev);

V=X;

for j=1:nLev

    start=-(lFilt-1)*2^(j-1);

    temp=[V(end+start+1:end);V];

    K=1;

    xtemp=temp(K:end+start+K-1);

    for k=2:lFilt

        K=K+2^(j-1);

        xtemp=[temp(K:end+start+K-1) , xtemp];

    end

    W(:,j) = xtemp*h';

    V = xtemp*g';

    boundCond(-start+1:T0,j)=1;

end
```

EXAMPLE

Let's consider a variable y, whose dynamics is primarily driven by an AR(1) process and five cyclical components:

$$y_t = 0.90y_{t-1} + \sum_{s=3}^{5} 5\cos(\frac{2\pi t}{s}) + \varepsilon_t, \tag{33}$$

ε_t is a i.i.d. Gaussian process with mean zero and unit variance and $t = 1, ..., 10.000$. In the absence of seasonalities and noise, the autocorrelation function of y should take value 0.90^k at lag k. Wavelets can be used to remove the impact of both noise and seasonalities. From equation

(33), one may notice that the cyclical components have a period length of 3 to 5 periods. Hence they have an impact on frequencies between $1/5$ and $1/3$.

At scale j, the wavelet detail D_j captures frequencies $1/2^{j+1} \le f \le 1/2^j$ and the wavelet smooth S_j captures frequencies $f < 1/2^{j+1}$. If we use a level 2 multiresolution analysis, the wavelet smooth S_2 will thus capture the components of the time-series, which have a frequency $f < 1/8$. This means that S_2 will take into account changes in y that are associated with a period length of at least 8 units of time. Therefore, S_2 should keep the AR(1) dynamics of y, while removing its cyclical behavior and noise.
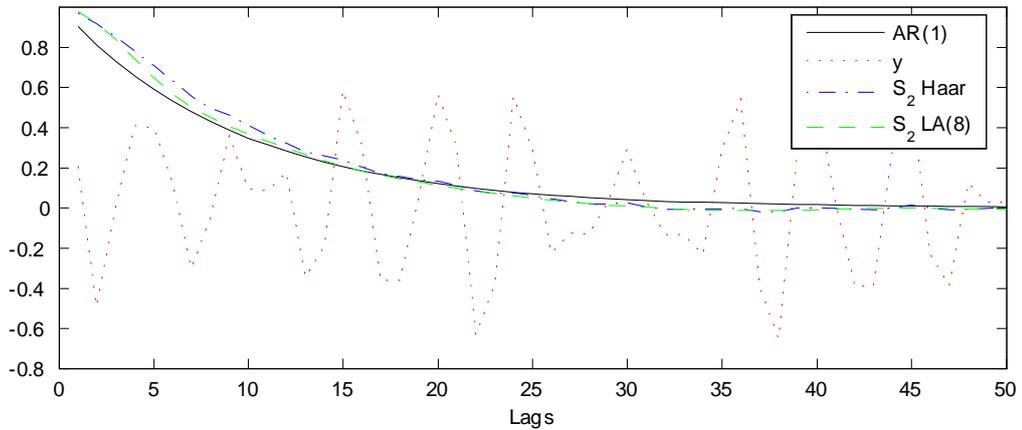
Figure 14: Autocorrelations estimated before and after having removed some specific frequency componenents of the original time-series. Autocorrelation coefficients for (i) a theoretical AR(1) process, (ii) for a process y that is affected by noise and cyclical components, and (iii-iv) for the wavelet smooth from a level-2 MRA with a Haar and a LA(8) wavelet.

Figure 14 reports the autocorrelation coefficients for the original time-series y, for the theoretical AR(1) process and for the wavelet smooth S_2. In order to assess the impact of choosing a different wavelet filter, we use both the Haar and the LA(8) wavelets. The results are very similar even if for short lags the LA(8) seems to provide some improvements over the Haar. All in one, this example demonstrates the ability of a wavelet filter to deal with a complex cyclical structure and with noise.

3.3   Illustration: Home price indices for different US-cities

DESCRIPTIVE STATISTICS

In this section, we study the evolution of home prices in twelve U.S. cities from January 1987 to May 2008 (257 months).[14] We employ the Case-Shiller home price indices. The data has been

---

[14] Los Angeles (LA), San Francisco (SF), Denver (De), Washington (Wa), Miami (Mi), Chicago (Chi), Boston (Bos), Las Vegas (LV), New York (NY), Portland (Po), Charlotte (Cha), Cleveland (Cl).

gathered from the Standard & Poors website. $P_j(t)$ denotes the price index level for the $j-$th city in period t and it has been standardized so that it has a value of 100 in the first month of the sample, i.e. $P_j(t=0) = 100$ $(\forall j)$. We compute the index returns as

$$r_j(t) = \log[P_j(t)] - \log[P_j(t-1)].$$

Table 1 reports some descriptive statistics for the returns on each of the 12 home price indices. There are huge disparities in performance. The largest price increases are to be found in Portland (+328%), San Francisco (+249%) and Los Angeles (+234%). Cleveland (+103%), Charlotte (+110%) and Boston (+129%) are the worst performers. The volatility of home price changes has been much larger in cities like San Francisco, Los Angeles, Las Vegas and Miami than in north-east cities (plus Portland and Denver). This higher volatility is, to a large extent, the result from the severe price downturn in these four cities during the last two years of the sample. It is worth mentioning that some cities have been much less affected by the home price crash than others. Portland is the most striking example, it has achieved a strong performance of 328% cumulated return over the sample period and, since its highest level, it has experienced a limited price drop of about 6%. From a risk-return perspective, such an investment seems to remain fairly attractive. Nevertheless, the large excess kurtosis and negative skewness of some indices indicate that the volatility may be an incomplete measure of risk in this context.

	LA	SF	De	Wa	Mi	Chi	Bos	LV	NY	Po	Cha	Cl
Mean	5.66	5.86	4.45	5.31	4.86	4.83	3.88	4.16	4.49	6.81	3.48	3.33
Std.	4.24	4.18	2.17	3.13	3.58	2.40	2.88	4.30	2.48	2.32	1.55	2.17
Skew	-0.54	-0.43	-0.60	-0.01	-1.33	-0.20	-0.15	0.06	0.11	0.26	0.14	-0.75
Kurt	4.65	5.53	4.39	3.95	7.86	5.10	2.96	10.27	2.49	4.70	3.85	8.12
$P(T)$	334.72	349.07	258.41	310.76	282.03	280.17	228.94	242.68	260.52	427.60	210.06	203.40
$\max[P]$	461.72	468.50	279.42	391.62	410.03	314.85	260.49	353.80	290.02	454.35	214.36	230.69
$\min[P]$	100.00	100.00	94.04	100.00	100.00	100.00	89.86	98.16	97.14	99.78	100.00	99.94

Table 1. Descriptive statistics for the 12 home price indices. We report the mean (in %), the annualized standard deviation ("Std." in the table, in %), skewness ("Skew") and kurtosis ("Kurt") of $r_j(t)$ for $j = 1, ..., 12$. We also report the index level in the last month of the sample ($P(T)$) as well as the maximum ("max[P]") and minimum ("min[P]") level achieved by each index over the period 1987-2008.

AUTOCORRELATIONS

Figure 15 shows the autocorrelations (up to 24 lags, i.e. 2 years) of the index returns. The full lines are for the original series, while the dotted lines show the autocorrelations computed from the smooth coefficients obtained using the MRA from a partial MODWT. We employ a LA(8) filter with $J_p = 3$. Hence the wavelet smooth (S_3) should capture the frequency components that are associated with a period length of at least 16 months and should therefore be free of seasonal effects. In order to deal with the boundary conditions, we use the reflection method.
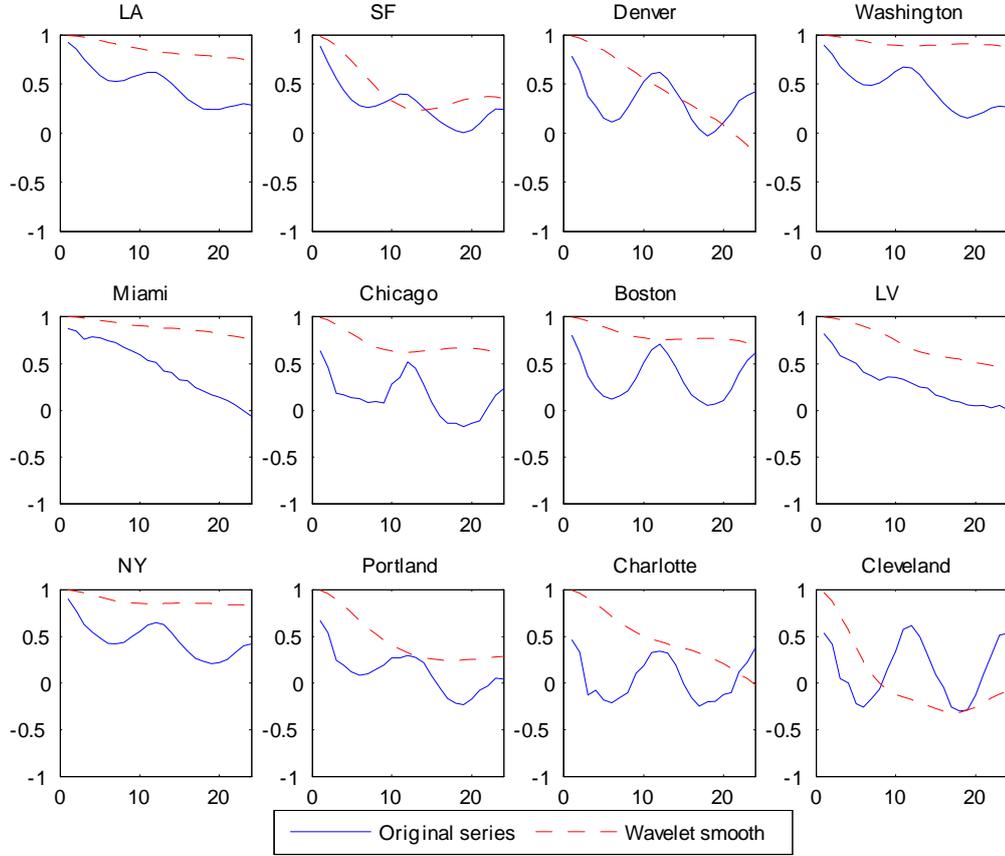
Figure 15: Autocorrelations of $r_j(t)$**.** This figure shows the autocorrelations for both the original returns series and the wavelet smooth series for each home price index.

Some indices are very affected by seasonal effects (Denver, Washington, Chicago, Boston, Portland, Charlotte and Cleveland), while the other indices are much less affected (Los Angeles, San Francisco and New York) or even unaffected (Miami and Las Vegas). In general the indices that display the least significant seasonal patterns are also those that have been the most affected by the recent crisis. This observation may suggest that the (quasi) absence of these patterns is the result of the predominant role of speculation on price changes in these cities.

One may again observe the ability of wavelets to remove seasonal patterns. The autocorrelations estimated from the wavelet smooth show the long-run temporal dynamics of home prices. In contrast to financial markets, whose evolution is almost unpredictable, home prices are strongly autocorrelated. The autocorrelation remains positive even after 2 years.

VARIANCE AND CROSS-CORRELATIONS

Next, we analyze the variance of each index returns series and the correlations between the different indices at a variety of frequencies. We employ a partial MODWT with $J_p = 5$. Hence at the largest scale, the wavelet coefficients contain information regarding price changes over
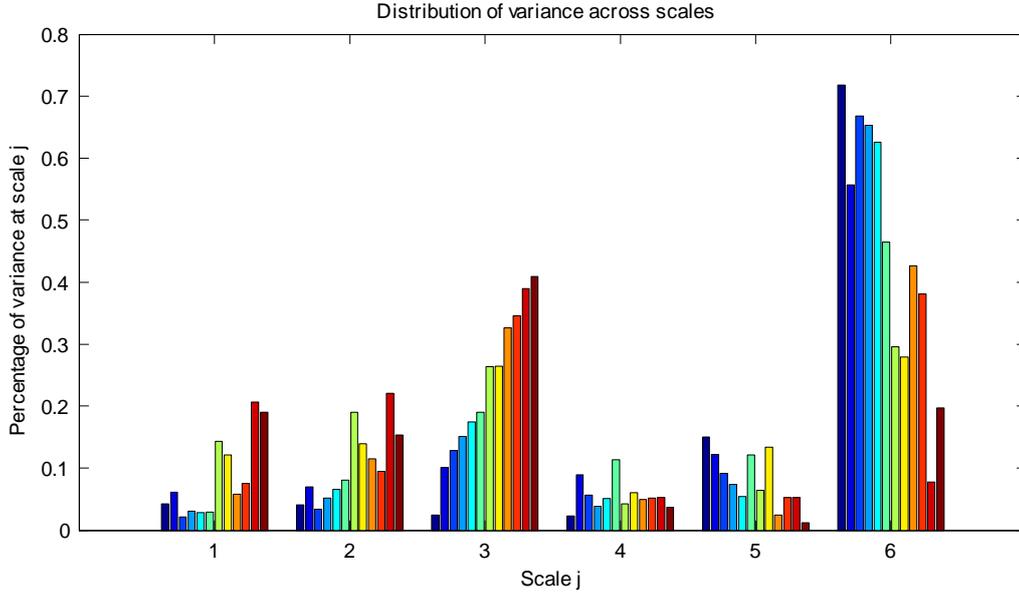
Figure 16: Distribution of the wavelet variance across scales. Each bar corresponds to the variance recorded at a specific scale j and for a particular city. Cities are ordered as follows (from left to right): Miami, Las Vegas, Los Angeles, Washington, New York, San Francisco, Chicago, Portland, Boston, Denver, Charlotte, Cleveland.

an horizon of 32 to 64 months. Similarly, the scaling coefficients capture the evolution for periods longer than 64 months. Figure 16 shows the distribution of variance across the different frequencies.

From figure 16, one may notice that most of the variance is due to frequency components associated with scales 3 and 6. On the other hand, variance at scales 1 and 2 is low and may be due to noise. Scales 4 to 5 are not related to important frequency components, neither to seasonal patterns (scale $j = 3$) nor to business cycle (scales $j > 5$) components. Thus it comes at no surprise that they do not have much information content.

Scale 3 corresponds to periods of 8 to 16 months and as such the variance observed at this scale reflects the importance of seasonal patterns in the dynamics of the various home price indices. As before, we observe that seasonal effects have a very limited impact on prices in Miami, they are also much less important in cities like Las Vegas, Los Angeles, Washington and New-York than in Boston, Denver, Charlotte and Cleveland. Interestingly the ordering is reversed when considering the variance at scale 6. That is, the index returns series on which seasonalities have a weak impact exhibit the largest percentage of long-term volatility.

Figure 17 reports the correlation between the various city indices at different scales. Dark colors correspond to correlations in the range -0.15 to 0.15., while light colors are associated with significantly positive correlations. The lighter the cell, the larger is the correlation between two indices. The diagonals are white as the correlation of an index with itself is obviously equal to 1. The largest correlations are observed at scales larger than 2. In particular at scale 3, the correlations are very significant. This is because seasonalities affect all indices simultaneously.
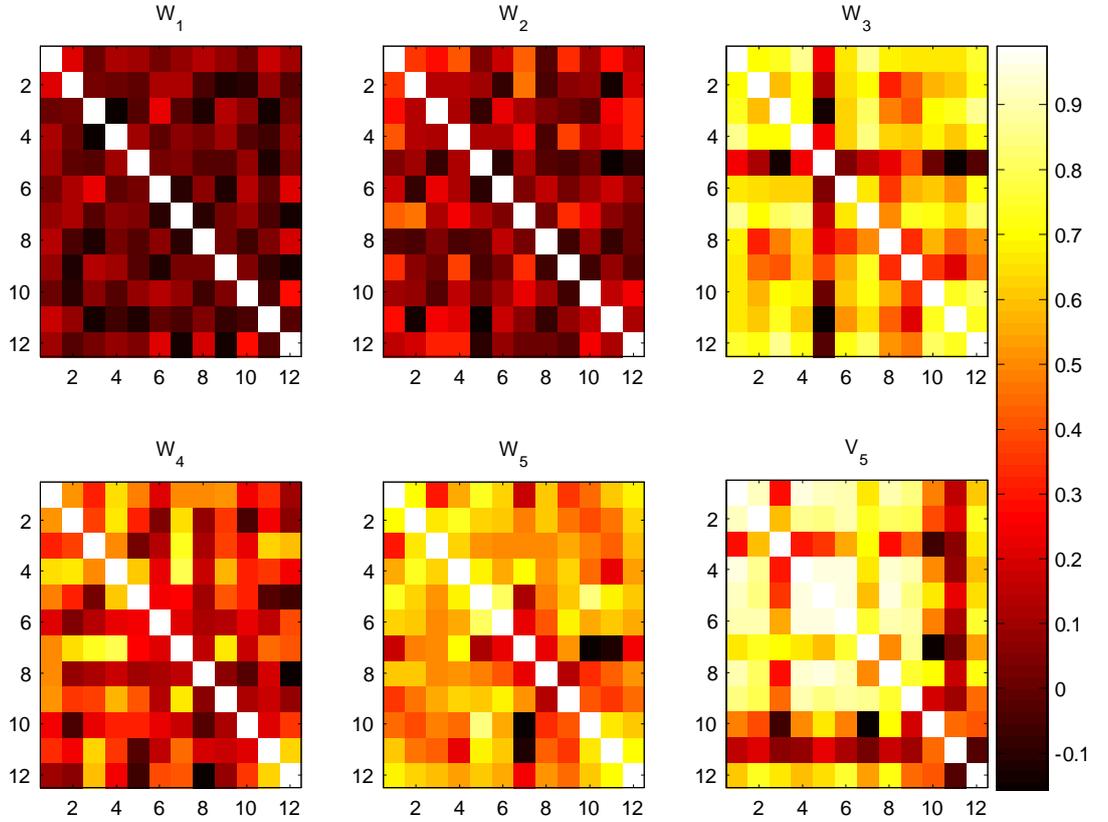
Figure 17: Correlations between home price indices at different frequencies. The panel denoted by W_1 to W_5 show the correlations between the wavelet coefficients at scales $j = 1, ..., 5$, while panel denoted V_5 show the correlations between the scaling coefficients at scale $j = 5$. The 12 cities are order as follows: Los Angeles, San Francisco, Denver, Washington, Miami, Chicago, Boston, Las Vegas, New York, Portland, Charlotte, Cleveland.

One may notice that the fifth index (Miami) shows less correlation with the other indices. At scales 5 and 6, the correlations become highly significant. This demonstrates that the series tend to behave very similarly in the long-run. There are, however, some exceptions. Charlotte is weakly correlated with the other indices; this is maybe because it is a medium size city. Cleveland, Portland and to a lesser extent Denver and Boston also display less correlation as compared to the other cities. In particular, Portland has a correlation that is nearly equal to zero with Denver, Boston and New-York.

An interesting extension to this correlation analysis is to study how many (economic) factors are important to explain the structure of the correlation matrix at each scale. In order to adress this question, one may resort to Random Matrix Theory (RMT). A very nice introduction is provided by Bouchaud/Potters (2004). A recent literature review on RMT can be found in Sharifi et al. (2004). Here, we concentrate on the main premisses of RMT. That is, this approach should enable one to (1) assess if the correlation coefficients have a genuine information content

or if they are merely due to the noise inherent in the data, and, (2) estimate the number of factors that are necessary to "explain" the correlation matrix. This is done by comparing the eigenvalues of the empirical correlation matrix with those from a theoretical distribution.

Let's consider a dataset with N time-series of length T. We assume that the theoretical random matrix belongs to the ensemble of Wishart matrices. On this basis, we can derive the theoretical distribution of the eigenvalues of the correlation matrix. Under the null of pure randomness, the eigenvalues must be confined within the bounds:[15]

$$\lambda_{\max} = (1 + \frac{1}{q} + 2\sqrt{\frac{1}{q}}) \quad \text{and} \quad \lambda_{\min} = (1 + \frac{1}{q} - 2\sqrt{\frac{1}{q}}), \tag{34}$$

where $q = T/N$. If the correlation matrix is random, then the probability that any of its eigenvalues lies outstide the bounds defined by $[\lambda_{\min}, \lambda_{\max}]$ is zero. Hence the presence of eigenvalues larger than the upper bound can be taken as an evidence that the structure of the correlation matrix is not due to chance; i.e. that there are deviations from RMT. Furthermore, the number of such eigenvalues can be interpreted as corresponding to the number of factors underlying the evolution of the N time-series. A potential problem with the RMT approach is that it requires $N \to \infty$ and $T \to \infty$. In our case, these conditions are evidently not fullfilled as $T = 256$ and $N = 12$. To account for this, we also estimate $\lambda_{\max}$ from the empirical distribution of the correlation matrix eigenvalues. To this aim, we resample (without replacement) the original time-series of returns and then apply the MODWT on the resampled data and calculate the correlation matrix and its eigenvalues at each scale of the MODWT. This procedure is done 100.000 times. In figure 18, we report $\lambda_{\max}$ as computed on the basis of equation (34) and the empirical values of $\widehat{\lambda}_{\max}$ obtained from our simulations. In the latter case, we consider two different values for $\widehat{\lambda}_{\max}$, which correspond to the $5\%-$ and, respectively, $1\%-$quantile of the empirical cumulative distribution of λ. The theoretical (RMT) maximum eigenvalue is always much lower than its empirical counterparts. This is probably due to the small sample size.

In Table 2, we report the eigenvalues of the correlation matrix for each scale of the MODWT. Comparing the eigenvalues λ_k, $k = 1, ..., 12$ with $\widehat{\lambda}_{\max}$ demonstrates that there is a single factor underlying the evolution of home prices. At each scale, the eigenvalue that is attached to this factor is significant at the 99% level. This unique factor can be interpreted as a sort of national-wide home price index.

---

[15] See Sharifi et al. (2004) and Kwapien et al. (2007).
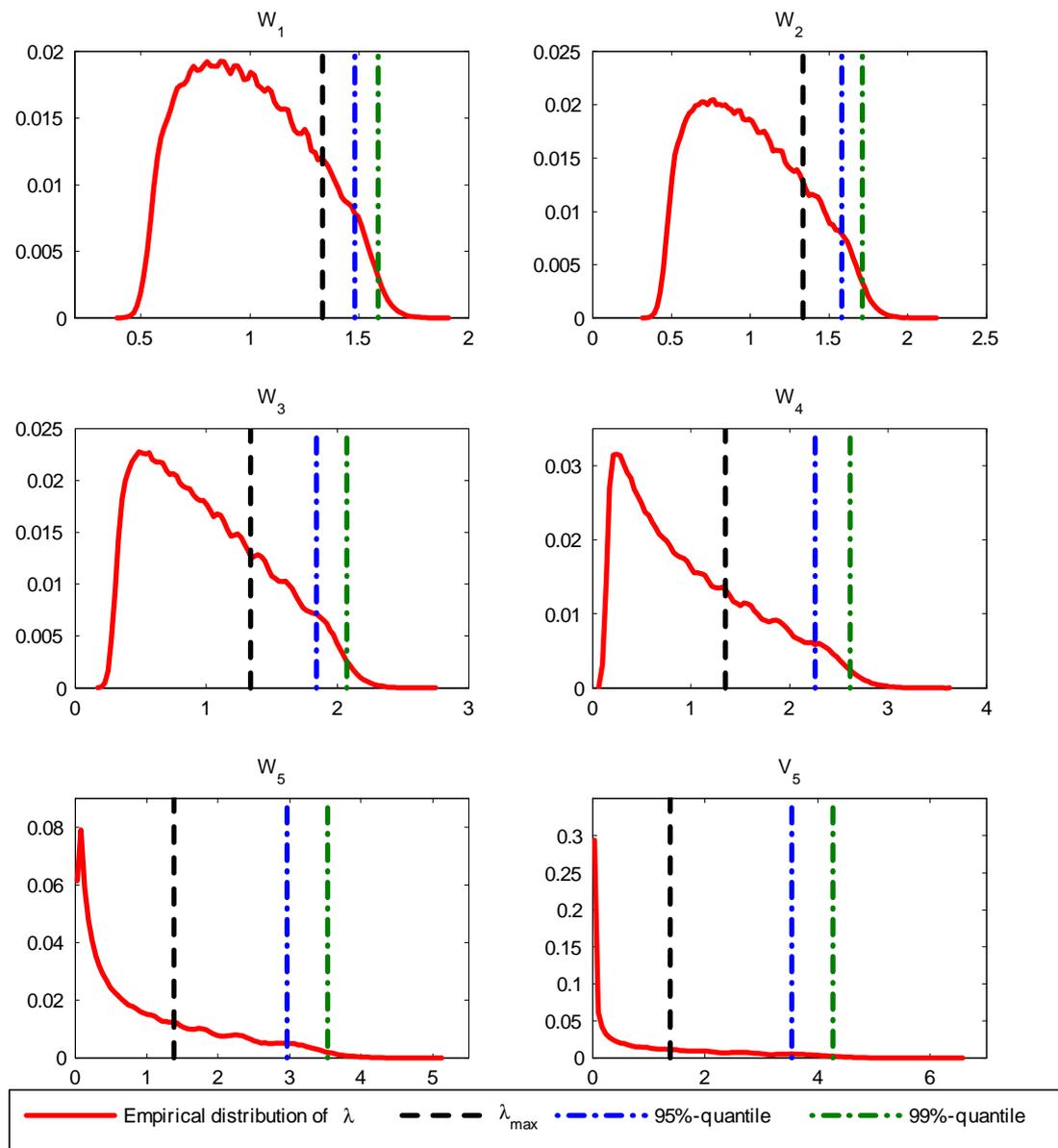
Figure 18: Distribution of the empirical eigenvalues of the correlation matrix.

	w_1	w_2	w_3	w_4	w_5	v_5
λ_1	0.50	0.32	0.05	0.07	0.01	0.00
λ_2	0.58	0.40	0.08	0.15	0.04	0.00
λ_3	0.66	0.42	0.09	0.19	0.11	0.00
λ_4	0.73	0.51	0.17	0.21	0.16	0.00
λ_5	0.85	0.66	0.22	0.33	0.20	0.01
λ_6	0.90	0.81	0.23	0.40	0.30	0.03
λ_7	1.01	0.88	0.34	0.59	0.43	0.12
λ_8	1.13	0.97	0.40	0.72	0.56	0.24
λ_9	1.27	1.04	0.51	1.08	0.72	0.87
λ_{10}	1.33	1.21	0.81	1.25	1.11	1.00
λ_{11}	1.43	1.46	1.48	1.72	1.89	1.80
λ_{12}	1.61**	3.33**	7.60**	5.30**	6.48**	7.92**
$\lambda_{\max}$	1.33	1.33	1.34	1.35	1.38	1.38
$\widehat{\lambda}_{\max}$ (5%)	1.48	1.58	1.84	2.26	2.96	3.54
$\widehat{\lambda}_{\max}$ (1%)	1.59	1.71	2.07	2.62	3.53	4.27

Table 2. Eigenvalues of the correlation matrix at each scale of the MODWT. We report the 12 eigenvalues of the correlation matrix at each scale j as well as the theoretical (RMT) maximum eigenvalue ("$\lambda_{\max}$" in the table) and the empirical 95%− and 99%−quantile ("$\widehat{\lambda}_{\max}$ (5%)" and "$\widehat{\lambda}_{\max}$ (1%)") of the empirical (bootstrapped) cumulative distribution of eigenvalues at scale j. ** and * denote significance at the 95% and 99% levels.

4 Conclusion

This paper discusses spectral and wavelet methods. It aims at being an easy-to-follow introduction and it is structured around conceptual and practical explanations. It also offers many supporting examples and illustrations. In particular, the last section provides a detailed case study, which analyzes the evolution of home prices in the U.S. over the last 20 years using wavelet methodology.

References

Bouchaud, J.-P./Potters, M. (2004), *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management*, Cambridge University Press.

Childers, D. (1978), *Modern Spectrum Analysis*, New York, IEEE Press.

Crowley, P. (2005), *An Intuitive Guide to Wavelets for Economists*, Working Paper, Bank of Finland.

Crowley, P. (2007), A guide to wavelets for economists, *Journal of Economics Surveys* **21**, pp. 207–264.

Fan, J./Fan, Y./Yazhen, W. (2007), Multiscale jump and volatility analysis for high-frequency financial data, *Journal of American Statistical Association* **102**, pp. 618–631.

Fernandez, V./Lucey, B. (2007), Portfolio management under sudden changes in volatility and heterogeneous investment horizons, *Physica A: Statistical Mechanics and its Applications* **375**, pp. 612–624.

Gençay, R./Selçuk, F./Whitcher, B. (2002), *An introduction to wavelets and other filtering methods in finance and economics*, Academic Press, San Diego.

Gençay, R./Selçuk, F./Whitcher, B. (2003), Systematic risk and timescales, *Quantitative Finance* **3**, pp. 108–116.

Gençay, R./Selçuk, F./Whitcher, B. (2005), *Information Flow Between Volatilities Across Time Scales*, Working Paper, Simon Fraser University.

Gu, K. (2002), The predictability of house prices, *Journal of Real Estate Research* **24**, pp. 213–234.

Hamilton, J. (1994), *Time Series Analysis*, Princeton.

Helder, J./Jin, H. (2007), Long memory in commodity futures volatility: a wavelet perspective, *The Journal of Futures Markets* **27**, pp. 411–437.

Hodrick, R./Prescott, E. (1997), Postwar u.s. business cycles: An empirical investigation, *Journal of Money, Credit, and Banking* **29**, pp. 1–16.

Iacobucci, A. (2003), *Spectral Analysis for Economic Time Series*, Documents de Travail de l'OFCE.

Kuo, C. (1996), Autocorrelation and seasonality in the real estate market, *Journal of Real Estate Finance and Economics* **12**, pp. 139–162.

Kwapien, J./Oswiecimka, P./Drozdz, S. (2007), Empirics versus rmt in financial cross-correlations, *Acta Physica Polonica B* **38**, pp. 4027–4039.

Lynch, P./Zumbach, G. (2003), Market heterogeneities and the causal structure of volatility, *Quantitative Finance* **3**, pp. 320–331.

Muller, U. A./Dacorogna, M. M./Dave, R. D./Pictet, O. V./Olsen, R. B./Ward, J. (1995), *Fractals and Intrinsic Time - a Challenge to Econometricians*, Working Papers, Olsen and Associates.

Nielsen, M./Frederiksen, P. (2005), Finite sample comparison of parametric, semiparametric, and wavelet estimators of fractional integration, *Econometric Reviews* **24**(4), pp. 405–443.

Oswiecimka, P./Kwapien, J./Drozdz, S. (2006), Wavelet versus detrended fluctuation analysis of multifractal structures, *Physical Review E* **74**, pp. 016103.

Percival, D./Walden, A. (2000), *Wavelet Methods for Time Series Analysis*, Cambridge University Press.

Ramsey, J. (2002), *Wavelets in Economics and Finance: Past and Future*, Working Paper, New York University.

Schleicher, C. (2002), *An Introduction to Wavelets for Economists*, Bank of Canada Working Paper.

Sharifi, S./Crane, M./Shamaie, A./Ruskin, H. (2004), Random matrix theory for portfolio optimization: a stability approach, *Physica A* **335**, pp. 629–643.

Struzik, Z. (2001), Wavelet methods in (financial) time-series processing, *Physica A* **296**, pp. 307–319.

Subbotin, A. (2008), *A multi-horizon scale for volatility*, Working Paper, University of Paris-1.

Vuorenmaa, T. (2005), *A wavelet analysis of scaling laws and long-memory in stock market volatility*, Bank of Finland Research Discussion Paper.

Welch, P. (1967), The use of fast fourier transforms for the estimation of power spectra: A method based on time averaging over short modified periodograms, *IEEE Transactions on Audio and Electroacoustics* **15**, pp. 70–73.