Analysis of high-frequency financial data over different timescales: *a Hilbert-Huang transform approach*

Noemi Nava

Supervisor: Prof. Tomaso Aste
Dr. Ioannis Andreopoulos

Department of Computer Science

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of University College London.

June 2016

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

Noemi Nava
June 2016

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my academic advisor, Professor Tomaso Aste, for his continuous support, patience, motivation and inspiration. I appreciate his ideas and contributions to make this experience productive and exciting. I will always be grateful for all his advice, including the uncomfortable truth and the encouraging words which were timely and conveniently provided.

I also want to thank Professor Tiziana Di Matteo, her wisdom, knowledge and commitment to the highest standards inspired and motivated me. My sincere thanks also goes to professor Philip Treleaven, for providing me with the opportunity to embark upon a PhD in the first place and for all his persistent remarks that motivated me to write this thesis, including a letter in Spanish to my parents.

Thanks to my second supervisor Dr. Ioannis Andreopoulos for all the discussions which enriched this research. I also want to thank Marzieh Saeidi for always being there to cheer, sharing this experience with her made this journey more enjoyable and the memories will last forever.

I wish to thank my entire family for providing moral support and a home where I always feel loved. Thank you for letting me find my own way. Finally, thank God I finished.

Abstract

This thesis provides a better understanding of the complex dynamics of high-frequency financial data. We develop a methodology that successfully and simultaneously characterizes both the short and the long-term fluctuations latent in a time series. We extensively investigate the applications of the empirical mode decomposition (EMD) and the Hilbert transform to the analysis of intraday financial data. The applied methodology reveals the time-dependent amplitude and frequency attributes of non-stationary and non-linear time series. We uncover a scaling law that links the amplitude of the oscillating components to their respective period. We relate such scaling law to distinctive properties of financial markets.

This research is relevant because financial data contain patterns specific to the observation frequency and are thus, of interest to different type of market agents (market traders, intraday traders, hedging strategist, portfolio managers and institutional investors), each characterized by a different reaction time to new information and by the frequency of its intervention in the market. Understanding how the investment horizons of these agents interact may reveal significant details about the physical processes that generate or influence financial time series.

We use the EMD to estimate volatility, generalising the idea of the popular realised volatility estimator by decomposing financial time series into several timescales components which are related to different investment horizons. We also investigate the dynamic correlation at different timescales and at different time-lags, revealing a complex structure of financial signals.

Following the multiscale analysis approach, we propose a novel empirical method to estimate a time-dependent scaling parameter in analogy to the scaling exponent for self-similar processes. Using numerical simulations, we investigate the robustness of our estimator to heavy-tailed distributions. We apply the scaling estimator to intraday stock market prices and uncover scaling properties which differ from what would be expected from a random walk.

We also introduce a novel entropy-like measure which estimates the regularity of a time series. This measure of complexity can be used to identify periods of high and low volatility

which could help investors to choose the appropriate time for investment. Finally, we propose a multistep-ahead forecasting framework based on EMD combined with support vector regression. The originality of our models is the inclusion of a coarse-to-fine reconstruction step to analyse the forecasting capabilities of a combination of oscillating functions. We compare our models with popular benchmark models which do not use the EMD as a preprocessing tool, obtaining better results with our proposed framework.

Part of the research developed on this thesis is published in Physica A: Statistical Mechanics and its Applications [137] and in the European Physical Journal, Special Topics [136]. It was also presented at international conferences, including the 20th annual workshop on the Economic Science with Heterogeneous Interacting Agents (WEHIA) 2015 and the 21st Computing in Economics and Finance (CEF) conference 2015.

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

ARFIMA	Autoregressive Fractionally Integrated Moving Average
ARIMA	Autoregressive Integrated Moving Average
BM	Brownian Motion
CWT	Continuous Wavelet Transform
DFA	Detrended Fluctuation Analysis
DWT	Discrete Wavelet Transform
EMD	Empirical Mode Decomposition
EMH	Efficient Market Hypothesis
FBM	Fractional Brownian Motion
FGN	Fractional Gaussian Noise
FMH	Fractal Market Hypothesis
FT	Fourier Transform
HHT	Hilbert-Huang Transform
HMH	Heterogeneous Market Hypothesis
HT	Hilbert Transform
IMF	Intrinsic Mode Function
MODWT	Maximal Overlap Discrete Wavelet Transform

RV	Realised Volatility
SLM	α-stable Lévy motion
S&P	Standard and Poor's
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
SVR	Support Vector Regression
TSRV	Two-Scale Realised Volatility

Chapter 1

Introduction

Over the last few years financial markets have witnessed the availability and widespread use of data sampled at high frequencies. The use of high-frequency data allows to identify the intraday structure of financial markets [57]. Data generated at these frequencies have properties which are not caused by a single process but by several components that are superimposed onto each other in a hierarchical form. These components are not immediately apparent, but once identified, they can be meaningfully categorized as noise, cycles at different timescales and trends [57].

Scaling behaviour in financial data was first studied by Mandelbrot [124, 126] and it is found across financial markets with complex properties that are significantly related to economic and financial characteristics of markets [64]. Self-similarity or scale-invariance is an attribute of many natural laws and it is the underlying concept of fractals. It is related to the occurrence of similar patterns at different timescales. In this sense, probabilistic properties of self-similar processes remain invariant when the process is viewed at different resolutions [41].

A classic example of self-similar process is fractional Brownian motion (FBM), a Gaussian process characterized by a positive scaling exponent $0 < H < 1$ [127]. When $0 < H < \frac{1}{2}$, FBM is said to be anti-persistent with negatively autocorrelated increments. For the case $\frac{1}{2} < H < 1$, FBM reflects a persistent behaviour and its increments exhibit long-range dependence. When $H = \frac{1}{2}$, FBM is reduced to a process with independent increments known as Brownian motion.

For FBM, all timescales contribute proportionally and there is a specific relation that links statistical properties at different timescales [73]. However, real financial time series have more complex scaling patterns, with some timescales contributing disproportionally; these patterns characterize multiscaling processes whose statistical properties vary at each timescale [63, 64]. Two different sources of multifractality have been documented in the literature: the heavy-tailed probability distribution and the autocorrelation structure of the data [27, 106], considering stochastic processes which reproduce these characteristics may improve the modelling of financial data.

Multiscale properties of financial time series gave rise to two behavioural market theories, the heterogeneous market hypothesis (HMH) [135] and the fractal market hypothesis (FMH) [146]. The former theory considers multiple investment horizons, affirming that market participants are not a single homogeneous group of investors all possessing the same information and objectives, and therefore they react to the market at different times. Market participants are highly heterogeneous and have investment horizons that vary from seconds to years (market makers, noise traders, hedge funds). As consequence, financial markets became highly mixed and irregular systems, with complex financial time series exhibiting non-linearity, non-stationarity and long memory.

Similarly, the FMH asserts that financial participants are heterogeneous and market stability exists if there are investors with different time-horizons whose interactions create liquidity. These investors treat the arriving information differently and affect the price dynamics in various ways depending on their trading timescale [174]. The HMH and the FMH challenge the classic approach known as the efficient market hypothesis (EMH) [70], a theory which states that financial markets are efficient if they reflect all available information, thus arbitrage conditions are non-existent or quickly eliminated. Based on the strength of its assumptions, the EMH is stated in three forms: the weak-form affirms that historical prices cannot be used to gain excess of returns on the stock market. The semi-strong form refers not just to the historical prices but to all publicly available information, affirming that all these data are already reflected in prices. The strong-form of the EMH implies that all information, including inside information, is already reflected in prices, meaning that no information can give the investor consistent excess of returns on the stock market.

Financial prices exhibit some universal characteristics persistent across various time periods, markets, assets, etc., and known as stylized facts. These characteristics include volatility clustering, autocorrelation, fat tails, high kurtosis, and could be interpreted as a contradiction to the prevalent EMH [44].

Non-stationarity and non-linearity in financial time series are also considered stylized facts [1, 93, 94, 154, 161, 177]. Testing for non-linearity in financial time series verifies the adequacy for linear models. If the underlying generating process is non-linear in nature, it would be inappropriate to employ linear models including Black–Scholes–Merton option pricing model, autoregressive models and stochastic volatility models [93, 95]. Chen et al [47] provide an overview and a critique of the performance of different tests for non-linearity in detecting different types of artificially generated non-linear structures.

The purpose of this research is to investigate the interactions between the different market participants and their investment horizons. We aim to reveal the underlying forces driving financial time series. The presence of fluctuations with larger magnitudes than the ones accounted for by the Gaussian distribution creates the need for new models and tools to study financial time series. Time-frequency representations, including Fourier and wavelet transforms, provide a powerful way to analyse time series by determining which frequencies are present, how strong the frequencies are, and how they vary over time. In particular, spectral analysis allows to infer information about the length of a cycle or to filter noise to reveal the data generating process.

The most common tool for spectral analysis is the Fourier transform, which represents the frequency components contained in a stationary time series through an orthogonal basis of sine and cosine functions. However, as these functions are analysed over the whole time domain at once, any time-related localization is lost. In this way, Fourier analysis is effective to study periodic and stationary time series whose properties do not change much over time.

Spectral analysis has been an important tool for time series analysis. Traditional methods are based on the assumption of second-order stationarity. However, this assumption is rarely fulfilled in real data or it is only approximately valid for time series of very short duration. The stationarity assumption can be lessen to a locally stationary property in which the spectrum is assumed to be changing slowly over time, and the time series can be approximated by a piecewise stationary time series. This approach can be considered as a local Fourier basis approximation using a time-varying transfer function to estimate the spectrum [3, 58, 59]. Several methods have been derived to model some stylized facts of financial log-returns by locally stationary process, see references [82, 121, 161]. For example, Fryzlewicz et al. [82] propose a non-stationary model for log-returns in which the time-varying volatility is a piecewise-constant function of time, allowing the modelling for abrupt changes in the return distribution.

In order to preserve the temporal local properties of the data, wavelet analysis was introduced to finance and econometrics. Wavelet analysis offers a non-parametric and concise way of studying the heterogeneity of financial markets under non-stationary conditions. By using wavelets, a decomposition into scales of different resolutions (the so called multiscale decomposition) is obtained. The basis functions used for the time-frequency representation are oscillations which decay rapidly with time and are termed as wavelets [61].

Although wavelet analysis has been widely used in finance, see for example [29, 71, 86], the main drawback of this technique is the need for an explicit selection of the mother wavelet able to achieve a meaningful decomposition which does not influence the interpretation of the results. Furthermore, the Heisenberg uncertainty principle limits the resolution that can be attained to the time-frequency representation of the Fourier and the wavelet transforms [120].

The Hilbert-Huang transform (HHT) [96] was introduced as a data decomposition tool aiming to eliminate the need for an a-priori basis selection. This transform does not assume any knowledge of the underlying dynamics and uses an adaptive basis which is extracted from the data itself. The HHT consists of two main steps:

1. Empirical mode decomposition (EMD). An algorithm which separates data into a small number of independent and nearly periodic functions called intrinsic mode functions (IMFs). These functions, being based on local characteristic scales defined as the distance between two successive local extrema, provide a tool to perform a more extensive and detailed analysis.

2. Hilbert transform. After obtaining the IMFs, their Hilbert transformation allows to obtain a localized time-frequency spectrum with physically meaningful instantaneous frequencies and amplitudes. These instantaneous attributes could be used to identify hidden structures embedded in the data.

1.1 Research objectives

Our research is motivated by the heterogeneous and the fractal market hypotheses and aims to understand market dynamics when different investment horizons interact. We question if the Hilbert-Huang transform, a completely adaptive and empirical decomposition framework, could be used to reveal hidden patterns in high-frequency financial data. We aim to identify the impact of the different investment horizons as explanatory factors driving the unknown generating process of financial time series. We also question if prices dynamics are generated by fluctuations at various timescales and how the variance of those fluctuations behaves with respect to the oscillating period.

In this thesis, we introduce a novel non-parametric analysis which consists of decomposing high-frequency financial data into a finite set of IMFs. These components are characterized by an oscillating frequency and are dominated by simpler generating processes. By studying the statistical properties of the flow of information between frequencies, we aim to identify properties that propagate across different timescales and to reveal the heterogeneous market structure.

Given the non-stationary nature of financial time series [161], time-dependent statistics which describe the changing dynamics of the data are needed, the HHT provides a new approach to achieve such localization. Among the studied stylized facts are: volatility clustering, dependence structures, anomalous scaling laws [137] and multifractal properties [136].

The main distinguishing feature of multifractal time series is the non-linear variation in the scaling behaviour of its moments. Multifractal structures appear to be a very practical and promising way to capture the hierarchical multicomponent structure of financial time series. Furthermore, multifractality seems to be able to account for rare events and to describe the intermittent properties of time series which display periods of high and volatile activity mixed with relatively calm periods. Using the HHT, we aim to describe the changing oscillatory behaviour and the heterogeneity of financial time series. We test if the level of fluctuation differs from the usual self-similar behaviour exhibited by a Gaussian distribution, where the fluctuations at fine scales are uniform and all of them follow the same scaling law.

By focusing exclusively on given timescale, we cannot explain the nature of the data generating process. A model which successfully explains daily price changes, is unable to characterize the nature of hourly price changes [57]. On the other hand, statistical properties of monthly price changes are often not fully covered by a model based on daily price changes [157].

In this work, we also investigate whether the extracted fluctuations can be used to construct non-linear models which could target short and long-term horizon, improving in this way forecasting results and producing above average risk-returns.

Publications

Part of this thesis is published in Physica A: Statistical Mechanics and its Applications [137] and in the European Physical Journal, Special Topics [136]. Part of the research was also presented at international conferences, including the 20th annual workshop on the Economic Science with Heterogeneous Interacting Agents (WEHIA) 2015 and the 21st Computing in Economics and Finance (CEF) conference 2015.

1.2 Data description

Across this thesis, we use a data set consisting of intraday observations of 22 different stock market indices which are reported in Table 5.3. We also use intraday data of the S&P 500 implied volatility index known as the VIX index. This volatility index is the trade mark of the Chicago Board Options Exchange and it is the markets' expectation of the future market volatility over a 30 day horizon [52].

All observations are recorded at 30-second intervals, with the exception of the Warsaw stock exchange index which observations were only available at every minute frequency. The data were obtained from Bloomberg and cover a period from July 2013 to November 2014. The number of working days and the number of observations for every trading day depend on the opening hours of each stock market exchange. Different subsets of the complete data set were considered for the studies reported in the different chapters of this thesis. Furthermore, we generally apply the proposed methodology to intraday observations of a single trading day, to subsequently repeat the analysis on each day of the data set and make deductions from the average behaviour of the analysed data.

1.3 Thesis structure

In each chapter, we propose, describe and discuss an application of the HHT to the analysis of high-frequency financial data. We used the HHT implementation in MATLAB available online at http://perso.ens-lyon.fr/patrick.flandrin/emd.html.

In **Chapter 2**, we present the literature review and the theoretical background of this research. We review some relevant theories and concepts used in the timescale analysis of financial time series. We discuss some attributes of financial time series, including volatility clustering, intraday seasonalities, correlation patterns and scaling properties. We also discuss the most commonly used spectral methods, describing in detail the Hilbert-Huang transform. Moreover, we provide a comparative example between the studied spectral methods, summarizing advantages and disadvantages.

In **Chapter 3**, we propose an alternative estimator of realised volatility which is based on the different oscillating components latent in a time series and obtained via the EMD. The scale-by-scale study of volatility assumes that market data contain patterns specific to some frequencies of observations and are thus of interest for different types of market agents. The proposed estimator provides information on the contribution of the different frequencies to the total variance of the underlying process.

In **Chapter 4**, we propose two approaches to investigate the dynamic correlation between a pair of time series. The multiscale analyses provided by the EMD allows to study timescale dependent correlations. The dependencies are quantified via the Pearson correlation applied to the IMFs. The time varying characteristics of these correlations are investigated by using a rolling-window approach. This method results in the estimation of both the strength of the correlation and the time-lag when the maximum correlation occurs. Under the EMH such relationships should not exist as all the information is incorporated in the prices, however, in real financial markets such dependencies exist for short periods of time.

In **Chapter 5**, we uncover a scaling law that relates the variance of the IMFs to a power law of the oscillating period. The scaling exponent is related to the Hurst exponent. We verify the scaling relationship with numerical simulations of well known long memory process. As an application of the proposed methodology, we investigate the scaling properties of intraday stock market indices. Analysing 22 different stock market indices, we observe deviations from the FBM and Brownian motion scaling behaviours.

In **Chapter 6**, we study the relative weight of the oscillating components present in financial time series coupled with their characteristic timescale. These components are extracted via the HHT. We propose two novel time-dependent measures of complexity: 1) an amplitude scaling exponent and 2) an entropy-like measure. The proposed measures are tested on simulations of fractional Brownian motion and α-stable Lévy motion. By using these measures on high-frequency financial data, we are able to identify intraday cycles, trends and intermittent behaviour. Our measures do not assume any particular parameter, the temporal behaviour and variations are found directly from the data, considering only the timescales obtained via the EMD.

In **Chapter 7**, we propose some multistep-ahead forecasting models based on EMD and support vector regression. The obtained IMFs are less noisy and conform more closely to the assumptions made by dominant forecasting models. Every IMF can be treated as a particular pattern, forecasting techniques based on different IMFs could be viewed as modelling different investment horizons. The novelty of our proposed method is the inclusion of a coarse-to-fine reconstruction step to analyse the forecasting capabilities of a combination of IMFs. We use our models for intraday multistep-ahead forecast. The fast-frequency components are used for the immediate steps (short-term horizons), while low-frequency components forecast long-term horizons.

Lastly, **Chapter 8** is the overall conclusion of this research along with remarks for future work. We develop a methodology that successfully and simultaneously characterizes both the short and the long-term horizons of a time series. The research of this thesis shows that many issues previously studied in financial time series may gain new insight with the HHT by separating processes into different timescales and repeating the traditional analysis on each of them. The characteristics of the HHT fit the features of financial time series which are attributed to the interaction of multiple processes at different timescales.

Chapter 2

Literature Review and Theoretical Background

In this chapter, we present an overview of several studies relevant to our work which consider the properties of financial data the result of the interactions of many heterogeneous participants. We discuss attributes of financial time series, including volatility clustering, intraday seasonalities, correlation patterns and scaling properties.

The first part of this chapter summarises some time-frequency methods, focusing on the Hilbert-Huang transform, the methodology that is extensively used in this thesis. We include a comparison example which discusses some advantages and drawbacks of the presented methods. We continue by reporting several conventional estimators of realised volatility and revising dependency structures found in financial time series, including long-memory and self-similarity. We finalise the chapter with a summary of various models and strategies to forecast financial time series.

2.1 Time-frequency methods applied to the analysis of financial time series

The general objective of spectral analysis is the decomposition of a time series into its frequency components in order to detect and investigate any cyclical behaviour of its generating process. A well established concept in the study of financial time series is the existence of multiscaling patterns, an observed time series may be produced by several interacting processes, each occurring on a different timescale. The presence of heterogeneous agents with different investment horizons may generate very complex patterns [135]. In this way, studying the properties of a time series using only a single frequency (the sampling frequency) can be misleading when trying to understand the market structure. Important information could be lost due to the naive aggregation of the different frequency components into a single component [131].

The most common tool for spectral analysis is the Fourier transform, which represents the latent frequency components contained in the time series through sine and cosine functions. However, as these functions are analysed over the whole time series at once, any time-related information is lost. Various attempts to preserve the temporal locality property of the data introduced wavelet analysis into finance. Wavelet analysis offers an approach to study the heterogeneity of financial markets under non-stationary conditions. By using wavelets, one obtains a decomposition into scales of different resolutions (the so called multiscale decomposition). The analysing basis functions are oscillations that decay rapidly with time and are termed as wavelets. These functions are very attractive as they possess the unique ability to provide a complete representation of a time series from both the time and the frequency domains. Wavelet analysis has been adopted in many studies in the finance literature, for example: estimation of volatility and jump variation [71], analysis of foreign exchange [148], analysis of scaling properties of foreign exchange volatility [85], multi-resolution forecasting of futures prices [186], to mention a few.

Aiming to develop a complete adaptive time-frequency representation, the Hilbert-Huang transform (HHT) was introduced [96]. The HHT is a two-step algorithm which combines the empirical mode decomposition and the Hilbert spectral analysis. In contrast to Fourier and wavelet transform, the HHT uses an empirical basis extracted from the analysed data itself. Furthermore, it provides time-varying amplitude and frequency attributes. Some applications of the HHT to financial data include: a measure of changeability as a proxy for volatility [97], phase correlation of foreign exchange time series [181], analysis of oil prices using EMD [187]. More recently, the EMD has been used to identify fluctuation tendencies that simplify the forecasting task into several simple forecasting subtasks [103].

The Hilbert transform as an independent tool has also been successfully applied for pricing different types of derivative contracts (plain vanilla, single and double barrier, lookback options) when the underlying asset process evolves according to a Lévy process [74, 129].

2.2 Review of time-frequency methods

2.2.1 Fourier analysis

The Fourier analysis is a spectral method attributed to the mathematician Joseph Fourier [81] which provides a link between the time-domain and the frequency-domain of a function. This analysis includes both the Fourier series and the Fourier transform (FT). The former is used to analyse periodic functions while the latter is applicable to functions that are defined on the real set [87].

Fourier Series

The Fourier series representation of a real-valued periodic function $f(t)$ is given by:

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos(k\omega_0 t) + \sum_{k=1}^{\infty} b_k \sin(k\omega_0 t), \tag{2.1}$$

where $\omega_0 = \frac{2\pi}{T}$ is the fundamental frequency [87]. The real quantities a_0, a_k and b_k are defined as:

$$a_0 = \frac{2}{T}\int_0^T f(t)dt, \tag{2.2}$$

$$a_k = \frac{2}{T}\int_0^T f(t)\cos(k\omega_0 t)dt, \tag{2.3}$$

$$b_k = \frac{2}{T}\int_0^T f(t)\sin(k\omega_0 t)dt, \tag{2.4}$$

with $k = 1, 2, \dots, \infty$.

Fourier transform

The Fourier transform of a function $f(t) \in L^2(R)$ of a real variable t is defined by the integral:

$$F(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt. \tag{2.5}$$

This equation states that for a frequency ω_1, the function $\hat{f}(\omega_1)$ represents the component of $f(t)$ at ω_1 [87]. If we can determine all the frequency components of $f(t)$, a superposition of all these components should reconstruct the original function:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\omega) e^{i\omega t} d\omega. \tag{2.6}$$

Previous integral is referred to as the inverse Fourier transform. If the variable t represents time, $\hat{f}(\omega)$ is called the spectrum of $f(t)$ [87]. The weakness of this transform is that the Fourier spectrum does not provide any time-domain information, a local analysis is needed to combine both the time and the frequency representation. A possible solution is the application of the short-time Fourier transform, where the transformation is locally calculated [87].

Short-time Fourier transform

The short-time Fourier transform (STFT) is composed of two steps, firstly, the time series is divided into segments and then, the spectrum of each segment is obtained via the Fourier transform [87]. The STFT of a function $f(t)$ with respect to the window function $\phi(t)$ and evaluated around the location b is defined as:

$$SF(\omega, b) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t)\phi(t-b) e^{-i\omega t} dt. \tag{2.7}$$

The signal can be reconstructed from its transform by the formula:

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} SF(\omega, b)\phi(t-b) e^{i\omega t} db\, d\omega, \tag{2.8}$$

where the window function $\phi(t)$ is allowed to be complex and must have a non-zero spectrum at $\omega = 0$, behaving like a low-pass filter [87]. Classical choices for $\phi(t)$ include the rectangular, the Hanning, the Hamming or the Gaussian windows [87]. The length of the window determines the time and the frequency resolution of the representation and this resolution is kept constant over the time-frequency plane. A short window leads to a representation which is fine in time but coarse in the frequency domain. Conversely, a long window leads to a representation which is coarse in time but fine in the frequency domain. This time-frequency resolution trade-off is formalized by the Heisenberg-Gabor uncertainty principle, stating that we can not obtain precision in both the time and the frequency domain simultaneously [87]. As a way to improve time-frequency localization, wavelet analysis was introduced [89] .

2.2.2 Wavelet analysis

Wavelets can be loosely described as oscillatory basis functions, constructed with some attractive features not possessed by other global functions such as sines and cosines waves. The term wavelet refers to an oscillatory vanishing wave with time-limited duration and with the advantage of representing a variety of functions in a sparse manner with simultaneous localisation in the time and the frequency domain [89].

Mathematically, a wavelet function $\psi(t) \in L^2(\mathbb{R})$ has an average value equal to zero, $\int_{-\infty}^{+\infty} \psi(t)dt = 0$ and the square of $\psi(t)$ integrates to unity, $\int_{-\infty}^{+\infty} \psi^2(t)dt = 1$ [172]. The function $\psi(t)$ is also known as the "mother wavelet". A family of wavelet functions can be derived from a mother wavelet by translation of a factor k and dilation of scale λ, that is:

$$\psi_{k,\lambda}(t) = \frac{1}{\sqrt{\lambda}}\psi\left(\frac{t-k}{\lambda}\right), \tag{2.9}$$

where $k, \lambda \in \mathbb{R}$. In this way, the wavelet transform of a time series evolving in time depends on two variables, scale and time.

Time localization is achieved by using translated versions of the mother wavelet and frequency localization is accomplished by scaled versions of it. The scaled and translated version of the mother wavelet are used to measure the correlation with the time series to be analysed. When the signal correlates to large scales, the coarse features of the input time series are highlighted. Contrary, high correlation with small scales disclose the fine features of the input time series.

The difference between the wavelet and the STFT lies in the shapes of the analysing functions, the STFT uses functions with the same width, contrary to the wavelet transform which uses width adapted functions. High-frequency wavelets are very narrow while low frequency wavelets are much broader [172]. For the sake of clarity, let us start the wavelet transform study with the description of the continuous wavelet transform.

Continuous wavelet transform

The continuous wavelet transform (CWT) of a function $f(t) \in L^2(\mathbb{R})$ is defined as a function of two variables, k and λ, time and scale respectively [172] and it is defined as:

$$W_{\psi,f}(k,\lambda) = \frac{1}{\lambda}\int_{-\infty}^{+\infty} f(s)\,\bar{\psi}_{k,\lambda}(t)\,ds, \tag{2.10}$$

where $\bar{\psi}$ denotes the complex conjugate of the wavelet function ψ. When λ is increased, the wavelet function is dilated and when k is varied, the wavelet is translated in time. Hence, by changing (k, λ), the function $W_{\psi,f}$ can be computed on the entire time-frequency plane [172].

In order to reconstruct the function $f(t)$ from its continuous wavelet transform, the mother wavelet should satisfy the admissibility condition [172]. A wavelet is said to be admissible if its Fourier transform $\Psi(\omega)$ satisfies:

$$C_{\psi} = \int_{-\infty}^{+\infty} \frac{\mid \Psi(\omega) \mid^2}{\omega} d\omega < \infty. \tag{2.11}$$

In this way, the function $f(t)$ can be reconstructed from its wavelet transform as:

$$f(t) = \frac{1}{C_{\psi}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\lambda^2} W_{\psi,f}(k, \lambda) \bar{\psi}_{k,\lambda}(t)\, dk d\lambda. \tag{2.12}$$

The parameters λ and k vary continuously over $\mathbb{R}$ (with the constraint $\lambda \neq 0$), making the continuous wavelet transform a redundant transformation. In order to minimize the amount of correlation information, discrete values of k and λ can be selected, introducing the discrete wavelet transform.

Discrete wavelet transform

The discrete wavelet transform (DWT) can be thought of as a subsampled version of the CWT. The DWT selects values of the parameters λ and k using a critical sampling that keeps the transformation invertible. The critical sampling will provide the minimal basis by selecting λ of the form 2^{-j}, and $k = m2^{-j}$, $j, m \in \mathbb{Z}$. Any coarser discretization will not produce a unique inverse transformation [172].

Mallat, S. [119] originally demonstrated that the computation of the discrete wavelet transform can be accomplished with a hierarchical structure of filter banks which produce two types of wavelet coefficients: approximation coefficients and detailed coefficients. The approximation coefficients describe the global features of the data and are obtained by filtering the input time series with a low-pass filter. The detail coefficients describe the local features of the data and are obtained by high-pass filtering. The approximation coefficients are used in the filter structure as the input for the next iteration. Each decomposition level corresponds to a specified resolution which decreases with the number of decomposition levels [172].

The DWT is a shift-variant transform, meaning that the DWT of a function shifted in time is quite different from the transform of the input function. The maximum overlap discrete wavelet transform (MODWT) is a non-orthogonal and shift-invariant modification of the DWT [145]. Unlike the DWT which down-samples the approximation coefficients and detail coefficients at each decomposition level, the MODWT does not perform a down-sampling [145]. Therefore, the approximation coefficients and the detail coefficients at each level have the same length as the input time series. Another advantage of the MODWT is that it can be used to analyse a time series of any arbitrary size.

The main drawback of the wavelet transform is that its performance depends on the explicit and a priori selection of the mother wavelet. This selection may influence the frequency analysis. Aiming for a completely adaptive tool, the Hilbert-Huang transform was proposed by Huang et al. [96].

2.2.3 Hilbert-Huang analysis

The Hilbert-Huang transform was designed to analyse non-linear and non-stationary time series. It was originally developed to study water-wave evolution, but it has proven to be a useful tool for other complex signals. The HHT consists of two steps: firstly, the empirical mode decomposition (EMD) and secondly, the Hilbert transform (HT) [96]. The EMD separates the time series into a set of narrow-band functions and the Hilbert transformation of these functions provides local frequency and amplitude attributes. The ability to capture non-linear characteristics with respect to amplitude and frequency and the fact that the HHT is a decomposition based on the local characteristics of the data, have made this transformation very appealing to many research areas.

Empirical mode decomposition

The EMD is a fully data-driven decomposition that can be applied to non-stationary and non-linear data [96]. Differently from the Fourier and the wavelet transforms, the EMD does not require any a priori filter function [143]. The purpose of the method is to identify a finite set of oscillations with scale defined by the local maxima and the local minima of the data itself. Each oscillation is empirically derived from the data and is referred to as an intrinsic mode function (IMF). An IMF must satisfy two criteria [96]:

1. The number of extrema and the number of zero crossings must either be equal or differ at most by one.

2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The first condition forces an IMF to be a narrow-band signal with no riding waves. The second condition ensures that the instantaneous frequency will not have fluctuations arising from an asymmetric wave form [96]. In Figure 2.1, we show an example of an IMF.

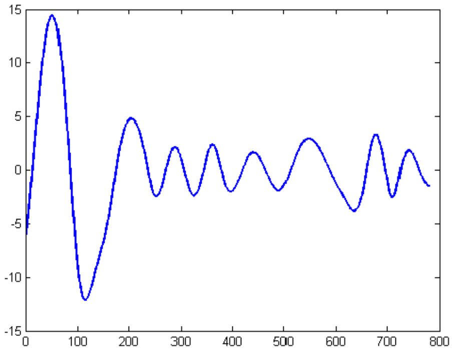


Fig. 2.1 Example of an IMF.

The IMFs are obtained through a process called sifting process which uses local extrema to separate oscillations starting with the highest frequency. Given a time series $x(t)$, $t = 1,2,...,N$, the process decomposes it into a finite number of functions, denoted as $IMF_k(t)$, $k = 1,...,n$, and a residue $r_n(t)$. The residue is the non-oscillating drift of the data. If the decomposed data consist of uniform scales in the frequency space, the EMD acts as a dyadic filter and the total number of IMFs is approximately equal to n= $\log_2(N)$ [80]. At the end of the decomposition process, the original time series can be reconstructed as:

$$x(t) = \sum_{k=1}^{n} IMF_k(t) + r_n(t). \tag{2.13}$$

The EMD comprises the following steps [96]:

1. Initialize the residue to the original time series $r_0(t) = x(t)$ and set the IMF index $k = 1$.

2. To extract the k^{th} IMF:

   (a) initialize $h_0(t) = r_{k-1}(t)$ and the iteration counter $i = 1$;

   (b) find the local maxima and the local minima of $h_{i-1}(t)$, see Figure 2.2(a);

   (c) create the upper envelope $E_u(t)$ by interpolating between the local maxima (lower envelope $E_l(t)$ by interpolating the local minima, respectively), refer to Figure 2.2(b);

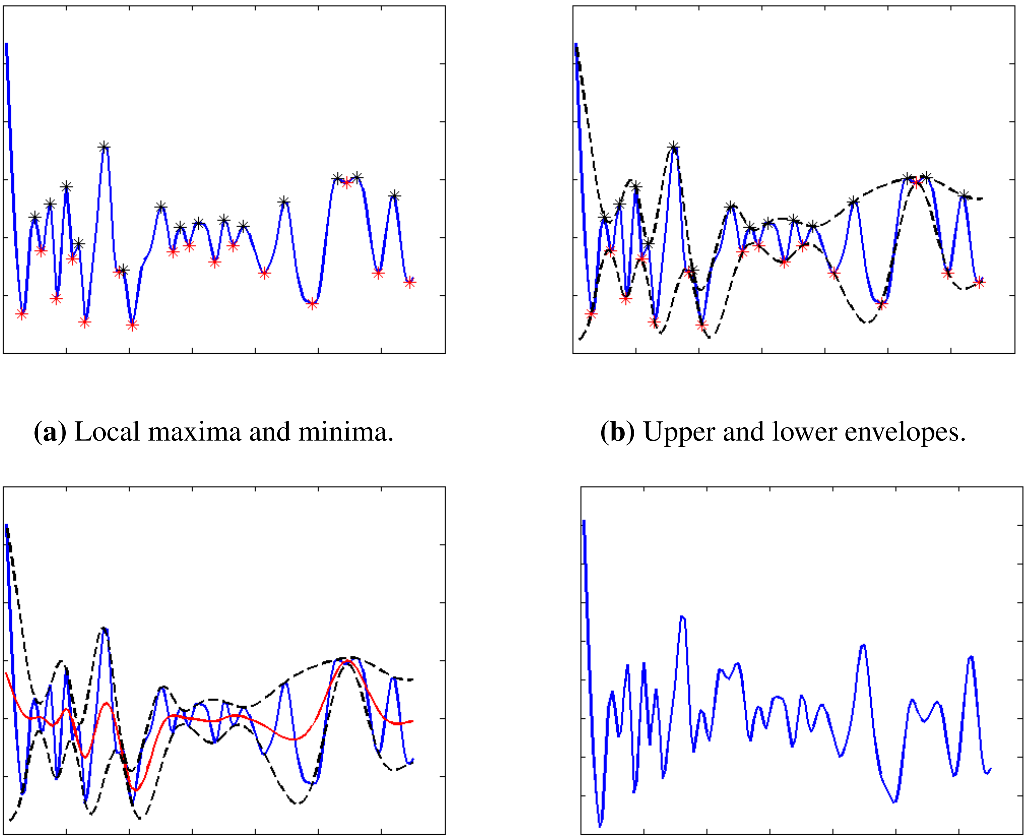
   (d) calculate the mean of both envelopes as $m_{i-1}(t) = \frac{E_u(t)+E_l(t)}{2}$, see Figure 2.2(c);

   (e) subtract the envelope mean from the input time series, obtaining $h_i(t) = h_{i-1}(t) - m_{i-1}(t)$, see Figure 2.2(d);

   (f) verify if $h_i(t)$ satisfies the IMF's conditions:

- if $h_i(t)$ does not satisfy the IMF's conditions, increase $i = i+1$ and repeat the sifting process from step (b), Figure 2.2(d);
- if $h_i(t)$ satisfies the IMF's conditions, set $IMF_k(t) = h_i$ and define $r_k(t) = r_{k-1}(t) - IMF_k(t)$, see Figure 2.3(d).

3. When the residue $r_k(t)$ is either a constant, a monotonic slope or contains only one extrema stop the process, otherwise continue the decomposition from step 2, setting $k = k+1$.



(a) Local maxima and minima.

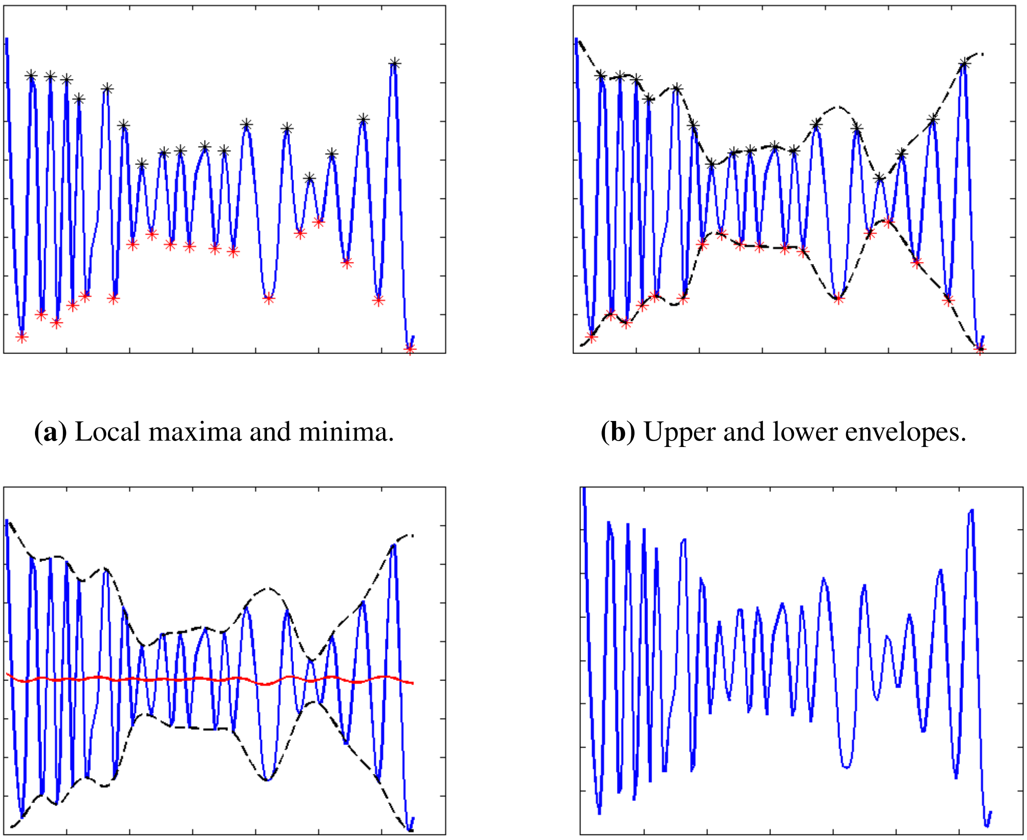
(b) Upper and lower envelopes.

(c) Envelope mean.

(d) Time series after one sifting step.

Fig. 2.2 Sifting process. (a) Input time series highlighting the local maxima and the local minima. (b) Time series with the interpolated upper and lower envelopes. (c) Time series with the envelopes and the mean of both envelopes. (d) First iteration of the sifting process. The extracted function does not satisfy the IMF's conditions.

In Figure 2.2, we exemplify some steps of the sifting process. After one iteration of the sifting process, the function $h_1(t)$ observed in Subfigure 2.2(d) is obtained. The resulting function is not symmetric and does not have zero mean, hence it is not an IMF yet. More iterations of the sifting process need to be applied to extract the first IMF of the input time series. Figure 2.3(a), 2.3(b) and 2.3(c) illustrate the last sifting iteration which extracts the first IMF displayed in Subfigure 2.3(d).



(a) Local maxima and minima.

(b) Upper and lower envelopes.

(c) Envelope mean.

(d) IMF example.

Fig. 2.3 Sifting process. (a) Input time series highlighting the local maxima and the local minima. (b) Input time series with the interpolated upper and lower envelopes. (c) Input time series with the envelopes and the mean of both envelopes. (d) Last iteration of the sifting process, the extracted function is the first IMF.

The IMFs are nearly orthogonal to each other and the variance of the input time series is approximated by the sum of the variance of the components plus the variance of the residue. However, it must be noted that the EMD is based on the timescale separation and not on the demand of orthogonality. For some non-linear data, orthogonality may not be satisfied implying that in general the sum of the variance of the components and the residue differs from the variance of the input time series [96].

The sifting process eliminates the riding waves and smooth uneven amplitudes to obtain meaningful values of instantaneous frequency [96]. This process terminates when the local mean of the extracted IMF is zero. The difficulty is that this condition can only be approximated and in order to avoid over-sifting and converting meaningful IMFs into meaningless fluctuations with constant amplitude, a stopping criterion needs to be implemented.

Stopping criterion for the sifting process

Several stopping criteria have been adopted, the original work of Huang et al. [96] proposes a stopping criterion based on a Cauchy type convergence test, making the sifting process to terminate when the normalized difference between two consecutive iterations is smaller than a predetermined threshold δ, specifically:

$$\sigma_k^2 = \sum_{t=1}^{N} \frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)} < \delta. \tag{2.14}$$

Huang et al. [96] proposed values of δ between 0.2 and 0.3 which guarantee that the IMFs components retain enough information of both amplitude and frequency modulations. However, this criterion does not depend on the definition of the IMFs since σ_k^2 might be small, but there is no guarantee that the function will have the same numbers of zero crossings and extrema [99].

Rilling et al. [151] proposed a variation of the stopping criterion by taking into consideration the local mean and the local amplitude of the envelope functions. These authors introduced a new criterion based on two thresholds parameters, θ_1 and θ_2, and aimed at guaranteeing globally small fluctuations in the mean while taking into account locally large amplitudes. Denoting by E_u and E_l the upper and the lower envelope functions respectively, a new function $\sigma(t)$ is defined as:

$$\sigma(t) = \frac{|E_u(t) + E_l(t)|}{|E_u(t) - E_l(t)|}. \tag{2.15}$$

The sifting process is iterated until $\sigma(t) < \theta_1$ for some prescribed fraction $(1-\alpha)$ of the total duration, while $\sigma(t) < \theta_2$ for the remaining fraction. Rilling et al. [151] proposed parameters values of $\alpha \approx 0.05$, $\theta_1 \approx 0.05$ and $\theta_2 \approx 10\theta_1$.

This stopping criterion assumes that the variations of an IMF, i.e., $(E_u(t) - E_l(t))$ are large compared to its mean envelope value $(E_u(t) + E_l(t))$ for most of the time, while are roughly of the same order for the remaining time. This stopping criterion is implemented in all the analyses presented in this thesis.

Properties of the EMD

The EMD is an algorithm completely determined by the input data and no mathematical formulation exists to describe the extracted functions. However, this decomposition empirically satisfies some major requirement for a time series decomposition method, specifically, completeness and orthogonality.

- Completeness. As indicated by Equation (2.13), the IMFs are sufficient to describe and to recover the input time series. The difference between the sum of the IMFs and the input data is considered the reconstruction error. Huang et al. [96] found this error to be of the order of:

$$\varepsilon = x(t) - \sum_{k=1}^{n} IMF_k - r_n < 10^{-14}. \tag{2.16}$$

- Orthogonality. Although orthogonality is not theoretically guaranteed, it is satisfied in practical terms and it can be numerically estimated a-posteriori [96]. Including the residue as the last component and rewriting Equation (2.13) as $x(t) = \sum_{i=1}^{n+1} C_i(t)$, the square of the values of $x(t)$ can be expressed as:

$$x^2(t) = \sum_{i=1}^{n+1} C_i^2(t) + \sum_{j\neq i}^{n+1} \sum_{i=1}^{n+1} C_i(t)C_j(t). \tag{2.17}$$

If the decomposition is orthogonal, the cross-terms should be zero. In this way, an index of orthogonality can be defined as:

$$IO = \sum_{t=1}^{N} \frac{\sum_{j\neq i}^{n+1} \sum_{i=1}^{n+1} C_i(t)C_j(t)}{x^2(t)}. \tag{2.18}$$

Orthogonality prevents energy leakage between the IMFs, that is, it prevents the problem of mistakenly identified latent frequencies. In some cases, for example, when analysing non-linear time series, the orthogonality condition cannot be guaranteed since it is not a necessary criterion for the basis selection. The principle of the IMF selection is merely based on the physical timescales that characterize the initial time series [96].

Shortcomings and proposed improvements of the EMD

The EMD is a completely adaptive method which has been widely applied in many research areas, however its theoretical foundations and limitations remain uncertain. The main deficiencies of this method can be listed as:

- Uniqueness. Given the lack of analytical representation, uniqueness in the decomposition cannot be guaranteed. Depending on the set of parameters applied to the sifting process (interpolation, boundaries, stopping criterion) the extracted IMFs may differ.

  Improvement: Huang et al. [98] studied the effect of the stopping criterion by calculating an ensemble mean and a standard deviation of an IMF set obtained using different stopping criteria. The sample mean was taken as the final IMFs and a confidence limit for the EMD was defined as a range of standard deviations. A shortcoming for the uncontrolled sifting process is that for each stopping criterion, the number of obtained IMFs might be different, making an average value a complicated solution.

- Stopping criterion. The sifting process creates a trade-off between producing an incomplete and incorrectly defined set of IMF due to insufficient sifts (under-sifting) and producing less physically meaningful IMFs with almost constant amplitudes (over-sifting). Thus, the challenge is to propose a reasonable stopping criterion.

  Improvement: Various stopping criteria have been proposed, for more details refer to Section 2.2.3.

- Spline interpolation. Generally, the envelope estimation is implemented by interpolating the local maxima and the local minima using cubic splines. The extracted IMFs are highly dependent on the interpolation outcome. Although the spline algorithm seems to produce acceptable results [80, 96], an overshoot problem can occur, shifting the mean value of the upper and the lower envelopes and degenerating the IMFs [141]. This deficiency may be magnified by the iterative nature of the sifting process.

  Improvement: A modification of the EMD replaces the local mean obtained by the difference of two cubic spline interpolations by a local mean obtained as the moving average of B-splines [46]. This approach gives a more analytical representation to the EMD since B-splines may lead to a proof of convergence [99].

- End effects. The first step of the EMD is to obtain the local maxima and the local minima of the analysed time series. The first and last point of the time series cannot be determined to be a maximum or a minimum, making the envelope to diverge and affecting the decomposition process. Moreover, these errors propagate to the next iteration creating some false IMFs. The end effects will also manifest as spectral leakage in the Hilbert transform.

Improvement: In order to minimize error propagations due to finite observations, the end points of the time series have to be treated differently and the data have to be extended beyond the existing range. The first technique for dealing with the end conditions was proposed by [96] and it consists of padding the beginning and the end of the time series with additional "characteristic waves" which are defined by the two consecutive extrema. Flandrin et al. [151] offer one of the simplest yet very robust method that uses a mirror symmetry with respect to the extrema closest to the end.

- Mode mixing. This problem appears when a single IMF either consists of widely variant frequencies, or a similar frequency resides in different IMFs. Mode mixing is a consequence of signal intermittency, which occurs when a component of a particular frequency either comes into existence or disappears completely in an inconsistent time series.

  Improvement: Wu et al. [183] proposed the ensemble empirical mode decomposition (EEMD), a noise-assisted data analysis method. Essentially, this method adds white noise of finite amplitude to the original time series. Thus, noise-adjusted time series are decomposed into IMFs by the EMD. The means of the corresponding IMFs generated from each time series are subsequently treated as the IMFs of the EEMD method. If the amplitude of the added noise is too small relative to the original signal, the noise may not affect the extrema that the EMD method relies on. As a result, no effect on mode mixing prevention can be achieved. On the other hand, if the amplitude of the added noise is too large, it would result in redundant IMFs. Thus, while the EEMD could eliminate the problem of mode mixing, how to choose an appropriate amplitude for the added noise and how to determine the number of ensemble trials is still a topic of discussion. Moreover the computational cost of the decomposition is highly increased.

Certainly, the most serious drawback of the EMD is its lack of theoretical foundation. In our view, the proposed improvements increase performance only marginally and tend to reduce the advantage of the EMD which is a completely adaptive method with no further assumptions imposed to the analysed data. Furthermore, it is difficult to measure the impact of the improvements, since there is no way to determine which set of IMFs is the best. Some enhancement may only work for specific data and can create an overcomplicated method that may require some previous knowledge before it can be used. For this reason, through the analysis presented in this thesis, we did not implement the improvements, taking the EMD as proposed by [96] with the modified stopping criterion proposed by [151].

Hilbert transform

The Hilbert spectral analysis is an alternative method to represent a time series in the time-frequency domain. Some instantaneous attributes of a time series (amplitude, phase and frequency) are obtained via the Hilbert transform. This one dimensional integral transformation convolves the time series $x(t)$ with the filter $1/(\pi t)$ to obtain the function $y(t)$ [108], that is:

$$y(t) = \frac{1}{\pi t} * x(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}\, d\tau, \tag{2.19}$$

where the integral has a singular point at $\tau = t$ and it is defined as a Cauchy principal value, i.e., defined via symmetric limits:

$$\begin{aligned} y(t) &= \frac{1}{\pi} \mathrm{PV} \int_{-\infty}^{+\infty} \frac{x(\tau)}{t-\tau}\, d\tau & (2.20) \\ &= \frac{1}{\pi} \lim_{\varepsilon \to 0} \left(\int_{-1/\varepsilon}^{t-\varepsilon} \frac{x(\tau)}{t-\tau}\, d\tau + \int_{t+\varepsilon}^{1/\varepsilon} \frac{x(\tau)}{t-\tau}\, d\tau \right). & (2.21) \end{aligned}$$

The Fourier transform of the filter $1/(\pi t)$ is given by:

$$F\left((\pi t)^{-1}\right) = -i\,\mathrm{sgn}(f) = \begin{cases} -i & \text{if } f > 0 \\ i & \text{if } f < 0 \end{cases} \tag{2.22}$$

where f is the frequency. In this way, the positive frequencies of the spectrum of $x(t)$ are shifted by 90° and the negative frequencies are shifted by 90°. The Hilbert transform can then be viewed as a filter of amplitude unity and phase 90° depending on the sign of the frequency of the input signal spectrum.

The time series $x(t)$ and its Hilbert transform $y(t)$ form an analytic signal in the complex plane. This analytical function is denoted as $z(t)$ and has the same positive frequency spectrum as $x(t)$ but has zero negative frequencies,

$$z(t) = x(t) + iy(t). \tag{2.23}$$

Theoretically, there are infinitely many ways of defining the imaginary part, but the Hilbert transform provides a unique way so that the result is an analytic signal [96].

In Figure 2.4, we show an example of a complex analytical signal of a periodic time series with increasing amplitude. The projection of the analytic signal onto the plane defined by the real axis and the time axis is the initial time series. The projection onto the plane defined by the imaginary axis and the time axis is the Hilbert transform of the time series. The projection onto the complex plane is the geometrical representation of a rotating phasor.

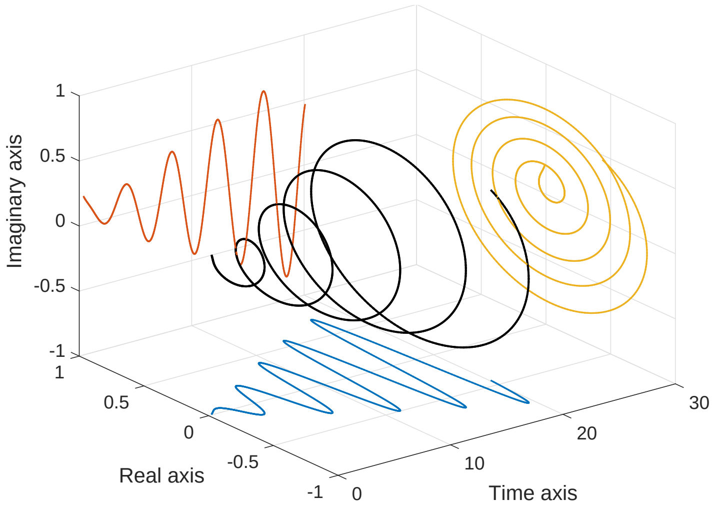


Fig. 2.4 Complex analytical representation. The real plane contains the original time series. The imaginary plane shows its Hilbert transform, and the complex plane is the geometrically representation of a rotating phasor.

The function $z(t)$ can be re-expressed in its exponential form, representing the time series $x(t)$ as an harmonic fluctuation modulated by the amplitude and the phase of its oscillations,

$$z(t) = a(t)\exp^{i\theta(t)}, \tag{2.24}$$

where the instantaneous amplitude is given by:

$$a(t) = \sqrt{x^2(t) + y^2(t)}, \tag{2.25}$$

and the instantaneous phase is defined as:

$$\theta(t) = \tan^{-1}\frac{y(t)}{x(t)}. \tag{2.26}$$

The shape of the variation of the instantaneous amplitude is called the wave envelope [72]. This envelope takes the overall shape of the time series $x(t)$, taking the maxima and the minima but never crossing the time series itself. Moreover, the instantaneous amplitude function represents the combined oscillations of all frequencies involved in the component but removes all the frequency information. The instantaneous phase $\theta(t)$ is the angle relative to a fixed direction of the rotating phasor in the complex plane [72].

The instantaneous frequency $\omega(t)$ is defined as the derivative of the unwrapped phase function with respect to time, that is:

$$\omega(t) = \frac{d\theta(t)}{dt}. \tag{2.27}$$

In this way, the instantaneous frequency is just the varying speed of the rotating phasor in the complex plane. In the time domain, if some negative instantaneous frequencies take place, they correspond to the appearance of complicated riding cycles of an alternating signal [72].

Mathematically, it is correct to define instantaneous frequencies for any signal, but in practice, it is only appropriate for mono-component time series which can reveal slow varying instantaneous characteristics. Other types of wide-band time series or a composition of several oscillating components will result in complicated fast varying instantaneous characteristics that could be more difficult to analyse than the input time series itself [72]. The EMD was proposed as a way to pre-process time series before applying the Hilbert transform. The EMD generates components of the time series whose Hilbert transformation could lead to meaningful definitions of instantaneous amplitude and frequency. Hence, the combination of the EMD and the Hilbert transform gave origin to the Hilbert-Huang transform.

The analytic signal for a time series $x(t)$ has a one-sided Fourier transform, i.e. the transform is zero for negative frequencies. It can be approximated by calculating the fast Fourier transform of $x(t)$, replacing the Fourier coefficients corresponding to negative frequencies with zero and calculating the inverse of the result [130]. A more accurate algorithm is based on a sinc function expansion, which is able to provide exponential convergence of the error [162], while the first “naive” method only achieves quadratic convergence. In finance the sinc function algorithm for the Hilbert transform has been used e.g. to price discretely monitored exotic options [74, 83], where the error was reduced to machine precision in order to validate and compare the pricing methods. In our case, where empirical data is used, such a high accuracy becomes superfluous.

Other applications of the Hilbert transform in finance include the detection of log-periodicity preceding crashes in financial markets [189]. The Hilbert transform has been widely used also in signal processing tasks, including envelope detection and demodulation, analytical signal construction, magnitude and phase identification of the time varying spectrum [108, 112].

Hilbert-Huang transform

After applying the EMD to a time series $x(t)$ and obtaining its respective IMFs, the Hilbert transform is applied to each individual IMF for a time frequency analysis. More precisely, each IMF is associated with its Hilbert transform via:

$$\widehat{IMF_k}(t) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{IMF_k(\tau)}{t-\tau}\, d\tau, \tag{2.28}$$

and the combination of $IMF_k(t)$ and $\widehat{IMF_k}(t)$ gives the analytical representations $z_k(t) = IMF_k(t) + \widehat{IMF_k}(t) = a_k(t)e^{i\theta_k(t)}$, where $a_k(t)$ denotes the instantaneous amplitude and $\theta_k(t)$ the instantaneous phase for the $k^{\text{th}}IMF$. The input time series can be expressed in its analytical form as:

$$x(t) = \sum_{k=1}^{n} a_k(t)e^{i\theta_k(t)} + r_n(t). \tag{2.29}$$

The residue, $r_n(t)$, is not expressed in terms of its amplitude and phase since it is a function with only one extrema not containing enough information to confirm whether it is an oscillatory component whose frequency is physically meaningful [96].

The Hilbert-Huang spectrum represents the amplitudes latent in a time series as a function of time and frequency, eliminating the restriction of the Fourier transform to have constant amplitude and fixed frequencies. A common method to display the Hilbert spectrum is to generate a two-dimensional plot with time and frequency axes. The amplitude is then plotted as a colour spectrogram in the time-frequency plane. By plotting the Hilbert spectra of all the IMFs together, one can obtain a complete time-frequency representation of the input time series [96].

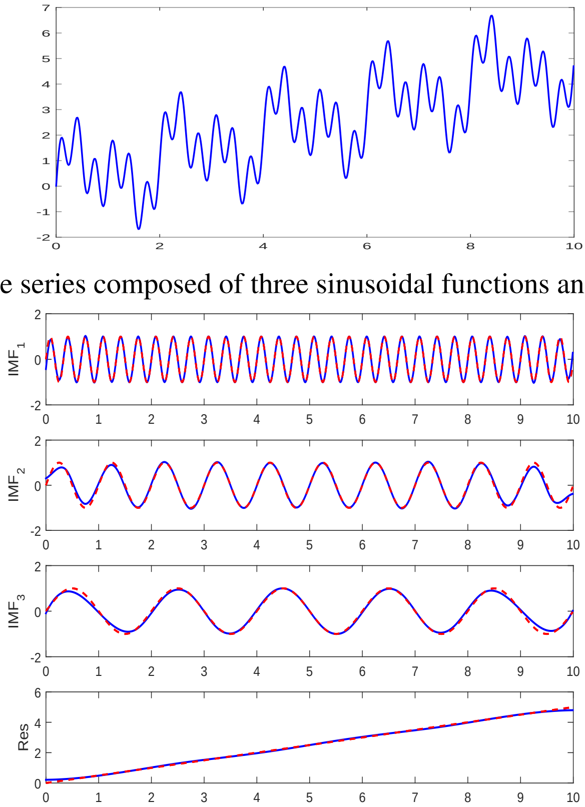
2.3 Example of a time series decomposition

In order to illustrate the functionality and adaptiveness of the EMD, let us here exemplify this algorithm by using a simple harmonic time series. For all the Hilbert-Huang transform analysis performed in this thesis we used the MATLAB implementation available in [139].

Denote by $x(t)$ a time series created as the sum of three sinusoidal components with different frequencies, $f_1 = 0.5$, $f_2 = 1$ and $f_3 = 3$ and a linear trend,

$$x(t) = \sin(2\pi t f_1) + \sin(2\pi t f_2) + \sin(2\pi t f_3) + 0.5t. \tag{2.30}$$

Setting the sampling frequency Fs= 100, and $t \in [0, 10]$, the length of $x(t)$ is $N = 1,000$. In Figure 2.5(a), we illustrate the harmonic time series $x(t)$.



(a) Input time series composed of three sinusoidal functions and a linear trend.

(b) Extracted IMFs. The blue lines represent the extracted IMFs and the red dotted lines indicate the original sinusoidal components. Some distortions created by the end effects of the EMD are observed in the extracted IMFs.

Fig. 2.5 EMD of the time series $x(t) = \sin(2\pi t f_1) + \sin(2\pi t f_2) + \sin(2\pi t f_3) + 0.5t$.

By applying the EMD to the time series $x(t)$, we are able to recover the sinusoidal components and the linear trend. Figure 2.5(b) illustrates the output of the EMD. The solid blue lines represent the extracted IMFs or the residue. For comparison reasons, we also plotted the pure sinusoidal components which are represented by the red dotted lines. The first IMF extracts the highest frequency component, $f_3 = 3$. The second IMF describes the oscillations with frequency $f_2 = 1$. Finally, the third IMF refers to sinusoidal components with the lowest frequency $f_1 = 0.5$. Using the EMD, we are able to recover the linear trend $0.5t$ which is represented by the residue. Some disturbances can be observed at the boundaries of the IMFs.

We compared the obtained decomposition against its wavelet and Fourier transform counterparts. For the wavelet transform, a Daubechies (db6) wavelet basis and a decomposition level $L = 3$ were selected. The decomposition level determines the scale of the extracted components. As can be seen from Figure 2.6, with $L = 3$, only the high-frequency details, denoted as D_1, D_2, D_3, are extracted and most of the information is kept in the approximation coefficients which are denoted by $A3$.

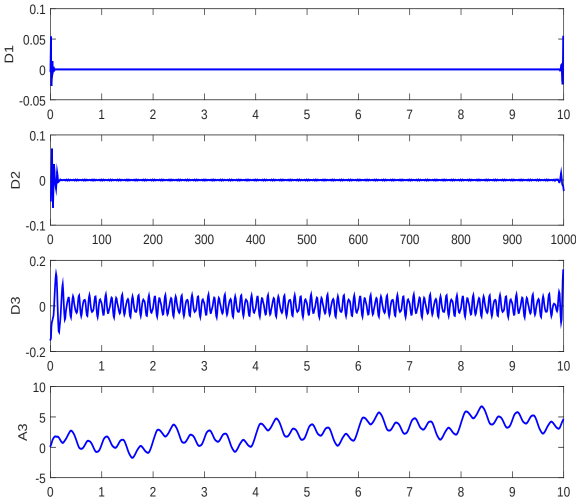
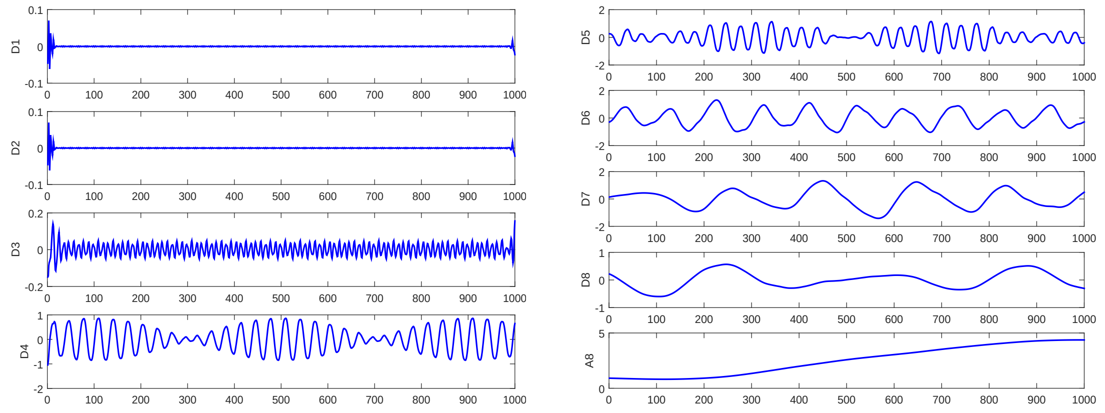


Fig. 2.6 Wavelet decomposition of the time series $x(t) = \sin(2\pi t f_1) + \sin(2\pi t f_2) + \sin(2\pi t f_3) + 0.5t$ using Daubechies (db6) wavelet basis and decomposition level $L = 3$.

A larger decomposition level which includes the lower frequency components is required. In Figure 2.7, we show the wavelet decomposition with $L = 8$. As can be observed from Figures 2.6 and 2.7, the wavelet transform does not recover the input oscillating components, but instead reveals the correlation between the input data and the selected mother wavelet. We observe that many of the resultant components are physically meaningless oscillations. Furthermore, the disadvantage of the wavelet transform is that its performance depends on an a-priori selection of the mother wavelet and the decomposition level.

On the other hand, the Fourier transform of the time series $x(t)$ could identify the sinusoidal components, but the presence of the linear component $0.5t$ contaminates the fre-

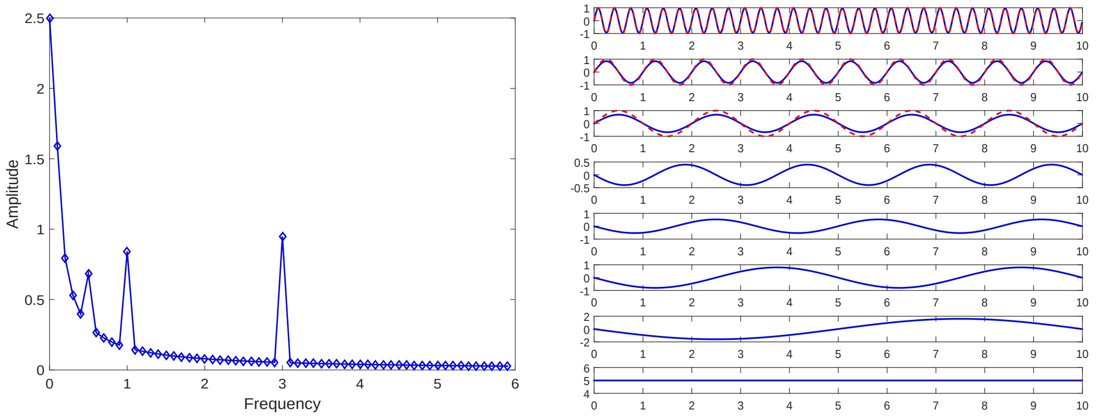


(a) .

(b)

Fig. 2.7 Wavelet decomposition of the time series $x(t) = \sin(2\pi t f_1) + \sin(2\pi t f_2) + \sin(2\pi t f_3) + 0.5t$ using Daubechies (db6) wavelet basis and decomposition level $L = 8$.

quency spectrum and creates false components. Figure 2.8(a) displays the Fourier amplitude spectrum which shows peaks not only at the expected frequencies, but also at some lower frequencies. These low frequency components are attributed to the linear trend that cannot be considered a periodic component. Figure 2.8(b) shows the inverse of the Fourier transform for the eight components with the largest amplitudes. The red dotted lines in this figure represent the original sinusoidal components of the analysed times series.



(a) Fourier amplitude spectrum.

(b) Inverse Fourier transform of the eight components with the largest amplitudes. The red dotted lines indicate the original sinusoidal components.

Fig. 2.8 Fourier decomposition of the time series $x(t) = \sin(2\pi t f_1) + \sin(2\pi t f_2) + \sin(2\pi t f_3) + 0.5t$.

Let us now illustrate the completeness of the EMD. By summing up the IMFs shown in Figure 2.5(b), we are able to recover the initial time series $x(t)$. Figure 2.9 illustrates the reconstruction of the input time series by adding to the residue, denoted by R, the IMFs from the lowest frequency to the highest frequency.

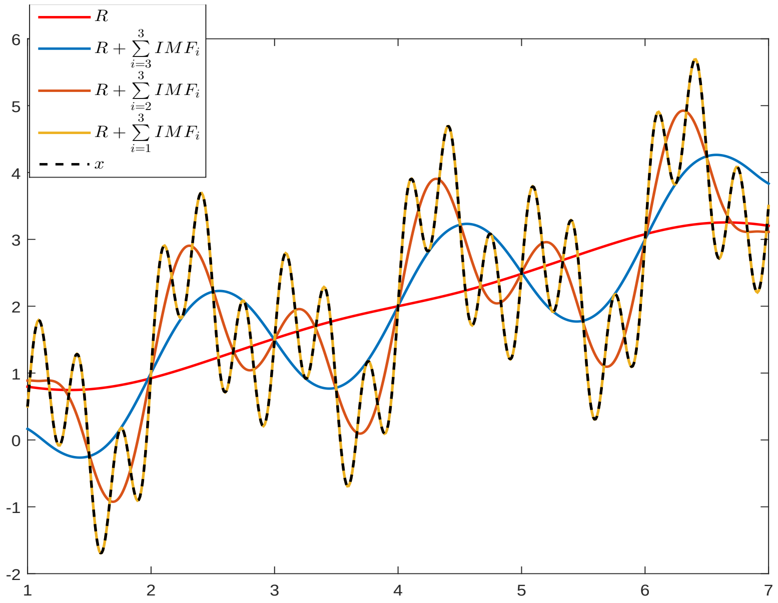


Fig. 2.9 Reconstruction of the input time series.

2.4 Integrated variance estimators

Following Andersen et al. [13], we consider a stochastic process X_t which describes the logarithmic price of a financial asset and evolves in continuous time over the interval $[0,T]$. According to the standard assumptions that the return process has finite instantaneous mean and does not allow arbitrage opportunities [13], the logarithmic price process X_t follows a special semi-martingale process that is uniquely decomposed into a local martingale and a predictable finite variation process [147]. The stochastic process X_t is described by the stochastic differential equation

$$dX_t = \mu_t dt + \sigma_t dW_t \qquad 0 \leq t \leq T, \tag{2.31}$$

where W_t is a standard Brownian motion, the drift term μ_t is locally predictable and of finite variation. The variable σ_t^2 is a càdlàg process such that $\int\limits_0^t \sigma_s^2 ds < \infty$ a.s. for any $t > 0$. In particular, the continuously compound return over a time Δt is

$$r_t = X_t - X_{t-\Delta t} = \int\limits_{t-\Delta t}^{t} \mu_s ds + \int\limits_{t-\Delta t}^{t} \sigma_s dW_s. \tag{2.32}$$

The integrated variance $\int_{t-\Delta t}^{t} \sigma_s^2 ds$ is not directly observable but can be estimated through the quadratic variation of the log price process. We define the partition $t - \Delta t = t_0, \ldots, t_N = t$ of the interval $[t - \Delta t, t]$ and we consider the discretely sampled process X_{t_i} whose quadratic variation is given by

$$[X]_t = \sum_{0 \leq i \leq N} (X_i - X_{i-1})^2 \tag{2.33}$$

where $X_i = X_{t_i}$ are observations of the return process in the interval $[t - \Delta t, t]$. Under the specifications of model (2.31), the integrated variance can be approximated by an ex-post volatility measure known as realised volatility [11].

We assumed the process described in Equation (2.31) does not include jumps, however the observed prices contain jumps and are contaminated with noise that is not Gaussian white noise as described in model (2.31). There is an extensive literature considering jump processes where the quadratic variation is decomposed into two parts: integrated variance of the latent price process and a jump variation part, see for example references [10, 23, 25, 29].

2.4.1 Realised volatility

The concept of realised volatility has been developed as a result of the availability of intraday transaction data. Given the discretely sampled returns $r_{t_i} = X_{t_i} - X_{t_{i-1}}$, the corresponding realised variance is defined by the summation of the N intraperiod squared returns [11, 13]:

$$RV_{t,\Delta t} = \sum_{i=1}^{N} r_{t_i}^2, \tag{2.34}$$

This estimator is simply the second sample moment of the return process over a fixed interval. Semi-martingale theory ensures realised variance is a consistent estimator of quadratic variation when enough data are accessible, i.e., when $N \to \infty$ [22]. Realised volatility is defined as the square root of the realised variance.

It is generally accepted that the return process is contaminated by micro-structure noise and that the realised variance does not converge as the sampling frequency increases [111]. In order to mitigate the impact of this noise, an estimator that considers a more sparse sampling of the return process was considered in [111]. This estimator is known as two-scale realised volatility estimator.

2.4.2 Two-scale realised volatility

Zhang et al. [111] proposed to model the micro-structure noise observed in financial markets as observational errors, specifically, as an i.i.d. Gaussian white noise, independent of the process X_t. In this way, the observed price process Y_t is of the form:

$$Y_t = X_t + \varepsilon_t. \tag{2.35}$$

The process X_t is the true or efficient logarithm of the price process that follows Equation (2.31) and the $\varepsilon_t' s$ are the independent errors around the true log-price process [111].

The two-scale realised volatility (TSRV) estimator is based on a sub-sampling and averaging procedure. This method takes advantage of the tick-by-tick data but corrects the adverse effect of micro-structure noise by combining the sum of squared estimators from two different timescales. The first one, $RV_{t,h}^{\text{all}}$ is the realised volatility of Equation (2.34) calculated from returns on a fast timescale. Similarly, $RV_{t,h}^{\text{sparse}}$ is calculated from the returns on a slow timescale.

The idea is to partition the original grid of observation times, $\mathcal{G} = t_1, t_2, \ldots, t_N$ into non-overlapping sub-grids $\mathcal{G}^k$, $k = 1, \ldots, K$ of size $\bar{N} = N/K$. For example $\mathcal{G}^1$, starts at the first observation and takes the next ones every s observations; $\mathcal{G}^2$ starts at the second observation and takes the next ones every s observations, etc., where s is any slow sampling scale. The TSRV estimator uses all the available data defining the average estimator as:

$$RV_{t,h}^{\text{average}} = \frac{1}{K} \sum_{k=1}^{K} RV_{t,h}^{\text{sparse},k},$$

where $RV_{t,h}^{\text{sparse},k}$ is the realised variance obtained on the k grid. The estimator $RV_{t,h}^{\text{average}}$ is still a biased estimator, though the bias increases with the average size of the sub-samples $\bar{N} = N/K$ [6]. The bias adjusted TSRV estimator is defined as:

$$TSRV_{t,h} = RV_{t,h}^{\text{average}} - \frac{\bar{N}}{N} RV_{t,h}^{\text{all}}. \tag{2.36}$$

2.4.3 Wavelet realised volatility

Wavelet decomposition has the ability to separate the energy of a time series across scales and it offers a multiscale approach to estimate realised volatility, see Section 2.2.2 for a detailed description of the wavelet transform. The MODWT is a conserving energy transform which provides an asymptotically efficient wavelet variance [144] .

By applying the MODWT to the return time series, we obtain the wavelet coefficients $W_{j,i}$, $i = 1, 2, \ldots, N$ at scale j and the detail coefficients $V_{J_{n,i}}$, where $J_n \leq \log_2 N$ denotes the maximum level of decomposition. The wavelet realised variance is defined as:

$$RV_{t,h}^{\mathrm{Wav}} = \sum_{i=1}^{N} (r_{t-1+ih}^{(h)})^2 = \sum_{j=1}^{J_n} \|W_j\|^2 + \|V_{J_n}\|^2, \tag{2.37}$$

where $\|W_j\|^2 = \sum_{i=1}^{N} W_{j,i}^2, \quad \|V_{J_n}\|^2 = \sum_{i=1}^{N} V_{J_{n,i}}^2$ [145] . Each j^{th} summand of equation (2.37) can be regarded as the contribution to the total energy due to variations at scale 2^{j-1}. The inclusion of the boundary coefficients can bias the variance estimator [144].

If one assumes that the return process is square-integrable, that there is no micro-structure noise and that the drift term μ, in equation (2.32) is equal to zero, then the RV^{Wav} is an unbiased estimator of realised variance. However, given all the micro-structure effects present in high-frequency financial data, the wavelet estimator only decomposes the realised variance into different timescales and it is still a biased estimator of quadratic variation when $N \to \infty$ [29].

2.5 Self-similarity

Self-similarity or scale invariance is an attribute of many laws of nature and it is the underlying concept of fractals. It is related to the occurrence of similar patterns at different timescales. A stochastic process $X(t)$ is statistically self-similar, with scaling exponent $H > 0$, if for any real $a > 0$ it follows the scaling law:

$$X(at) \overset{\mathrm{d}}{=} a^H X(t) \qquad t \in \mathbb{R}, \tag{2.38}$$

where the equality ($\overset{\mathrm{d}}{=}$) is in probability distribution [41].

Each of the properties of self-similar process is controlled by a single exponent H, however for some of the observed data, it is unlikely that a single parameter can convey all the information to describe the dynamics of the process, instead a continuous spectrum of exponents is required. A stochastic process $X(t)$ is called multiscaling if it has stationary increments and satisfies the scaling relation:

$$E(|X^q(t)|) = c(q)t^{H(q)+1} \qquad \text{for all } \; t \in T, \; q \in Q, \tag{2.39}$$

where T and Q are intervals on the real line such that $0 \in T$, $[0,1] \in Q$. The functions $H(q)$ and $c(q)$ have domain in Q [125]. The scaling function $H(q)$ is non-linear and must be concave, it takes into account the influence of time on the absolute moment of order q and relates all the information about the rate of growth of the moments of $X(t)$ as t varies.

2.6 Long-range dependence

The concept of long-range dependence is closely related to self-similarity. If a self-similar process has stationary increments, these increments form a stationary time series which can display long-range dependence [67]. Conversely, a central limit type theorem affirms that a stationary time series with long-range dependence yields a self-similar process with stationary increments. The intensity of the long-range dependence is related to the scaling exponent of the self-similar process [67].

A common definition of long-range dependence is the slow, power-law like decrease at large lags of the autocovariance function [31]. Let $Y(t)$ be a stationary stochastic process, denoting by $\gamma(k)$ its autocovariance function and assuming $c > 0$, the long-range dependence condition is given by:

$$\gamma(k) \sim ck^{-\beta} \quad \text{as} \quad k \to \infty \quad \beta \in (0,1). \tag{2.40}$$

Equivalently, long-range dependence can be defined as a divergence at the origin of its spectral density function f, that is:

$$f(\lambda) \sim c_f \lambda^{\beta-1} \quad \text{as} \quad \lambda \to 0 \quad c_f > 0. \tag{2.41}$$

The parameter $H = 1 - \frac{\beta}{2}$ is sometimes used instead of β [31]. Fractional Brownian motion with $H > 1/2$ is a typical example of self-similar process whose increments exhibit long-range dependence [31]. When $H = \frac{1}{2}$, FBM is reduced to a process with independent increments known as Brownian motion.

For self-similar process with stationary increments and infinite variance, the scaling exponent $H > 1/2$ does not necessarily imply that the increments manifest long-range dependence. There are so called α-stable processes with $0 < \alpha < 2$ that are self-similar with parameter $H = 1/\alpha > 1/2$ and independent increments, see [156].

2.6.1 Long-range dependence estimators

Let $\{Y(t), 1 \leq t \leq N\}$ denote the observations of a stationary time series with finite second moments. There exist numerous methods to detect long memory in a time series, refer to [31, 67] for an extensive review, here we briefly describe some of the classical approaches:

- Rescaled Range [31]. Estimate the partial sum of $Y(t)$ as $X_k = \sum_{i=1}^{k} Y(i)$, $k > 1$ and its sample variance as $S^2(k) = \frac{1}{k}\sum_{i=1}^{k}(Y_i - k^{-1}X_k)^2$. The R/S statistic is defined as:

$$\frac{R}{S}(k) = \max_{0\leq t\leq k}\left(X_t - \frac{t}{k}X_k\right) - \min_{0\leq t\leq k}\left(X_t - \frac{t}{k}X_k\right). \tag{2.42}$$

Many naturally occurring time series seem to present the relation $E\left[\frac{R}{S}(k)\right] \sim ck^H$, $c > 0$ as $k \to \infty$, with typical values of the Hurst parameter H in the interval $(0.5, 1)$ [128]. If the observations come from a short-range dependent model $E\left[\frac{R}{S}(k)\right] \sim ck^{0.5}$ as $k \to \infty$ [14].

The R/S statistic is robust to heavy-tailed distributions, in this way, if $Y(t)$ were a time series with long-range dependence and with infinite variance, the R/S statistic would still estimate the autocorrelation in the process [18, 28].

- Variance plot [31]. Let k be an integer, for different integers k in the range $2 \leq k \leq N/2$ and m_k subseries of length k, calculate the sample means $\bar{Y}_1(k), \bar{Y}_1(k), \ldots, \bar{Y}_{m_k}(k)$ and the overall mean as:

  $$\bar{Y}(k) = \frac{1}{m_k}\sum_{j=1}^{m_k}\bar{Y}_j(k). \tag{2.43}$$

  For each k, calculate the sample variance of the sample means $\bar{Y}_j(k)$:

  $$S^2(k) = \frac{1}{m_k}\sum_{j=1}^{m_k}\left(\bar{Y}_j(k) - \bar{Y}(k)\right)^2. \tag{2.44}$$

  The long-memory parameter H can be obtained from the proportionality between the window size and the sample variance by plotting $\log\left(S^2(k)\right)$ against $\log(k)$. For large values of k, the values of the plot lie around a straight line with negative slope $2H - 2$. In the case of short-range dependence or independence, the slope is equal to $2H - 2 = -1$. For infinite variance processes, this method provides an estimate of the long-memory parameter H [166].

- Detrendred fluctuation analysis (DFA) [142]. Define a profile time series as $X(t) = \sum_{t=1}^{N} Y(t)$ which is divided into N_s non-overlapping segments of equal length s. For each segment, the local trend is subtracted by a polynomial least-square fit which could be linear (DFA1), squared (DFA2), etc. The variance of each segment is calculated and denoted as $F_s^2(\nu)$. The average over all the segments is computed and a fluctuation function is defined as:

$$F(s) = \sqrt{\frac{1}{N_s}\sum_{v=1}^{N_s} F_s^2(v)}. \tag{2.45}$$

This fluctuation function follows the scaling law, $F(s) \sim s^{\zeta}$, where the exponent $\zeta = 1-\beta/2$ is called the scaling exponent and represents the correlation properties of the time series. Power law behaviour with $\zeta > 0.5$ indicates long-range dependence in the time series.

This method has been extended to consider different moments of the fluctuation function and can thus be applied to multifractal data [106]. When analysing stochastic processes with heavy tails, the exponent obtained from the multifractal detrended fluctuation satisfies:

$$H(q) \approx \begin{cases} \frac{1}{q} & \text{for} \quad q > \alpha \\ \frac{1}{\alpha} & \text{for} \quad q \leq \alpha, \end{cases} \tag{2.46}$$

where α is the parameter of the analysed stable distribution [106].

- Wavelet method [67]. Denote by W_j the wavelet coefficients at scale j extracted from the time series $Y(t)$, see section 2.2.2 for more details about wavelet analysis. A power-law relationship between the variance of the wavelet coefficients and the scale parameter j exists, such that:

$$\mathrm{Var}(W_j) = c2^{j(2H-1)}, \tag{2.47}$$

with $c > 0$. Taking the logarithm of both sides of previous equation results in a linear function of j with slope $2H-1$. Moreover, if the decomposed time series is FBM-like (non-stationary), a similar relationship can be obtained but with a different exponent

$$\mathrm{Var}(W_j) = c2^{j(2H+1)}. \tag{2.48}$$

As reported in [2], the wavelet estimator can be applied to heavy-tailed distributions, obtaining an estimate of the self-similarity parameter of the time series.

2.6.2 Example of self-similar and long-range dependent processes

Fractional Brownian motion

A Gaussian self-similar continuous process $X_H(t)_{t\in\mathbb{R}}$ with stationary increments and $0 < H < 1$ is called fractional Brownian motion (FBM) [67]. When $H = 1/2$ the FBM is the usual Brownian motion. The paths of FBM are characterized by anti-persistent behaviour when $0 < H < 1/2$, and persistent behaviour when $1/2 < H < 1$. The increment process of FBM, i.e., $Y(k) = X(k+1) - X(k)$, $k \in \mathbb{Z}$ is called fractional Gaussian noise (FGN) and it is a stationary process which exhibits long-range dependence if $1/2 < H < 1$.

Fractional Brownian motion is the only Gaussian self-similar process with stationary increments. There are other self-similar processes with stationary increments and infinite variance, for example, α-stable Lévy motion.

α-stable Lévy motion

A stochastic process $X(t)$, $t > 0$ is called (standard) α-stable Lévy motion (SLM) if:

1. $X(0) = 0$ a.s.

2. X has independent increments.

3. $X(t) - X(s) \sim S_\alpha\left((t-s)^{1/\alpha}, \beta, 0\right)$ for any $0 \leq t < \infty$ and for some $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$. With S_α denoting an α-stable distribution with scale parameter $(t-s)^{1/\alpha}$, skewness parameter β and shift parameter equal to zero.

The process X has stationary increments and it is $1/\alpha$-self-similar, i.e., $H = 1/\alpha$. The case $\alpha = 2$ corresponds to Brownian motion. For more details about stable distributions refer to [156].

ARFIMA process

The autoregressive fractionally integrated moving average (ARFIMA) process $X(t)$, $t \in \mathbb{Z}$, is usually denoted as ARFIMA(p,d,q) where $p, q \in \mathbb{N} \cup 0$, $-1/2 < d < 1/2$ and defined as:

$$\phi_p(B)\Delta^d X(t) = \Theta_q(B)\varepsilon(t), \tag{2.49}$$

where ϕ_p and Θ_q are polynomials of order p and q respectively, B denotes the backward operator and $\varepsilon(t)$ are i.i.d. random variables with either finite or infinite variance [67]. This model is an extension of the ARIMA(p,d,q) model, refer to section 2.7, allowing the differencing exponent d to take fractional values, $-1/2 < d < 1/2$. The correspondence between the parameters H and d is given by $H = d + 1/2$. The interval $0 < d < 1/2$ of long-range dependence corresponds to $1/2 < H < 1$ [67].

2.7 Autoregressive models

Autoregressive moving average (ARMA) models [37] are used to predict future values of a stationary time series using a linear regression between consecutive observations. However, most financial time series are non-stationary, reducing applicability of such models. For non-stationary time series, the autoregressive integrated moving average (ARIMA) model [37] can be used instead. The assumption to apply an ARIMA (p,d,q) model is that after differencing d times the input time series x_t, the obtained values $y_t = \Delta^d x_t$ form a stationary time series with zero mean. It is also assumed that the future values of the time series y_t are a linear function of p past observations $y_{t-1}, y_{t-2}, \ldots, y_{t-p}$ and q random errors $z_t, z_{t-1}, \ldots, z_{t-q}$, which are i.i.d. white noise with zero-mean and constant variance σ^2. The ARIMA model is thus expressed as:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \ldots + \theta_p y_{t-p} + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_q z_{t-q}. \tag{2.50}$$

The order of the model is defined by the values of p and q which are identified using patterns in the autocorrelation function and the partial autocorrelation function of the time series y_t [38]. After selecting p and q, the model parameters, θ and ϕ, are estimated using the maximum likelihood estimation method. Model selection criteria such as the Akaike information criterion (AIC) and the Schwarz's Bayesian information criterion (BIC) can be used to select the best fitting model [133].

2.8 Support vector regression

Support vector machines (SVMs) were originally developed to solve classification problems in pattern recognition. The introduction of the insensitive loss function allowed its use in non-linear regression estimation problems and the formulation of SVR [171]. The main advantage of SVR is its global and unique solution, while classical neural networks suffer from local minima problems [165]. Moreover, SVR has a simple geometric interpretation and a sparse solution that is obtained implementing the structural risk minimization principle and aims to minimize an upper bound of the generalization error [167].

A brief description of SVR is introduced as follows. A regression problem can be defined as to determine a function for approximating the output from a set of training data $X = (x_1,y_1),(x_2,y_2),\dots,(x_l,y_l)$ where x_i is the input value, $x_i \in X \subseteq \mathbb{R}^n$, y_i is the target value, $y_i \in Y \subseteq \mathbb{R}$, and l is the total number of training samples. SVR approximates the given observations by a linear function $f(x)$ and a non-linear map $\phi(x)$, from $\mathbb{R}^n$ to a high dimensional feature space $\mathscr{F}$ [171]:

$$y = f(x) = w\phi(x) + b, \tag{2.51}$$

The coefficients w and b are estimated by minimizing the regularized risk function:

$$R(C) = \frac{1}{2}\|w\|^2 + C\frac{1}{l}\sum_{i=1}^{l} L_\varepsilon(x_i,y_i), \tag{2.52}$$

with insensitive loss function given by:

$$L_\varepsilon(x_i,y_i) = \begin{cases} \|x_i - y_i\| - \varepsilon & \text{if } \|x_i - y_i\| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}. \tag{2.53}$$

The constant C is the regularization term and it determines the trade-off between the flatness of $f(x)$ and the approximation accuracy required on the training data which is specified by the ε parameter [160]. The insensitivity parameter ε acts together with C as a safeguard against over-fitting. Not penalizing small errors avoids increasing the model complexity. Both C and ε are parameters determined by the user.

After the quadratic optimization problem is solved, the parameter vector w of Equation (2.51) is given by:

$$w = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*)\phi(x_i), \tag{2.54}$$

where α_i and α_i^* are the Lagrange multipliers [171]. Only a certain number of coefficients $(\alpha_i - \alpha_i^*)$ will assume non-zero values. The data points associated with them have approximation errors larger than ε and are referred to as support vectors. These are the data points lying on or outside the ε-bound of the decision function, and are the only elements of the data points that are used to determine the decision function. Generally, the larger the ε, the fewer the number of support vectors and thus the sparser the representation of the solution. However, a larger ε can also reduce the approximation accuracy placed on the training data. In this sense, ε is a trade-off between the sparseness of the representation and closeness to the data.

Finally, the SVR function is given by:

$$f(x) = \sum_{i=1}^{l} (\alpha_i - \alpha_i^*) K(x, x_i) + b, \tag{2.55}$$

where $K(x, x_i)$ is the kernel function. The type of kernel function implicitly defines the non-linear map from the input space to some high-dimensional feature space.

2.8.1 Kernel function

The advantage of using a kernel function is that one can deal with feature spaces of arbitrary dimensionality without having to compute the map $\phi(x)$ explicitly. The most widely used kernel function is the Gaussian radial basis function (RBF), defined as $K(x, y) = e^{-\gamma(x-y)^2}$, where γ denotes the width of the RBF [50].

2.8.2 Parameter selection

Parameter selection plays an important role in obtaining a function that produces robust and accurate estimates. SVR performance depends on a good setting of parameters, such as regularization constant C, insensitive coefficient ε and kernel width parameter γ. A relatively simple parameter optimization is a grid-search [30]. The optimal parameters are based on which combination of parameters performs the best; the grid point that achieves the smallest average validation error is chosen as the model parameters. The main drawback of this method is that high accuracy requires a fine grid, making the method more computational expensive.

2.8.3 Cross-validation

Training an algorithm and assessing its statistical performance on the same data may create over-optimistic results. Cross-validation is a re-sampling technique for model selection which uses multiple training and validation subsamples to avoid such optimistic results [84]. When applying cross-validation on time series data, the time dependence structure of the data needs to be considered in order to prevent the use of future observations in the forecasting process. A moving cross-validation scheme [105] involves dividing the data into a series of overlapping training and validation sets. Each set is moved forward through the time series k times (the folds), keeping constant the length of both sets, see Figure 2.10. The k results from the folds are averaged to produce a single estimation. The variance of the resulting estimate is reduced as k is increased. The disadvantage of this validation is that the algorithm has to be trained k times [16].

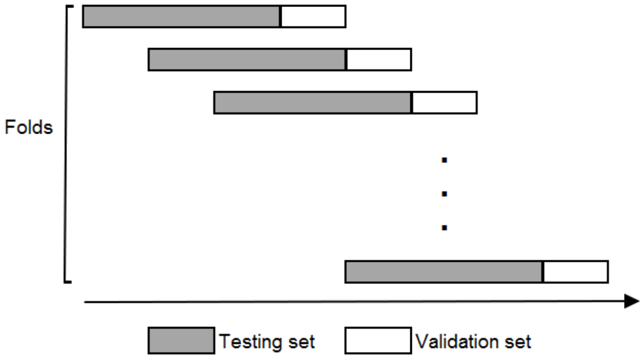


Fig. 2.10 Moving cross-validation scheme.

2.9 Multistep-ahead forecasting strategies

Multistep-ahead forecast can either be produced recursively, by iterating a one-step-ahead model or directly, by estimating a different model for each forecast horizon. We denote by x_t the input time series, and by h the target time-horizon. In the following section, we describe in some detail both strategies.

Recursive strategy

This strategy constructs a prediction model $f(\cdot)$ which minimizes the in-sample one-step-ahead prediction error [51]:

$$\hat{x}_{t+1} = f\left(x_t, x_{t-1}, x_{t-2}, \dots, x_{t-(m-1)}\right), \tag{2.56}$$

where m is the maximum embedding order, i.e., the number of past values taken into consideration to predict the future value. The next forecasted value is obtained using the same model $f(\cdot)$:

$$\hat{x}_{t+2} = f\left(\hat{x}_{t+1}, x_t, \dots, x_{t-(m-2)}\right). \tag{2.57}$$

The forecasted value of $\hat{x}_{t+1}$ is used instead of the true value which is unknown. For the h-step-ahead forecast, the values $\hat{x}_{t+1}$ to $\hat{x}_{t+h-1}$ are forecasted recursively,

$$\hat{x}_{t+h} = f\left(\hat{x}_{t+h-1}, \hat{x}_{t+h-2}, \dots, x_{t-(m-h)}\right). \tag{2.58}$$

When h becomes larger than m, all the input values are outputs of the forecasting model, a factor which may deteriorate the accuracy of the prediction. The main advantage of this forecasting strategy is that only one model has to be trained.

Direct strategy

The direct strategy improves the forecasting accuracy but increases the complexity of the algorithm. With this strategy, a different model $f_h(\cdot)$ has to be trained for each forecast horizon. The various forecasting models are independently estimated, for the h-step-ahead forecast, the forecasting model is expressed as:

$$\hat{x}_{t+h} = f_h\left(x_t, x_{t-1}, x_{t-2}, \ldots, x_{t-(m-1)}\right).$$

The previous forecasted values are not used as inputs, therefore the errors are not accumulated to the next step.

Chapter 3

Volatility Estimation at Different Timescales

The central point of this chapter is the scale-by-scale analysis of variance. Market data contain patterns specific to the observation frequency and are thus, of interest to different type of market agents (intraday speculators, daily traders, portfolio managers and institutional investors), each having their characteristic period of reaction to news and frequency of intervention to the market. In this chapter, we propose an EMD-based realised volatility estimator which identifies the oscillating components with the largest contributions to the total volatility. We apply the proposed estimator to intraday data of the S&P 500 index and we compare the results with the wavelet realised volatility estimator.

3.1 Estimation of realised volatility using high-frequency data

Estimation and prediction of volatility are key factors in finance and they are essential to the theory and practice of asset pricing, portfolio selection, hedging strategies and risk management. The main difficulty is that volatility is not an observable quantity; therefore it has to be estimated. Common parametric approaches to estimate volatility include stochastic volatility models [159] and the (generalized) autoregressive conditional heteroskedasticity model, commonly referred as (G)ARCH model [33, 68]. These autoregressive models were proposed to capture the observed properties of the distribution of returns, such as heavy tails and temporal dependencies in its second moment. However, these models fail to capture the asymmetry in volatility [33].

The recent availability of high-frequency financial data has expanded the volatility modelling literature. In the early work of Merton, R. [132], the asset's volatility over a fixed period of time was estimated as the sum of squared returns for sufficiently finely sampled observations. High-frequency financial data have promoted the formulation of realised volatility estimators, which are non-parametric estimators based, for example, on squared returns over a relevant time-horizon [11], refer to Section 2.4 for more details about realised volatility estimators.

Realised volatility has gained a lot of popularity due to its practical estimation without the assumption of any model and its effectiveness to analyse the market with all available information. Unfortunately, squared returns are contaminated by micro-structure noise, which could be attributed to many factors, including bid-ask bounces, discreteness of price changes, trades occurring across different markets, non-uniformity in trade sizes, gradual responses of prices to large block trades, etc. The higher the frequency at which the prices are sampled at, the larger the effect of micro-structure noise, causing a higher volatility estimation [5, 111].

Zhou, B. [188] was one of the pioneers to estimate realised volatility using high-frequency data and to correct for the bias by explicitly subtracting the autocorrelation of the returns sampled at high frequencies. Other volatility estimators were proposed trying to overcome the problem of micro-structure noise by moderating the sampling frequency and choosing an optimal one [5, 19, 90]. The drawback of the previous estimators is that they do not make use of all available data, which could be considered a failure for a robust model.

Zhang et al. [111] suggested the first consistent estimator of integrated variance using all available data in the sample. Their two-scale realised volatility estimator (TSRV), combines two measures of realised volatility and reduces the bias created by micro-structure noise. A detailed analysis of the accuracy of realised kernels as estimators of quadratic variation was provided by Barndorff et al. [21]. The realised kernel estimators reduce the autocorrelation observed in prices by using a weighted average. These estimators proved to be robust to time dependent noise and to asynchronous sampling [140].

Some studies admit the presence of jumps or discontinuities in the definition of the stochastic volatility model. It appears that many log-price processes are best described by a combination of a continuous and mean reverting process and a much less persistent jump component [4, 10]. In this case, the realised volatility is the sum of the integrated volatility and the jump component. Barndorff-Nielsen [23, 24] proposed the bipower variation estimator which separates both components. Bipower variation compares two measures of the integrated variance, one containing the jump variation and the other being robust to jumps and containing only the integrated variation component.

Introducing time-frequency methods to the estimation of realised volatility, Malliavin et al. [26, 122] introduced a Fourier based estimator which reconstructs the instantaneous volatility as a series expansion with coefficients obtained from the Fourier coefficients of the price variation. This Fourier estimator uses all the available observations eliminating the need for equally spaced data and is robust to microstructure noise [123]. Fan and Wang [71] introduced the concept of wavelet realised volatility. Their method is based on the definition of the TSRV estimator and is robust to both jumps in the price and to market micro-structure noise in the observed data. Gençay et al. [85] proposed another application of wavelets showing that volatility is asymmetric across timescales. By using wavelets, these authors described a model which explained the information flow between volatilities across timescales. Baruník, J. and Vácha, L. [29] decomposed realised variance into different investment horizons and jumps, generalizing the approach of [71] by using the maximal overlap discrete wavelet transform (MODWT), as authors prior to this were merely incorporating the standard discrete wavelet transform (DWT).

In the following section, we propose a basic estimator of realised volatility that is based on the EMD. The proposed estimator provides a perspective on the influence of different timescales on volatility estimation.

3.2 EMD-based realised volatility

Let X_t, be the logarithm of a price asset at time $t = 1,2,\dots,N$. Realised variance over a time interval is calculated as the summation of the squared returns on that interval, see section 2.4.1 for further details about realised variance estimators.

In this section we propose an estimator of realised variance using the EMD. Applying this decomposition method to the log-price time series X_t, we obtain a set of n IMFs and a residue, such that:

$$X_t = \sum_{j=1}^{n} IMF_j(t) + r_n(t). \tag{3.1}$$

A measure of variability for the log-price time series at timescale j can be estimated as the sum of the squared returns of each IMF_j, that is:

$$RV\left(IMF_j^{\mathrm{ret}}\right) = \sum_{i=1}^{N-1} IMF_j^{\mathrm{ret}}(t_i)^2 = \sum_{i=1}^{N-1} \left(IMF_j(t_{i+1}) - IMF_j(t_i)\right)^2. \tag{3.2}$$

A similar measure of variability is calculated for the residue:

$$RV\left(r_n^{\mathrm{ret}}\right) = \sum_{i=1}^{N-1} r_n^{\mathrm{ret}}(t_i)^2 = \sum_{i=1}^{N-1} \left(r_n(t_{i+1}) - r_n(t_i)\right)^2 . \tag{3.3}$$

The total realised variance of the input time series is estimated as the sum of all the variances at different timescales:

$$RV^{\mathrm{EMD}} = \sum_{j=1}^{n} RV\left(IMF_j^{\mathrm{ret}}\right) + RV\left(r_n^{\mathrm{ret}}\right). \tag{3.4}$$

The proposed realised variance estimator RV^{EMD} does not take into account the micro-structure noise in the data or the jumps in the price process. This simple estimator provides a variance decomposition into different investment horizons. Contrary to the wavelet estimator given in Equation (2.37), the time-horizons are not predetermined but extracted from the data itself considering only the local maximum and minimum of the data.

Ignoring the potential presence of jumps in the log-price process may cause a large bias in the estimation of realised volatility [9]. This neglect can also bias the estimation of the periodic components. Similarly, accounting for periodicity improves the accuracy of intraday jump identification. It increases the power to detect the relatively small jumps occurring at times when volatility is periodically low and reduces the number of false jump detections when volatility is high [10, 36]

With wavelet transform, the information on jump locations and jump sizes is stored at high-resolution wavelet coefficients, whereas useful information for integrated volatility is stored at the low-resolution ones [29, 71]. In the same way, the EMD could be used to localize the jumps in the log-price process and to allocate them to the highest frequency IMFs without the corruption of subsequent IMFs which represent longer cycles.

3.3 Realised volatility analysis of the S&P 500 index

In order to estimate the daily integrated variance, we applied the proposed realised volatility estimator to intraday observations of the S&P 500 index. The complete data set consists of 178 intraday time series, from July 11$^{\mathrm{th}}$ 2013 to March 31st 2014. The observations are collected from the starting of the trading session 9:30 a.m. to the closing time 4:00 p.m. EST. Observations are sampled at every 30 seconds. This sampling frequency provides a reasonable balance between the effects of market micro-structure frictions at the highest sampling frequencies on the one hand, and the analytics on a wide frequency range on the other.

At higher sampling frequencies, market micro-structure effects have a larger impact. At the tick timescale, the data differ from the theoretical diffusion process, and the volatility computed with very short time intervals is no longer an unbiased and consistent estimator of the integrated variance. This effect can be observed by studying the realised volatility, Equation (2.34), as a function of the sampling frequency [12, 55, 90]. In Figure 3.1, we show the realised volatility signature plot for the analysed S&P 500 data, the microstructural factors cause a positive serial correlation at high frequencies, resulting in a smaller estimate of realised volatility.

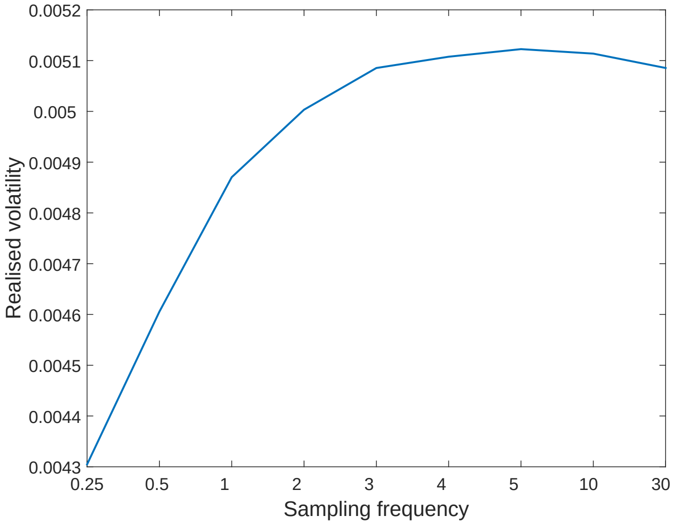


Fig. 3.1 Realised volatility signature plot. The vertical axis is the RV estimator averaged over all trading days. The horizontal axis is the sampling frequency expressed in minutes.

3.3.1 Intraday realised volatility, example on a single time series

For the sake of clarity, let us first consider the decomposition and realised variance analysis of a single intraday time series, bearing in mind that the same analysis was performed on all the 178 intraday time series. Figure 3.2 shows the S&P 500 prices for the day taken as an example, December 9$^{\text{th}}$ 2013. In Figure 3.3, we display the timescale decomposition obtained via the EMD on the logarithm of the previous time series. Note that the sifting process produces six IMFs and a residue.

We estimated the realised variance for a period of one day. The relative variance contribution of each IMF and the residue to the total realised variance is reported in Table 3.1. The second column of this table shows the IMF oscillating period expressed in minutes and calculated by dividing the total number of points by the number of peaks of each IMF. The third column shows the variance associated to each IMF and finally, the fourth

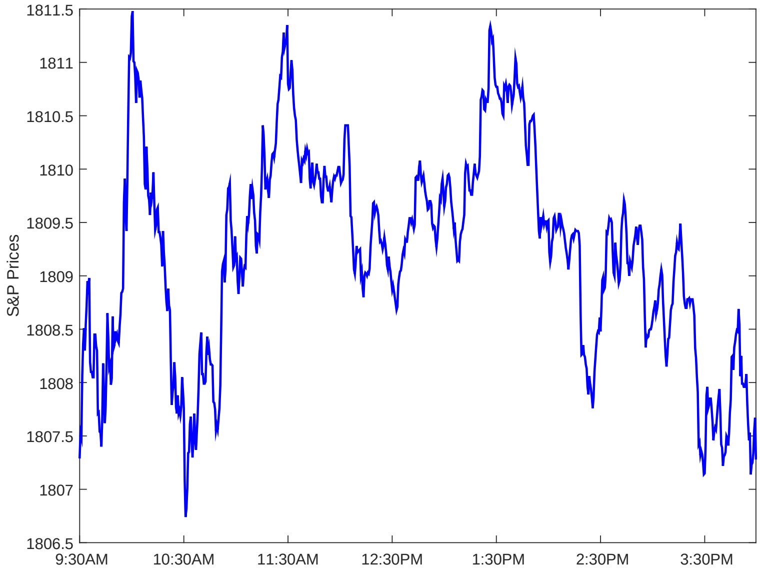


Fig. 3.2 Example of a intraday time series of the S&P 500 index. December 9th 2013.

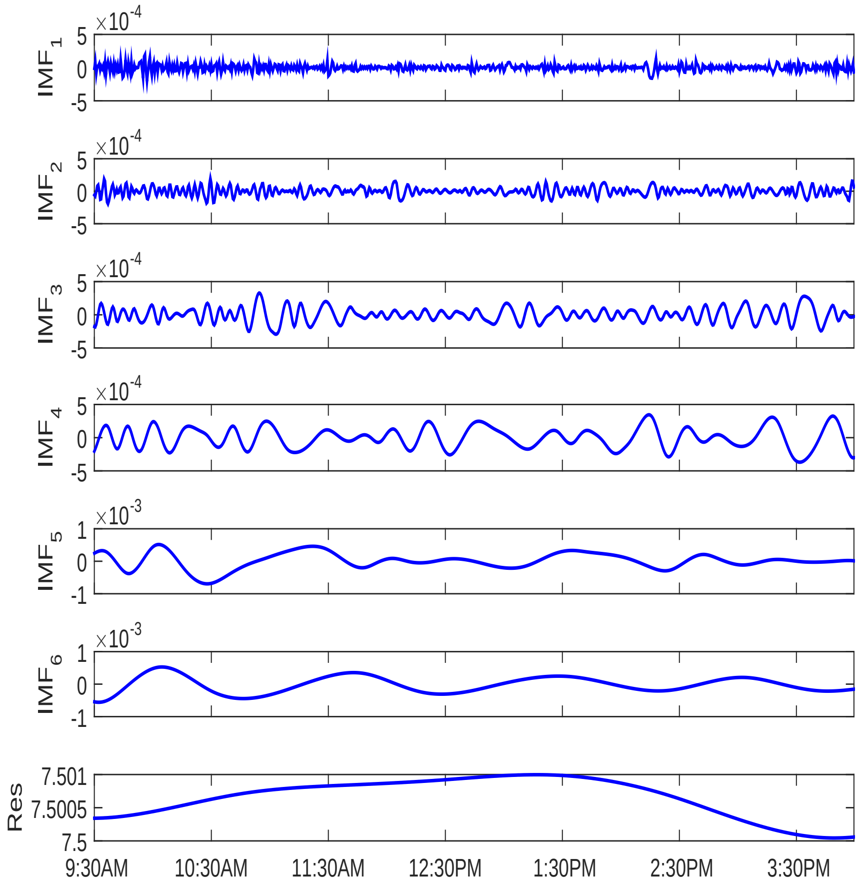


Fig. 3.3 IMFs and residue for the logarithm of the S&P 500 prices taken as an example.

column shows the relative contribution (as a percentage) with respect to the total variance. The residue is defined as the non-oscillating part of the input time series, hence, we do not calculate an oscillating period.

IMF	Period	RV^{EMD}	Variance Contribution
IMF_1	1.70	3.3×10^{-6}	51.8%
IMF_2	3.68	1.6×10^{-6}	24.9%
IMF_3	8.67	8.9×10^{-7}	14.0%
IMF_4	21.67	3.7×10^{-7}	5.8%
IMF_5	43.33	1.5×10^{-7}	2.4%
IMF_6	97.50	6.3×10^{-8}	1.0%
Residue	–	5.3×10^{-9}	0.1%
Total EMD realised variance		6.4×10^{-6}	100%

Table 3.1 Oscillating period expressed in minutes, variance and contribution to the total variance for the IMFs and the residue extracted from the S&P 500 prices taken as an example.

The highest frequency component, IMF_1, has an oscillating period of 1.7 minutes and contributes to the total variance by 51.8%. The second component, IMF_2 has a period of 3.68 minutes and it adds 24.9% to the total variance, etc. The three highest frequency components account for 90% of the total variance.

Before moving to the realised variance analysis of the complete data set, let us briefly illustrate the completeness of the EMD, which guarantees a perfect reconstruction of the input time series. Figure 3.4 shows the reconstruction process. The residue of the time series is represented by a red line. We continue adding the IMFs from the lowest to the highest frequency until we include all the IMFs, this sum is denoted as $R + \sum_{i=1}^{6} IMF_i$. The sum of all the components and the residue completely overlaps with the initial S&P 500 time series.

The decomposition into different frequencies allows to create a partial reconstruction of the input time series, either by creating a denoised version of it or by identifying its time-varying trend. The denoised time series can be obtained by excluding the components with the highest frequencies or by setting a threshold parameter that minimizes the contribution of some IMFs. On the other hand, the trend of a time series is represented by the lowest frequency components, specifically, the residue or the residue plus one or more of the low frequency IMFs.

Subfigure 3.5(a) provides a simple example of a denoised time series that was obtained by excluding the two IMFs with the highest frequencies. Subfigure 3.5(b) illustrates a possible trend for the input time series which is constructed by the residue of the sifting process.

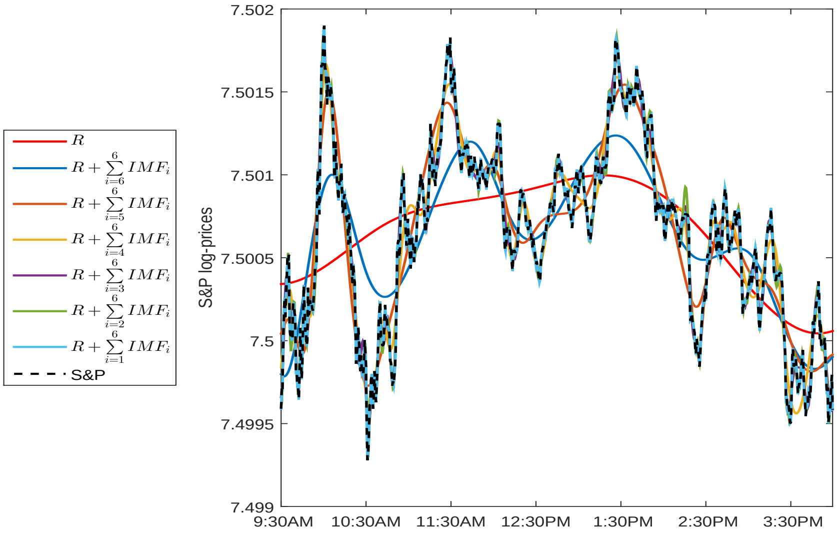
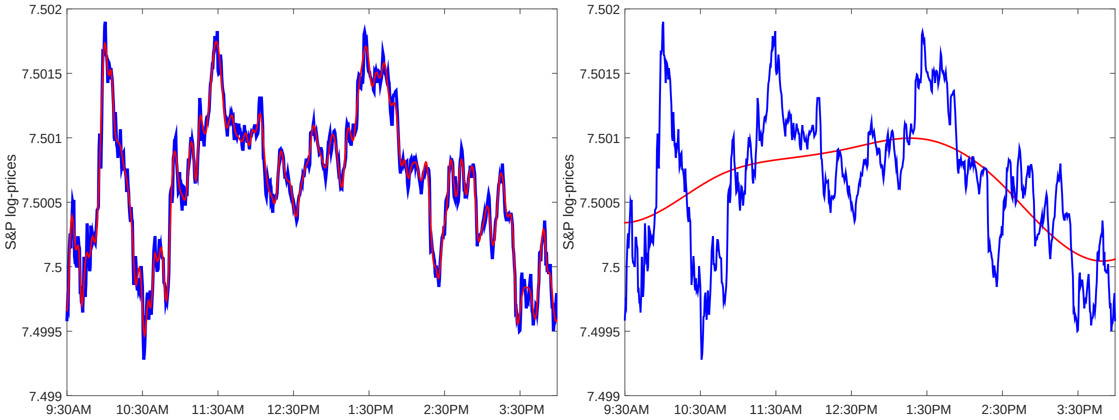


Fig. 3.4 Reconstruction of the logarithm of the S&P 500 prices taken as an example.



(a) Denoised time series obtained by excluding the two highest frequency IMFs.

(b) Trend of the time series obtained by the residue of the EMD.

Fig. 3.5 Partial reconstruction of the logarithm of the S&P 500 prices.

3.3.2 Intraday realised volatility, analysis on the complete data set

We repeated the realised variance analysis described in the previous section on the remaining 177 intraday time series of the S&P 500 index. The average variance contribution of each IMF over the 178 days is reported in Table 3.2(a). The average oscillating period (in minutes) is also reported in this table.

The results reaffirm that the highest frequency components contribute the most to the total realised variance. We compared the EMD variance decomposition against its wavelet counterpart using the MODWT with the Daubechies family of wavelets, specifically the D4 wavelet basis with filter length L=4. We decomposed each time series into the maximum level of decomposition $J_n = 8$. [1]

In Table 3.2(b), we report the average of the variance contribution obtained with the wavelet transform. We also report the scale and oscillating period associated to each level of decomposition $j = 1, 2, \ldots, 8$. We observe differences in the detected oscillating periods identified by the EMD and the wavelet methods. These difference could be attributed to the fundamental nature of the decompositions. The wavelet period is preselected and the EMD period is derived intrinsically from the data.

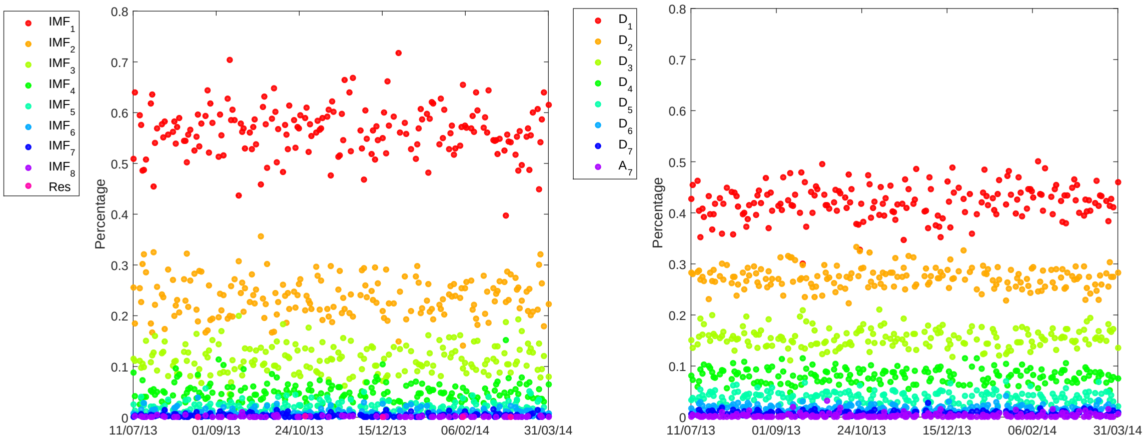
IMF	Period	% Variance
IMF_1	1.7	56.4%
IMF_2	4.1	23.7%
IMF_3	9.3	11.5%
IMF_4	21.4	5.0%
IMF_5	51.3	2.0%
IMF_6	137.9	0.9%
IMF_7	311.5	0.4%
IMF_8	375.4	0.1%
Residue	—	0.04%

(a) EMD variance contribution.

Component	Scale	Period (min)	% Variance
D_1	1	1-2	42.16%
D_2	2	2-4	27.31%
D_3	4	4 -8	15.56%
D_4	8	8-16	8.14%
D_5	16	16-32	4.02%
D_6	32	32-64	1.74%
D_7	64	64-128	0.76%
A_7	128	> 128	0.31%

(b) Wavelet variance contribution.

Table 3.2 Average oscillating period (minutes) and average contribution of components to the total variance.



(a) EMD variance contribution. **(b)** Wavelet variance contribution.

Fig. 3.6 Daily contribution of variance for the period July 2013 to March 2014.

In Figure 3.6(a), we show the graphical representation of the daily variance contribution

[1] The maximum number of possible level of decomposition is bounded by $log_2 N$ and was selected by considering the level for which there exists at least one interior wavelet coefficient, i.e., a coefficient not subject to circular filter operations.

obtained via the EMD. The x axis indicates the analysed date and the y axis the percentage of the contribution for the different IMFs and the residue. The coloured dots represent each IMF. For comparison, Figure 3.6(b) displays the variance distribution obtained using wavelet transform. Both methods agree on the large contribution of the high-frequency components to the total variance. However, the EMD identifies a higher contribution of the fastest component. This difference could be attributed to the difference in timescales. A similar structure in the variance distribution is also observed when decomposing white noise, see reference [182]. In Chaper 5, we continue the analysis of the variance distribution and introduce a method to test the information content of the IMFs obtained from financial time series. Specifically, we compare the variance distribution between the IMFs obtained from Brownian motion and the IMFs obtained from financial time series. This approach identifies the component which do not follow the Brownian motion scaling behaviour.

To further analyse the proposed EMD-based realised variance estimator, we compared the RV^{EMD} estimator, with some other estimators described in section 2.4.1. In Figure 3.7 we report the daily volatility estimation using: Realised volatility, RV, Equation (2.34); two-scale realised volatility[2], $TSRV$, Equation (2.36); wavelet realised volatility RV^{WAV}, Equation (2.37).

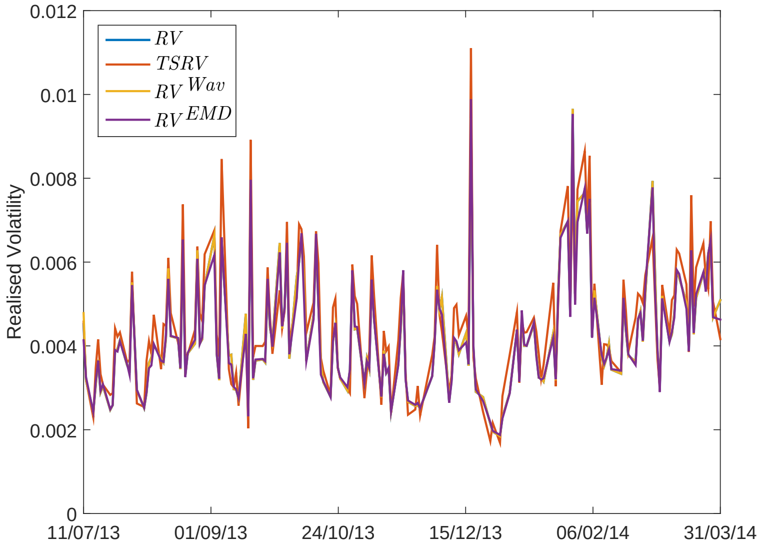


Fig. 3.7 Realised volatility estimators for the S&P 500 index for the period July 2013 to March 2014.

Since the MODWT is an energy preserving transform [145], the RV and the RV^{WAV} coincide. However for the EMD, orthogonality is not guaranteed and some energy leakage can occur [96]. Nevertheless, as one can note from the closeness of the RV^{EMD} estimator to

[2] The $TSRV$ is estimated using a slow timescale of 5 minutes and a fast scale of 30 seconds.

the *RV* estimator, the energy leakage is small. This is confirmed by the index of orthogonality, described in Section 2.2.3, which has an average value (average over the 178 days) of $IO = 2.51 \times 10^{-5}$, corroborating that the IMFs and the residue are close to orthogonal.

The RV^{WAV} and RV^{EMD} estimators are biased and they simply replicate the values of the naive *RV* estimator. The advantage of these estimators is that they provide information about the variance distribution across the oscillating components.

3.3.3 Summary

In this chapter, we proposed a realised volatility estimator based on the EMD. With the scale-by-scale analysis, we are able to identify the time-horizons that lead the intraday variance of the considered S&P 500 time series. The high-frequency oscillations extracted from the EMD are attributed to actions of fast traders, in a similar way that low-frequency oscillations are accredited to longer term investors. Our results demonstrate that the shortest time-horizons (average 1.7 minutes) produce more than 50% of the total variation. In general volatility can be attributed to the fastest investors, indicating that higher volatility is a reflection of faster trading activity.

We compared the EMD against the wavelet decomposition, despite their theoretical differences, both methods agree that the high volatility in financial time series can be attributed to the highest frequency components. The main difference between the methods is that the EMD performs an adaptively time series decomposition with oscillating periods extracted from the data, whereas wavelet transform uses a set of predefined filters.

Chapter 4

EMD-Based Correlation Estimators

The multiscale analyses provided by the EMD allows to study the temporal dependence of financial time series. In this chapter, we propose two approaches to estimate the correlation between a pair of time series. The first approach uses the Pearson correlation to estimate frequency-dependent correlation. The second approach considers a time-dependent correlation based on the rolling-window approach of Chen et al. [48] . The motivation behind the window approach is the assumption that time series are relatively stationary on the timescale of the window length. When applied to high-frequency financial data, the time-dependent estimator captures the intraday correlation and uncovers lead-lag relationships which could be attributed to different levels of trading activity. We apply the proposed estimators to two well known correlated pairs of stock market indices, revealing the time-varying correlation patterns at different frequencies.

4.1 Correlation structures in financial time series

Correlation between financial time series is an important characteristic which describes financial market dynamics. A deep understanding of correlation is of vital relevance for portfolio risk assessment. When estimating correlation, it is necessary to obtain not just a global static measure but a more robust estimator which considers the different timescales and oscillating components latent in a time series. It has been documented that the correlation of stock returns varies over time [116, 153, 170]. Hence, we have to propose time-dependent estimators capable of assessing the dynamic risk exposures. Furthermore, if the degree of correlation between two assets varies across frequencies, short and long-term market participants will have different risk exposures and will analyse different parts of the correlation spectrum.

The study of time-dependent correlation provides some understanding on the collective behaviour of traders with varying strategies [116]. Papadimitriou et al. [138] proposed a time-dependent correlation by comparing the local autocovariance matrices of each time series via its eigenvectors. A rolling-window to localize correlation is proposed, although it is not clear how to choose the appropriate size of the window.

Another technique to measure time-varying correlation is the wavelet coherence [153, 170]. Wavelet analysis allows to study co-movements in the time-frequency domain. The wavelet squared coherence is essentially the ratio of the squared cross-wavelet spectrum to the product of two wavelet spectra. Smoothing the spectral estimates allows the coherence to vary in the range $[0,1]$, with a bias related to the degree of smoothing performed. However, it is not straightforward which type of smoothing should be implemented and whether or not this should be done in both the time and the frequency domains [169].

Recently, Chen et al. [48] proposed to use the EMD to estimate the time-dependent intrinsic correlation (TDIC). In this approach, two time series are first decomposed into IMFs, and the Pearson correlation is calculated in an adaptive window whose length depends on the instantaneous period of the correlated IMFs. In particular, through wavelets and EMD we can assess simultaneously the strength of the comovement at different frequencies and how such strength evolves over time. In this way it is possible to identify regions in the time–frequency space where the comovement is higher are the risk diversification is essential [153].

A common method for estimating the association between two time series is the lagged correlation, the Pearson correlation between two time series shifted in time relative to one another [56]. Lagged correlation is important in studying the relationship between time series since one series may have a delayed response to the other time series or the response of one time series may propagate in time. The existence of lead-lag relations in intraday financial time series have been documented in for example [77, 101]. These authors described how fast one time series reflects new information with respect to the other and concluded that less liquid assets follow the behaviour of more liquid ones.

4.2 Frequency-dependent correlation

Denote by X_t and Y_t, $t = 1,2,\ldots,N$ two time series of equal size N and with equal intervals of time s between observations. The proposed frequency-dependent correlation uses the IMF_i^X, IMF_j^Y, $i, j = 1,\ldots,n$ obtained from the decomposition of the time series X_t and Y_t, respectively.

For the oscillating frequencies i, j, we define the frequency-dependent correlation estimator $\rho_{i,j}$, as the Pearson correlation between IMF_i and IMF_j, specifically:

$$\rho_{i,j} = \frac{1}{N}\sum_{t=1}^{N} \frac{\left(IMF_i^X(t) - \overline{IMF_i^X}\right)\left(IMF_j^Y(t) - \overline{IMF_j^Y}\right)}{\sigma_i^X \sigma_j^Y} \tag{4.1}$$

where $\overline{IMF_i^X}$ denotes the sample mean over time and σ_i^X denotes the standard deviation of IMF_i^X.

Although the IMFs are not theoretically stationary, the IMFs satisfy the condition of having local mean equal to zero and can then be considered to be at least locally stationary [96]. Contrary, the residue does not need to satisfy the IMF conditions, and particularly, for an initial non-stationary time series, the extracted residue will contain the trend of the time series, making it a non-stationary component. Thus, a correlation coefficient between residues is just a measure of linear dependency of the trends indicating if they move in the same direction. This correlation coefficient is likely to be high, and could give misleading results for the interpretation of the dependence structure.

4.3 Time-dependent correlation

We define the time-dependent correlation between a pair of IMFs with the same index (or equivalent frequency) as the Pearson correlation on overlapping windows of size W and lagged by τ observations:

$$\rho_{i,\tau}^T = \frac{1}{W-\tau}\sum_{t=1}^{W-\tau} \frac{\left(IMF_i^X(t) - \overline{IMF_i^X}\right)\left(IMF_i^Y(t+\tau) - \overline{IMF_i^Y}\right)}{\sigma_i^X \sigma_j^Y} \tag{4.2}$$

where $\overline{IMF_i^X}$ denotes the sample mean and σ_i^X and σ_i^X denotes the standard deviation of IMF_i^X over the analysed window.

The time-lag τ is measured in units of the sampling frequency and it is calculated as $\tau = \max\left(P_{X_i}, P_{Y_i}\right)$, with P_{X_i} and P_{X_i} denoting the oscillating period of IMF_i^X and IMF_i^Y, respectively. Choosing τ larger than the oscillating period will result in repetitive patterns in the correlation structure. On the other hand, a shorter time-lag may not reveal any synchronization.

The window size is calculated as $W = \tau \geq 20$ and corresponds to the timescale that we are interested in, but requiring at least 20 observations to calculate a correlation statistic. The window approach has the advantage of only assuming local stationarity rather than stationarity over the entire time series. Although this method is based on a simple measure of correlation (Pearson correlation), it adapts to the nature of the data and provides a measure of correlation in the time frequency space.

4.4 Correlation analysis of high-frequency financial data

For the application of the two proposed correlation measures, we consider intraday data sampled at 30-second intervals for two stock market indices and a volatility index, namely, the S&P 500 index (USA), the IPC index (Mexico) and the VIX index (implied volatility index). The observation period includes 184 days, ranging from September 2014 to July 2015 and it only considers the trading days available for the three indices, see Figure 4.1.

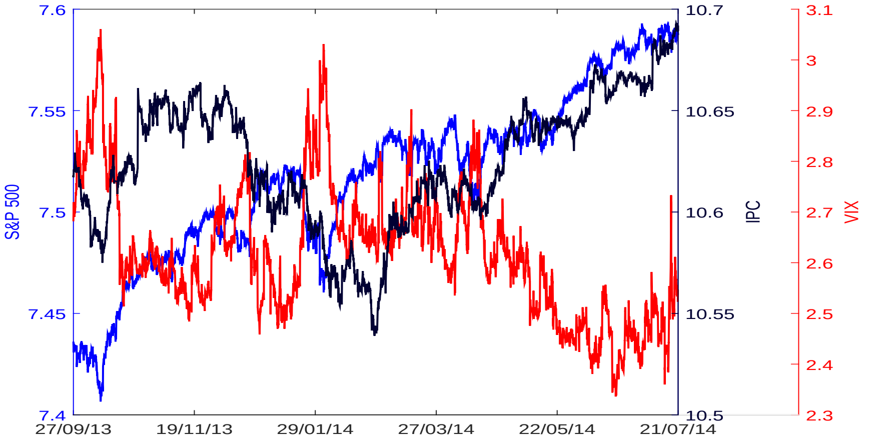


Fig. 4.1 Intraday observations (sampled at 30-second intervals) for the S&P 500, the IPC and the VIX indices for the period September 2014 to July 2015.

We calculated intraday correlation between the S&P 500 and the IPC indices and between the S&P 500 and the VIX indices. As shown in Figure 4.1 and as documented for example in [15, 17], the S&P 500 and the IPC indices are positively correlated. Contrary, the risk-price relationship between the S&P 500 and the VIX indices shows negative correlation [175], a behaviour which could be attributed to the leverage effect [155].

4.4.1 Intraday analysis of correlation, example on a single time series

Let us exemplify the intraday analysis of correlation on a randomly chosen day, July 18^{th} 2014. Figure 4.2 displays the logarithm of prices for the three indices. Applying the EMD to each time series, we obtained five IMFs and a residue, see Figure 4.3. The oscillating period of each IMFs is calculated by dividing the total number of points by the number of peaks, the rounded values are reported in Table 4.1.

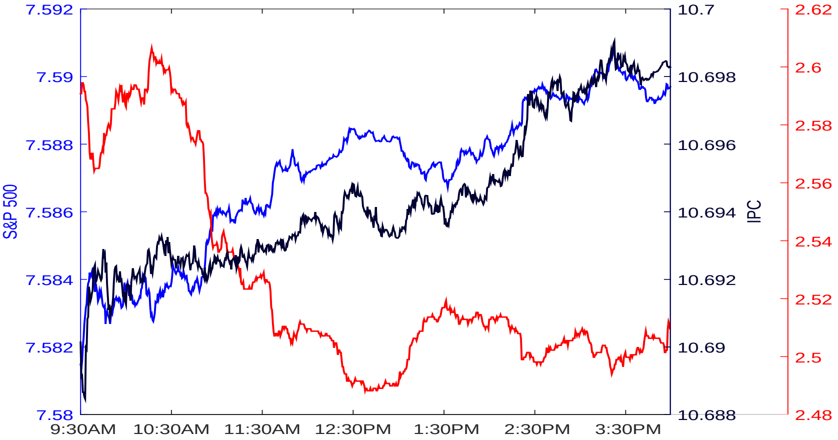
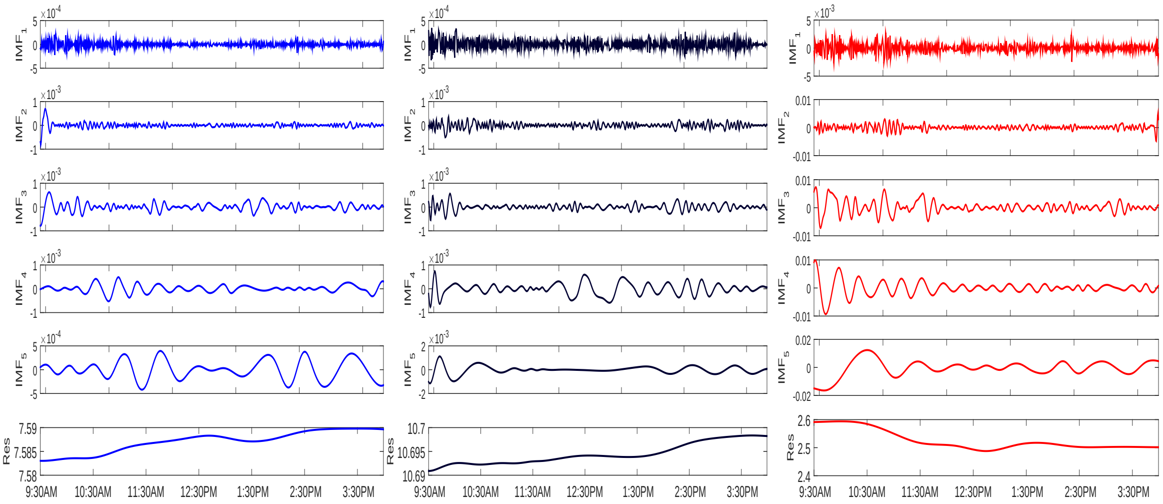


Fig. 4.2 Intraday log-prices for the S&P 500, the IPC and the VIX indices, July 18th 2014.



(a) S&P 500 index.

(b) IPC index.

(c) VIX index.

Fig. 4.3 IMFs of the stock market indices and the volatility index.

Frequency-dependent correlation, example

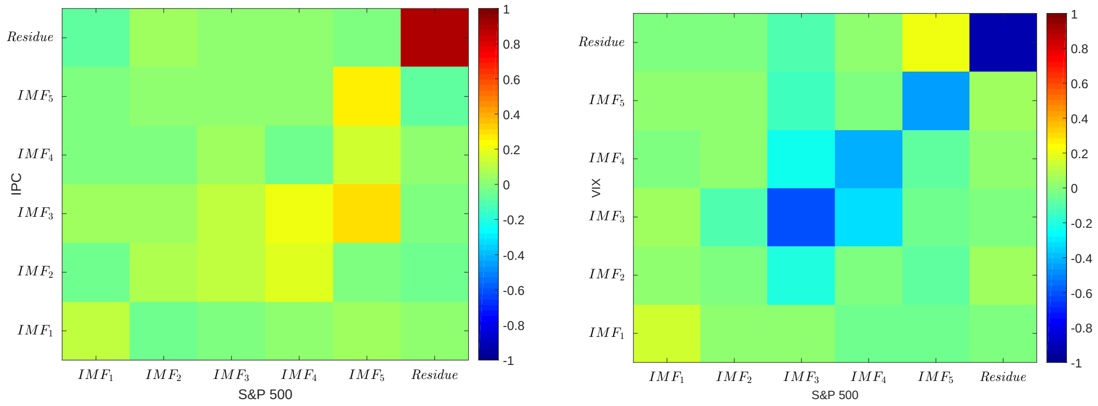
The frequency-dependent correlation obtained from Equation (4.1) can be represented as a matrix of pairwise correlations between the IMFs where the magnitude of the correlation is represented by colours. Figure 4.4(a) shows the correlation matrix between the S&P 500 and the IPC indices. We observe almost zero correlation for the IMFs outside the diagonal, but positive correlation across the diagonal elements whose magnitude decreases as the frequency of the IMF increases, see Figure 4.4(a).

On the other hand, a negative correlation between the S&P and the VIX indices is obtained for the diagonal IMFs, see Figure 4.4(b). We also observe that the correlation decreases as the frequency of the IMFs increases. In this way, although previous time series

Index	***IMF*$_1$**	***IMF*$_2$**	***IMF*$_3$**	***IMF*$_4$**	***IMF*$_5$**	**Residue**
S&P	4	8	20	44	88	–
IPC	4	8	16	40	88	–
VIX	4	8	20	40	88	–

Table 4.1 Oscillating period for the IMFs shown in Figure 4.3 and calculated by dividing the total number of points by the number of peaks.

are known to be highly correlated, they are less correlated at higher frequencies.



(a) S&P 500 index versus IPC index.

(b) S&P 500 index versus VIX index.

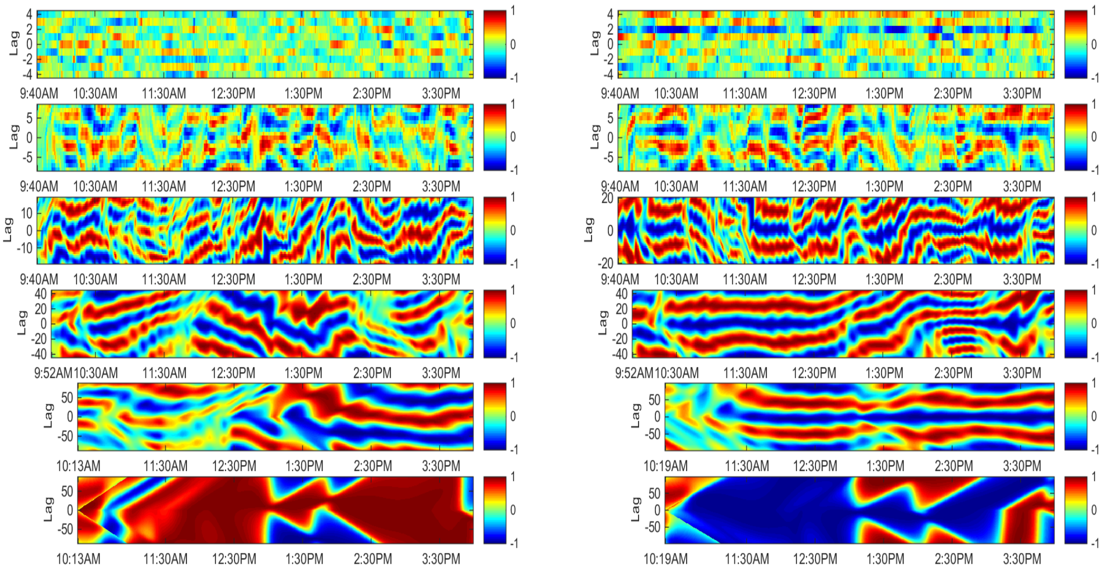
Fig. 4.4 Frequency-dependent correlation.

Time-dependent correlation, example

We also estimate the time-dependent correlation between the S&P 500 and the IPC indices using Equation (4.2), see Figure 4.5(a). The correlation is represented as a coloured matrix in which each column represents a successive window and each row represents a specific time-lag. The magnitude of the correlation is indicated by colours. The intraday correlation values are reported after W observations, with W the size of the rolling-window. In this way, the size of the correlation matrix is reduced according to the applied window.

From Figure 4.5(a), it is difficult to identify correlations patterns for the highest frequency IMFs. However, for IMFs with lower frequency, $IMF_2,\dots,IMF_5$, we observe intervals of stronger correlation characterized by the nature of the oscillating IMFs, i.e, we observe lapses of positive correlation lagged in time by negative values of correlation, making the lead-lag relation between the IMFs almost symmetric with respect to the zero lag.

Figure 4.5(b) shows the correlation matrices for the S&P 500 and the VIX indices. Contrary to the correlation between the S&P 500 and the IPC indices, the correlation between the S&P and the VIX indices is negative at all frequencies and during the entire trading day. At the highest frequency, IMF_1, we observe a clear pattern of negative correlation at lag $\tau = 2$ (1 min), indicating that the S&P 500 leads the VIX index by 1 minute. When correlating the residue components, we observe a dominant blue band (a similar red band is observed for the correlation between the S&P 500 and the IPC indices) which could be attributed to the linear and non-stationary characteristics of the residues.



(a) S&P 500 index versus IPC index.

(b) S&P 500 index versus VIX index.

Fig. 4.5 Intraday time-dependent correlation.

4.4.2 Intraday correlation, analysis on the complete data set

In order to identify clearer patterns in the intraday correlation, we estimated the frequency-dependent correlation and the time-dependent correlation for each of the 184 days available in the data set. We decomposed each daily time series into five IMFs and a residue. In the following sections, we report the average patterns of the proposed estimators.

Average frequency-dependent correlation

We report the frequency dependent correlation for each of the trading days using histograms and only considering the diagonal elements of the correlation matrix, i.e., the IMFs with the same sub-index and which also display the highest correlation. Figure 4.6 shows the histogram for the frequency-dependent correlation between the S&P 500 and the IPC indices.

For most of the IMF pairs, we observe distributions centred around zero and correlations statistically non different from zero, except for the residue function which is highly skewed, given its non-stationarity and the fact that at values close to one the distribution of the correlation is generally skewed [78]. The phenomenon of observing smaller correlation at higher frequencies is well documented and known as the Epps effect [69]. It could be attributed to microstructure effects that are more evident at higher frequencies, including lagged correlations, asynchronous trading and decimalization of the tick-size (lowest possible price change) [35, 150].

Figure 4.7 shows the corresponding histograms for the correlation between the S&P 500 and the VIX indices. We observe a negative correlation which depends on the analysed frequency. There are two popular theories associated with the reported negative return - risk correlation, namely the leverage hypothesis and the volatility feedback hypothesis [35]. The leverage effect occurs when the value of a firm drops, then the debt-to-equity radio increases. Since the assets of the firm restrain the risk of the firm, the volatility of the equity increases. The other major explanation for the return-risk relation considers a time varying risk premium, or volatility feedback effect where an anticipated increase of future volatility by market operators triggers sell orders, which therefore decreases the asset price [34].

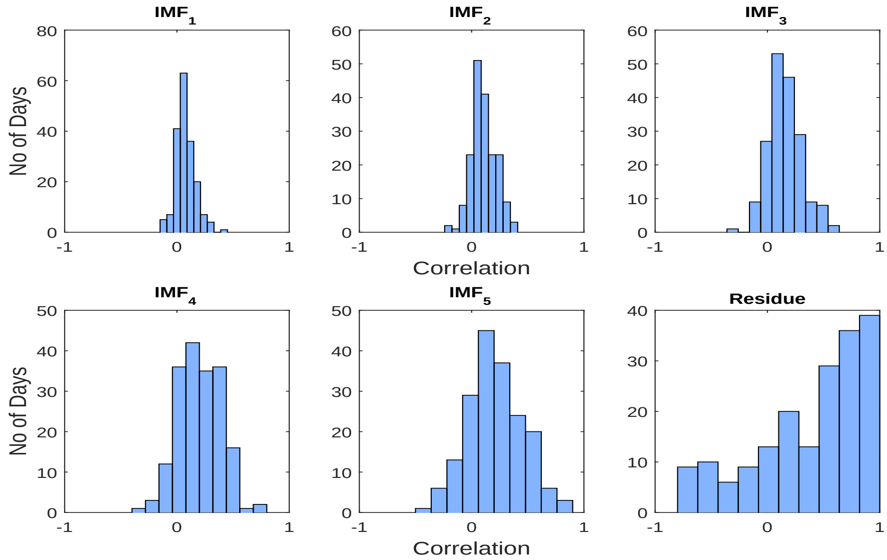


Fig. 4.6 Distribution of the frequency-dependent correlation between the IMFs of the S&P 500 index and the IMFs of the IPC index.

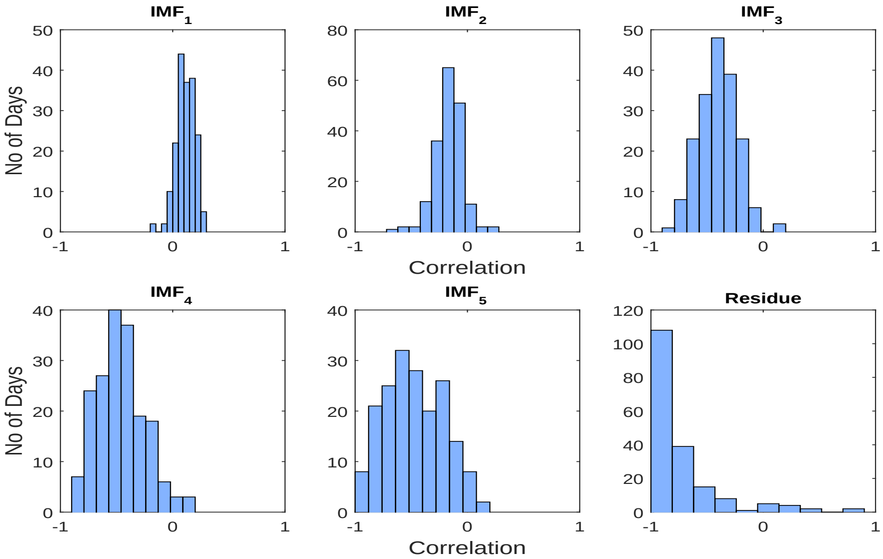


Fig. 4.7 Distribution of the frequency-dependent correlation between the IMFs of the S&P 500 index and the IMFs of the VIX index.

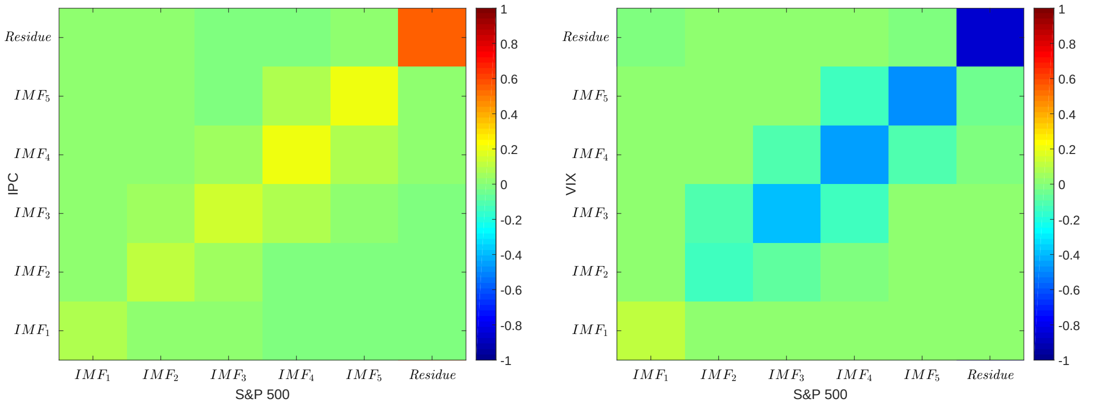
To analyse the average behaviour of the frequency dependent correlation, we propose to use the sample median of the distribution since this statistic is not influenced by outliers. In Figure 4.8, we present the sample median correlation matrix for the analysed time series. The case S&P and IPC is reported in Figure 4.8(a), where we observe an almost zero correlation for the non-diagonal elements. And although from the histograms displayed in Figure 4.6, we observe that for most of the IMFs the correlation is statistically non different from zero, the correlation for the residue is indeed more significant. We observe the same pattern for the S&P and the VIX correlation with non-homogeneous correlations across different frequencies, as reported in other studies including [118, 153, 170].

Average time-dependent correlation

We analysed the median of the time-dependent correlation matrices. In order to have compatible correlation intervals, each intraday matrix was calculated using the same value for the window size and the time-lag. The applied values were computed as the sample mean of the parameters over the 184 days. The obtained values are reported in Table 4.2.

The median correlation matrix between the S&P 500 and the IPC indices is displayed in Figure 4.9(a). This matrix shows clearer patterns for the intraday dynamics, it does not show lead-lag relationships between the IMFs, but it emphasizes some stronger patterns of correlation at the beginning and at the end of the trading session.

For the correlation between the S&P 500 and the VIX indices, Figure 4.9(b), we found stronger correlations at high frequencies (IMF_1, IMF_2 and IMF_3) which are lagged by one



(a) S&P 500 index and the IPC index. **(b)** S&P 500 index and the VIX index.

Fig. 4.8 Sample median of the frequency-dependent correlation matrices over the period September 2014 to July 2015.

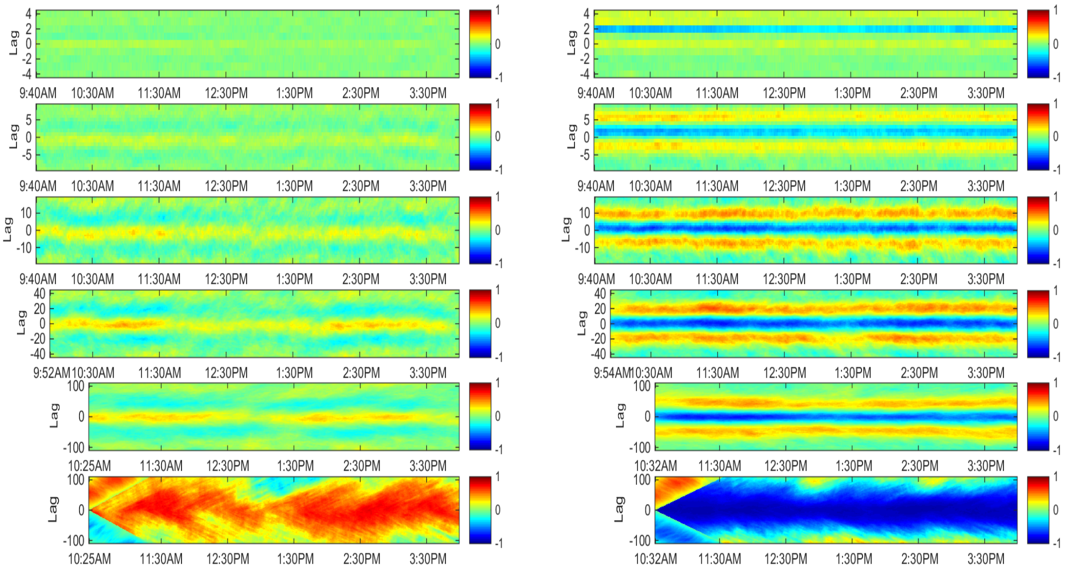
Component	**S&P vs IPC**		**S&P vs VIX**	
	Lag	**Window**	**Lag**	**Window**
$\boldsymbol{IMF}_1$	4	20	4	20
$\boldsymbol{IMF}_2$	9	20	9	20
$\boldsymbol{IMF}_3$	19	20	21	21
$\boldsymbol{IMF}_4$	44	44	48	48
$\boldsymbol{IMF}_5$	110	110	124	124
Residue	110	110	124	124

Table 4.2 Average of the number of lags and the size of the rolling-window used for the time-dependent correlation analysis.

minute ($\tau = 2$) and are stable across the day. Our results show that at higher frequencies, changes of the S&P 500 index are more likely to lead changes of the VIX index.

4.5 Summary

The multiscale analyses provided by the EMD allows to study the dynamic correlation between two non-stationary time series. In this chapter, we proposed two approaches to estimate the frequency and time dependent correlation. With the proposed estimators, we studied the intraday co-movements between two pairs of time series. Our empirical findings confirmed the positive correlation between the S&P 500 index and the IPC index [15], and the negative correlation between the risk-price relation between the S&P 500 index and the VIX index [75, 92]. More importantly, the results show that the correlation patterns are time and frequency dependent. Similarly to the coherence measure obtained using wavelet trans-



(a) S&P 500 index versus IPC index.

(b) S&P 500 index versus VIX index.

Fig. 4.9 Sample median of the time-dependent correlation matrices over the period September 2014 to July 2015.

form [170], our proposed correlation measures could be used to analyse transient dynamics of a time series and distinguish between long-run trends and cycles. Given the adaptiveness and the simplicity of the EMD, the proposed correlation measures offer a more comprehensible analysis.

In order to generalize the results for our intraday analysis, we considered the distribution of correlations and the sample average over the analysed days. The time-dependent correlation shows clearer patterns for the intraday dynamics of correlation. For the correlation between the S&P and the VIX indices, we found strong correlations at high frequencies that are lagged by one minute and that are stable across the day. Our results show that at higher frequencies, the change of S&P 500 index is more likely to lead the change of the VIX index.

The lead-lag analysis produces more insight into the dynamics of co-movements and could be used in forecasting models, for example in regression models. It offers a better understanding about the speed of different assets processing and reflecting information, and the degree to which the information contained in one time series could be used to make predictions on the other.

Chapter 5

Anomalous Volatility Scaling in Stock Market Indices

In this chapter, we study the scaling properties of intraday stock market indices computed at various time horizons. We demonstrate that when EMD is applied to fractional Brownian motion and to α-stable Lévy motion (a heavy-tailed stochastic process), we retrieve a scaling law that relates the variance of the components to a power law of the oscillating period. The obtained scaling exponent is comparable the long-range dependence exponent obtained with the wavelet methodology. In contrast, when analysing 22 different stock market indices, we observe deviations from the fractional Brownian motion and Brownian motion scaling behaviour. Part of this chapter is published in the paper "Anomalous volatility scaling in high-frequency financial data" [137].

5.1 Self-similarity and long-range dependence in financial time series

The concepts of self-similarity, scaling behaviour, fractional processes and long memory have been widely used to describe properties of financial time series [54]. The idea that stock returns could exhibit long-range dependence was first suggested by Mandelbrot [124], who observed self-similar behaviour in the logarithm of returns and proposed the use of stable distributions to model them [124]. Empirical finance research has emphasized the presence of long memory in stock markets, which is observed across various time periods and in different markets, FX [86, 134], equity [11, 178], stock indices [64, 65]. The autocorrelation of returns is typically insignificant at lags between a few minutes and days, however at higher frequencies, some autocorrelation patterns can be identified and are sometimes attributed to the micro-structure noise latent in high-frequency financial data [54]. Contrary, squared returns or absolute returns are highly autocorrelated processes exhibiting long-range dependence [53].

Another argument is whether the observed long-memory property is constant over time. Time-dependent scaling behaviour has been reported in for example [8, 32], the local variations of roughness can be described by allowing the Hurst exponent to vary with time [39]. Carbone et al., [43] calculated the local scaling exponent over partially overlapping subsets of the analysed time series, finding a much more pronounced time-variability in the local scaling exponent of financial time series than in Brownian motion.

Entropy has been proposed as an alternative measure to capture not just linear dependencies in financial time series [60] and it is also consider a measure of complexity [109, 191]. In general, a low value of entropy indicates the presence of more predictable patterns which are therefore associated with periods of financial inefficiency. Conversely, when the time series exhibit more irregular and less predictable patterns, the uncertainty level is higher and such periods are described by larger values of entropy.

5.2 EMD-based scaling exponent

Although the EMD is a completely adaptive technique which makes no assumptions of the true data generating process, when applied to Gaussian noise, the EMD acts as a wavelet filter bank able to isolate different frequency bands [80]. Flandrin et al. [79] empirically showed that when decomposing fractional Gaussian noise (fGn), the EMD could be used to estimate its scaling exponent H, if $H > \frac{1}{2}$. The authors ascertained that the variance progression across the IMFs satisfies, $\mathrm{Var}(IMF_k^{fGn}) \propto \tau_k^{2(H-1)}$, where the function τ_k denotes the period of the k^{th} IMF [1].

In this chapter, we follow a similar approach to [79], but instead of applying the EMD to FGN, we consider its integrated process, FBM. We empirically show that a similar scaling law holds for the variance of the IMFs:

$$\mathrm{Var}(IMF_k^{FBM}) \propto \tau_k^{2H}. \tag{5.1}$$

The EMD estimator of H can be determined by the slope of a linear regression fit on the logarithm of the variance as a function of the logarithm of the period

$$\log\left(\mathrm{Var}(IMF_k^{FBM})\right) = 2H\log(\tau_k) + \log(c_0), \tag{5.2}$$

[1] The oscillating periods τ_k can be approximated as the total number of data points divided by the total number of zero crossings of each IMF.

where c_0 is the intercept constant of the linear regression.

To visualize the linear relationship of Equation (5.2), we explicitly show the relation between $\log(\text{Var}(IMF_k^{FBM}))$ and $\log(\tau_k)$ for an FBM simulation of scaling exponent $H = 0.6$ and length $N_1 = 10,000$ points, see Figure 5.1. In this example, the resulting estimator is $H^* = 0.593$ which accurately approximates the scaling exponent of the simulated process.

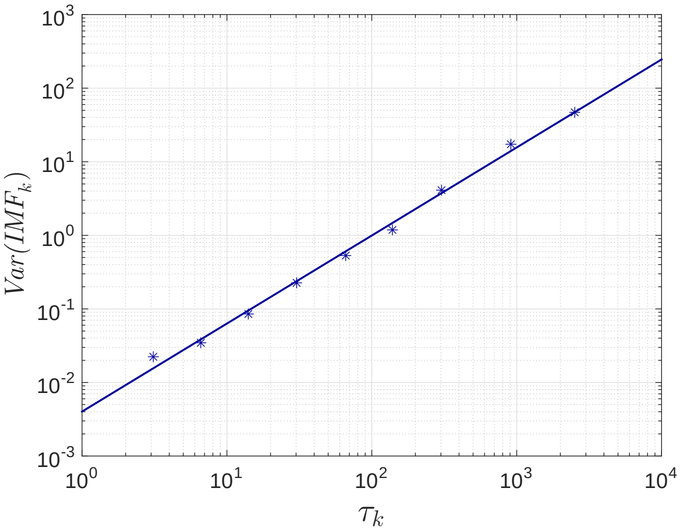


Fig. 5.1 Log-log plot of the IMF variance as a function of period for an FBM of $H = 0.6$ and length $N_1 = 10,000$. The blue line represents the least-square fit. The scaling exponent $H^* = 0.593$ can be recovered from half the slope of the least-square linear fit.

5.3 Numerical study of long-memory processes

In order to verify the filter bank properties of the EMD and the scaling law of Equation (5.1), we extended the simulation set of the FBM and we also considered α-stable Lévy motion (SLM), a self-similar process with independent increments and heavy-tailed distribution, for more details about these processes refer to Section 2.6.2. By estimating a scaling exponent for models with known scaling laws, we demonstrate that our measure varies consistently around the expected values set in the models.

5.3.1 FBM simulation analysis

We generated $M = 100$ FBM paths for the following values of the scaling exponent $H = 0.1, 0.2, \ldots, 0.9$. The simulated processes have two different lengths, $N_1 = 10,000$ and $N_2 =$

$100,000.$[2]

We applied the EMD to each FBM simulation and we calculated its respective H^* exponent. In Table 5.1, we report $\langle H^* \rangle$, the mean over the 100 estimators. We also report the root mean square error (RMSE) of the estimators, $RMSE = \sqrt{\frac{\sum_{i=1}^{M}\left(H_i^* - H\right)^2}{M}}$. We observe that the longer the analysed time series, the better the estimation of H is. For length $N_2 = 100,000$, $\langle H^* \rangle$ is indeed very close to the scaling exponent H (for all values of H). In Figure 5.2, we plot the mean values of the H^* exponent as presented in Table 5.1. The error bars represent the RMSE of the estimator. For comparison, we also estimated the Hurst exponent using the generalized exponent approach with $q = 2$ [63], obtaining consistent results. Table 5.1 reports the mean and the RMSE of this estimator which is denoted as H_G.

	10,000				**100,000**			
H	$\langle H^* \rangle$	$RMSE_{H^*}$	$\langle H_G \rangle$	$RMSE_{H_G}$	$\langle H^* \rangle$	$RMSE_{H^*}$	$\langle H_G \rangle$	$RMSE_{H_G}$
0.1	0.05	0.06	0.15	0.05	0.11	0.03	0.15	0.05
0.2	0.15	0.07	0.22	0.03	0.21	0.03	0.22	0.03
0.3	0.26	0.06	0.31	0.01	0.31	0.03	0.31	0.01
0.4	0.38	0.05	0.40	0.01	0.41	0.02	0.40	0.01
0.5	0.49	0.05	0.50	0.01	0.51	0.03	0.50	0.01
0.6	0.59	0.04	0.60	0.01	0.60	0.03	0.60	0.01
0.7	0.70	0.04	0.70	0.01	0.69	0.03	0.70	0.01
0.8	0.80	0.04	0.79	0.01	0.78	0.04	0.79	0.02
0.9	0.90	0.05	0.88	0.03	0.87	0.04	0.87	0.03

Table 5.1 Confirmation that the empirical scaling law of Eq. 3 retrieves the expected scaling exponent for FBM. Mean of the scaling exponent H^* over 100 simulations of FBM with parameter $H = 0.1, 0.2, \ldots, 0.9$ and length, left: $N_1 = 10,000$ and right: $N_2 = 100,000$. For comparison, we included the mean and the RMSE of the generalized Hurst exponent estimator with $q = 2$ and denoted as H_G.

5.3.2 α-stable Lévy motion simulation analysis

To perform the simulation analysis, we generated stable random variates using the Chambers algorithm [104] and from these we construct sample trajectories of the SLM. We used the toolbox provided by [164], sample paths are of length $N = 10,384$, with parameters for the generation $m = 128$ and $L = 6000$, making $m(L+N)$ to be a power of 2, see [164] for more details. We considered the case $H = 1/\alpha$ for values of $H = 0.5, 0.55, \ldots, 0.95$.

In Table 5.2, we report $\langle H^* \rangle$, the mean over the 100 estimators. We also report the RMSE of these estimators. We observe that for all the values of $H = 1/\alpha$, the mean of H^* consistently estimates the independence of the increments, providing values around 0.5. For

[2] All the FBM paths were generated using MATLAB® wavelet toolbox.

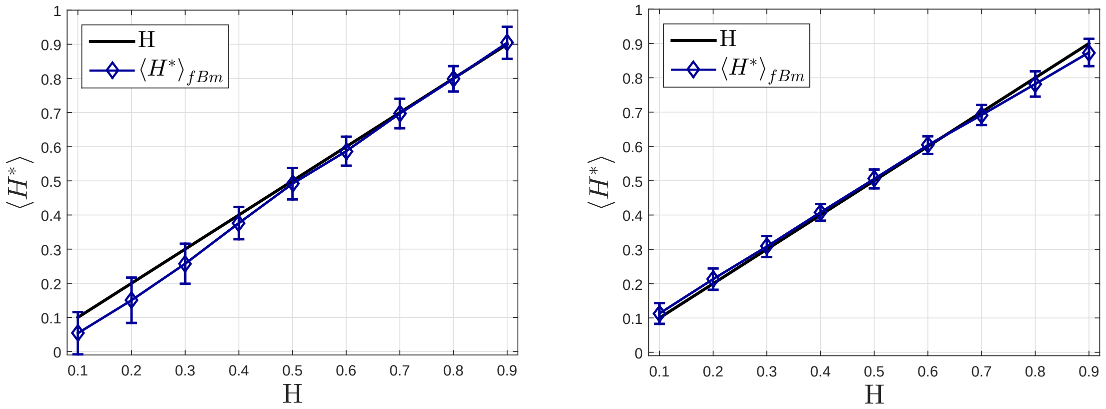


Fig. 5.2 Demonstration that the empirical scaling law of Eq. 3 retrieves the expected scaling exponent for FBM. Mean of the scaling exponent H^* over 100 simulations of FBM with parameter $H = 0.1, 0.2, \ldots, 0.9$ and length, left: $N_1 = 10,000$ and right: $N_2 = 100,000$. The error bars denote the RMSE of the estimator.

comparison, we also estimated the Hurst exponent using the generalized exponent approach [63] with $q = 2$, obtaining consistent results. In Table 5.2, we include the mean and the RMSE of this estimator denoted as H_G.

Let us emphasize that we do not propose the EMD as a way to estimate the Hurst exponent that can be more accurately estimated by other methods like the ones reported in [20, 106]. The advantage of the proposed estimator is its ability to analyse the interactions between the different timescales latent in the data.

H	$\langle H^* \rangle$	$RMSE_{H^*}$	$\langle H_G \rangle$	$RMSE_{H_G}$
0.50	0.49	0.05	0.50	0.01
0.55	0.48	0.05	0.50	0.01
0.60	0.48	0.05	0.50	0.01
0.65	0.48	0.05	0.50	0.01
0.70	0.48	0.06	0.50	0.02
0.75	0.48	0.06	0.50	0.01
0.80	0.48	0.06	0.50	0.01
0.85	0.48	0.07	0.50	0.01
0.90	0.47	0.07	0.50	0.01
0.95	0.46	0.07	0.50	0.01

Table 5.2 Demonstration that the empirical scaling law of Eq. 3 retrieves the expected scaling exponent for SLM. Mean of the scaling exponent H^* over 100 simulations of SLM with parameter $H = \frac{1}{\alpha} = 0.5, 0.55, \ldots, 0.95$ and length $N = 10,384$.

5.4 Variance scaling in intraday financial data

We analysed intraday prices for 22 different stock market indices. The data set covers a period of 6 months from May 5th, 2014 to November 5th, 2014. Prices are recorded at 30 second intervals. The list of analysed stock market indices is reported in Table 5.3.

Country	Index	Length
Brazil	BOVESPA	105,000
China	SSE	60,480
France	CAC 40	136,080
Greece	ASE	106,470
Hong Kong	HSI	98,154
Hungary	BUX	122,880
Italy	FTSE MIB	133,056
Japan	NIKKEI 225	75,600
Malaysia	KLSE	115,320
Mexico	IPC	100,620
Netherlands	AEX	130,680
Poland	WIG	64,680
Qatar	DSM	52,080
Russia	RTSI	133,120
Singapore	STI	123,840
South Africa	JSE	117,500
Spain	IBEX	135,527
Turkey	XU 100	91,760
UAE	UAED	60,000
UK	FTSE	130,560
USA	S&P 500	99,840
USA	NASDAQ	100,620

Table 5.3 Studied stock market indices, including the length of the time series.

We applied the EMD to the logarithm of each stock market index. For the sake of clarity, in this section we only focus on the decomposition of the S&P 500 index, but a similar analysis has been done for the other stock market indices. For the S&P 500 log-price time series, we extracted 17 IMFs and a residue which describe the local cyclical variability of the original signal and represent it at different timescales. The original log-price time series and its IMFs are displayed in Figure 5.3. In this figure, we observe temporary clusters of volatility that characterize some of the components, for example the high volatility at the end of the time series can evidently be seen in components 2,3,4,6 and 7.

The IMF periods, calculated as the total number of data points divided by the total number of zero crossings, are reported in Table 5.4. These periods are converted into minutes, hours and days. The fastest component has a cycle of 1.6 minutes, contrasting the slowest cycle of 11.6 days. Notice that the first 12 IMFs represent the intraday activity (6.5 hours of trading), while the remaining IMFs (from 13$^{\text{th}}$ to 17$^{\text{th}}$) are associated with the inter-day cycles. The last component is the residue of the EMD.

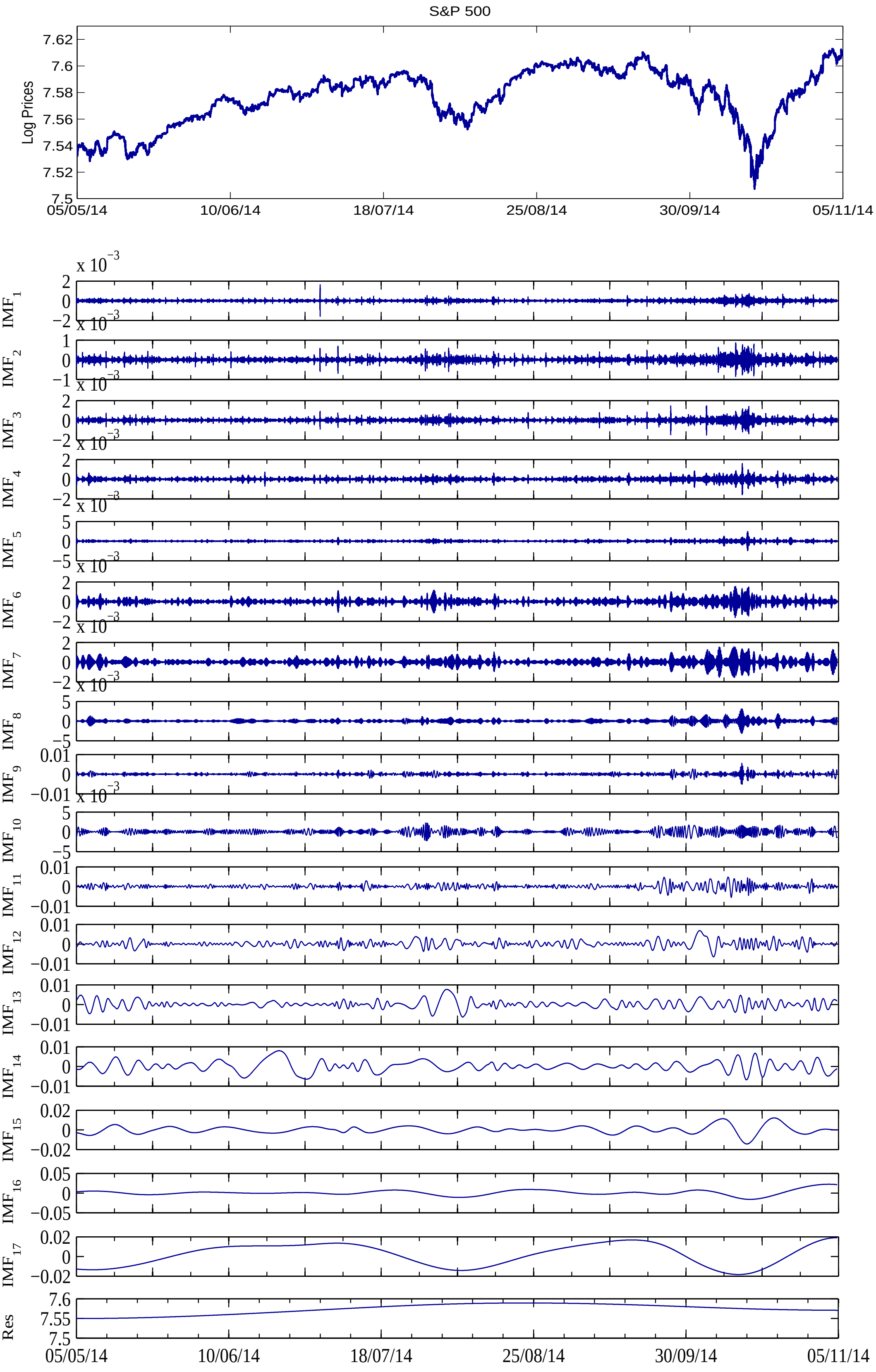


Fig. 5.3 Top: log-price time series of the S&P 500 index for the period 05/05/2014 to 05/11/2014. Bottom: the 17 IMFs and the residue obtained through EMD of the log-prices.

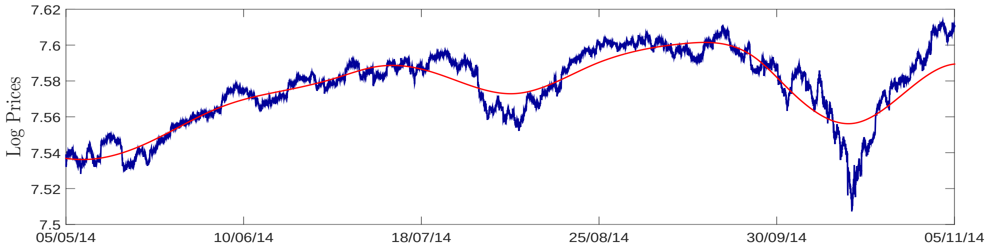


Fig. 5.4 Log-price time series of the S&P 500 index (blue line). The red line represents the 'trend' of the data calculated as the sum of the residue plus the last IMF.

IMF	**Period/min**	***IMF***	**Period/hr**	***IMF***	**Period/days**
1	1.6	**9**	1.1	**14**	1.1
2	2.8	**10**	1.9	**15**	2.2
3	4.9	**11**	3.0	**16**	4.3
4	8.4	**12**	5.9	**17**	11.6
5	13.0	**13**	11.7	**18**	Residue
6	19.3				
7	28.8				
8	41.7				

Table 5.4 Oscillating period of the IMFs obtained from the S&P 500 index.

The overall trend of the time series is given by the residue, and each component can be seen as an oscillating trend of the previous component on a shorter timescale. The effectiveness of EMD as a detrending and smoothing tool is illustrated in Figure 5.4. In this Figure, the original time series (blue line) is compared with a 'trend' (red line), calculated as the sum of the residue plus the last component.

In previous section we discussed that for FBM, the EMD produces a linear relationship between the logarithm of the variance of the IMFs and their respective period of oscillation (Equation (5.2)). We tested whether this relationship also holds for financial time series. In Figure 5.5, we show the log-log plot of the variance as a function of the period for the IMFs obtained from the S&P 500 index (red diamonds). The estimated scaling exponent has a value of $H^* = 0.55$. The goodness of the linear fit was estimated by the coefficient of determination[3] which is $R^2 = 0.992$. We can conclude that this stock market index satisfies the linear relationship of Equation (5.2).

We performed the same analysis for the other stock market indices, finding both signif-

[3] This coefficient of determination is the square of correlation between the dependent and independent variable. Values of this coefficient range from 0 to 1, with 1 indicating a perfect fit between the data and the linear model, see for example [149].

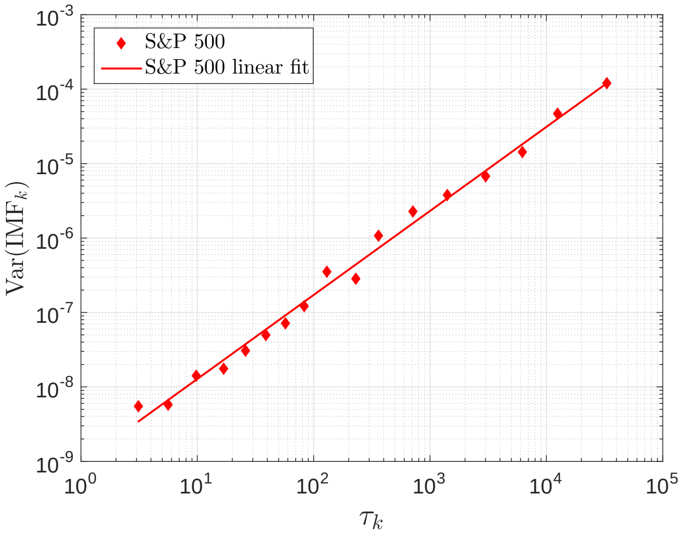


Fig. 5.5 Log-log plot of the IMF variance as a function of period for the EMD of the S&P 500 index. The red line represents the best least-square fit. The goodness of the linear fit is $R^2 = 0.992$.

icant deviations from Brownian motion ($H^* \neq 0.5$) and deviations from the scaling law of Equation (5.2). We note that given the heuristic nature of our proposed EMD estimator, values of $H^* \neq 0.5$ do not necessarily imply deviations from Brownian motion. It is necessary to consider the standard errors of the estimator. As reported in Table 5.1, the RMSE of the EMD estimator could be as high as 0.3. In the following section we provide a test using some confidence interval for the Brownian motion scaling behaviour. In this way, we can identify which stock market indices deviate the most from the Browniam motion scaling behaviour.

In Table 5.5, we report further details about the decomposition and scaling law found on each stock market index. We include the number of IMFs and the index of orthogonality which is described by Equation (2.18). We observe small values of the IO, indicating an almost orthogonal decomposition. Furthermore, we report the estimated scaling exponent and the goodness of the linear fit. Although, we note that the coefficient of determination for all the stock market indices is above 0.94, we shall discuss shortly that significant deviations from linearity (FBM behaviour) are observed, especially in less developed markets. We also note that the S&P 500 index follows more closely the scaling properties of FBM.

Deviations from Brownian motion

Let us now discuss the deviations of the scaling laws found in stock market indices from the scaling expected in Brownian motion (BM). With this aim, we generated $M = 100$ paths of

Index	**# IMFs**	**IO** $\times 10^4$	R^2	H^*
S&P 500	17	3.7	0.992	0.564
BOVESPA	18	6.3	0.989	0.561
FTSE MIB	18	8.6	0.987	0.571
XU 100	19	3.5	0.985	0.563
RTS	20	11	0.985	0.581
CAC 40	17	8.8	0.984	0.564
UAED	16	6.3	0.978	0.616
FTSE	23	2.9	0.977	0.529
ASE	15	29	0.974	0.587
IBEX	18	5.7	0.973	0.531
WIG	16	8.2	0.973	0.591
SSE	14	1.6	0.971	0.534
DSM	18	7.6	0.971	0.618
IPC	18	0.68	0.971	0.555
BUX	19	4.2	0.970	0.542
HSI	19	1.0	0.969	0.554
AEX	21	13	0.968	0.558
NASDAQ	20	5.1	0.960	0.530
NIKKEI 225	22	8.4	0.959	0.544
JSE	19	2.9	0.956	0.518
KLSE	22	0.77	0.943	0.540
STI	21	2.6	0.942	0.522

Table 5.5 Stock market indices including the number of IMFs obtained when applying EMD to the logarithm of the price. The second column report the index of orthogonality ($\times 10^4$). Stock market indices are reported in descending order of R^2, which represents the goodness of the linear fit of Equation (5.2). Last column reports the estimated exponent H^* of the same equation.

BM with length N equal to the analysed stock market index (see Table 5.3). We applied the EMD to each simulation and obtained its respective intrinsic oscillations denoted as $IMF_k^{BM_i}$, $i = 1,2,\dots,100$, $k = 1,2,\dots,n_i$, with n_i the number of IMFs for each BM simulation. In order to compare the variance of the IMFs extracted from the financial index X, $\mathrm{Var}(IMF_k^X)$, against the $\mathrm{Var}(IMF_k^{BM_i})$, we rescaled the latter as:

$$\widehat{\mathrm{Var}(IMF_k^{BM_i})} = c_i \,\mathrm{Var}(IMF_k^{BM_i}), \tag{5.3}$$

where

$$c_i = \frac{\frac{1}{n}\sum\limits_{k=1}^{n} (\mathrm{Var}(IMF_k^X)/\tau_k^X)}{\frac{1}{n_i}\sum\limits_{k=1}^{n_i} (\mathrm{Var}(IMF_k^{BM_i})/\tau_k^{BM_i})}. \tag{5.4}$$

In Figure 5.6, we present all the 100 linear fits as light blue lines. In the same figure, we plotted the variance of the IMFs extracted from the S&P 500 index, same as reported in

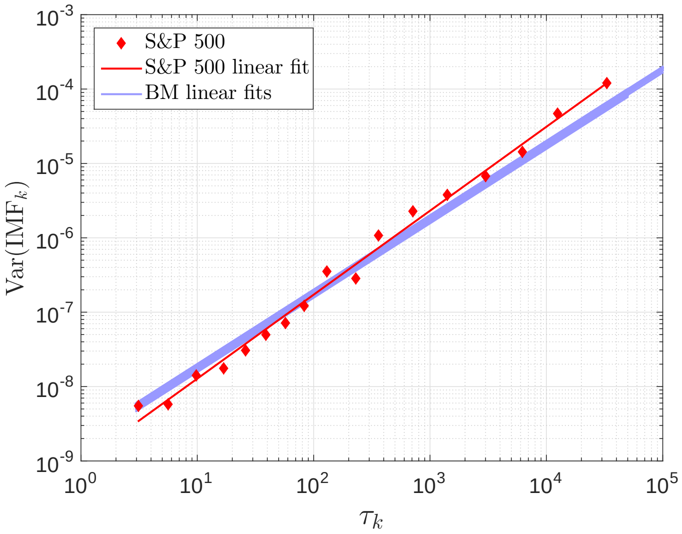


Fig. 5.6 Log-log plot of variance as a function of period for the S&P 500 IMFs (red diamonds) compared with 100 rescaled BM linear fits of slope $H^* = 0.5$ (blue lines).

Figure 5.5. We observe that the Brownian motion linear fits (blue lines) and the linear fit of the S&P 500 index (red line) are close to each other, suggesting this stock market index exhibits a scaling behaviour similar to the one observed in Brownian motion.

The goodness of the linear fit between the financial data points (red diamonds) and each of the Brownian motion linear fit (blue lines) was calculated as follow:

$$R^2_{BM_i} = 1 - \frac{\sum\limits_{k=1}^{n} \left[\log\left(\mathrm{Var}\left(IMF_k^X\right)\right) - \log\left(c_i\, c_{0_i} \tau_k^X\right)\right]^2}{\sum\limits_{k=1}^{n} \left[\log\left(\mathrm{Var}\left(IMF_k^X\right)\right) - \left\langle \log\left(\mathrm{Var}\left(IMF_k^X\right)\right)\right\rangle\right]^2}. \tag{5.5}$$

where $\langle \cdot \rangle$ indicates the mean over the n IMF variances obtained from the stock market index. The deviations from Brownian motion were calculated by the mean over the goodness of the linear fits, i.e., we calculated:

$$\left\langle R^2_{BM} \right\rangle = \frac{1}{n} \sum_{i=1}^{100} R^2_{BM_i}. \tag{5.6}$$

For the S&P 500 index, we obtained a coefficient equal to $\left\langle R^2_{BM} \right\rangle = 0.979$, demonstrating the similarity between the scaling properties of this stock market index and Brownian motion.

5.5 Observed scaling properties of stock market indices

In previous section, we introduced two measures that quantify the deviations from the scaling behaviour of fractional Brownian motion and Brownian motion. These measures are given by:

1. R^2, coefficient of determination (square of correlation) between the logarithm of period and logarithm of variance of IMFs obtained from stock market indices.

2. $\left\langle R^2_{BM} \right\rangle$, mean of the relative squared residuals between the IMF variances obtained from financial data and each of the linear fits for Brownian motion simulations.

In Table 5.6, we report the values of $\left\langle R^2_{BM} \right\rangle$ for the 22 stock market indices. For comparison purposes, we repeated the R^2 values reported in Table 5.5. The last column of Table 5.6 indicates the ordering of the markets if R^2 were used as the ranking measure.

The S&P 500 index is ranked the highest in both scales. Developed markets tend to be at the top of the table, with some exceptions that may arise from the specific characteristics of the analysed period of time, May 5th, 2014 to November 5th, 2014. In Figure 5.7, we plot the $R^2_{BM_i}$ ranking of the stock market indices. The horizontal bars represent the 5th and the 95th percentiles of the $R^2_{BM_i}$ distribution. The blue dot inside each bar indicates the mean value $\left\langle R^2_{BM} \right\rangle$ as reported in Table 5.6. Despite the fact that some financial stock indices have similar values of $R^2_{BM_i}$, we can recognize statistically significant differences between developed and emerging markets, observing a clear tendency for the developed markets to present larger values of $\left\langle R^2_{BM} \right\rangle$ with narrower distributions.

In order to visualize the anomalous scaling in some stock markets indices and to understand the origin of the differences in the results, we compare the cases of the NASDAQ (USA), BOVESPA (Brazil), NIKKEI 225 (Japan) and DSM (Qatar) indices in more detail.

For the NASDAQ index (USA), we obtained 20 IMFs and a residue. In Figure 5.8(a), we display the log-prices (blue line) and the 'trend' consisting of the residue plus the last IMF (red line). In Figure 5.8(b), we observe that the deviation from the linear relationship of Equation (5.2) is significant. Thus, the log-log relationship between period and variance is not completely satisfied. The resultant coefficient of determination is $R^2 = 0.960$, ranking this stock market index at the 18th position. Moreover, when compared with BM, we identify that most of the components deviate from the BM linear fits (blue lines). We also note that the total number of components (21) is considerable larger than what would be expected from a process with uniform scales, i.e., $\log_2(100620)$=16.6. The presence of these extra oscillations with reduced variance suggests a more complex structure than BM. The

Country	**Index**	$\langle R^2_{BM} \rangle$	$Rank_{\langle R^2_{BM} \rangle}$	R^2	$Rank_{R^2}$
USA	S&P 500	0.979	**1**	0.992	**1**
Brazil	BOVESPA	0.977	**2**	0.989	**2**
UK	FTSE	0.973	**3**	0.977	**8**
Turkey	XU 100	0.972	**4**	0.985	**4**
Italy	FTSE MIB	0.971	**5**	0.987	**3**
France	CAC 40	0.970	**6**	0.984	**6**
Spain	IBEX	0.969	**7**	0.973	**10**
China	SSE	0.967	**8**	0.971	**12**
Russia	RTSI	0.964	**9**	0.985	**5**
Hungary	BUX	0.963	**10**	0.970	**15**
Mexico	IPC	0.960	**11**	0.971	**14**
Hong Kong	HSI	0.958	**12**	0.969	**16**
USA	NASDAQ	0.954	**13**	0.960	**18**
Netherlands	AEX	0.953	**14**	0.968	**17**
South Africa	JSE	0.952	**15**	0.956	**20**
Japan	NIKKEI 225	0.949	**16**	0.959	**19**
Greece	ASE	0.948	**17**	0.974	**9**
Poland	WIG	0.947	**18**	0.973	**11**
UAE	UAED	0.939	**19**	0.978	**7**
Singapore	STI	0.934	**20**	0.942	**22**
Malaysia	KLSE	0.933	**21**	0.943	**21**
Qatar	DSM	0.928	**22**	0.971	**13**

Table 5.6 Stock market indices ranked in descending order of $\langle R^2_{BM} \rangle$. The last column indicates the ordering of the markets with respect to R^2.

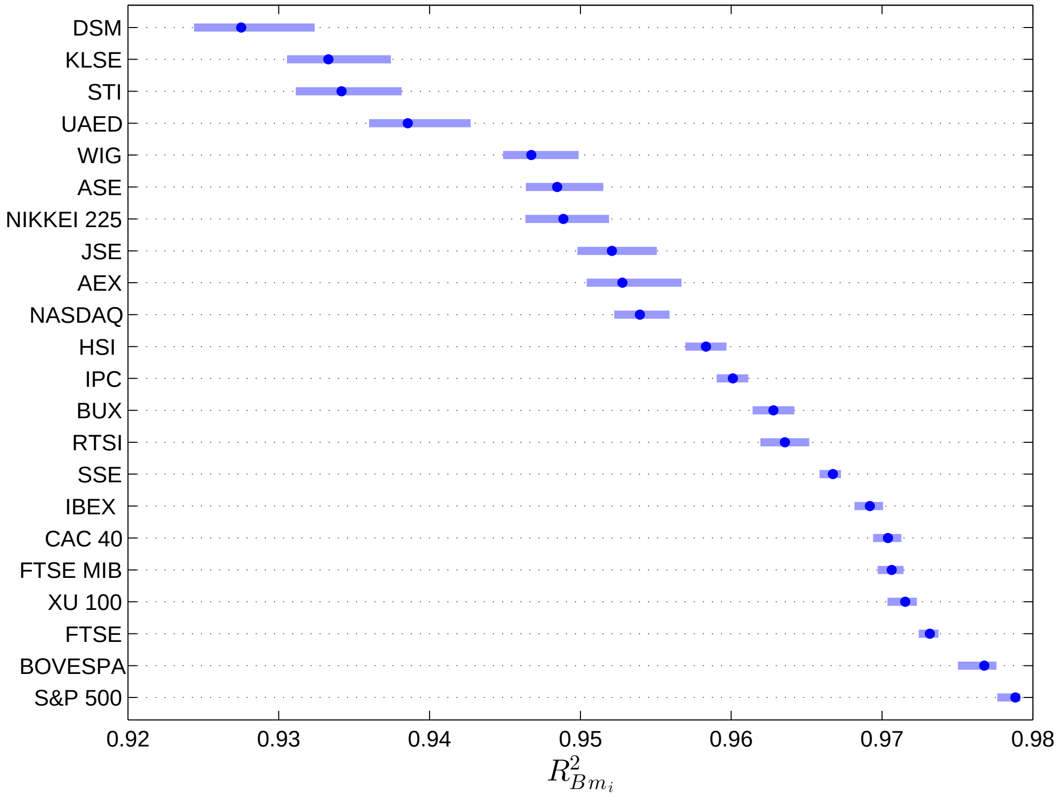
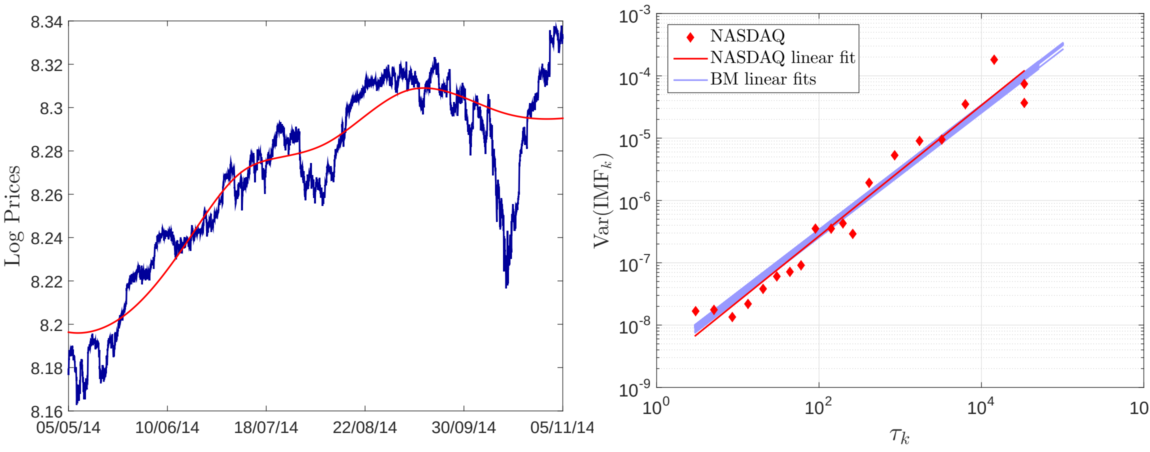


Fig. 5.7 Percentiles 5th and 95th of the $R^2_{BM_i}$ distribution for the analysed stock market indices. The blue dot inside each bar indicates the value of $\langle R^2_{BM} \rangle$ used for the stock market index ranking.

deviations from the BM scaling behaviour, quantified by the coefficient $\langle R^2_{BM} \rangle = 0.958$, rank this index at the 13th position.

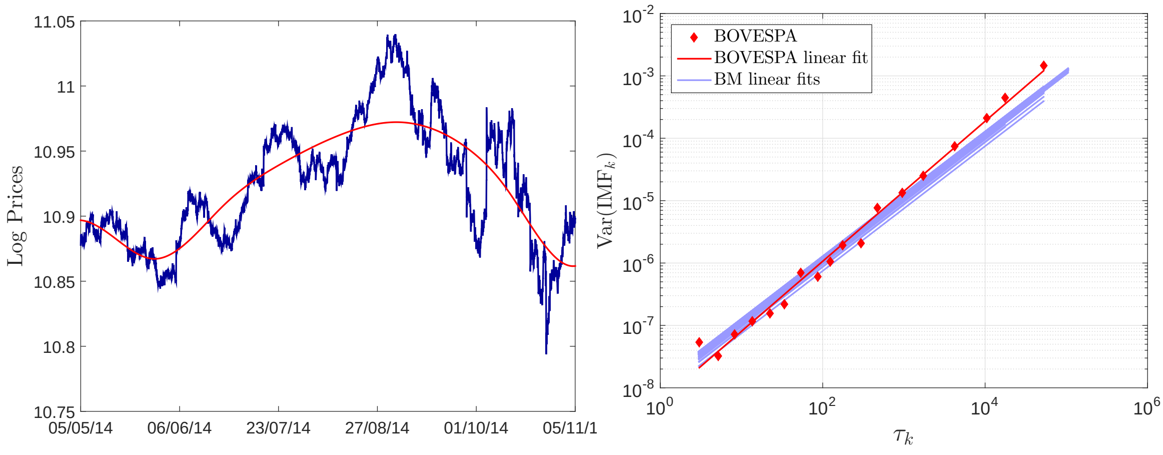


(a) **(b)**

Fig. 5.8 EMD analysis for the NASDAQ index. Captions for figures (a) and (b) are the same as captions for figures 5.4 and 5.6 respectively.

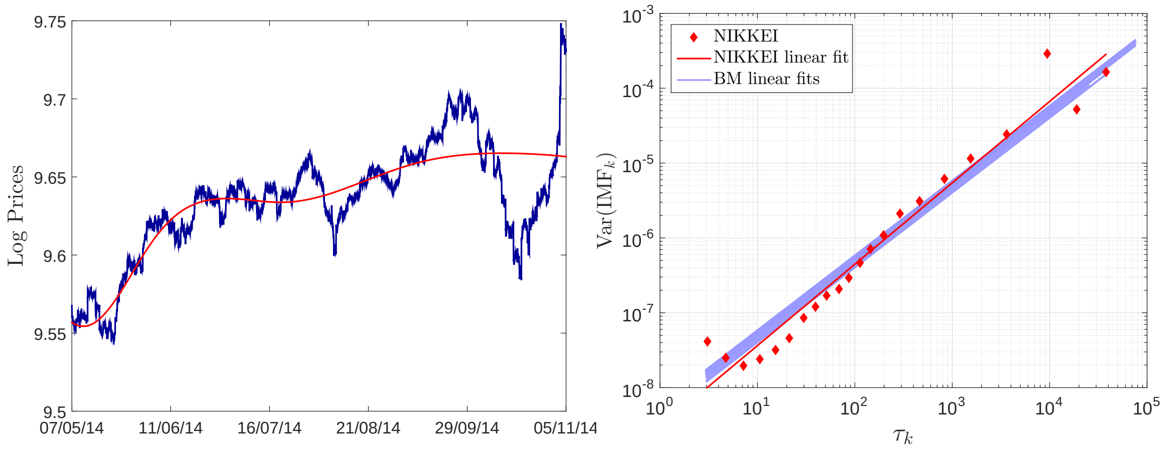
The variance scaling properties of the BOVESPA index (Brazil) are presented in Figure 5.9. For this stock market index, the EMD identifies long-period cycles with larger variance than what would be expected from BM, see Figure 5.9(b). However, the linear fit between the logarithm of IMF variances and periods is in general good with $R^2 = 0.989$. The goodness of the linear fit between the BM simulations is $\langle R^2_{BM} \rangle = 0.977$, placing this index at the second position. Such a good ranking for this market may be unexpected, but we must stress that it only reflects the six-month period of observations. From Figure 5.9(a), we can see that this was a rather random but calm period.

For the NIKKEI 225 index (Japan), we obtained 22 IMFs and a residue. Similar as the NASDAQ index, the number of components is considerable larger than what would be expected from BM, i.e., $\log_2(75600)$=16.2. These many oscillations, specially the high-frequency components, generate a non-linear behaviour that deviates from BM. Given the anomalous scaling behaviour of this stock market index, see Figure 5.10(b), we obtained $\langle R^2_{BM} \rangle = 0.949$, ranking it at the 16th position.



(a) **(b)**

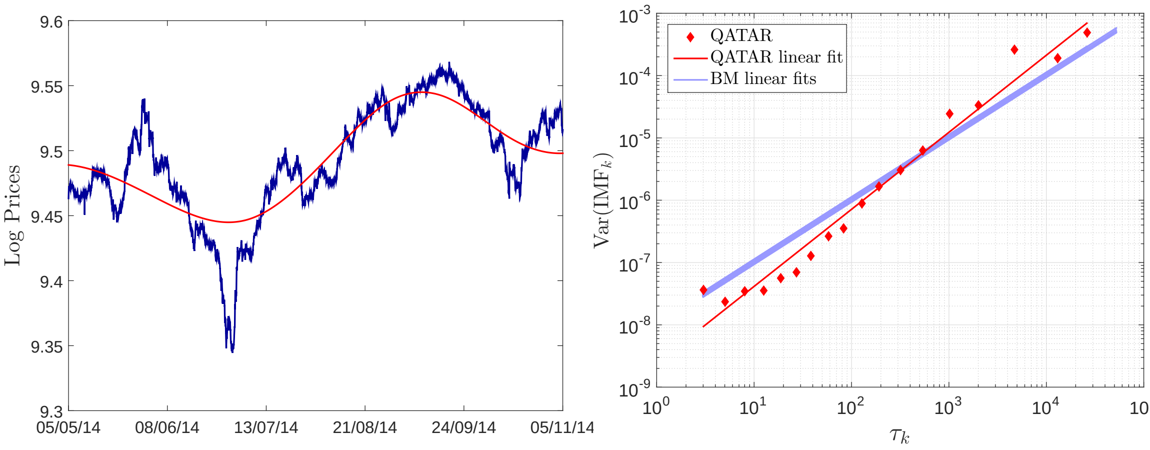
Fig. 5.9 EMD analysis for the BOVESPA index. Captions for figures (a) and (b) are the same as captions for figures 5.4 and 5.6 respectively.



(a) **(b)**

Fig. 5.10 EMD analysis for the NIKKEI 225 index. Captions for figures (a) and (b) are the same as captions for figures 5.4 and 5.6 respectively.

Finally, the DSM index (Qatar) is displayed in Figure 5.11. The log-price time series and its respective ‘trend’ are displayed in Figure 5.11(a). In Figure 5.11(b), we observe the poor liner fit of Equation (5.2) that is characterized by a considerable steep slope. We obtained $R^2 = 0.971$, ranking this index at the lowest position. Furthermore, if we compare its IMF variances against the BM linear fits, we observe that most of the variance values (red diamonds) follow outside the band expected from BM. The large variance of the low frequency components suggests the presence of important long-period cycles. Given its deviations from BM behaviour, this index is also ranked the lowest with respect to the measure $\langle R^2_{BM} \rangle = 0.928$.



(a) **(b)**

Fig. 5.11 EMD analysis for the DSM index. Captions for figures (a) and (b) are the same as captions for figures 5.4 and 5.6 respectively.

5.6 Summary

We empirically showed that FBM and α-stable Lévy motion obey a scaling law that relates linearly the logarithm of the variance and the logarithm of the period of the IMFs. We demonstrated that the extracted scaling exponent equals the Hurst exponent H multiplied by two. The proposed estimator is robust to heavy-tailed distributions and can be used to estimate the long-range dependence of a time series.

When applied to stock market indices, the EMD reveals instead different scaling laws that can deviate significantly from both Brownian motion and fractional Brownian motion behaviour. In particular, we noted that the EMD of high-frequency financial data results in a larger number of IMFs than what would be expected from Brownian motion. These many components, specially high-frequency components, create a curvature that disobeys the linearity found in the log-log relation between the IMF variance and the period of FBM. This is a direct indication of anomalous scaling that reveals a more complex structure in financial time series than in self-similar processes.

In this study, we applied the EMD to 22 different stock market indices and observed that developed markets (European and North American markets) tend to have scaling properties closer to Brownian motion properties. Conversely, larger deviations from uniscaling laws are observed in some emerging markets such as Malaysian and Qatari.

These findings are in agreement with the discernible characteristics of developed and emerging markets, the former type being more likely to exhibit an efficient behaviour, see for example [64, 65]. Compared to previous approaches, the EMD method has the advantage to directly quantify the cyclical components with strong deviations, giving a further instrument to understand the origin of market inefficiencies.

Our results confirm the need for multifractal models which have recently been introduced as a new type of data-generating process [41]. The main distinguishing property of these models is the non-linear variation in the scaling behaviour of its moments. Calvet et al. [40, 42] proposed a multi-frequency model in which innovations in dividend volatility depend on shocks that decay with different frequencies. We proposed a method using high-frequency data which identifies the interactions of these frequencies and reveals scaling laws in financial time series. Essentially, the multifractal structure appears to be a very parsimonious and robust way to capture a hierarchical multicomponent structure of the price process.

Chapter 6

Time-Dependent Scaling Properties of Stock Market Indices

In this chapter, we investigate the influence of different timescales on the dynamics of financial market data. This is obtained by decomposing financial time series into a set of simple oscillations associated with distinctive timescales and with time-dependent attributes, such as amplitude and period. In the first part of this chapter, we introduce a novel time-dependent scaling exponent that quantifies the relative hierarchical variations of the amplitudes of the components with respect to their associated timescales. The proposed exponent is related to the scaling properties of self-similar processes. In the second part of this chapter, we propose an entropic measure which quantifies the dispersion of the amplitudes of the components. We apply the time-dependent measures to four different stock market indices, aiming to reveal their intraday scaling behaviour.
Part of this chapter is published in the paper: "Time-dependent scaling patterns in high-frequency financial data" [136].

6.1 Time-dependent scaling exponent

In Section 5.1, we discussed some literature related to the importance of time-dependent scaling parameters. The first step to estimate the proposed time-dependent scaling exponent is to apply the HHT to the input time series X_t, $t = 1, 2, \ldots, N$, to obtain its time time-varying amplitude and period attributes.

The proposed estimator, denoted as $H^*(t)$, is constructed by observing the way the local amplitudes $a_k(t)$, Equation (2.25), change with respect to the local periods $\tau_k(t) = \omega_k(t)^{-1}$, Equation (2.27), for all $k = 1, 2, \ldots, n$.

We first applied the method to FBM and observed that the time-dependent amplitude follows a power-law behaviour with respect to the instantaneous period:

$$a_k(t) \propto \tau_k^{H^*(t)}(t). \tag{6.1}$$

where the exponent $H^*(t)$ describes the local scaling properties of the IMF amplitudes and takes values distributed around the self-similar exponent H of FBM. The exponent $H^*(t)$ is based on the scaling properties of the absolute value of the fluctuations, therefore comparisons with other estimators such as the generalized Hurst exponent [63] must be done for the first order moments i.e. $q = 1$ [107].

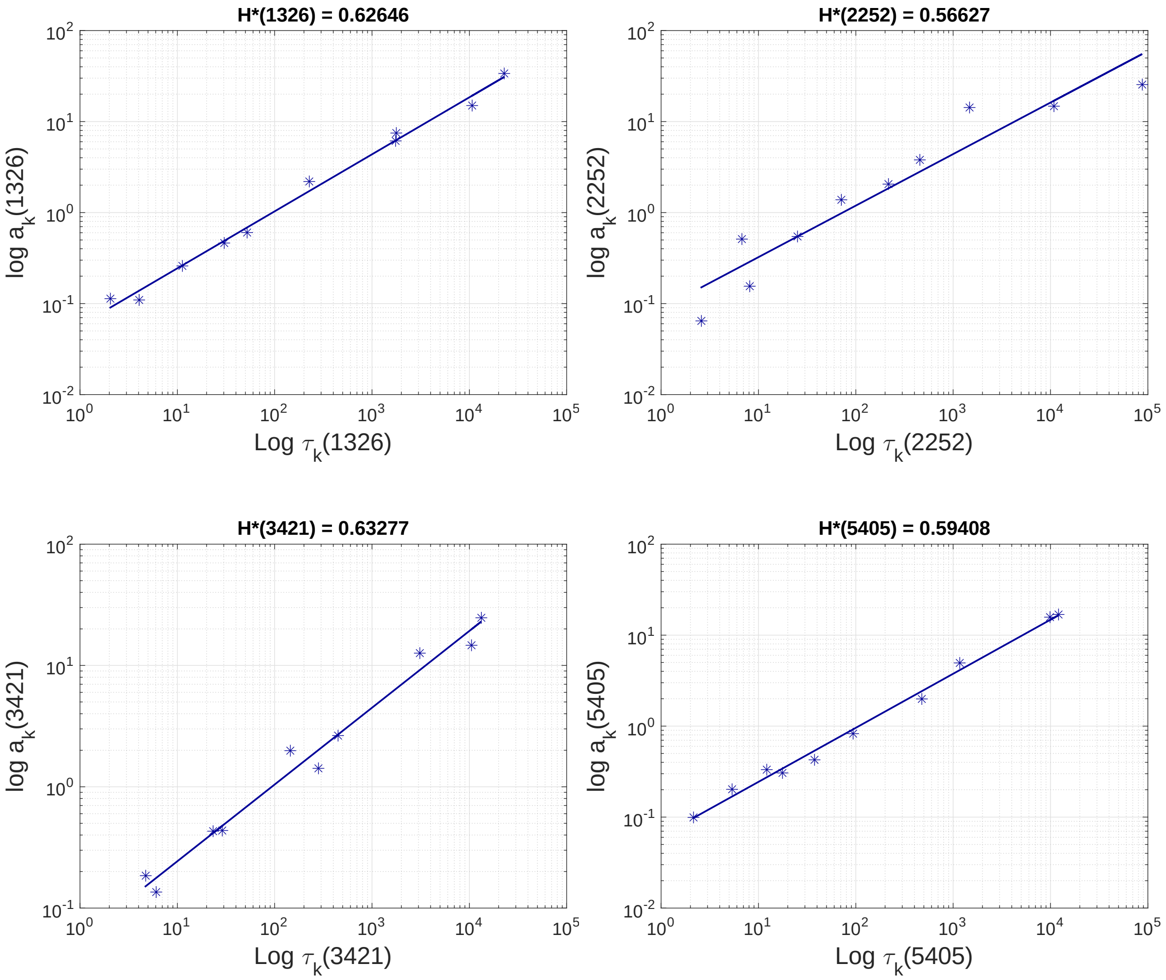


Fig. 6.1 Illustration that for FBM the local amplitudes $a_k(t)$ and the periods $\tau_k(t)$ follow Equation (6.1): $a_k(t) \propto \tau_k^{H^*(t)}(t)$. Plots report instantaneous amplitude as a function of period for the following four randomly chosen times: $t = 1326, 2252, 3421, 5405$. The simulated process is an FBM with self-similar exponent $H = 0.6$ and length $N = 10,000$ points. The straight lines represent the best-fit linear regressions.

In Figure 6.1, we report a particular instance of Equation (6.1) showing the linear fit between $\log a_k(t)$ and $\log \tau_k(t)$ for four randomly chosen times of an FBM with self-similar exponent $H = 0.6$ and length $N = 10,000$. The values of $H^*(t)$ reported in the plots are obtained from the slope of the regression fit. We observe that they are all consistently close to the self-similar value $H = 0.6$. For the chosen values of t, we calculated the goodness of the linear fit by estimating the coefficient of determination R^2 [149] (values of this coefficient range from 0 to 1, with 1 indicating a perfect fit between the data and the linear model). Results for the four randomly chosen times: $t = 1326, 2252, 3421, 5405$ are as follow: $R^2(1326) = 0.99$, $R^2(2252) = 0.90$, $R^2(3421) = 0.98$, $R^2(5405) = 0.99$, indicating therefore that for those time instances, the data are well represented by the log-linear model of Equation (6.1). Similar linear scaling results are obtained across all times, but the scaling exponent is different at each time step, making $H^*(t)$ a time-dependent estimator.

A value of $H^*(t) > 0.5$ is obtained when around time t, the amplitude of long cycles is larger than in a pure random walk. This can be interpreted as a persistent behaviour in the amplitudes of the process, meaning that in a neighbourhood around time t the process is in a cycle indistinguishable from a trend. On the contrary, values of $H^*(t) < 0.5$ represent a rougher and more chaotic behaviour around time t. These processes are composed of oscillations with more similar amplitudes across timescales than in Brownian motion, creating a complex and uncertain behaviour. In this case, high-frequency components are more active and their contribution to the total variance is more significant than in a random walk process.

6.2 Numerical study of self-similar and long-memory processes

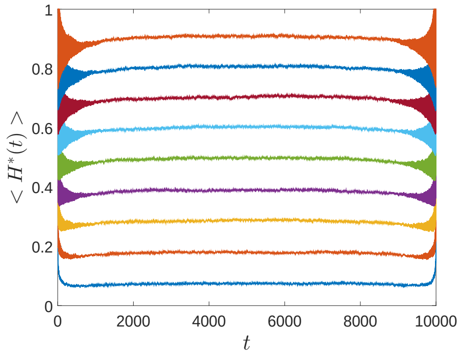
In order to test the power-law relation of Equation (6.1), we extended the simulation set of FBM and we considered other two different self-similar processes, namely α-stable Lévy motion (SLM) [156] and autoregressive fractionally integrated moving average (ARFIMA) processes [88], refer to Section 2.6.2 for more details about these stochastic processes.

For each stochastic process, we simulated $M = 1,000$ paths of length $N = 10,000$ points[1]. We estimated $H^*(t)$ and calculated the time-dependent sample mean over the number of simulations, i.e., we calculated $\langle H^*(t)\rangle = \frac{1}{M}\sum_{i=1}^{M} H_i^*(t)$.

We also estimated the sample mean of $H^*(t)$ over time and over the number of simulations, $\langle\langle H^*\rangle\rangle = \frac{1}{N}\sum_{t=1}^{N}\langle H^*(t)\rangle$. The standard deviation of $H^*(t)$ is calculated as: $\sigma_{H^*} = \sqrt{\sum_{t=1}^{N}\sum_{i=1}^{M}\left(H_i^*(t) - \langle\langle H^*\rangle\rangle\right)^2 / (NM-1)}$.

For each process, the average over time and over the number of simulations of the coeffi-

[1] The length of the SLM is set to $N = 2^{14} - 6000$, following the algorithm proposed in [164]



(a)

H	H^*			**GHE(1)**	
	$\langle\langle H^* \rangle\rangle$	σH^*	$\langle\langle R^2 \rangle\rangle$	$\langle H_G \rangle$	σ_G
0.1	0.07	0.08	0.20	0.15	0.01
0.2	0.18	0.08	0.39	0.22	0.01
0.3	0.28	0.08	0.59	0.31	0.01
0.4	0.38	0.09	0.73	0.40	0.01
0.5	0.50	0.09	0.81	0.50	0.01
0.6	0.60	0.09	0.86	0.60	0.01
0.7	0.70	0.10	0.89	0.70	0.01
0.8	0.80	0.10	0.92	0.79	0.01
0.9	0.90	0.11	0.93	0.87	0.01

(b)

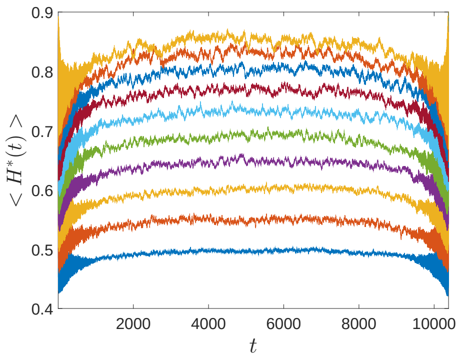
Fig. 6.2 a) Illustration that for FBM the scaling exponent $H^*(t)$ is on average close to the self-similar exponent H. The plot reports the sample mean of $H^*(t)$, denoted as $\langle H^*(t) \rangle$ and computed over $M = 1{,}000$ simulations of FBM with self-similar exponent $H = 0.1, 0.2, \ldots, 0.9$ (bottom to top) and length $N = 10{,}000$ points.
b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^* \rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2 \rangle\rangle$. The values of GHE(1) denote the generalized Hurst exponent with $q = 1$.

cient of determination is also estimated and denoted as $\langle\langle R^2 \rangle\rangle$. We compared the estimated $H^*(t)$ with the generalized Hurst exponent [63] with $q = 1$, here denoted as H_G.

- **Fractional Brownian motion**. Stochastic processes with scaling exponent varying from $H = 0.1, 0.2, ..., 0.9$ were simulated. All the simulations were done using the Matlab® wavelet toolbox. Results for different values of H are reported in Figure 6.2(a). We observe that $\langle H^*(t) \rangle$ consistently varies around the input self-similar value of H.

  In Table 6.2(b), we report $\langle\langle H^* \rangle\rangle$ and the standard deviation of $H^*(t)$. We observe a good agreement with the self-similar parameter H, but large values for σ_{H^*} that could be attributed to the local characteristics of $H^*(t)$. We notice that for FBM with scaling exponent $H < 0.3$, the $\langle\langle R^2 \rangle\rangle$ coefficients are small, indicating significant deviations from the scaling law of Equation (6.1). We obtained consistent results when comparing the estimated $H^*(t)$ with the generalized Hurst exponent.

  Moreover, we also considered FBM paths with shorter length, $N = 1{,}000$ and $N = 500$ points, we do not report these results, but we noted that the longer the time series, the better the estimation of the scaling exponent H. Likewise, the standard deviation and the goodness of the linear fit improve with the length of the time series. However, all results are consistent with the ones reported here for length $N = 10{,}000$.



(a)

H	H^*			**GHE(1)**	
	$\langle\langle H^*\rangle\rangle$	σ_{H^*}	$\langle\langle R^2\rangle\rangle$	$\langle H_G\rangle$	σ_G
0.5	0.50	0.09	0.81	0.50	0.01
0.55	0.54	0.11	0.82	0.55	0.01
0.6	0.59	0.13	0.82	0.60	0.02
0.65	0.64	0.15	0.82	0.65	0.03
0.7	0.68	0.17	0.82	0.69	0.04
0.75	0.72	0.18	0.82	0.74	0.04
0.8	0.75	0.20	0.81	0.78	0.05
0.85	0.79	0.21	0.80	0.82	0.05
0.9	0.81	0.23	0.78	0.85	0.05
0.95	0.83	0.24	0.77	0.88	0.04

(b)

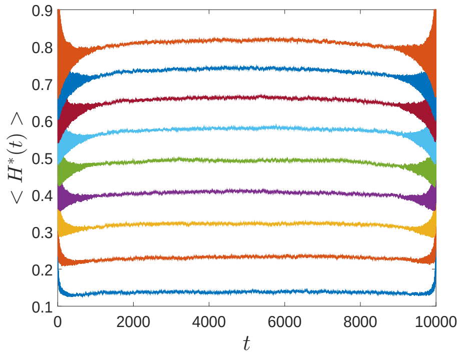
Fig. 6.3 a) Illustration that for SLM the scaling exponent $H^*(t)$ is on average close to the value $H = \frac{1}{\alpha}$. The plots report the sample mean of $H^*(t)$, denoted as $\langle H^*(t)\rangle$ and computed over $M = 1{,}000$ simulations of SLM with self-similar exponent $H = 0.5, 0.55, \ldots, 0.95$ (bottom to top) and length $N = 10{,}384$ points.
b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^*\rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$.

- **α-stable Lévy motion (SLM)**. We generated SLM processes using the toolbox provided by [164], sample paths are of length $N = 10{,}384$, with parameters for the generation $m = 128$ and $L = 6000$, making $m(L+N)$ to be a power of 2, see [164] for more details. We considered the case $H = 1/\alpha$ for values of $H = 0.5, 0.55, \ldots, 0.95$.

  The time-dependent sample mean over the number of simulations is displayed in Figure 6.3(a). We observe a noisier estimator than the one obtained for FBM. In Table 6.3(b), we report $\langle\langle H^*(t)\rangle\rangle$, noticing a fair approximation to the self-similar parameter H, with better results for processes with $H < 0.7$. The means of the coefficient of determination suggest that the scaling relation of Equation (6.1) is indeed satisfied. We compared the proposed estimator with the generalized Hurst exponent with $q = 1$, obtaining consistent results, i.e., $H_G = 1/\alpha$ [106].

- **ARFIMA**. We tested the log-linear relationship of Equation (6.1) in ARFIMA(p,d,q) processes with Gaussian innovations, $p, q \in \mathbb{N}$, autoregressive and moving average coefficients respectively [88]. We considered the simple case of ARFIMA(0,d,0) with fractional order $d = -0.4, -0.3, \ldots, 0.4$ and length $N = 10{,}000$. We calculated $H^*(t)$ for the integrated ARFIMA time series.

  From Figure 6.4(a), we observe that the estimator $\langle H^*\rangle$ is a good approximation of the



(a)

H	H^*			GHE(1)	
	$\langle\langle H^*\rangle\rangle$	σ_{H^*}	$\langle\langle R^2\rangle\rangle$	$\langle H_G\rangle$	σ_G
0.1	0.14	0.09	0.29	0.21	0.01
0.2	0.23	0.09	0.48	0.27	0.01
0.3	0.32	0.09	0.64	0.34	0.01
0.4	0.40	0.09	0.74	0.42	0.01
0.5	0.49	0.09	0.81	0.50	0.01
0.6	0.57	0.09	0.85	0.59	0.01
0.7	0.65	0.09	0.88	0.68	0.01
0.8	0.73	0.10	0.90	0.77	0.01
0.9	0.81	0.10	0.91	0.84	0.01

(b)

Fig. 6.4 a) Illustration that for ARFIMA(0,1,0) the scaling exponent $H^*(t)$ is on average close to the value $H = d + 0.5$. The plots report the sample mean of $H^*(t)$, denoted as $\langle H^*(t)\rangle$ and computed over $M = 1,000$ simulations of ARFIMA(0,d,0) with self-similar exponent $H = 0.1, 0.2, \ldots, 0.9$ (bottom to top) and length $N = 10,000$ points.
b) Sample mean of $H^*(t)$ over time and over the number of simulations, denoted as $\langle\langle H^*\rangle\rangle$, standard deviation of $H^*(t)$ and sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$.

exponent H. In Table 6.4(b), we report $\langle\langle H^*\rangle\rangle$, the standard deviation of $H^*(t)$ and the sample mean of the coefficient of determination, $\langle\langle R^2\rangle\rangle$. Similarly to the FBM case, the estimation is more accurate for larger H with larger coefficient of determination.

From the analysis of these three different stochastic processes, we observe that our proposed method produces a fair estimate of the self-similar parameter H. The chosen models concern two different properties which contribute to the self-similarity of the processes. The first property is the long-range autocorrelation of the increments. The second property is the high variability or the heavy tails in the distribution of the increments of the α-stable Lévy motion. We have demonstrated empirically that the analytics of the relative amplitude of the oscillating components of the signal capture both of these properties. Let us remark that this scaling exponent is not intended as an alternative method to estimate H, which can instead be obtained with more reliable tools [65, 76, 163]. The aim of this method is instead to compute the time-dependent amplitude contribution of the prevalent fluctuations present in a time series, distinguishing between periods when high or low frequencies are contributing more than what could be expected from Brownian motion.

As observed from Figures 6.2(a), 6.3(a) and 6.4(a) some boundary effects of the EMD affect the estimation of the time-dependent scaling exponent. These effects emerge when the end points are not extrema. The interpolated envelope diverges and there are wild swings which could propagate trough the time series. Furthermore, the Hilbert transform is based in on the Fourier transform, which may produce some errors due to the Gibbs phenomena

[72].

In order to minimize error propagations due to finite observations, the end points of the time series have to be treated differently and the data have to be extended beyond the existing range. The first technique for dealing with the end conditions was proposed by [96] and it consists of padding the beginning and the end of the time series with additional "characteristic waves" which are defined by the two consecutive extrema. Flandrin et al. [151] offer one of the simplest yet very robust method that uses a mirror symmetry with respect to the extrema closest to the end. Recently, some forecasting methods based on machine learning algorithms have been proposed to extend the original time series and reduce the impact of the end effects [91, 114]

6.3 Time-dependent complexity measure

We define a time-dependent Shannon entropy-like measure based on the square of the amplitude of the IMFs. This measure provides a time-varying quantification of complex spectrum which offers an alternative to the scaling exponent to measure the strength of the cycles present in financial time series. Making use of the functions a_k, described in Equation (2.25), we define a timescale relative distribution of amplitudes as:

$$p_k(t) = \frac{a_k^2(t)}{\sum\limits_{k=1}^{n} a_k^2(t)}, \tag{6.2}$$

where n is the number of IMFs excluding the residue. Similarly to Shannon entropy [158], we define the time-dependent complexity measure as:

$$C^*(t) = -\sum_{k=1}^{n} p_k(t) \ln p_k(t). \tag{6.3}$$

Equation (6.3) provides a measure of the distribution of amplitudes between the oscillating components. If the total amplitude at time t is concentrated in one oscillation mode, we observe a low complexity value, implying that around time t the process is following a prevalent trend. On the contrary, if at time t all the oscillation modes have similar amplitudes, we obtain a large complexity value that indicates a more erratic and unpredictable behaviour.

Thus, $C^*(t)$ provides a time-varying estimation of disorder and adapts closely to our visual perception of complexity. Moreover, Equation (6.3) offers a more general measure of uncertainty than the variance since the latter measures the dispersion around the mean, while $C^{*}(t)$ measures the dispersion of energy around the different IMFs. Similar to an entropy measure, the value of the proposed complexity at time t varies between zero, if one IMFs dominates the energy of the process, and $\log(n)$ if the energy is uniformly distributed between the n IMFs.

The choice of the weights equal to the square of the amplitudes in Equation (6.2) is arbitrary, although it is in agreement with other measures of entropy that have been defined, for example in [152]. We tested alternative choices, such as the linear weight $a_k(t)$, obtaining analogous results to the ones reported here.

Let us note that, although not independent, the two measures convey different information. The estimator $C^{*}(t)$ is an information quantifier of uncertainty, it is obtained from the distribution of the amplitudes regardless of their timescales and it only quantifies the homogeneity of the components. On the other hand, the scaling exponent, $H^{*}(t)$, measures the change in the amplitudes across timescales, testing the scaling law of Equation (6.1). In this respect, $H^{*}(t)$ is a more restrictive measure that assumes a log-linear relationship between amplitudes and periods.

6.4 Time-dependent scaling in financial markets

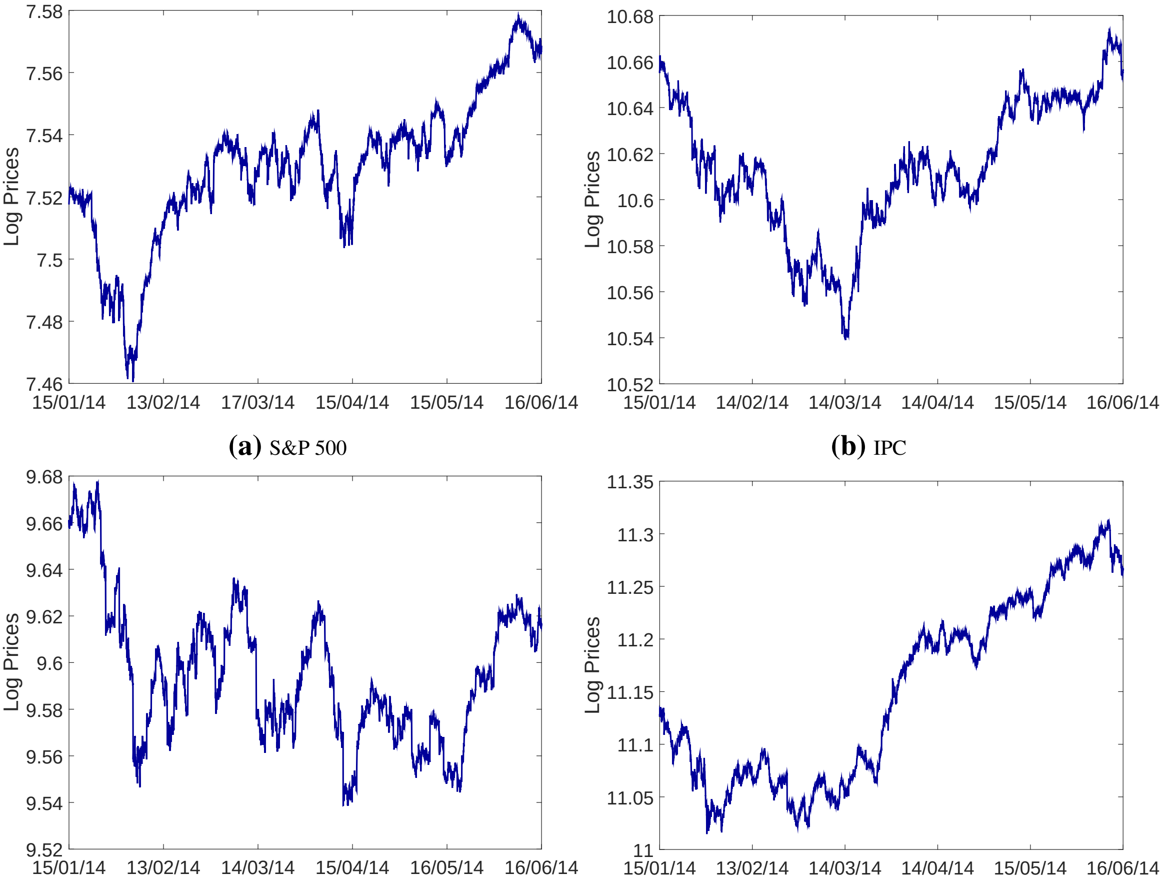
We applied the proposed measures to intraday prices of four stock market indices: (1) S&P 500 (USA), (2) IPC (Mexico), (3) Nikkei 225 (Japan) and (4) XU 100 (Turkey). We intentionally chose two financial markets that are classified as developed (USA and Japan) and two emerging markets (Mexico and Turkey) with the additional feature that the Japanese and Turkish stock exchanges have two trading sessions separated by a lunch break.

The data set consists of prices recorded at 30-second intervals. It covers a period of 5 months, from January 15th, 2014 to June 16th, 2014. The logarithm of the prices for the stock market indices are plotted in Figure 6.5. Table 6.1 shows the number of days and the length of each analysed time series.

Country	Index	No. of Days	Length
USA	S&P 500	105	81,900
Japan	Nikkei 225	104	62,400
Mexico	IPC	101	78,780
Turkey	XU 100	106	78,440

Table 6.1 Number of days and length of each financial time series.

The time evolution of $H^{*}(t)$ over the 5-month period for the four financial indices is



(a) S&P 500

(b) IPC

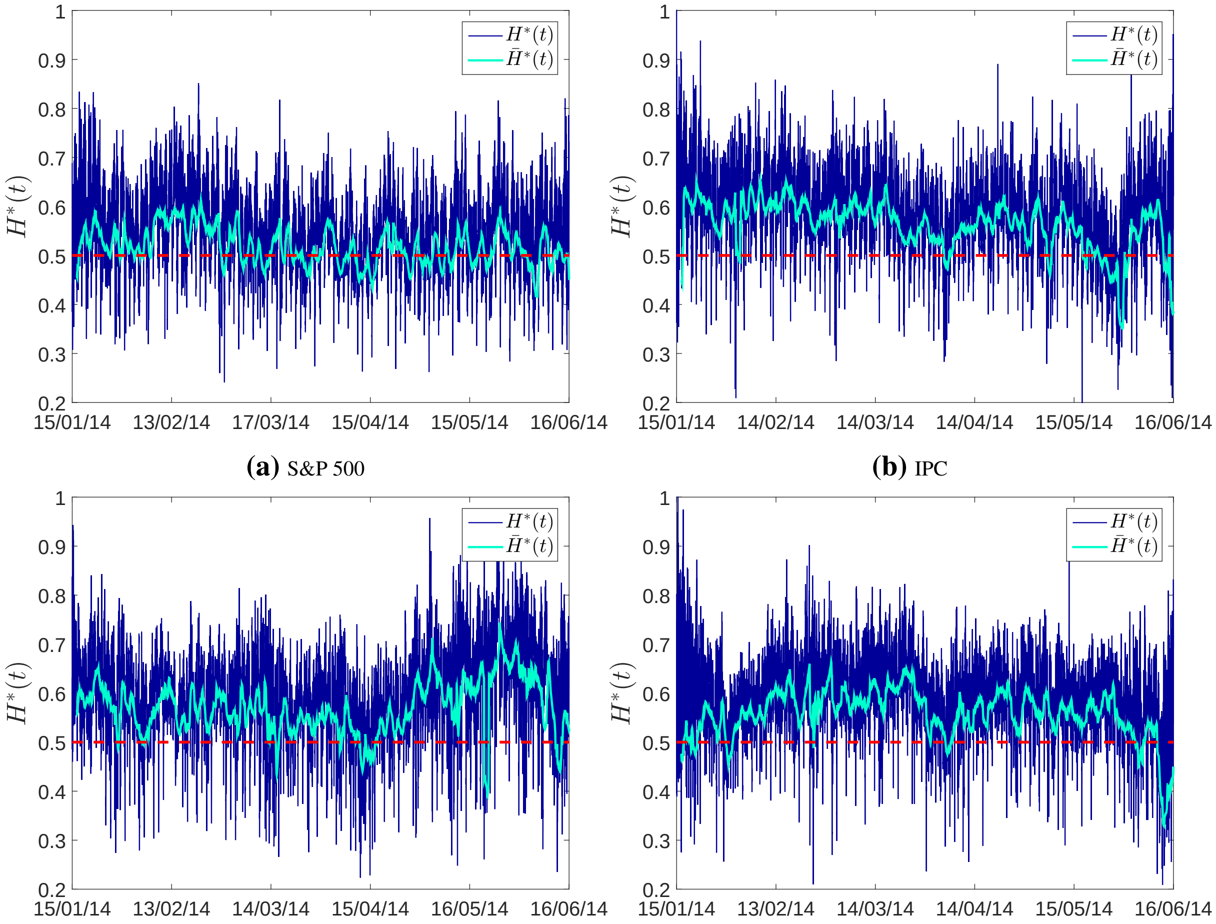
(c) Nikkei 225

(d) XU 100

Fig. 6.5 30-second sampled log-prices for different stock market indices for the period January 15th, 2014 to June 16th, 2014. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

shown in Figure 6.6 (dark-blue line). We note that $H^*(t)$ has large intraday variations which obscure any possible trend over longer periods. For this reason, we report a moving average version of $H^*(t)$ denoted as $\bar{H}^*(t)$ (light-blue line in the same Figure). More specifically, $\bar{H}^*(t)$ is calculated from the relation, $\bar{a}_k(t) \propto \bar{\tau}_k^{\bar{H}^*(t)}(t)$, where $\bar{a}_k(t)$ and $\bar{\tau}_k(t)$ are the averages over a rolling window of the size of a trading day. The dashed red line in this Figure indicates the value $H = 0.5$.

By comparing the values of $H^*(t)$ and $\bar{H}^*(t)$ for the four different stock market indices, we observe that the S&P 500 index is the one closest to the value $H = 0.5$ as expected for Brownian motion, with values of $\bar{H}^*(t)$ fluctuating around 0.5 exposing only some brief departures from it, see Figure 6.6(a). For instance, we can detect a period around February 2014 where the scaling parameter results in significantly larger values. In this period of time, the S&P 500 index was indeed in a rising trend, see Figure 6.5(a). This suggests that the identified persistent behaviour could be attributed to a long timescale cycle with larger



(a) S&P 500

(b) IPC

(c) Nikkei 225

(d) XU 100

Fig. 6.6 Time-dependent scaling exponent for different stock market indices for the period January 15th, 2014 to June 16th, 2014. The scaling exponent $H^*(t)$ is depicted by a dark-blue line. The light-blue line represents $\bar{H}^*(t)$, a rolling-window average over the length of a trading day. The red line indicates the value $H = 0.5$. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

amplitudes than the ones found in a pure random walk.

In Figure 6.6(c), we report the scaling dynamics for the Nikkei 225 index. We observe that $\bar{H}^*(t)$ has values constantly above 0.5, specially at the end of the analysed period. It should be noted that this market has lunch breaks that affect the intraday values of $\bar{H}^*(t)$.

For the IPC index the values of $H^*(t)$ and $\bar{H}^*(t)$ are consistently closer to $H = 0.6$ than to the Brownian motion value $H = 0.5$, see Figure 6.6(b). This suggests that the IPC index shows intervals where the amplitude displays a persistent behaviour. Similarly, the Turkish scaling exponents take values larger than $H = 0.5$, see Figure 6.6(d).

We tested the validity of Equation (6.1) when applied to financial data by computing for every time t, the coefficient of determination $R^2(t)$. The mean over the whole period is reported in the second column of Table 6.2. We also considered the three cases: $H^*(t) < 0.45$, a window around $0.45 < H^*(t) < 0.55$ and $H^*(t) > 0.55$. We observe that the goodness-of-fit is generally better for $H^*(t) > 0.5$, the interesting case when financial data show trending behaviour, see Table 6.2.

Index	$\langle R^2 \rangle_{All}$	$\langle R^2 \rangle_{H^*<0.45}$	$\langle R^2 \rangle_{0.45<H^*<0.55}$	$\langle R^2 \rangle_{H^*>0.55}$
S&P 500	0.8753	0.825	0.8716	0.8916
IPC	0.8812	0.7971	0.8703	0.8915
NIKKEI 225	0.8072	0.7345	0.7829	0.8198
XU100	0.9196	0.7987	0.8737	0.9209

Table 6.2 Average goodness-of-fit coefficient (R^2) for the amplitude versus period log-linear model. First, the average is calculated for all the times t, then, it is calculated separately for those time instances where $H^*(t) < 0.45$, $0.45 < H^*(t) < 0.55$ and $H^*(t) > 0.55$.

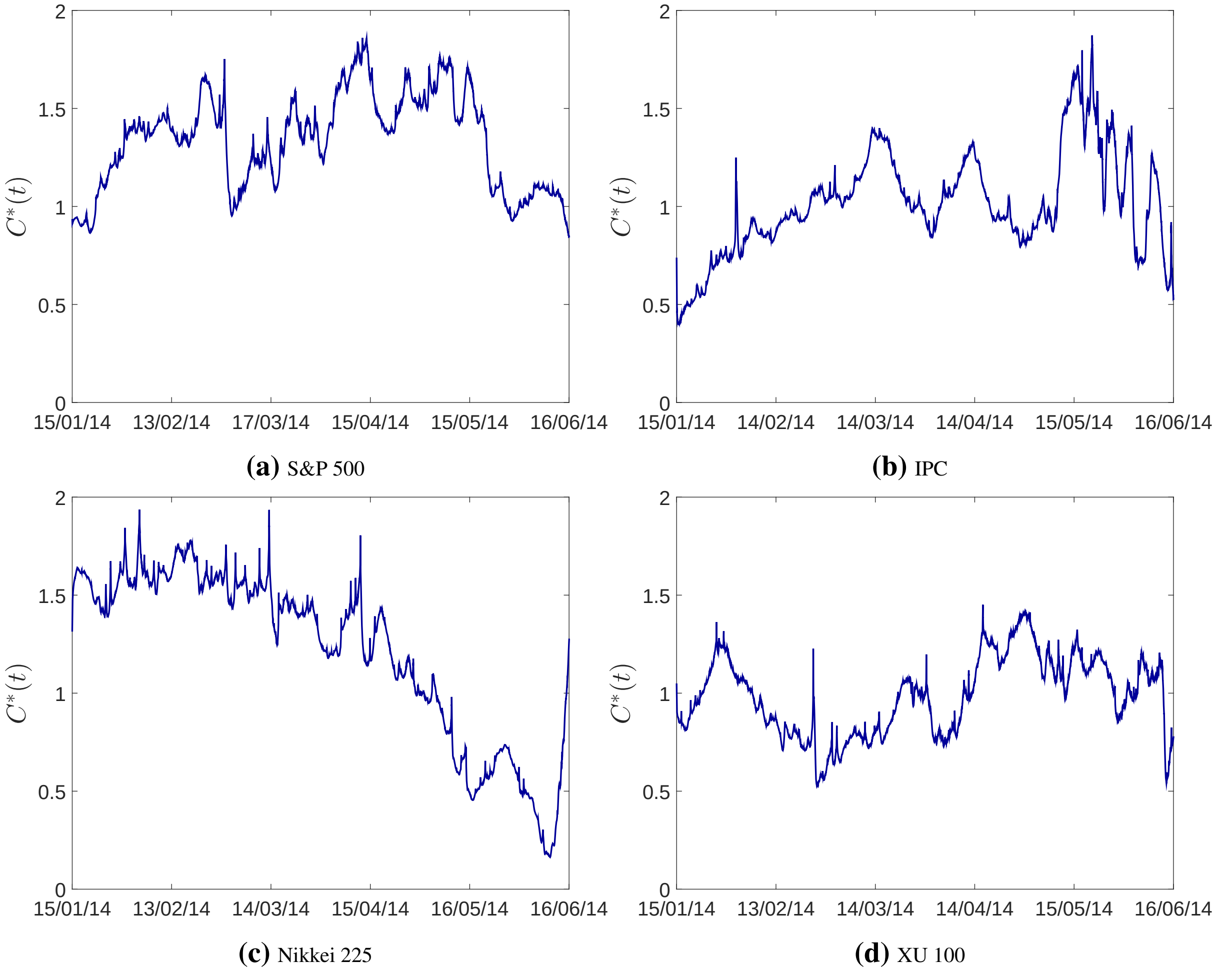
For a comparative analysis, we calculated the complexity measure $C^*(t)$ described by Equation (6.3). The obtained values for each stock market index are illustrated in Figure 6.7. We observed that the complexity values for the S&P 500 index, Figure 6.7(a), are overall the largest among the four indices.

The IPC index shows an increasing evolution of $C^*(t)$, suggesting a more uniform distribution of amplitudes at the beginning of 2014, Figure 6.7(b). On the contrary, the Nikkei 225 index presents a decreasing measure of complexity, indicating a period of higher complexity at the beginning of the sample period, Figure 6.7(c). Finally, the XU100 index presents alternate intervals of high and low complexity, displaying regularly large values in the last two months of the sample period. This higher randomness is also visible from the scaling exponent which displays relatively lower values of $\bar{H}^*(t)$, Figure 6.7(d).

Overall, the functions $H^*(t)$ and $C^*(t)$ vary in opposite directions. For each stock market index, the correlation between these two measures is negative with values $\rho_{S\&P} = -0.21$, $\rho_{IPC} = -0.23$, $\rho_{Nikkei} = -0.28$, $\rho_{XU} = -0.15$. We note these correlations values are small, indicating a weak linear dependence between these variables. This is to be expected as the underlying measures are associated with rather different properties.

6.4.1 Intraday analysis of scaling patterns

We investigated the intraday patterns by separating the paths of $H^*(t)$ and $C^*(t)$ into daily windows. Taking for example the time series $H^*(t)$ for the S&P 500 index, Figure 6.6(a), we separated this time series into the 105 days which compose the data set, see Table 6.1. In Figure 6.8(a), we display these daily time series (one day on top of the other). The colour



(a) S&P 500

(b) IPC

(c) Nikkei 225

(d) XU 100

Fig. 6.7 Time-dependent complexity measure, $C^*(t)$, for four different stock market indices for the period January 15th, 2014 to June 16th, 2014. (a) S&P 500, (b) IPC, (c) Nikkei 225 and (d) XU 100.

bar represents the value of $H^*(t)$. This graphical representation allows us to compare the trading sessions dynamics and to identify patterns at specific times of the day.

We estimated the statistical mean of $H^*(t)$ across the days, resulting in an average value for each time t of the trading session. This average, denoted as $\langle H^*(t)\rangle_{\text{days}}$, describes the regular behaviour of $H^*(t)$ on a trading session, see Figure 6.8(b).

In order to validate that the observed dynamics of $\langle H^*(t)\rangle_{\text{days}}$ are statistically significant, we compared these dynamics with scaling exponents obtained from several simulations of Brownian motion of length equal to the analysed financial time series, see Table 6.1. We denote the Brownian motion scaling exponents as $H^*_{BM}(t)$. The time series of $H^*_{BM}(t)$ were fragmented into l windows of equal length and equal to a trading day of the analysed stock market index. The mean over these l windows is denoted as $\langle H^*_{BM}(t)\rangle$. The pink band reported in Figure 6.8(b) corresponds to the 5th and the 95th percentiles of the empirical distribution of $\langle H^*_{BM}(t)\rangle$ computed from 100 simulations.

We compared the values of $H^*(t)$ obtained for each stock market index with the $\langle H^*_{BM}(t)\rangle$ band. At each time t during the trading session, we estimated the relative fraction of $H^*(t)$ values that falls outside the pink band. In Figure 6.8(c), we report these results as a ratio of number of days outside the band divided by the total number of days. This ratio is labelled as likelihood. The colour bar of this figure represents the value of the average scaling exponent, i.e., $\langle H^*(t)\rangle_{\text{days}}$ (value plotted in Figure 6.8(b)). From this Figure, we observe that across the day there are periods of time with very high empirical probability of observing values of the scaling exponents significantly different from the corresponding values extracted from pure Brownian motion.

The mean of the complexity measure, $C^*(t)$, at each time t of the trading session is shown in Figure 6.8(d). Equally as with the $H^*(t)$ exponent, the mean of $C^*(t)$ is computed across the 105 days and it is denoted as $\langle C^*(t)\rangle_{\text{days}}$. The same daily analysis for the remaining three stock market indices is reported in Figures 6.9, 6.10 and 6.11, respectively.

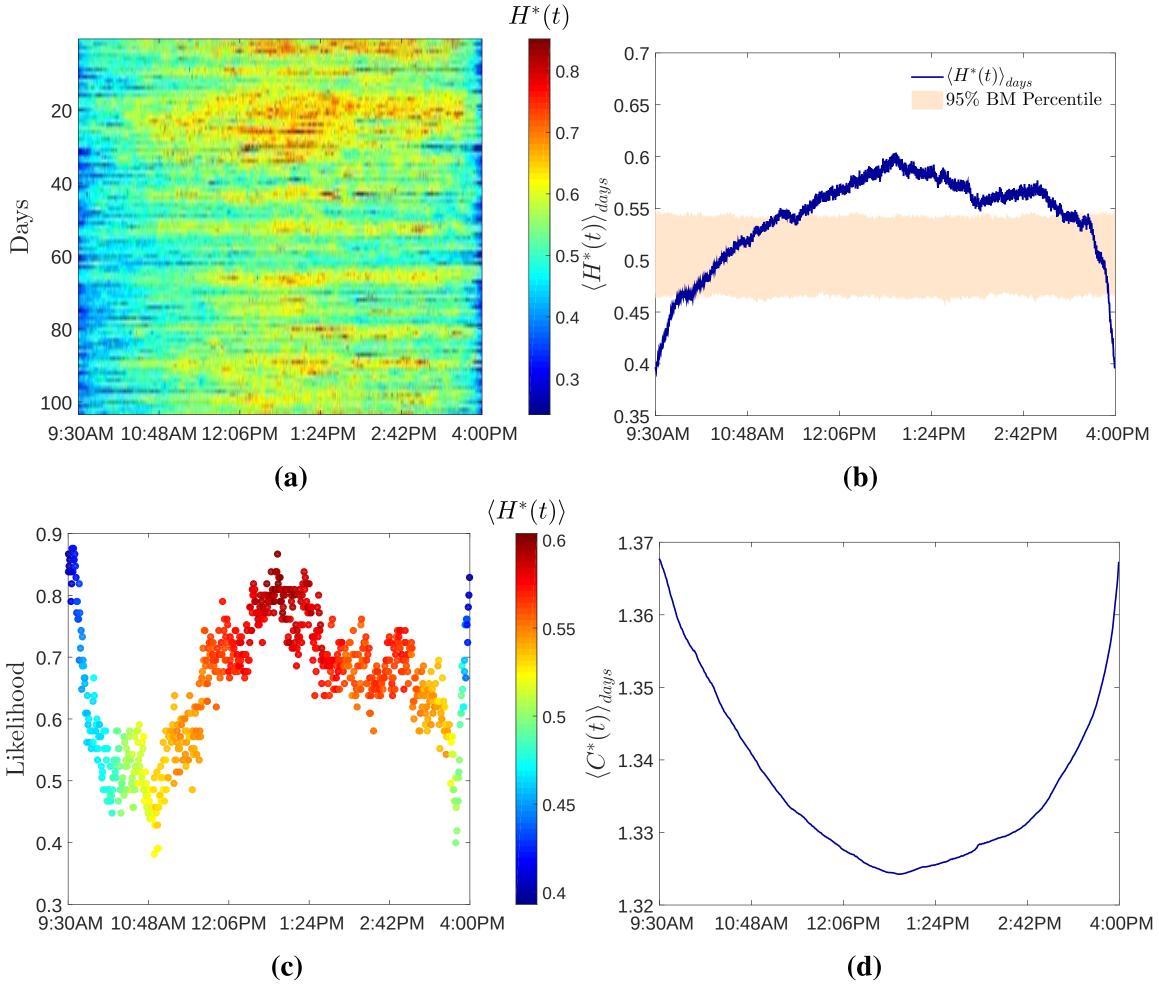
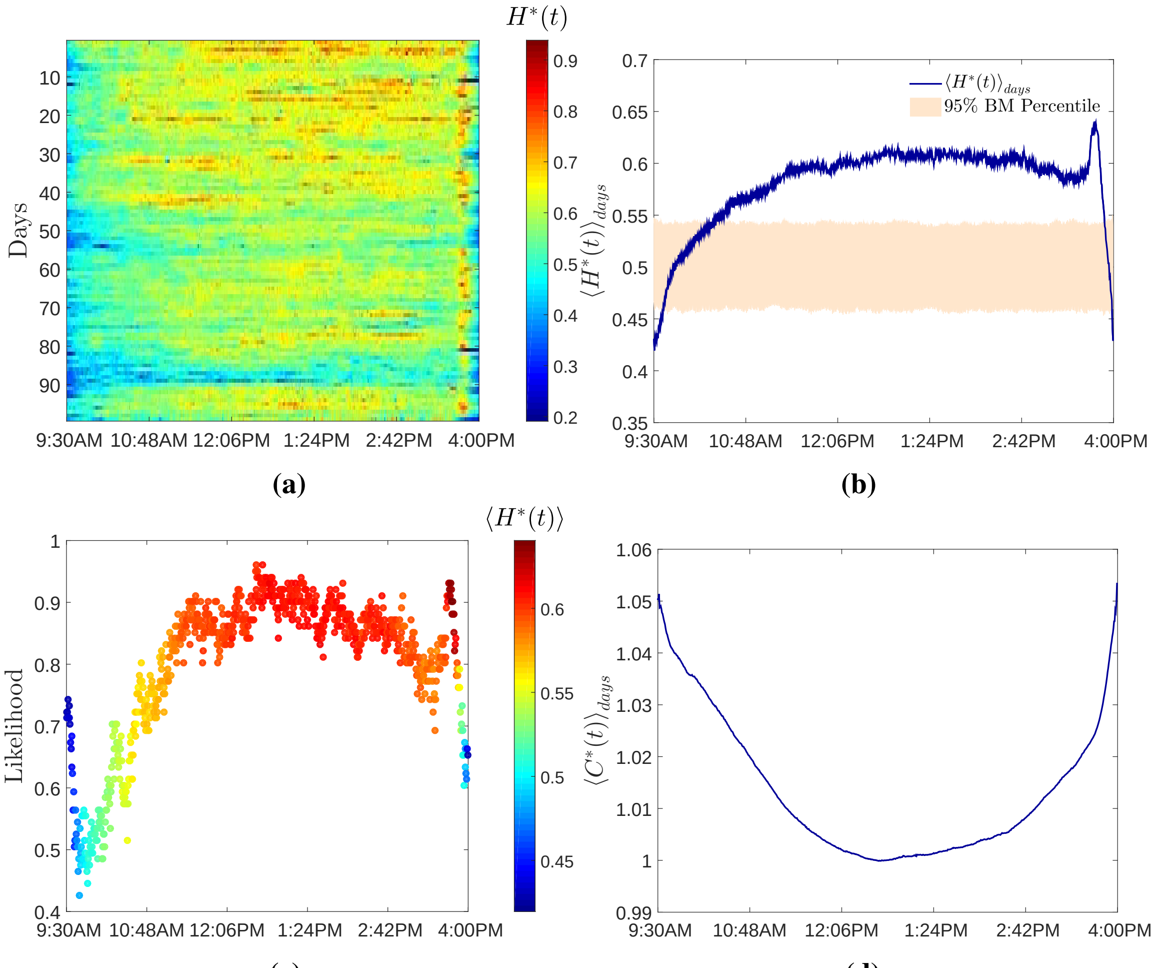


Fig. 6.8 Intraday analysis of the S&P 500 index.
(a) Intraday dynamics of the scaling exponent $H^*(t)$ as a function of day and time. The colour bar indicates the value of $H^*(t)$.
(b) Mean of $H^*(t)$ over the 105 days, denoted as $\langle H^*(t)\rangle_{\text{days}}$. The pink band corresponds to the 5th and 95th percentile of the distribution of $\langle H^*_{BM}(t)\rangle$. The distribution is estimated using 100 simulations of Brownian motion.
(c) Likelihood of $H^*(t)$ to fall outside the 5th and 95th percentile band for Brownian motion (pink band of Figure (b)). The colour bar indicates $\langle H^*(t)\rangle_{\text{days}}$, the value shown in Figure (b).
(d) Mean of the windowed complexity measure, denoted as $\langle C^*(t)\rangle_{\text{days}}$.



(a) **(b)**

(c) **(d)**

Fig. 6.9 Intraday analysis for the IPC index. Caption for sub-figures (a), (b), (c) and (d) same as Figure 6.8.

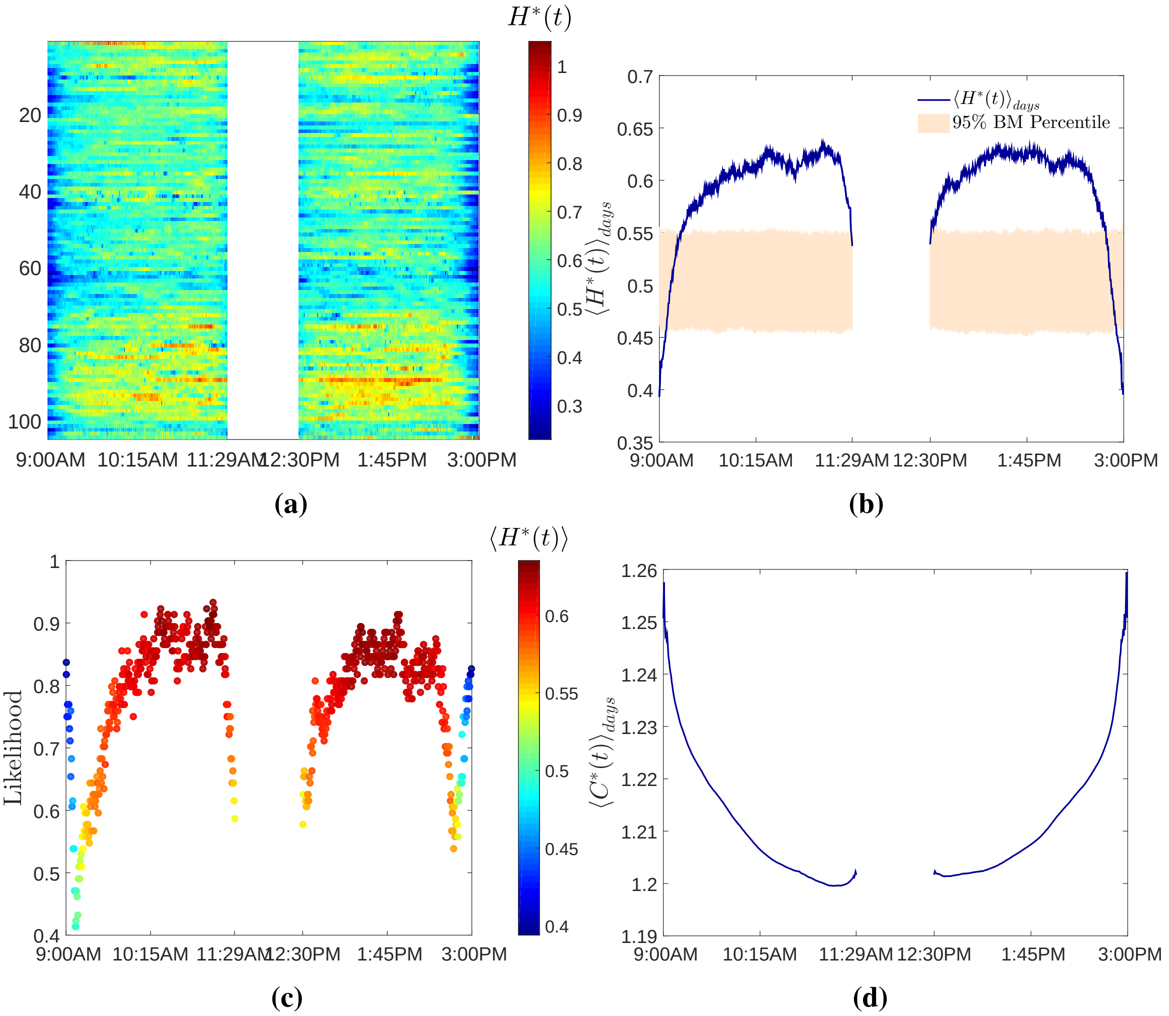
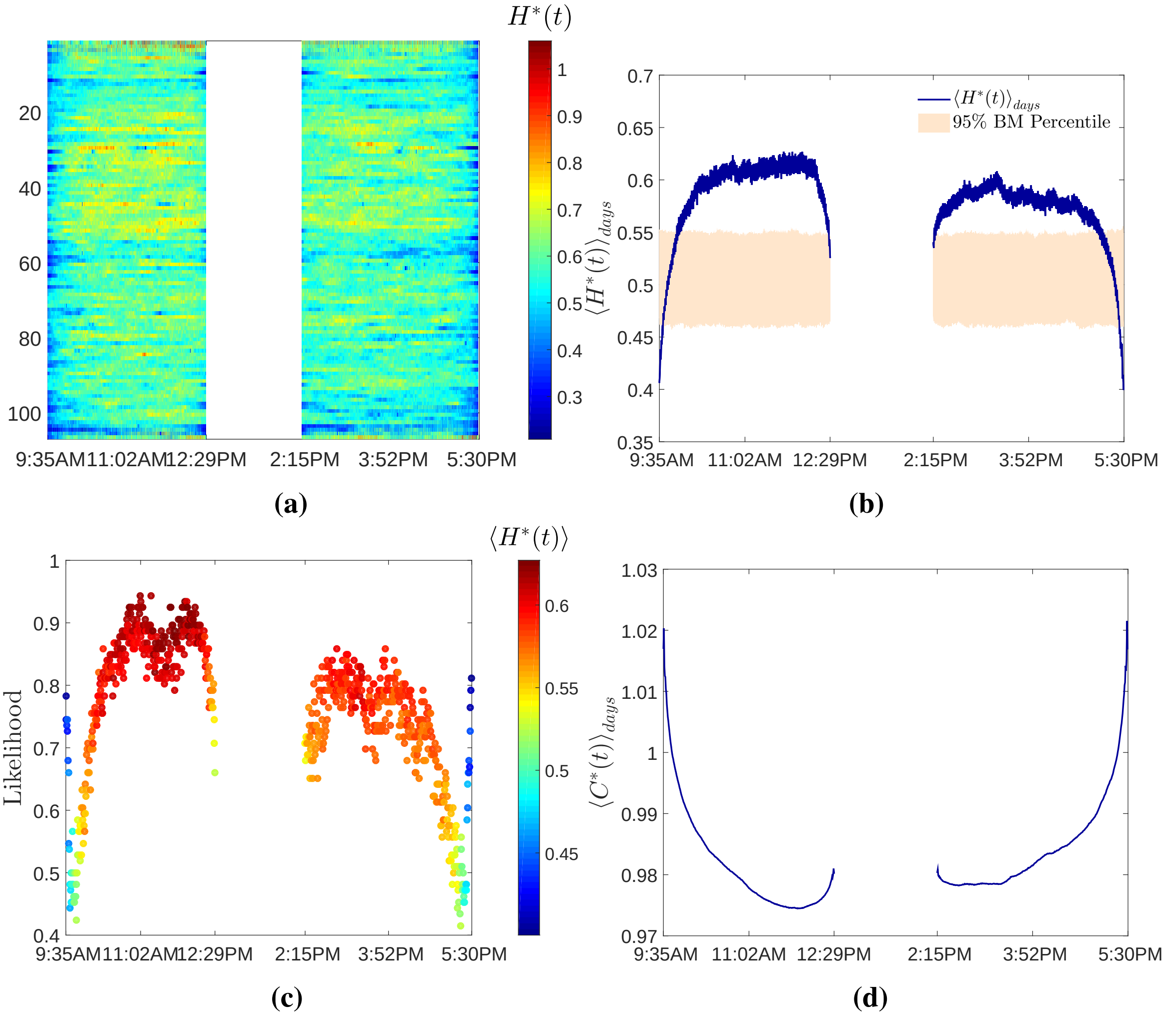


Fig. 6.10 Intraday analysis for the Nikkei 225 index. Caption for sub-figures (a), (b), (c) and (d) same as Figure 6.8. The white vertical band in each sub-figure corresponds to the lunch break in this stock exchange.



(a) (b)

(c) (d)

Fig. 6.11 Intraday analysis for the XU 100 index. Caption for sub-figures (a), (b), (c) and (d) same as Figure 6.8. The white vertical band in each sub-figure corresponds to the lunch break in this stock exchange.

Overall, from the intraday scaling and complexity measures we observed the following patterns:

- For each stock market index, the daily average $\langle H^*(t)\rangle_{\text{days}}$ displays an inverted U-shaped form that reflects a more chaotic behaviour at the beginning and at the end of the trading session. The opposite behaviour is observed for the complexity measure, $\langle C^*(t)\rangle_{\text{days}}$, which reveals a U-shaped form.

- The S&P 500 index displays the largest values of $H^*(t)$ (a stronger amplitude persistent behaviour) during the middle of the trading session. From Figure 6.8(a), we observe that most of the days present large values of $H^*(t)$ around midday. At this time, the average exponent $\langle H^*(t)\rangle_{\text{days}}$ reaches a value of 0.6, see Figure 6.8(b). These values are significantly different from what would be expected from Brownian motion with more than 80% of the observations outside the 5^{th} and the 95^{th} percentiles, see Figure 6.8(c). Consistently, the complexity measure reaches its minimum at the same time of the trading session, see Figure 6.8(d).

- The Mexican stock exchange is characterized by some large scaling values at the middle of the day. However the most noticeable pattern is the large values just before the end of the trading session, see Figure 6.9(a). The mean of the windowed values reaches a maximum of 0.65, creating an upswing shape at this time, see Figure 6.9(b). This could be associated with an increase of the trading activity in the last few minutes of the session, creating an extreme change in the amplitudes. This pattern is only present in the Mexican stock exchange.

  From Figure 6.9(c), we observe that some minutes before the closing of the market, more than 90% of the local scaling exponents fall outside the 5^{th} and the 95^{th} percentile band for Brownian motion. The complexity measure also reflects a steep increase of disorder at the end of the trading session, see Figure 6.9(d).

- The Japanese and Turkish stock exchanges display two regions of large values for the scaling exponent. These regions are separated by the lunch break, see Figures 6.10(a) and 6.11(a) respectively. The mean of the scaling exponent reflects a quasi-double inverted U-shaped form that is associated with the opening and closing of the morning and afternoon sessions, Figures 6.10(b) and 6.11(b). It is worth noting that the two trading sessions do not display exactly the same profile. For the Japanese stock market index, the inverted U-shaped form of the first trading session is slightly skewed to the right, in comparison with the more symmetric shape of the second trading session, see Figure 6.10(b).

- For the Turkish stock exchange, we observe that the first part of the trading session shows larger values of $H^*(t)$. More than 90% of the analysed days present local scaling exponents which exceed the value of 0.6, see Figure 6.11(c). The dominance of one IMF amplitude in the first trading session is also reflected in the lower values of the complexity measure, which reaches the lowest value when compared to the other stock market indices, see Figure 6.11(d).

Overall the intraday patterns of $H^*(t)$ and $C^*(t)$ confirm the well known fact that activity on financial markets is not constant throughout the day. The uncovered patterns corroborate the hectic buy and sell activity affecting different markets at the opening and closing of trading sessions [9, 44]. Smaller values of $H^*(t)$ (large values of $C^*(t)$) imply a non-persistent and rougher behaviour which is reflected in higher volatility. The exposed daily patterns of $\langle H^*(t)\rangle_{\mathrm{days}}$ and $\langle C^*(t)\rangle_{\mathrm{days}}$ are in agreement with the results which document the existence of a distinct U-shaped pattern in market activity and volatility over a trading day, i.e., volatility is higher at the opening and at the closing of the trading session and low in the middle of the day, see for example references [7, 36, 180].

By comparing the intraday complexity values across the four markets, we observe that the S&P 500 index displays the largest values across all the trading session. Nikkei 225 index is the second most complex, followed by the IPC index and lastly the XU100 index, which has the smaller complexity values. This is in agreement with the results reported for developed and emerging markets [65, 109, 190].

Part of the observed dynamics could be explained by the intraday volatility patterns related to the periodic arrival of information to the market, news arrive at precise moments during the trading session. This cyclical pattern has an obvious influence over the long memory displayed by the stock market indices. It is also well documented that financial time-series exhibit intermittent behaviour as well as other statistical properties collectively known as stylized facts, including multifractal behaviour. Calvet et al. [41] laid the theoretical and practical formulation of multifractal analysis applied to financial time-series and propose a model able to explain most of the stylized facts observed in financial time-series.

6.5 Summary

We proposed two new time-dependent measures: 1) an amplitude scaling exponent and 2) an entropy-like complexity measure. Our measures are non-parametric and they do not assume any a priori stochastic process. The scaling exponent only assumes the existence of a power-law relation between the instantaneous amplitudes and the instantaneous periods which was empirically shown to be present. By applying our methodology to different models with known scaling laws, we demonstrated that the time-dependent values of our scaling exponent vary consistently around the known exponent of the models. When applied to real financial data, our measures uncover significant variations of the scaling properties of the market during the trading day and provide evidences of non-stationary patterns. We verified that the scaling exponent and the complexity measure are related, with larger scaling exponents associated with lower complexity values. However, the proposed measures convey different information about the properties of the oscillating components of the signal. Specifically, we applied the scaling and the entropy-like measures to the study of four financial markets, two developed (US and Japan) and two emerging markets (Mexico and Turkey). We contrasted and compared the decomposition of their financial indices. With the use of intraday data, we recognized some patterns and identified periods of low and high complexity.

Compared to the other analysed stock market indices, the S&P 500 index results the most complex. The intraday analysis reveals a distinctive anti-persistent behaviour at the opening and at the closing of the trading session, contrasting with the persistent behaviour at the middle of the session. Similar intraday results are obtained for the other stock market indices. The variations observed in the scaling measure are well outside the 5^{th} and the 95^{th} percentiles expected for Brownian motion, suggesting strong deviations from this model that could be attributed to the presence of long-range dependence or/and heavy tails.

With the proposed measures, we are able to describe the dynamics of financial time series whose regularity changes over time. Our results suggest that financial time series have dynamic scaling properties that change during the trading day following identifiable patterns that are characteristic of each market. The origin of the scaling laws could be attributed to the autocorrelation of the process, the presence of heavy tails and the non-stationarity of the time series. It is beyond the purpose of this paper to investigate this aspect (which is discussed in [27] by using a different approach). Our aim here was to uncover non-stationary scaling patterns which we showed to be significant, reproducible and characteristic of specific stock markets. These non-stationary scaling patterns must be considered when modelling financial time series and building trading strategies.

Chapter 7

EMD-Based Forecasting Models

In this chapter, we introduce some multistep-ahead forecasting models based on EMD combined with support vector regression (SVR). By separating the input time series into a finite set of intrinsic mode functions, we reduce the complexity of the forecasting task. The novelty of the proposed models is the inclusion of a coarse-to-fine reconstruction step to analyse the forecasting capabilities of a combination of IMFs. Denoised time series are used to forecast short-term horizons and the trend of the time series is used to forecast long-term horizons. We include an example which provides specific details of how intraday forecasting is performed. We generalize the results for a data set consisting of six months of intraday data. Our models are compared with benchmark models commonly used in the literature, testing the null hypothesis of no difference in the accuracy of two competing models.

7.1 Forecasting financial time series

Forecasting financial data is regarded as a challenging task and it remains a very active research area. Over the last decades, efforts to improve forecasting techniques have included the EMD as a preprocessor tool. This timescale decomposition has proved to be an efficient methodology following the "divide and conquer" philosophy [184], which consists of three main steps:

1. The EMD is used to separate the initial time series into a finite set of IMFs and a residue.
2. Forecasting techniques such as autoregressive models, artificial neural networks (ANN), and support vector machines are applied to forecast each IMF and the residue.
3. The forecasted components are combined to forecast the initial time series.

The “divide and conquer” philosophy has been widely used in different areas, to mention a few: crude oil spot prices [184], foreign exchange rates [113], market stock indices [49], wind speed [173], computer sales [117], tourism demand [45]. Within the mentioned studies, it is common practice to divide the time series into a training set and a testing set. The former is used to calibrate the model, the latter to measure the performance of the model. However, in the cited studies, authors applied the EMD to the undivided data, including future observations in the training set. We consider this approach inadequate, due to the fact that the EMD is a data based algorithm which is highly influenced by the local maxima and the local minima. The use of future observations as the training set could explain the good performance of some of the proposed EMD-based models.

Liu et al. [115] proposed a hybrid EMD-ANN model to forecast one, two and three steps ahead of wind speed time series. The approach taken in this study is improved since the EMD is applied every time a new observation is included and a new forecasting model is retrained. This computationally more expensive model outperforms benchmark autoregressive models and ANN models which are trained on the initial time series without preprocessing the data using the EMD.

Zeng and Qu [185] showed the forecasting effectiveness of the EMD by analysing the Baltic Dry Index (BDI). The initial BDI time series is decomposed into several independent functions using the EMD. These functions are grouped into three components, namely, high-frequency, low-frequency and long-term trend component. An ANN is used to model each of the components and the prediction results of all the sets are combined to formulate the prediction output for the initial BDI time series. The proposed model is compared with existing ANN models and traditional econometric models such as vector autoregression (VAR), obtaining better results with the EMD-based method.

In the following sections we provide a brief review of two benchmarks models that are widely used in the forecasting literature: autoregressive models and support vector regression.

7.2 EMD-SVR forecasting models

Decomposition is a critical step to analyse complex data, information contained in each component is analysed separately to reduce the complexity and improve the forecasting accuracy. Literature related to the applications of the EMD as a forecasting tool was discussed in Section 7.1

Based on EMD and SVR, we introduce some novel multistep-ahead forecasting models using the direct and the recursive strategies. For more details about SVR and forecasting strategies refer to Section 7.1. Making use of the extracted IMFs, we propose two forecasting schemes: the univariate EMD-SVR and the multivariate EMD-SVR.

Given the time series X_t, $t = 1,2,\ldots,N$, our aim is to predict h steps ahead.

7.2.1 Univariate EMD-SVR

The proposed univariate EMD-SVR approach forecasts each IMF and the residue independently using SVR. The forecasted values are combined to obtain the final forecast for the input time series. This forecasting approach consists of the following steps:

1. Apply the EMD to the training time series X_t, $t = 1,,2,\ldots,N$ and extract a set of n IMFs and a residue.

2. Scale the IMFs and the residue. Linear transformation to adjust the data to the interval $[0,1]$ is used. Denote by y the input time series (each IMFs and the residue), the scaled version of it is calculated as:

$$y_s = \frac{y - y_{min}}{y_{max} - y_{min}}, \tag{7.1}$$

   where y_{min} and y_{max} denote the minimum and maximum values, respectively. This adjustment is a prerequisite for fast convergence of the algorithm. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. Another advantage is to prevent numerical difficulties during the calculation.

3. Select the input vector to train the SVR. The input vector consists of previous values of the time series to be forecasted, i.e., the n IMFs and the residue. In order to achieve a better model accuracy, we need to construct input vectors with the most information and capable of fitting the training data.

4. Train the $n+1$ SVR models. Consider both the direct and the recursive strategies when training an SVR model for each IMF_i, $i = 1,\ldots,n$ and for the residue. The regularization constant C, the insensitive coefficient ε and the kernel width parameter γ are selected using a grid search and cross-validation.

5. Forecast each IMF and the residue separately. Use both the direct or the recursive strategy.

6. Combine the forecasted IMFs and the residue to forecast the initial time series up to the value x_{t+h}, i.e., for the forecast horizon h. In this step, the forecasted IMFs are

used to generate multiple partially reconstructed models. We define a coarse-to-fine forecasting scheme by using the cumulative sum of sequential IMFs, i.e., adding more details to the low frequency modes. We define $n+1$ coarse-to-fine models as:

$$\begin{aligned} M_1(t) &= R(t), \\ M_2(t) &= R(t) + \sum_{i=n}^{n} IMF_i(t), \\ M_3(t) &= R(t) + \sum_{i=n-1}^{n} IMF_i(t), \\ &\vdots \\ M_{n+1}(t) &= R(t) + \sum_{i=1}^{n} IMF_i(t). \end{aligned}$$

7. Measure the performance of each model at the forecast horizon h.

7.2.2 Multivariate EMD-SVR

The proposed multivariate EMD-SVR scheme is very similar to the univariate EMD-SVR, the main difference resides in the construction of the input vector. For the multivariate EMD-SVR scheme, a vector combining historical data from all the IMFs and the residue is generated to train a single SVR model, hence, the forecasted value is a function of all the IMFs and the residue.

$$\begin{aligned} \hat{x}_{t+h} =& f(IMF_1(t),\dots,IMF_1(t-m_i+1),\dots,IMF_n(t),\dots, \\ & IMF_n(t-m_i+1),R(t),\dots,R(t-m_R+1)), \end{aligned} \tag{7.2}$$

The embedded dimension m_i is specific to each IMF_i and to the residue. The advantage of this method is that a single forecasting model needs to be trained, and therefore it is a faster algorithm. The algorithm steps can be summarized as follow:

1. Apply the EMD to the training time series X_t, $t = 1,,2,\dots,N$ and extract a set of n IMFs and a residue.

2. Scale the IMFs and the residue to the interval $[0,1]$.

3. Select the input vector. The input vector consists of a combination of previous values of all the IMFs and the residue.

4. Train an SVR model for each forecast horizon h using the direct strategy. For the multivariate scheme we can not easily implement the recursive strategy. This would imply knowing forecasted values of each IMF that can only be obtained by individually forecasting each IMF (univariate scheme).
5. Forecast the value x_{t+h}.
6. Measure the performance of this model at the forecast horizon h.

7.3 Forecasting intraday financial data

In order to test the proposed EMD-SVR forecasting models, we use two well known stock market indices, the S&P 500 index and the FTSE 100 index. For both indices, the complete data set consists of 128 days of intraday data sample at 30-second intervals. We forecast intraday data for each single day, analysing a training sample of $N = 500$ prices at the start of the trading session and forecasting up to $h = 50$ steps ahead of the last part of the trading session. Given the sampling frequency of the data, $h = 50$ steps ahead means the following 25 minutes.

7.3.1 Intraday forecasting, example on a single time series

For the sake of clarity, let us first exemplify our forecasting process on one randomly chosen time series of the S&P 500 index (August 7$^{\text{th}}$ 2014), keeping in mind that we performed the same analysis on the remaining time series of both stock market indices. Figure 7.1 illustrates the time series used as an example. We did not include the first 5 minutes of the daily prices.

We analysed a training sample of $N = 500$ prices and aim to forecast up to $h = 50$ steps ahead. The first step of the forecasting algorithms is the application of the EMD to the training data. We obtained five IMFs and a residue. To create the input vectors for the univariate EMD-SVR and the multivariate EMD-SVR schemes, we tested three input vectors of different length m, as follows:

1. $m = 1$ lagged values of each IMF and the residue.
2. $m = 5$ lagged values of each IMF and the residue.
3. $m = p + d$, where p denotes the number of autoregressive terms and d is the number of differentiations of an ARIMA model that was fitted to each of the IMFs and to the residue. For the implementation of the ARIMA models, the auto.arima function available in software R was used [102].

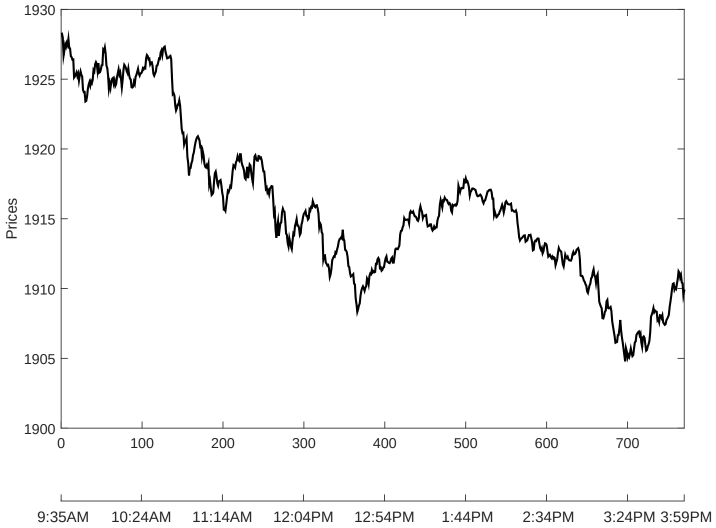


Fig. 7.1 S&P 500 index for the trading day, August 7th 2014

To perform all the SVR analysis, we used the LIBSVM software system[1]. We applied support vector regression with Gaussian kernel and performed grid search to find the optimal parameters: regularization parameter C, the Gaussian kernel's bandwidth γ and the precision parameter ε. The grid consisted of a three-dimensional parameter space $\log_{10}(C) \in (-4,4)$, $\log_{10}(\gamma) \in (-4,4)$, and $\log_{10}(\varepsilon) \in (-4,0)$. A six-fold moving validation is used in each iteration of the grid search to avoid over-fitting, see Section 2.8.3.

Let us describe separately the univariate EMD-SVR and the multivariate EMD-SVR frameworks.

Univariate EMD-SVR results

For the univariate EMD-SVR forecasting scheme, the five IMFs and the residue are independently forecasted. We considered three input vectors with different lengths m, but for this example we only report the results for the ARIMA-based input vector, $m = p + d$, which produced the best results. To estimate the values of p and d, we fitted an ARIMA model to each IMFs and to the residue. The obtained number of autoregressive terms, p and the number of differentiations, d are reported in Table 7.1.

The univariate EMD-SVR approach can be implemented using both the recursive and the direct strategies. For the recursive strategy, we forecasted $h = 1, 2, \ldots, 50$ steps ahead, but we only report eight values of h, $h \in [1, 2, 3, 5, 10, 20, 30, 50]$ in order to compare them with the direct strategy. Note that for the direct strategy a model is trained for each step ahead, thus, eight different models are trained for each IMF and eight models are trained for

[1] Available at https://www.csie.ntu.edu.tw/~cjlin/libsvm/.

	AR (p)	MA (q)	Differencing (d)	m
$\boldsymbol{IMF_1}$	2	1	0	2
$\boldsymbol{IMF_2}$	2	5	0	2
$\boldsymbol{IMF_3}$	5	1	0	5
$\boldsymbol{IMF_4}$	5	3	0	5
$\boldsymbol{IMF_5}$	0	3	1	1
Residue	2	2	2	4

Table 7.1 Order of the ARIMA models fitted to each IMF and to the residue. The number of lagged values $m = p + d$ is used to construct the input vectors for the EMD-SVR models.

the residue.

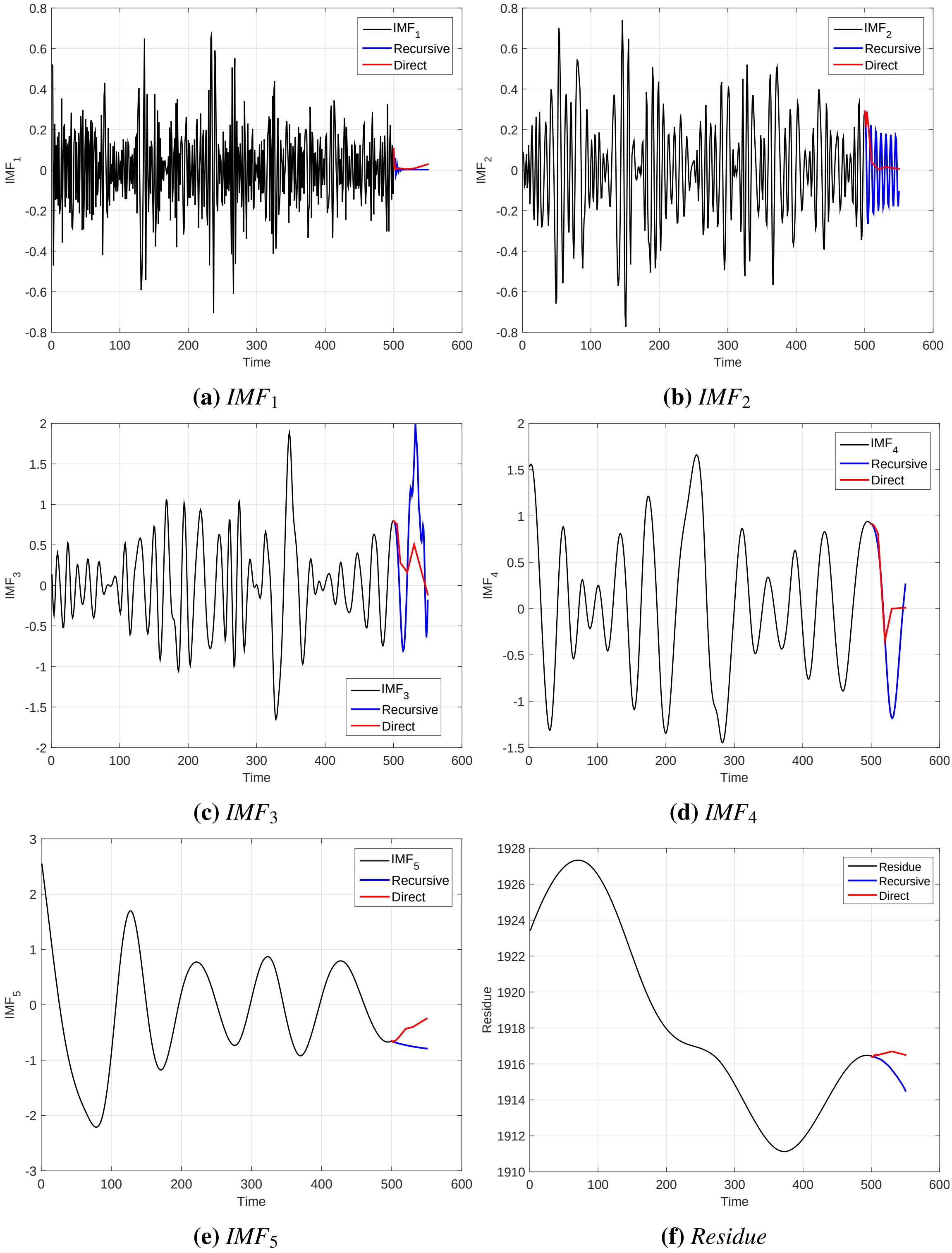
In Figure 7.2, we compare the forecasted IMFs using the recursive and the direct strategies. Subfigure 7.2(a) corresponds to the first IMF. Subfigure 7.2(b) illustrates the results of the second IMF, and so on, until Subfigure 7.2(f) which shows the forecasted residue. The black line in each plot represents the input IMF, the blue line represents the recursive strategy used to forecast up to 50 steps ahead. Finally, the red line corresponds to the direct strategy for steps $h \in [1,2,3,5,10,20,30,50]$. We observe that both strategies capture some of the oscillating patterns of the IMFs.

The forecasted values of the IMFs and the residue are used to generate a coarse-to-fine reconstruction which generates six forecasting models. The results of these models are illustrated in Figure 7.3. The first coarse model only considers the residue and it is denoted as R, see Figure 7.3(a). The second model uses the forecasting of the residue plus the fifth IMF ($R + IMF_5$), see Figure 7.3(b). We continue this process until we have included all the IMFs, ($R + \sum_{i=1}^{5} IMF_i$), see Figure 7.3(f). The observed stock market index is shown in a black line. The forecasted values are represented by a blue line, recursive strategy and by a red line, direct strategy.

Multivariate EMD-SVR results

For the multivariate EMD-SVR, we created an input vector combining the lagged values of all the IMFs and the residue. We tested three input vector with different lenghts $m = 1$, $m = 5$ and $m = p + d$, but in this example we only report the results for the ARIMA-based input vector, $m = p + d$ which produced the best results. Using the direct strategy, we trained an SVR model for each of the forecast horizons $h \in [1,2,3,5,10,20,30,50]$.

In Figure 7.4, we present the results of the multivariate EMD-SVR models. The black line represents the observed stock market index and the red line the forecasted values.



(a) *IMF*$_1$

(b) *IMF*$_2$

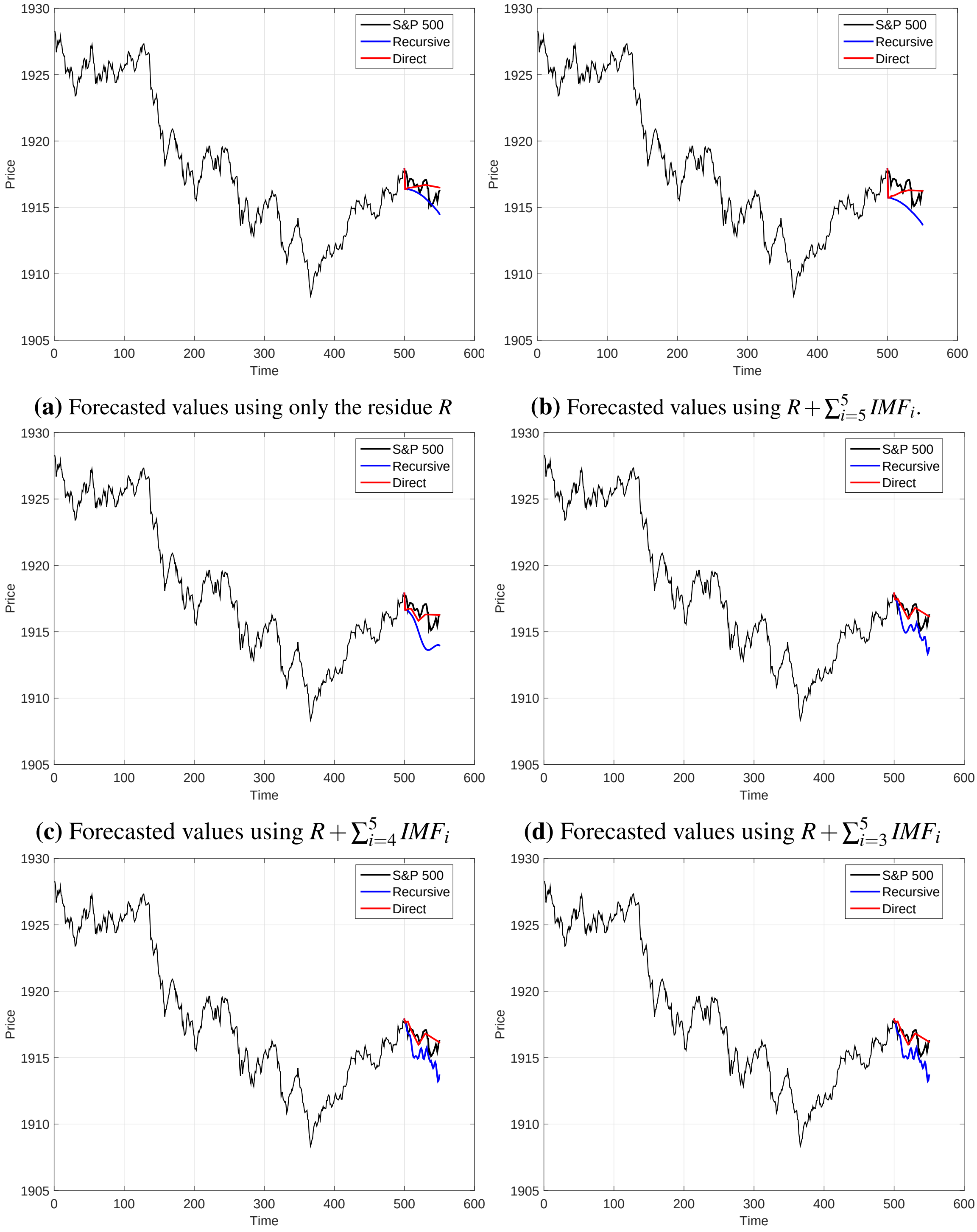
(c) *IMF*$_3$

(d) *IMF*$_4$

(e) *IMF*$_5$

(f) *Residue*

Fig. 7.2 Actual and forecasted IMFs and residue extracted from the S&P 500 index shown in Figure 7.1. The forecasted values were obtained using the univariate EMD-SVR model, both the recursive and the direct strategies.



(a) Forecasted values using only the residue R

(b) Forecasted values using $R+\sum_{i=5}^{5} IMF_i$.

(c) Forecasted values using $R+\sum_{i=4}^{5} IMF_i$

(d) Forecasted values using $R+\sum_{i=3}^{5} IMF_i$

(e) Forecasted values using $R+\sum_{i=2}^{5} IMF_i$

(f) Forecasted values using $R+\sum_{i=1}^{5} IMF_i$

Fig. 7.3 Actual and forecasted values for the S&P 500 index shown in Figure 7.1. Forecasted values were obtained using partial reconstructions of the univariate EMD-SVR model, both the recursive and the direct strategies.

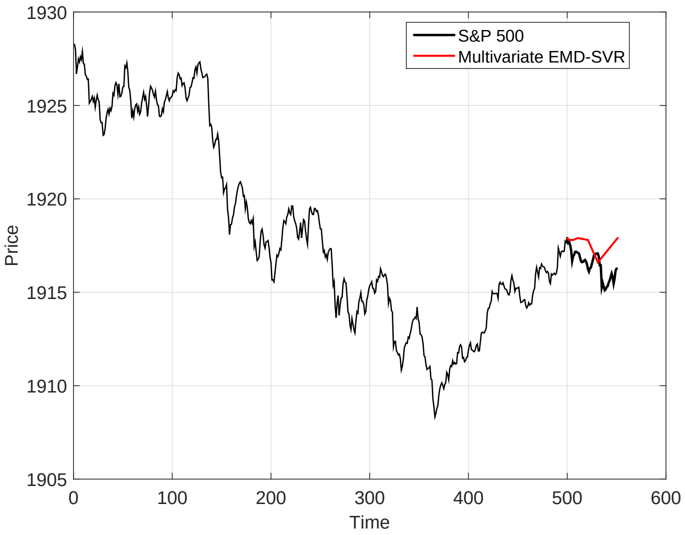


Fig. 7.4 Actual and forecasted values for the S&P 500 index shown in Figure 7.1. Forecasted values were obtained using the multivariate EMD-SVR model.

7.3.2 Intraday forecasting, analysis on the complete data set

We did the same intraday forecasting for each of the 128 daily time series of the S&P 500 index. The first step is the application of the EMD and in order to fairly compare the forecasting capabilities of the trend and the gradually reconstructed time series, each time series was decomposed into five IMFs and a residue. We fitted ARIMA and SVR models to each intraday time series and its respective IMFs. A summary of the parameters obtained for the ARIMA models is reported in Appendix A.

For the error valuation we used the mean absolute error (MAE), a conventional metric to measure forecasting performance of a model and which is defined as:

$$MAE = \frac{1}{M}\sum_{i=1}^{M} |\hat{x}_i - x_i|, \tag{7.3}$$

where x_i denotes the original measured data, $\hat{x}_i$ denotes the forecasted data, and M represents the number of data points. Another commonly used error measure is the root mean squared error, however, this measure is more sensitive to the occasional large errors due to the squaring process [179].

To estimate the performance of the proposed models at each forecast horizons $h \in [1,2,3,5,10,20,30,50]$, we calculated the MAE over the 128 time series. The proposed EMD-SVR models are compared with other commonly used benchmark models which are applied to the initial time series (without the EMD). The benchmark models can be listed as:

- Naive model. Keeps constant the last observed value in the time series.
- ARIMA model.
- SVR model. We used a RBF kernel, and a three-dimensional grid to look for the optimal parameters: $\log_{10}(C) \in (-4,4)$, $\log_{10}(\gamma) \in (-4,4)$, and $\log_{10}(\varepsilon) \in (-4,0)$. A six-fold moving validation is used in each iteration of the grid search for parameter tuning. In Appendix A, we report the obtained parameters for this SVR model.

In Table 7.2, we report the MAE and its standard deviation (parentheses) for the EMD-SVR models with input vector of length $m = 1$. We include the MAE for the naive, the ARIMA and the SVR models applied to the initial time series. The MAE for the EMD-SVR models with input vector of length $m = 5$ and $m = p + d$ are reported in Table 7.3 and Table 7.4, respectively. For comparison reasons, in each table we repeated the results of the benchmark models. For each forecast horizon, the smallest error across the compared models is set in boldface. The smallest error across the three tables is indicated with a dagger (†).

In Figures 7.5, 7.6, 7.7, we show the graphical representation of previous tables. The MAE versus the forecast horizon for the EMD-SVR models with $m = 1$ lagged values as input vector are displayed in Figure 7.5. This figure also includes the errors for the benchmark models.

Subfigures 7.5(a) and 7.5(b) show the MAE for the univariate EMD-SVR models, direct and recursive strategies, respectively. The dotted lines indicate the MAE of the coarse-to-fine reconstruction models. The light blue dotted line indicates the errors of the forecasting model which uses the residue only. The green dotted line indicates the error of the models using the residue plus the fifth IMF, $R + \sum_5^5 IMF_i$, and so on, until the model which includes the residue and all the IMFs, $R + \sum_1^5 IMF_i$.
Subfigure 7.5(c) illustrates the MAE for the multivariate EMD-SRV model.

Figures 7.6 and 7.7 illustrate the errors for models with input vector of length $m = 5$ and $m = p + d$, respectively. The MAE for naive and ARIMA models are merely repeated.

As for comparison across the forecasting models, the experimental results reveal some interesting facts that could be used for common forecasting practice.

- The MAE increases as the forecast horizon does.
- The direct strategy achieves more accurate forecasts than the recursive strategy in almost all the tested models. It is possible that the inferiority of the recursive strategy is due to the accumulation of errors, deteriorating the accuracy of the forecast.

- Providing a basic input vector of length $m = 1$ produces EMD-SVR models with limited capabilities. Unexpectedly, the recursive strategy outperforms the direct strategy, see Figure 7.5. We attribute the poor performance of the direct strategy to the limited information contained in the input vector with $m = 1$ lagged values. Differently, the recursive strategy keeps building up by using previous information to forecast the target horizon. The SVR model is badly trained and does not outperform the naive or the ARIMA models. None of the proposed EMD-SVR models consistently outperform the naive model, see Figure 7.5.

- For the input vector of length $m = 5$, the direct strategy outperforms the recursive strategy. For the recursive strategy, the EMD-SVR models that include all the IMFs and residue have smaller errors for one step ahead forecast, see Figure 7.6(a). However, better results are obtained with the univariate EMD-SVR, direct strategy. For short-term horizons (1-10 steps-ahead) using all the IMFs, $(R + \sum_{i=1}^{5} IMF_i)$, or denoising the time series by not considering the high-frequency component in the reconstruction process $(R + \sum_{i=2}^{5} IMF_i)$, produce the smallest MAE, see Figure 7.6(b). For long-term forecasting (20-50 steps ahead), the coarse approximations (R or $R + \sum_{i=5}^{5} IMF_i$) produce smaller errors than the benchmark models and than the models that include all the IMFs.

  The multivariate EMD-SVR model outperforms the benchmark models, see Figure 7.6(c).

- For an input vector of length $m = p + d$, we observe that the smallest errors (marked with a dagger in Table 7.4) are produced with this number of lagged values. Although similar results are obtained for models with input vector $m = 5$ since $m = p + d$ is close to five, refer to Appendix B.

  For the univariate EMD-SVR model, recursive strategy, smaller errors are produced by a model that includes all the IMFs and the residue but only for short-term horizons, see Figure 7.7(a). For the univariate EMD-SVR model, direct strategy, we observe the same pattern where the coarse reconstruction produces smaller errors for the long-term horizons and outperforms the benchmark models. For short-term forecasting, the complete reconstruction and the denoised time series are the most accurate models, see Figure 7.7(b).

  The multivariate EMD-SVR models produces similar errors as the pure SVR, see Figure 7.7(c).

- A coarse-to fine-reconstruction confirms that the trend possesses enough information to forecast the long-term horizons and that we may neglect some of the highest frequency IMFs.

	Steps ahead	1		2		3		5		10		20		30		50	
	Model	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
	Naive	0.147	(0.186)	0.256	(0.322)	0.328	(0.418)	0.423	(0.514)	0.651	(0.801)	0.991	(1.053)	1.143	(1.294)	1.559	(1.803)
	ARIMA	**0.145**	(0.185)	**0.247**	(0.303)	**0.313**	(0.374)	0.421	(0.477)	0.663	(0.783)	1.021	(1.035)	1.154	(1.304)	1.644	(1.789)
Direct	SVR	0.204	(0.200)	0.313	(0.395)	0.378	(0.473)	0.529	(0.610)	0.832	(1.040)	1.270	(1.380)	1.540	(1.587)	2.185	(2.336)
	Multivariate	0.162	(0.180)	0.264	(0.317)	0.330	(0.374)	0.417	(0.484)	0.650	(0.718)	**0.921**	(1.010)	1.229	(1.292)	1.642	(1.947)
	$R+\sum_{i=1}^{5} IMF_i$	0.196	(0.205)	0.308	(0.356)	0.378	(0.425)	0.491	(0.504)	0.748	(0.805)	1.180	(1.259)	1.498	(1.361)	1.903	(2.103)
	$R+\sum_{i=2}^{5} IMF_i$	0.196	(0.198)	0.304	(0.354)	0.376	(0.432)	0.489	(0.502)	0.747	(0.803)	1.181	(1.258)	1.498	(1.362)	1.903	(2.102)
	$R+\sum_{i=3}^{5} IMF_i$	0.243	(0.244)	0.310	(0.345)	0.384	(0.433)	0.490	(0.514)	0.749	(0.806)	1.180	(1.259)	1.505	(1.362)	1.904	(2.098)
	$R+\sum_{i=4}^{5} IMF_i$	0.380	(0.430)	0.436	(0.452)	0.478	(0.483)	0.543	(0.532)	0.799	(0.867)	1.198	(1.265)	1.488	(1.349)	1.919	(2.093)
	$R+\sum_{i=5}^{5} IMF_i$	0.616	(0.677)	0.657	(0.712)	0.691	(0.709)	0.730	(0.764)	0.869	(1.021)	1.218	(1.317)	1.506	(1.382)	1.915	(2.085)
	R	0.890	(0.971)	0.924	(0.989)	0.934	(0.946)	0.964	(0.966)	1.033	(1.076)	1.297	(1.358)	1.546	(1.505)	1.924	(2.036)
Recursive	SVR	0.204	(0.200)	0.283	(0.365)	0.324	(0.412)	0.439	(0.556)	**0.628**	(0.831)	0.986	(1.140)	**1.077**	(1.221)	1.599	(1.868)
	Multivariate	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} IMF_i$	0.196	(0.205)	0.278	(0.331)	0.333	(0.432)	0.416	(0.474)	0.633	(0.774)	0.939	(1.036)	1.143	(1.307)	**1.545**	(1.775)
	$R+\sum_{i=2}^{5} IMF_i$	0.196	(0.198)	0.279	(0.331)	0.333	(0.432)	**0.415**	(0.474)	0.631	(0.774)	0.940	(1.036)	1.143	(1.307)	1.545	(1.775)
	$R+\sum_{i=3}^{5} IMF_i$	0.243	(0.244)	0.301	(0.332)	0.354	(0.402)	0.437	(0.469)	0.653	(0.770)	0.950	(1.035)	1.158	(1.312)	1.557	(1.789)
	$R+\sum_{i=4}^{5} IMF_i$	0.380	(0.430)	0.431	(0.444)	0.458	(0.471)	0.514	(0.508)	0.710	(0.775)	1.010	(1.074)	1.209	(1.309)	1.621	(1.800)
	$R+\sum_{i=5}^{5} IMF_i$	0.616	(0.677)	0.647	(0.708)	0.675	(0.695)	0.685	(0.741)	0.816	(0.982)	1.076	(1.237)	1.295	(1.282)	1.644	(1.851)
	R	0.890	(0.971)	0.920	(0.987)	0.923	(0.940)	0.937	(0.951)	0.989	(1.049)	1.209	(1.237)	1.414	(1.389)	1.721	(1.848)

Table 7.2 MAE and std for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = 1$ lagged values. The smallest MAE of each forecast horizon is set in boldface.

	Steps ahead	1		2		3		5		10		20		30		50	
	Model	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
	Naive	0.147	(0.186)	0.256	(0.322)	0.328	(0.418)	0.423	(0.514)	0.651	(0.801)	0.991	(1.053)	1.143	(1.294)	1.559	(1.803)
	ARIMA	0.145	(0.185)	0.247	(0.303)	0.313	(0.374)	0.421	(0.477)	0.663	(0.783)	1.021	(1.035)	1.154	(1.304)	1.644	(1.789)
Direct	SVR initial	0.145	(0.146)	0.242	(0.289)	0.294	(0.362)	0.408	(0.450)	0.615	(0.723)	0.943	(0.981)	1.101	(1.201)	1.641	(1.733)
	Multivariate	0.144	(0.180)	0.222	(0.289)	0.284	(0.353)	0.379	(0.448)	0.585	(0.711)	0.866	(0.928)	1.017	(1.153)	1.379	(1.585)
	$R+\sum_{i=1}^{5} IMF_i$	**0.120**	(0.146)	**0.181**	(0.226)	**0.234**	(0.278)	**0.371**	(0.426)	0.557	(0.667)	0.874	(0.896)	1.024	(1.046)	1.430	(1.601)
	$R+\sum_{i=2}^{5} IMF_i$	0.124	(0.138)	0.182	(0.228)	0.234	(0.286)	0.371	(0.421)	0.557	(0.658)	0.879	(0.898)	1.023	(1.049)	1.433	(1.602)
	$R+\sum_{i=3}^{5} IMF_i$	0.173	(0.184)	0.217	(0.240)	0.247	(0.283)	0.373	(0.387)	0.556	(0.637)	0.886	(0.910)	1.029	(1.053)	1.441	(1.611)
	$R+\sum_{i=4}^{5} IMF_i$	0.277	(0.324)	0.316	(0.331)	0.334	(0.349)	0.389	(0.390)	**0.544**	(0.604)	0.795	(0.848)	0.916	(0.930)	1.298	(1.425)
	$R+\sum_{i=5}^{5} IMF_i$	0.449	(0.495)	0.471	(0.519)	0.491	(0.511)	0.503	(0.536)	0.594	(0.714)	0.788	(0.903)	0.927	(0.916)	1.258	(1.393)
	R	0.655	(0.709)	0.675	(0.725)	0.679	(0.694)	0.688	(0.707)	0.729	(0.782)	**0.788**	(0.811)	**0.908**	(0.876)	**1.167**	(1.234)
Recursive	SVR initial	0.145	(0.146)	0.257	(0.357)	0.393	(0.566)	0.577	(0.841)	1.080	(1.922)	1.228	(1.939)	1.338	(2.054)	1.627	(2.366)
	Multivariate	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} IMF_i$	0.120	(0.1460)	0.241	(0.321)	0.383	(0.463)	0.634	(0.813)	0.829	(0.915)	1.077	(1.138)	1.298	(1.391)	1.807	(1.949)
	$R+\sum_{i=2}^{5} IMF_i$	0.124	(0.138)	0.239	(0.302)	0.381	(0.460)	0.612	(0.820)	0.826	(0.921)	1.085	(1.153)	1.299	(1.389)	1.806	(1.950)
	$R+\sum_{i=3}^{5} IMF_i$	0.173	(0.184)	0.245	(0.299)	0.370	(0.442)	0.535	(0.703)	0.823	(0.909)	1.082	(1.143)	1.341	(1.410)	1.807	(1.939)
	$R+\sum_{i=4}^{5} IMF_i$	0.277	(0.324)	0.315	(0.312)	0.383	(0.380)	0.544	(0.560)	0.724	(0.791)	1.076	(1.123)	1.327	(1.390)	1.777	(1.906)
	$R+\sum_{i=5}^{5} IMF_i$	0.449	(0.495)	0.464	(0.501)	0.482	(0.497)	0.522	(0.552)	0.673	(0.773)	0.929	(1.070)	1.211	(1.329)	1.736	(1.896)
	R	0.655	(0.709)	0.666	(0.709)	0.657	(0.669)	0.651	(0.674)	0.736	(0.756)	0.962	(1.019)	1.134	(1.240)	1.523	(1.887)

Table 7.3 MAE and std for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = 5$ lagged values. The smallest MAE of each forecast horizon is set in boldface.

	Steps ahead	1		2		3		5		10		20		30		50	
	Model	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
	Naive	0.147	(0.186)	0.256	(0.322)	0.328	(0.418)	0.423	(0.514)	0.651	(0.801)	0.991	(1.053)	1.143	(1.294)	1.559	(1.803)
	ARIMA	0.145	(0.185)	0.247	(0.303)	0.313	(0.374)	0.421	(0.477)	0.663	(0.783)	1.021	(1.035)	1.154	(1.304)	1.644	(1.789)
Direct	SVR initial	0.137	(0.139)	0.214	(0.275)	0.268	(0.334)	0.372	(0.409)	0.594	(0.722)	0.869	(0.974)	1.017	(1.056)	1.485	(1.596)
	Multivariate	0.141	(0.178)	0.224	(0.281)	0.299	(0.356)	0.369	(0.445)	0.571	(0.692)	0.858	(0.916)	1.001	(1.120)	1.364	(1.577)
	$R+\sum_{i=1}^{5} IMF_i$	**0.116**†	(0.132)	0.178	(0.211)	0.232	(0.267)	**0.350**	(0.382)	**0.530**	(0.613)	0.829	(0.839)	0.990	(1.034)	1.362	(1.620)
	$R+\sum_{i=2}^{5} IMF_i$	0.118	(0.124)	**0.175**†	(0.210)	**0.229**†	(0.271)	**0.350**†	(0.379)	**0.530**†	(0.609)	0.826	(0.842)	0.989	(1.033)	1.363	(1.621)
	$R+\sum_{i=3}^{5} IMF_i$	0.162	(0.172)	0.204	(0.225)	0.235	(0.265)	0.353	(0.364)	0.538	(0.599)	0.826	(0.846)	0.994	(1.040)	1.368	(1.620)
	$R+\sum_{i=4}^{5} IMF_i$	0.261	(0.303)	0.298	(0.308)	0.315	(0.327)	0.369	(0.361)	0.516	(0.567)	**0.736**†	(0.801)	0.892	(0.890)	1.215	(1.399)
	$R+\sum_{i=5}^{5} IMF_i$	0.421	(0.463)	0.441	(0.486)	0.462	(0.479)	0.478	(0.506)	0.574	(0.678)	0.746	(0.853)	0.883	(0.869)	1.185	(1.363)
	R	0.614	(0.664)	0.632	(0.679)	0.636	(0.650)	0.646	(0.662)	0.684	(0.734)	0.739	(0.761)	**0.856**†	(0.830)	**1.097**†	(1.166)
Recursive	SVR initial	0.137	(0.139)	0.256	(0.496)	0.310	(0.450)	0.478	(1.174)	0.688	(1.023)	0.998	(1.566)	1.053	(1.490)	1.346	(1.725)
	Multivariate	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} IMF_i$	0.116	(0.132)	0.222	(0.251)	0.356	(0.457)	0.518	(0.636)	0.645	(0.692)	0.981	(1.177)	1.140	(1.192)	1.642	(1.960)
	$R+\sum_{i=2}^{5} IMF_i$	0.118	(0.124)	0.222	(0.246)	0.361	(0.465)	0.511	(0.637)	0.646	(0.693)	0.979	(1.178)	1.141	(1.193)	1.641	(1.962)
	$R+\sum_{i=3}^{5} IMF_i$	0.162	(0.172)	0.234	(0.253)	0.361	(0.459)	0.468	(0.533)	0.651	(0.731)	0.968	(1.191)	1.187	(1.271)	1.652	(1.910)
	$R+\sum_{i=4}^{5} IMF_i$	0.261	(0.303)	0.306	(0.304)	0.346	(0.333)	0.464	(0.438)	0.619	(0.680)	0.933	(0.992)	1.146	(1.183)	1.594	(1.822)
	$R+\sum_{i=5}^{5} IMF_i$	0.421	(0.463)	0.439	(0.481)	0.459	(0.464)	0.477	(0.493)	0.592	(0.674)	0.873	(0.980)	1.053	(1.225)	1.494	(1.820)
	R	0.614	(0.664)	0.627	(0.679)	0.622	(0.645)	0.617	(0.651)	0.693	(0.709)	0.935	(0.974)	1.083	(1.239)	1.377	(1.784)

Table 7.4 MAE and std for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = p + d$ lagged values, the same input vector as the ARIMA model. The smallest MAE of each forecast horizon is set in boldface. The values marked with a dagger (†) indicate the smallest MAE of each horizon for the models with different input vector.

7.3.3 Testing statistically significant differences between models

In order to further evaluate the performance of the proposed EMD-SVR models, we use a test for the null hypothesis of equal forecast accuracy. We used the Wilcoxon signed rank test [176], a non-parametric test which estimates the statistically significant difference among a pair of models. We applied Wilcoxon test to the rank of the difference of the absolute errors and evaluated the null hypothesis that the two related error samples have the same distribution.

We denote by $E_{1,i}$ and $E_{2,i}$ $i = 1,2,\dots,M$, the number of observed absolute errors of model 1 and 2 respectively. The Wilcoxon test uses the rank of $d_i = |E_{1,i} - E_{2,i}|$ to create the following statistic:

$$W = \sum_{i=1}^{M} [I_+(d_i)R_i], \tag{7.4}$$

where R_i is the rank from the smallest absolute difference to the largest absolute difference, and the function $I_+(d_i)$ is given by:

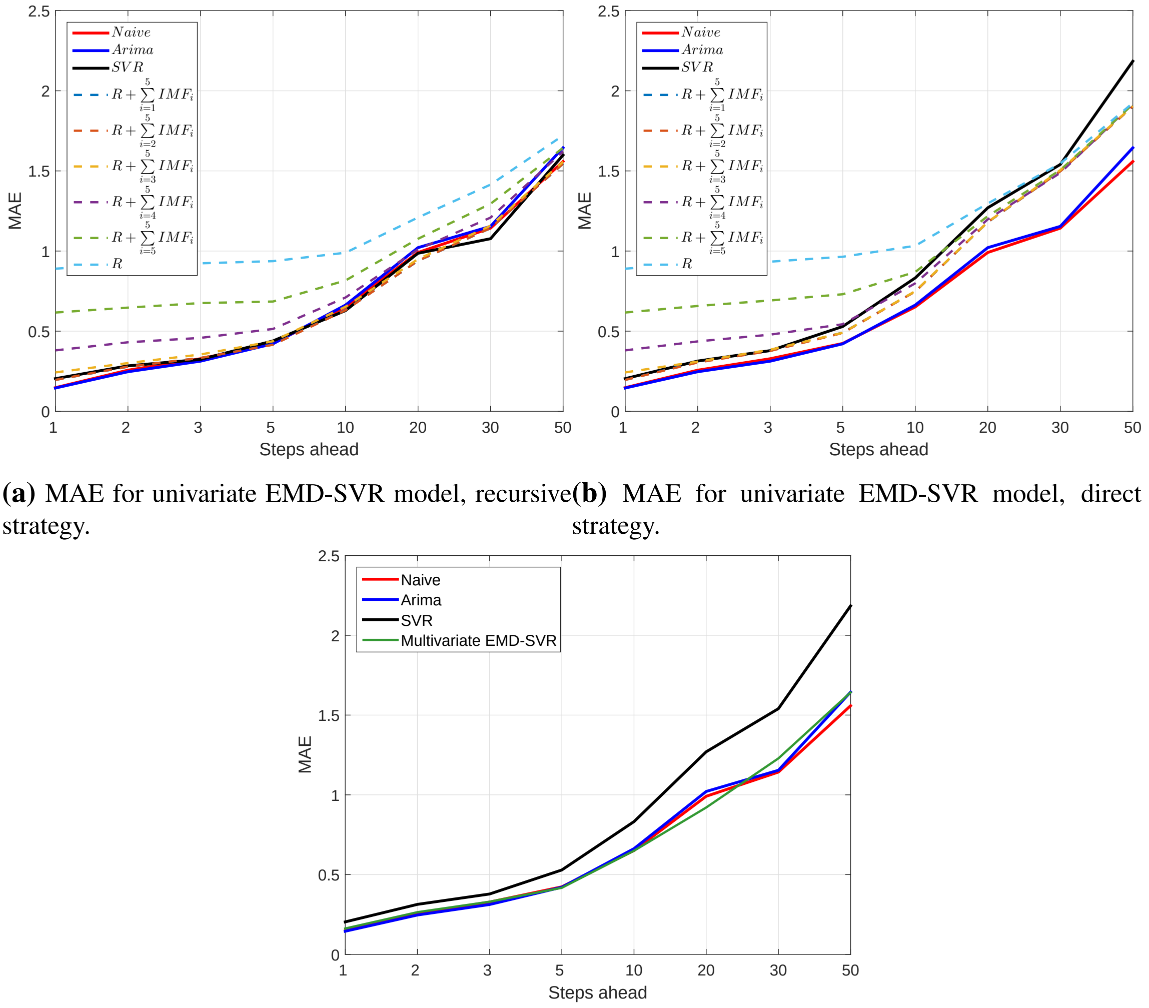
$$I_+(d_i) = \begin{cases} 1 & \text{if} \quad d_i > 0 \\ 0 & \quad \text{otherwise.} \end{cases} \tag{7.5}$$

The distribution of W has been tabulated for various values of $M < 25$, [176]. For larger values of M, the standard normal distribution provides a good approximation of the standardized version of the statistic W, which is referred as the Z statistic [66]:

$$Z = \frac{W - M(M+1)/4}{\sqrt{\frac{M(M+1)(2M+1)}{24}}}. \tag{7.6}$$

We applied the Wilcoxon test to each forecast horizon with $M = 128$, denoting the number of analysed time series. All tests were performed at the 5% and 1% significance level. The benchmark model is always the naive model, and it is compared against the proposed EMD-SVR models. In Tables 7.5, 7.6 and 7.7, we present the Z-statistic values of the two-tail Wilcoxon signed rank test. A positive value of the Z-statistic indicates that the tested model has smaller errors than the naive model. Contrary, a negative value indicates that the naive model outperforms the tested model. The values label with * and ** indicate the rejection of the null hypothesis at the 5% and 1% significance level, respectively.

In Table 7.5, we report the Z-statistic for the EMD-SVR models with input vector of length $m = 1$. With this input vector, we mainly observe negative numbers, implying that none of the proposed EMD-SVR models outperform the naive model.

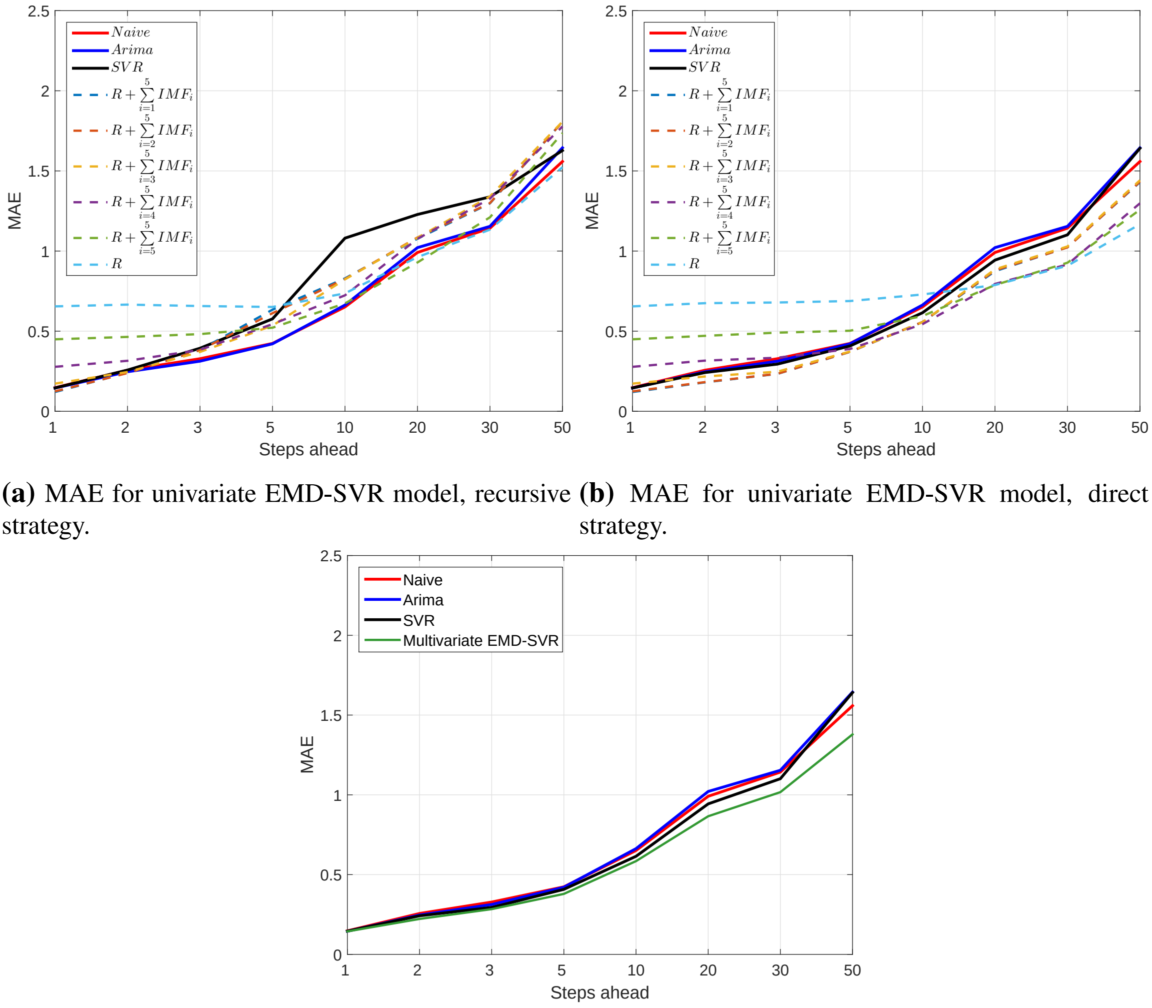


(a) MAE for univariate EMD-SVR model, recursive strategy.

(b) MAE for univariate EMD-SVR model, direct strategy.

(c) MAE for multivariate EMD-SVR model.

Fig. 7.5 MAE as a function of the forecast horizon for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = 1$ lagged values.

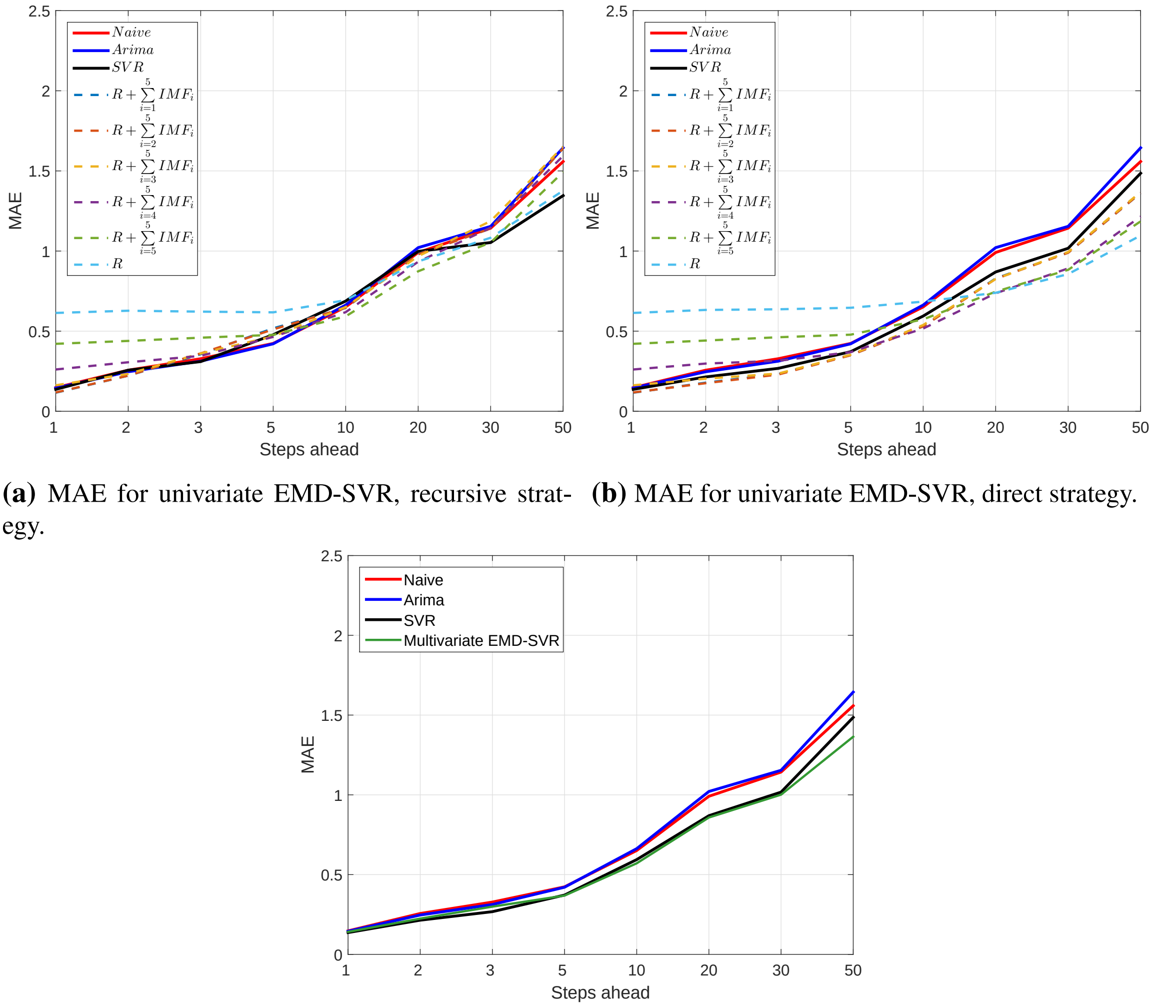


(a) MAE for univariate EMD-SVR model, recursive strategy.

(b) MAE for univariate EMD-SVR model, direct strategy.

(c) MAE for multivariate EMD-SVR model.

Fig. 7.6 MAE as a function of the forecast horizon for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = 5$ lagged values.



(a) MAE for univariate EMD-SVR, recursive strategy.

(b) MAE for univariate EMD-SVR, direct strategy.

(c) MAE for multivariate EMD-SVR.

Fig. 7.7 MAE as a function of the forecast horizon for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = p + d$ lagged values.

In Table 7.6, we show the Z-statistic for the EMD-SVR models with input vector of length $m = 5$. Results are very similar than for input vector $m = p + d$ since on average $m = p + d = 5$, refer to Appendix A.

Table 7.7 reports the Wilcoxon test results for the models with the input vector which produces the smallest MAE, $m = p + d$ lagged values. The EMD-SVR model with direct strategy applied to the denoised reconstruction $R + \sum_{i=1}^{5} IMF_i$ and $R + \sum_{i=2}^{5} IMF_i$, is significantly better than the naive model. For short-term forecast horizons, the naive model outperforms the coarse R model. However, for long-term horizons, the errors of the coarse model R are smaller, outperforming the naive model. We note that the direct strategy on SVR applied to the initial time series performs significantly better than the naive model but only for some of the forecasted horizons. The multivariate EMD-SVR model significantly outperforms the naive model, except for the 1-step ahead forecast.

7.3.4 FTSE 100 forecasting results

So as to verify the generalization of the proposed EMD-SVR models, we did the complete same analysis for the Financial Times Stock Exchange (FTSE) 100 index. A summary of the results is presented in Appendix B. We obtained the same conclusions with respect to the accuracy of the proposed models and we confirmed the good performance of the proposed EMD-SVR models.

7.4 Summary

In this chapter we introduced a multistep-ahead forecasting scheme for non-linear and non-stationary time series based on EMD and SVR. The EMD can fully capture the local fluctuations of the analysed time series and can be used as a preprocessor to decompose non-stationary data into a finite set of IMFs and a residue. The extracted IMFs have simpler structures and defined oscillating frequencies than can simplify the forecasting process.

We proposed a univariate and a multivariate EMD-SVR forecasting schemes. For the univariate scheme, we forecasted each IMFs and the residue separately and forecasted the input time series as the sum of the components. We defined coarse-to-fine reconstruction models using the cumulative sum of sequential IMFs, i.e., adding more details to the low frequency components. We used two multistep-ahead prediction strategies, the recursive and the direct strategies. Moreover, we proposed a multivariate EMD-SVR, that combines information of all the IMFs and the residue into one input vector. As benchmark models, we

	Model \ Steps ahead	1	2	3	5	10	20	30	50
	ARIMA	-0.22	0.16	0.89	0.30	-1.45	-1.59	-0.49	-1.47
Direct	**SVR**	-4.72**	-3.05**	-2.54*	-3.24**	-3.47**	-2.75**	-3.35**	-3.39**
	Multivariate	-1.83	-0.80	-0.39	0.94	0.22	1.37	-0.18	1.41
	$R+\sum_{i=1}^{5} \boldsymbol{IMF}_i$	-4.07**	-3.07**	-3.11*	-2.96**	-2.72**	-2.11*	-3.70**	-2.32*
	$R+\sum_{i=2}^{5} \boldsymbol{IMF}_i$	-4.21**	-3.12**	-2.84**	-2.88**	-2.73**	-2.12*	-3.71**	-2.32*
	$R+\sum_{i=3}^{5} \boldsymbol{IMF}_i$	-5.62**	-3.18**	-2.99**	-2.60**	-2.69**	-2.12*	-3.74**	-2.37*
	$R+\sum_{i=4}^{5} \boldsymbol{IMF}_i$	-7.25**	-5.73**	-4.65**	-3.48**	-3.37**	-2.33*	-3.51**	-2.45*
	$R+\sum_{i=5}^{5} \boldsymbol{IMF}_i$	-8.35**	-7.51**	-6.36**	-4.99**	-3.84**	-2.46*	-3.20**	-2.47*
	R	-9.06**	-8.14**	-7.84**	-7.00**	-5.51**	-3.10**	-3.66**	-2.66**
Recursive	**SVR**	-4.72**	-1.83	-0.44	-0.58	0.03	1.01	1.37	0.27
	Multivariate	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} \boldsymbol{IMF}_i$	-4.07**	-1.96	-0.41	-0.21	0.63	1.95	0.90	1.55
	$R+\sum_{i=2}^{5} \boldsymbol{IMF}_i$	-4.21**	-2.08*	-0.62	-0.13	0.76	1.93	0.92	1.51
	$R+\sum_{i=3}^{5} \boldsymbol{IMF}_i$	-5.62**	-2.82**	-1.76	-1.11	-0.09	1.44	0.43	1.34
	$R+\sum_{i=4}^{5} \boldsymbol{IMF}_i$	-7.25**	-5.55**	-4.48**	-3.38**	-2.17*	-1.09	-1.05	-0.10
	$R+\sum_{i=5}^{5} \boldsymbol{IMF}_i$	-8.35**	-7.33**	-6.28**	-4.46**	-2.95**	-1.18	-1.90	0.11
	R	-9.06**	-8.13**	-7.77**	-6.78**	-5.25**	-2.26*	-2.84**	-0.86

Table 7.5 Z-statistic for the Wilcoxon signed-rank test for the null hypothesis that the naive model is as accurate as the studied models: ARIMA, univariate and multivariate EMD-SVR with input vector $m = 1$. Top, direct strategy, bottom, recursive strategy.

* Statistically significant at the 5% confidence level

** Statistically significant at the 1% confidence level.

	Model \ Steps ahead	1	2	3	5	10	20	30	50
Direct	**ARIMA**	-0.22	0.16	0.89	0.30	-1.45	-1.59	-0.49	-1.47
	SVR	-0.45	1.08	1.54	0.72	0.57	1.24	0.67	0.14
	Multivariate	0.20	5.58**	4.70**	4.57**	5.67**	8.43**	7.70**	7.66**
	$R+\sum_{i=1}^{5} \boldsymbol{IMF}_i$	3.33**	6.71**	6.41**	3.17**	2.79**	2.79**	1.61	2.39**
	$R+\sum_{i=2}^{5} \boldsymbol{IMF}_i$	2.49*	5.94**	6.24**	3.21**	2.64**	2.72**	1.70	2.37*
	$R+\sum_{i=3}^{5} \boldsymbol{IMF}_i$	-1.94	1.88	3.93**	1.75	1.94	2.54*	1.66	2.31*
	$R+\sum_{i=4}^{5} \boldsymbol{IMF}_i$	-5.66**	-2.31*	-0.47	0.59	2.57*	4.02**	3.99**	3.55**
	$R+\sum_{i=5}^{5} \boldsymbol{IMF}_i$	-7.45**	-5.54**	-3.82**	-1.76	0.61	3.56**	3.42**	3.98**
	R	-8.59**	-7.12**	-6.28**	-4.12**	-2.12*	3.37**	3.19**	4.10**
Recursive	**SVR**	-0.45	1.53	1.15	0.87	0.36	1.79	1.59	2.29*
	Multivariate	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} \boldsymbol{IMF}_i$	3.33**	1.43	-0.51	-2.26*	-1.31	-0.15	-0.95	-1.24
	$R+\sum_{i=2}^{5} \boldsymbol{IMF}_i$	2.49*	1.45	-0.49	-1.98*	-1.23	-0.08	-1.02	-1.19
	$R+\sum_{i=3}^{5} \boldsymbol{IMF}_i$	-1.94	1.03	0.11	-1.32	-1.38	-0.22	-1.33	-1.31
	$R+\sum_{i=4}^{5} \boldsymbol{IMF}_i$	-5.66**	-2.30*	-1.72	-2.38*	-0.73	0.07	-1.18	-1.28
	$R+\sum_{i=5}^{5} \boldsymbol{IMF}_i$	-7.45**	-5.59**	-3.86**	-2.01*	-0.03	2.14*	0.34	-1.00
	R	-8.59**	-7.13**	-6.06**	-3.86**	-2.11*	1.55	1.04	1.22

Table 7.6 Z-statistic for the Wilcoxon signed-rank test for the null hypothesis that the naive model is as accurate as the studied models: ARIMA, univariate and multivariate EMD-SVR with input vector $m = 5$. Top, direct strategy, bottom, recursive strategy.

* Statistically significant at the 5% confidence level

** Statistically significant at the 1% confidence level.

	Model \ Steps ahead	**1**	**2**	**3**	**5**	**10**	**20**	**30**	**50**
Direct	**ARIMA**	-0.22	0.16	0.89	0.30	-1.45	-1.59	-0.49	-1.47
	SVR	0.57	2.81**	2.58**	1.94	1.37	2.41*	1.69	1.56
	Multivariate	1.27	4.27**	2.93**	5.43**	6.07**	8.31**	7.77**	8.52**
	$R+\sum_{i=1}^{5}\boldsymbol{IMF}_i$	3.40**	6.24**	6.00**	3.61**	3.14**	3.12**	2.51*	3.12**
	$R+\sum_{i=2}^{5}\boldsymbol{IMF}_i$	2.80**	6.05**	6.28**	3.67*	3.08*	3.16**	2.54**	3.06**
	$R+\sum_{i=3}^{5}\boldsymbol{IMF}_i$	-1.28	2.61**	4.51**	2.50*	2.54*	3.15**	2.51*	3.01**
	$R+\sum_{i=4}^{5}\boldsymbol{IMF}_i$	-5.34**	-1.67	0.20	1.30	3.41**	4.79**	4.08**	4.20**
	$R+\sum_{i=5}^{5}\boldsymbol{IMF}_i$	-7.22**	-5.10**	-3.32**	-1.28	0.93	4.16**	3.63**	4.53**
	R	-8.45**	-6.88**	-5.82**	-3.65**	-1.44	2.73**	2.38*	3.38**
Recursive	**SVR**	0.57	3.12**	2.58**	3.24**	2.66**	3.57**	3.67**	4.18**
	Multivariate	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5}\boldsymbol{IMF}_i$	3.40**	1.83	0.75	-1.47	0.49	1.22	0.18	-0.40
	$R+\sum_{i=2}^{5}\boldsymbol{IMF}_i$	2.80**	1.69	0.57	-1.33	0.45	1.26	0.19	-0.38
	$R+\sum_{i=3}^{5}\boldsymbol{IMF}_i$	-1.28	1.26	0.68	-0.41	0.57	1.55	-0.26	-0.57
	$R+\sum_{i=4}^{5}\boldsymbol{IMF}_i$	-5.34**	-1.93	-0.84	-1.23	0.78	2.02*	-0.09	-0.36
	$R+\sum_{i=5}^{5}\boldsymbol{IMF}_i$	-7.22**	-5.17**	-3.44**	-1.30	1.08	3.07**	1.87	0.93
	R	-8.45**	-6.89**	-5.71**	-3.25**	-1.41	2.10*	1.84	2.15*

Table 7.7 Z-statistic for the Wilcoxon signed-rank test for the null hypothesis that the naive model is as accurate as the studied models: ARIMA, univariate and multivariate EMD-SVR with input vector $m = p+d$. Top, direct strategy, bottom, recursive strategy.

* Statistically significant at the 5% confidence level

** Statistically significant at the 1% confidence level.

used a naive model, the ARIMA model and SVR applied to the initial time series (without a priori EMD).

We evaluated the performance of our multistep-ahead forecasting models on intraday data, considering two stock market indices, the S&P 500 and the FTSE 100 indices. The results suggest that for an input vector of length $m = 5$ or $m = p + d$ (p and d obtained from a fitted ARIMA model), the multivariate EMD-SVR models perform better than the benchmark models. However, the best results were obtained with the direct strategy applied to the univariate EMD-SVR. We notice that for short-term forecasting, the model using the full reconstruction (all IMFs) performs better. For long-term forecasting, a coarse-to-fine reconstruction confirms that we may neglect some of the highest frequency modes. The residue captures the most important features of the original data and adding on the higher, noisier frequency modes thereafter reduces generalisation and does not improve forecasting accuracy.

In most cases, the direct strategy outperforms the recursive strategy but it is more computationally expensive since a model needs to be trained for each forecast horizon. We conclude that the EMD improves the forecasting performance of the SVR method, either by including all the high-frequency IMFs to forecast short-term horizons or by using the trend and low-frequency IMFs to forecast long-term horizons. The limited improvement for short term horizons may be due to the boundary effects of the EMD which produce swings in the extreme of the IMFs and perturb the first forecasting steps. Although the proposed integrated system has a satisfactory predictive performance, certainly, there is scope for further improvement, for example, the selection of the input vector and the choice of parameters for the SVR.

Chapter 8

Conclusion

This chapter summarizes the main findings of this thesis and discusses some further research which could extend the proposed framework.

8.1 Summary

This thesis provides a framework to analyse high-frequency financial data based on the Hilbert-Huang transform (HHT), a technique which reveals the time-dependent characteristics of non-stationary and non-linear time series. We questioned if a totally adaptive decomposition as the HHT could shed light into the generating process of financial time series. The central argument of this thesis is that the HHT provides a tool to simultaneously characterize both the short and the long-term fluctuations latent in a time series.

We argued that financial market data contain patterns specific to the observation frequency and are thus, of interest to different type of market agents (market traders, intraday traders, hedging strategist, portfolio managers and institutional investors), each characterized by a reaction time to new information and by the frequency of its intervention to the market. The high-frequency oscillations extracted from the EMD are attributed to actions of fast traders, in a similar way that low-frequency oscillations are accredited to longer term investors. The time-dependent characteristics of the studied financial time series reveal a more complex structure than what would be expected from Brownian motion or fractional Brownian motion.

We compared the HHT against the Fourier and the wavelet transforms, discussing some drawbacks of the last two, for instance, the need for an a-priori basis selection (sinusoids and wavelets, respectively). The interpretation of the Fourier and the wavelet transforms depends on the match between the chosen basis and the input time series. Furthermore, these transforms are unable to reveal instantaneous attributes, such as amplitude and frequency.

Since most of the financial time series are non-stationary, time-varying spectral methods need to be considered.

The main advantages of the introduced framework include the non-parametric approach, the fully adaptiveness and the localization properties in both the time and the frequency domain. A clear disadvantage is its lack of a formal mathematical formulation, any study can only be validated by numerical experiments or by comparison between existing and similar methods. Nevertheless, the HHT has demonstrated to provide a meaningful analysis of data. In this research, we demonstrated some applications to high-frequency financial data which can be listed as:

Variance

We proposed a scale-by-scale analysis of variance that reproduces the results of a simple realised volatility estimator, but which allows to identify the time-horizons that dominate the total variance of the input time series. Our results demonstrated that the shortest time-horizons account for more than 50% of the total variation. In general, volatility can be attributed to the fastest investors, indicating that higher volatility is a reflection of faster trading activity. We compared the EMD against the wavelet decomposition, and despite their theoretical differences, both methods confirmed that most of the volatility can be attributed to the higher frequency components.

Correlation

The multiscale analyses provided by the EMD allows to study the dynamic correlation between two non-stationary time series. We proposed two approaches to study dependencies in high-frequency financial data. The first approach, a frequency-dependent correlation, allows the segregation of correlation at different frequencies, focusing only on the correlation occurring at a given timescale. We observed that the high-frequency IMFs tend to be less correlated than the low-frequency components, a result that could be explained by the Epps effect [69].

The second approach consists of a rolling-window analysis which estimates time-dependent correlation. This approach captures the intraday dynamics of the analysed time series and uncovers lead-lag relationships which could be attributed to different levels of trading activity. The time-dependent correlation offers a better understanding about the speed of different assets processing and reflecting information and the degree to which the information contained in one time series could be used to make predictions on the other. This analysis could be used in forecasting models, for example, in regression models.

Variance scaling patterns

We empirically showed that when EMD is applied to fractional Brownian motion and α-stable Lévy motion, we obtain a scaling law that relates linearly the logarithm of the variance and the logarithm of the period of the IMFs. The extracted scaling exponent equals the Hurst exponent multiplied by two. By estimating a scaling exponent for models with known scaling laws, we demonstrate that our measure varies consistently around the expected values set in the models. However, when applied to stock market indices, the EMD revealed instead different scaling laws that can deviate significantly from both Brownian motion and FBM behaviour. In particular, we noted that the EMD of high-frequency financial data results in a larger number of IMFs than what would be expected from Brownian motion. The anomalous scaling unveiled a more complex structure in financial data than in artificial self-similar processes. We observed that developed markets tend to have scaling properties closer to Brownian motion properties. Conversely, larger deviations from uniscaling laws are observed in some emerging markets. Compared to previous approaches, the EMD method has the advantage to directly quantify the cyclical components with strong deviations, giving a further instrument to understand the origin of market inefficiencies.

Time-dependent scaling and complexity measure

Using the localisation properties of the HHT, we proposed two novel time-dependent measures of complexity: 1) an amplitude scaling exponent and 2) an entropy-like measure. These measures allow to identify trends and intermittent behaviour in financial time series, they do not assume any particular parameter, the temporal behaviour and variations are found directly from the data, considering only the scales obtained via the EMD. With the proposed measures, we are able to describe the dynamics of financial time series whose regularity changes over time. Our results suggest that a time-varying scaling exponent could better reflect the scaling behaviour of financial data. The intraday analysis of some stock market indices revealed a distinctive anti-persistent behaviour at the opening and at the closing of the trading session, contrasting with the persistent behaviour at the middle of the session.

Forecasting models

Based on EMD and SVR, we proposed a multistep-ahead forecasting scheme. The advantage of using the EMD is the non-parametric approach which does not overcomplicate the forecasting process. The obtained IMFs have simpler structures and defined oscillating frequencies which can be forecasted with more accuracy. We defined a coarse-to-fine reconstruction by using the cumulative sum of sequential IMFs (adding more details to the low frequency components). We concluded that the EMD improves the forecasting performance of the SVR method. The best results were obtained with the direct strategy applied to the univariate EMD-SVR. We observed that for short-term forecasting, the model using all the IMFs (full reconstruction) produces better results. Though the limited improvement may be due to boundary effects of the EMD. These effects produce swings in the extreme of the IMFs perturbing the first forecasting steps.

On the other hand, for long-term forecasting, the residue captures the most important features of the input data and adding on the higher, noisier frequency modes thereafter reduces generalisation and does not improve forecasting accuracy.

8.2 Future research

Further research could extend the proposed framework by considering:

- An approach to make the EMD robust to outliers. If the input data have outliers, the resultant IMFs could be deformed due to the interpolation process involved in the construction of the envelopes. The impact of outliers could propagate not only to the high frequency modes but also to lower frequencies, corrupting the information of each oscillation. It could be advantageous to pre-process the data, to implement a method that assign the outliers to the highest frequency component or to smooth the impact of the outliers. This would enhance the properties of the IMFs when describing the true generating process of the input data, retaining the adaptiveness of the EMD.

- An approach for noise reduction and frequency filtering. The principles of hard and soft wavelet thresholding [145] could be adapted to develop denoising methods for the IMFs. A filtering scheme based on the partial reconstruction of the time series could be implemented. This scheme is premised on the fact that the first several IMFs consist of noise and sometimes not very significant information. Nevertheless, a more refined study to select the IMFs is needed. The denoised time series could be used to estimate volatility or to forecast the true data generating process.

- A method for trend extraction. The EMD offers a natural way to extract a time-varying mean of the input time series. The residue, the sum of the last IMFs or a sensible combination of the IMFs can generate a non-linear trend that could be used, for example, in forecasting tasks.

- A dependency measure which considers the time-varying phase and amplitude of the IMFs. We did investigate the topic of correlation, however, we think that more robust estimators could be achieved and further conclusions could be drawn. A coherence measure, similar to the wavelet coherence [168], could be obtained from the HHT. Moreover, by separating the amplitude and the frequency information, the dependency measures could be directed to estimate the amplitude or phase correlation. Dependency measures such as: phase-locking value [110] and Granger causality [62] could also be estimated using the HHT.

- A scaling parameter that considers multifractal process. We did a comprehensive study of the use of the HHT to identify scaling patterns, establishing its use for long-memory and self-similar processes. However, the research could be extended to multifractal processes. Huang et al. [100] proposed a method to characterize the scale-invariant properties of time series in the amplitude-frequency space. We could extend our research on the time dependent scaling exponent to investigate if a generalization to arbitrary order could recover the exponent which characterizes multiscaling processes.

- Improvements to the forecasting scheme proposed in this thesis:
  - Refine the implementation of SVR by further investigating the selection of the parameters: regularization constant, C; insensitive coefficient, ε; and kernel width, λ.
  - A better selection of the input vector. In our proposed forecasting models, the input vector consisted only of past observations of the input time series. It is worth to investigate if incorporating information from other time series could improve forecasting accuracy. In this sense, we could consider correlation, causality and co-integration between financial time series.
  - Consider other methods, such as, artificial neural networks, kernel regression, k-nearest neighbour regression, fuzzy time series, etc., which could capture the oscillating nature of the IMFs and improve forecasting accuracy.

Overall, the HHT proved to be an efficient tool to analyse high-frequency financial data. The EMD generates simpler oscillating components which preserve the information of the heterogeneous financial time series. Well-established methodologies can be improved if they are applied to these oscillating components. This research shows that new time-varying statistics obtained from the HHT contribute to a better understanding of the complex behaviour of financial time series.

References

[1] Abhyankar, A., Copeland, L. S., and Wong, W. (1997). Uncovering nonlinear structure in real-time stock-market indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100. *Journal of Business & Economic Statistics*, 15(1):1–14.

[2] Abry, P., Delbeke, L., and Flandrin, P. (1999). Wavelet based estimator for the self-similarity parameter of α-stable processes. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on,* 3, 1729–1732.

[3] Adak, S. (1998). Time-dependent spectral analysis of non-stationary time series. *Journal of the American Statistical Association*, 93(444):1488–1501.

[4] Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics*, 74(3):487–528.

[5] Aït Sahalia, Y., Mykland, P. A., and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18(2):351–416.

[6] Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2011). Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1):160–175.

[7] Allez, R. and Bouchaud, J.-P. (2011). Individual and collective stock dynamics: intra-day seasonalities. *New Journal of Physics*, 13 025010(2).

[8] Alvarez-Ramirez, J., Alvarez, J., Rodriguez, E., and Fernandez-Anaya, G. (2008). Time-varying Hurst exponent for US stock markets. *Physica A: Statistical Mechanics and its Applications*, 387(24):6159–6169.

[9] Andersen, T. G. and Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115–158.

[10] Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4):701–720.

[11] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

[12] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2000). Great realizations. *Risk*, 13:105–108.

[13] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

[14] Annis, A. and Lloyd, E. (1976). The expected value of the adjusted rescaled Hurst range of independent normal summands. *Biometrika*, 63(1):111–116.

[15] Araújo, E. (2009). Macroeconomic shocks and the co-movement of stock returns in Latin America. *Emerging Markets Review*, 10(4):331–344.

[16] Arlot, S., Celisse, A., et al. (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4:40–79.

[17] Arouri, M. E. H., Lahiani, A., and Nguyen, D. K. (2015). Cross-market dynamics and optimal portfolio strategies in Latin American equity markets. *European Business Review*, 27(2):161–181.

[18] Avram, F. and Taqqu, M. S. (2000). Robustness of the R / S statistic for fractional stable noises. *Statistical Inference for Stochastic Processes*, 3(1):69–83.

[19] Bandi, F. M. and Russell, J. R. (2008). Microstructure noise, realized variance, and optimal sampling. *The Review of Economic Studies*, 75(2):339–369.

[20] Barabási, A.-L. and Vicsek, T. (1991). Multifractality of self-affine fractals. *Physical Review A*, 44:2730–2733.

[21] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

[22] Barndorff-Nielsen, O. E. and Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied econometrics*, 17(5):457–477.

[23] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, 2(1):1–37.

[24] Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics*, 4(1):1–30.

[25] Barndorff-Nielsen, O. E. and Shephard, N. (2007). Variation, jumps, and high-frequency data in financial econometrics. In *Advances in Economics and Econometrics*. Cambridge University Press, 3, 328–372.

[26] Barucci, E. and Renó, R. (2002). On measuring volatility of diffusion processes with high frequency data. *Economics Letters*, 74(3):371–378.

[27] Baruník, J., Aste, T., Di Matteo, T., and Liu, R. (2012). Understanding the source of multifractality in financial markets. *Physica A: Statistical Mechanics and its Applications*, 391(17):4234–4251.

[28] Baruník, J. and Kristoufek, L. (2010). On Hurst exponent estimation under heavy-tailed distributions. *Physica A: Statistical Mechanics and its Applications*, 389(18):3844–3855.

[29] Baruník, J. and Vácha, L. (2015). Realized wavelet-based estimation of integrated variance and jumps in the presence of noise. *Quantitative Finance*, 15(8):1347–1364.

[30] Bennett, K., Hu, J., Ji, X., Kunapuli, G., and Pang, J.-S. (2006). Model selection via bilevel optimization. In *Neural Networks, 2006. IJCNN '06. International Joint Conference on,* 1922–1929.

[31] Beran, J. (1994). *Statistics for long-memory processes*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, Florida.

[32] Bianchi, S., Pantanella, A., and Pianese, A. (2013). Modeling stock prices by multifractional Brownian motion: an improved estimation of the pointwise regularity. *Quantitative Finance*, 13(8):1317–1330.

[33] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.

[34] Bollerslev, T., Litvinova, J., and Tauchen, G. (2006). Leverage and volatility feedback effects in high-frequency data. *Journal of Financial Econometrics*, 4(3):353–384.

[35] Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2):409–424.

[36] Boudt, K., Croux, C., and Laurent, S. (2011). Robust estimation of intraweek periodicity in volatility and jump detection. *Journal of Empirical Finance*, 18(2):353–367.

[37] Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall, Michigan, 3rd edition.

[38] Brockwell, P. J. and Davis, R. A. (2002). *Introduction to time series and forecasting*. Springer, New York, 2nd edition.

[39] Cajueiro, D. O. and Tabak, B. M. (2004). The Hurst exponent over time: testing the assertion that emerging markets are becoming more efficient. *Physica A: Statistical Mechanics and its Applications*, 336(3-4):521–537.

[40] Calvet, L. E. and Fisher, A. (2008). *Multifractal volatility: theory, forecasting, and pricing*. Academic Press advanced finance series. Academic Press, Amsterdam.

[41] Calvet, L. E. and Fisher, A. J. (2002). Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics*, 84(3):381–406.

[42] Calvet, L. E., Fisher, A. J., and Thompson, S. B. (2006). Volatility comovement: a multifrequency approach. *Journal of Econometrics*, 131(1-2):179–215.

[43] Carbone, A., Castelli, G., and Stanley, H. (2004). Time-dependent Hurst exponent in financial time series. *Physica A: Statistical Mechanics and its Applications*, 344:267–271.

[44] Chakraborti, A., Toke, I. M., Patriarca, M., and Abergel, F. (2011). Econophysics review: I. Empirical facts. *Quantitative Finance*, 11(7):991–1012.

[45] Chen, C.-F., Lai, M.-C., and Yeh, C.-C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, 26:281–287.

[46] Chen, Q., Huang, N., Riemenschneider, S., and Xu, Y. (2006). A B-spline approach for empirical mode decompositions. *Advances in Computational Mathematics*, 24(1-4):171–195.

[47] Chen, S.-H., Lux, T., and Marchesi, M. (2001). Testing for non-linear structure in an artificial financial market. *Journal of Economic Behavior & Organization*, 46(3):327–342.

[48] Chen, X., Wu, Z., and Huang, N. (2010). The time-dependent intrinsic correlation based on the empirical mode decomposition. *Advances in Adaptive Data Analysis*, 2(2):233–265. cited By 19.

[49] Cheng, C.-H. and Wei, L.-Y. (2014). A novel time-series model based on empirical mode decomposition for forecasting TAIEX. *Economic Modelling*, 36(0):136–141.

[50] Cherkassky, V. and Ma, Y. (2004). Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17(1):113–126.

[51] Chevillon, G. (2007). Direct multi-step estimation and forecasting. *Journal of Economic Surveys*, 21(4):746–785.

[52] Chicago Board Options Exchange (2015). The CBOE volatility index - VIX. https://www.cboe.com/micro/vix/vixwhite.pdf. Accessed 13/05/2015.

[53] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236.

[54] Cont, R. (2005). Long range dependence in financial markets. In *Fractals in Engineering*. Springer, 159–179.

[55] Corsi, F., Zumbach, G., Muller, U. A., and Dacorogna, M. M. (2001). Consistent high-precision volatility from high-frequency data. *Economic Notes*, 30(2):183–204.

[56] Curme, C., Tumminello, M., Mantegna, R. N., Stanley, H. E., and Kenett, D. Y. (2015). Emergence of statistically validated financial intraday lead-lag relationships. *Quantitative Finance*, 15(8):1375–1386.

[57] Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., and Pictet, O. V. (2001). *An introduction to high frequency finance*. Academic Press, San Diego.

[58] Dahlhaus, R. et al. (1997). Fitting time series models to nonstationary processes. *The annals of Statistics*, 25(1):1–37.

[59] Dahlhaus, R., Polonik, W., et al. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39.

[60] Darbellay, G. A. and Wuertz, D. (2000). The entropy as a tool for analysing statistical dependences in financial time series. *Physica A: Statistical Mechanics and its Applications*, 287(3):429–439.

[61] Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, Philadelphia.

[62] Dhamala, M., Rangarajan, G., and Ding, M. (2008). Estimating Granger causality from Fourier and wavelet transforms of time series data. *Physical Review Letters*, 100:018701.

[63] Di Matteo, T. (2007). Multi-scaling in finance. *Quantitative Finance*, 7(1):21–36.

[64] Di Matteo, T., Aste, T., and Dacorogna, M. (2003). Scaling behaviors in differently developed markets. *Physica A: Statistical Mechanics and its Applications*, 324(1):183–188.

[65] Di Matteo, T., Aste, T., and Dacorogna, M. (2005). Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance*, 29(4):827–851.

[66] Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3):253–63.

[67] Doukhan, P., Oppenheim, G., and Taqqu, M. (2003). *Theory and applications of long-range dependence*. Birkhäuser Boston, Boston.

[68] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987–1007.

[69] Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association*, 74:291–298.

[70] Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105.

[71] Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480):1349–1362.

[72] Feldman, M. (2011). *Hilbert transform applications in mechanical vibration*. Wiley, New Delhi.

[73] Feller, W. (1966). *An introduction to probability theory and its applications*, volume 2 of *Wiley publications in statistics*. John Wiley & Sons Inc., New York, 2nd edition.

[74] Feng, L. and Linetsky, V. (2008). Pricing discretely monitored barrier options and defaultable bonds in Lévy process models: a fast Hilbert transform approach. *Mathematical Finance*, 18(3):337–384.

[75] Fernandes, M., Medeiros, M. C., and Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40:1–10.

[76] Fernández-Martínez, M., Sánchez-Granero, M., and Segovia, J. T. (2013). Measuring the self-similarity exponent in Lévy stable processes of financial time series. *Physica A: Statistical Mechanics and its Applications*, 392(21):5330–5345.

[77] Fiedor, P. (2014). Information-theoretic approach to lead-lag effect on financial markets. *The European Physical Journal B*, 87(8):1–9.

[78] Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10(4):507–521.

[79] Flandrin, P. and Gonçalves, P. (2004). Empirical mode decompositions as data-driven wavelet-like expansions. *International Journal of Wavelets, Multiresolution and Information Processing*, 2(4):477–496.

[80] Flandrin, P., Rilling, G., and Goncalves, P. (2004). Empirical mode decomposition as a filter bank. *Signal Processing Letters, IEEE*, 11(2):112–114.

[81] Fourier, J. and Freeman, A. (1878). *The analytical theory of heat*. Cambridge University Press, Cambridge.

[82] Fryzlewicz, P., Sapatinas, T., and Rao, S. S. (2006). A Haar-Fisz technique for locally stationary volatility estimation. *Biometrika*, 93(3):687–704.

[83] Fusai, G., Germano, G., and Marazzina, D. (2016). Spitzer identity, Wiener-Hopf factorization and pricing of discretely monitored exotic options. *European Journal of Operational Research*, 251(1):124–134.

[84] Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350):320–328.

[85] Gençay, R., Gradojevic, N., Selçuk, F., and Whitcher, B. (2012). Asymmetry of information flow between volatilities across time scales. *Quantitative Finance*, 10(8):895–915.

[86] Gençay, R., Selçuk, F., and Whitcher, B. (2001). Scaling properties of foreign exchange volatility. *Physica A: Statistical Mechanics and its Applications*, 289(1–2):249–266.

[87] Goswami, J. and Chan, A. (2011). *Fundamentals of wavelets: theory, algorithms and applications*. Wiley Series in Microwave and Optical Engineering. Wiley, New Jersey.

[88] Granger, C. W. J. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29.

[89] Grossmann, A. and Morlet, J. (1984). Decomposition of hardy functions into square integrable wavelets of constant shape. *SIAM Journal on Mathematical Analysis*, 15(4):723–736.

[90] Hansen, P. R. and Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, 24(2):127–161.

[91] He, Z., Shen, Y., and Wang, Q. (2012). Boundary extension for Hilbert-Huang transform inspired by gray prediction model. *Signal Processing*, 92(3):685–697.

[92] Hibbert, A. M., Daigler, R. T., and Dupoyet, B. (2008). A behavioral explanation for the negative asymmetric return-volatility relation. *Journal of Banking & Finance*, 32(10):2254–2266.

[93] Hinich, M. J. and Patterson, D. M. (1985). Evidence of nonlinearity in daily stock returns. *Journal of Business & Economic Statistics*, 3(1):69–77.

[94] Hommes, C. (2001). Financial markets as nonlinear adaptive evolutionary systems. *Quantitative Finance*, 1(1):149–167.

[95] Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *The Journal of Finance*, 46(5):1839–1877.

[96] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995.

[97] Huang, N. E., Wu, M.-L., Qu, W., Long, S. R., and Shen, S. S. P. (2003a). Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Bussiness and Industry*, 19(3):245–268.

[98] Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., and Fan, K. L. (2003b). A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 459(2037):2317–2345.

[99] Huang, N. E. and Wu, Z. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2). RG2006.

[100] Huang, Y. X., Schmitt, F. G., Lu, Z. M., and Liu, Y. L. (2008). An amplitude-frequency study of turbulent scaling intermittency using empirical mode decomposition and Hilbert spectral analysis. *Europhysics Letters*, 84(4):40010.

[101] Huth, N. and Abergel, F. (2014). High-frequency lead/lag relationships-empirical facts. *Journal of Empirical Finance*, 26:41–58.

[102] Hyndman, R. J. and Khandakar, Y. (2008). Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, 26(3):1–22.

[103] Islam, M. R., Rashed-Al-Mahfuz, M., Ahmad, S., and Molla, M. K. I. (2012). Multi-band prediction model for financial time series with multivariate empirical mode decomposition. *Discrete Dynamics in Nature and Society*, 2012.

[104] J. M. Chambers, C. L. Mallows, B. W. S. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344.

[105] Kaastra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3):215–236.

[106] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H. (2002a). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316(1):87–114.

[107] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., and Stanley, H. (2002b). Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 316:87–114.

[108] King, F. W. (2009). *Hilbert transforms*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge.

[109] Kristoufek, L. and Vosvrda, M. (2014). Measuring capital market efficiency: long-term memory, fractal dimension and approximate entropy. *The European Physical Journal B*, 87(7):1–9.

[110] Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208.

[111] Lan Zhang, Per A. Mykland, Y. A.-S. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

[112] Lawrie, J. B. and Abrahams, I. D. (2007). A brief historical perspective of the Wiener–Hopf technique. *Journal of Engineering Mathematics*, 59(4):351–358.

[113] Lin, C.-S., Chiu, S.-H., and Lin, T.-Y. (2012a). Empirical mode decomposition based least squares support vector regression for foreign exchange rate forecasting. *Economic Modelling*, 29(6):2583–2590.

[114] Lin, D.-C., Guo, Z.-L., An, F.-P., and Zeng, F.-L. (2012b). Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression. *Mechanical Systems and Signal Processing*, 31:13–28.

[115] Liu, H., Chen, C., Tian, H.-Q., and Li, Y.-F. (2012). A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renewable Energy*, 48:545–556.

[116] Longin, F. and Solnik, B. (1995). Is the correlation in international equity returns constant: 1960-1990? *Journal of International Money and Finance*, 14(1):3–26.

[117] Lu, C.-J. and Shao, Y. E. (2012). Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. *Mathematical Problems in Engineering*, 2012.

[118] Madaleno, M. and Pinho, C. (2012). International stock market indices comovements: a new look. *International Journal of Finance & Economics*, 17(1):89–102.

[119] Mallat, S. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7):674–693.

[120] Mallat, S. (2008). *A wavelet tour of signal processing*. Academic Press, Amsterdam, 3rd edition.

[121] Mallat, S., Papanicolaou, G., and Zhifeng, Z. (1998). Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics*, 26(1):1–47.

[122] Malliavin, P. and Mancino, M. E. (2002). Fourier series method for measurement of multivariate volatilities. *Finance and Stochastics*, 6(1):49–61.

[123] Mancino, M. and Sanfelici, S. (2008). Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise. *Computational Statistics & Data Analysis*, 52(6):2966–2989.

[124] Mandelbrot, B. (1967). The variation of certain speculative prices. *The Journal of Business*, 40(4):393–413.

[125] Mandelbrot, B., Fisher, A., and Calvet, L. (1997). A multifractal model of asset returns. Cowles Foundation Discussion Papers 1164, Cowles Foundation for Research in Economics, Yale University.

[126] Mandelbrot, B. and Taylor, H. M. (1967). On the distribution of stock price differences. *Operations Research*, 15(6):1057–1062.

[127] Mandelbrot, B. B. and Ness, J. W. V. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4):422–437.

[128] Mandelbrot, B. B. and Wallis, J. R. (1969). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, 5(5):967–988.

[129] Marazzina, D., Fusai, G., and Germano, G. (2012). Pricing credit derivatives in a Wiener-Hopf framework. In *Topics in Numerical Methods for Finance*. Springer, 139–154.

[130] Marple Jr, S. L. (1999). Computing the discrete-time analytic signal via FFT. *Signal Processing, IEEE Transactions on*, 47(9):2600–2603.

[131] Masset, P. (2008). Analysis of financial time-series using Fourier and wavelet methods. *Available at SSRN 1289420.* Working paper.

[132] Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, 8(4):323–361.

[133] Montgomery, D., Jennings, C., and Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. Wiley Series in Probability and Statistics. Wiley, New York.

[134] Müller, U. A., Dacorogna, M. M., Dave, R. D., Olsen, R. B., Pictet, O. V., and von Weizsacker, J. E. (1997). Volatilities of different time resolutions – Analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2–3):213–239.

[135] Müller, U. A., Dacorogna, M. M., Davé, R. D., Pictet, O. V., Olsen, R. B., and Ward, J. R. (1993). Fractals and intrinsic time: A challenge to econometricians. *Unpublished manuscript, Olsen & Associates, Zürich.*

[136] Nava, N., Di Matteo, T., and Aste, T. (2016a). Time-dependent scaling patterns in high frequency financial data. *The European Physical Journal*. In press.

[137] Nava, N., Matteo, T. D., and Aste, T. (2016b). Anomalous volatility scaling in high frequency financial data. *Physica A: Statistical Mechanics and its Applications*, 447:434–445.

[138] Papadimitriou, S., Sun, J., and Yu, P. S. (2006). Local correlation tracking in time series. In *Sixth International Conference on Data Mining* 456–465.

[139] Patrick, F. (2007). http://perso.ens-lyon.fr/patrick.flandrin/emd.html. Accessed 26/02/2015.

[140] Patton, A. J. (2011). Data-based ranking of realised volatility estimators. *Journal of Econometrics*, 161(2):284–303.

[141] Pegram, G., Peel, M., and McMahon, T. (2008). Empirical mode decomposition using rational splines: an application to rainfall time series. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 464(2094):1483–1501.

[142] Peng, C.-K., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., and Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, 49:1685–1689.

[143] Peng, Z., Tse, P. W., and Chu, F. (2005). A comparison study of improved Hilbert-Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, 19(5):974–988.

[144] Percival, D. B. (1995). On estimation of the wavelet variance. *Biometrika*, 82(3):619–631.

[145] Percival, D. B. and Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University, Cambridge.

[146] Peters, E. E. (1994). *Fractal market analysis: applying chaos theory to investment and economics*. Wiley, New York.

[147] Protter, P. (2005). *Stochastic integration and differential equations*. Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, New York.

[148] Ramsey, J. B. and Zhang, Z. (1997). The analysis of foreign exchange data using waveform dictionaries. *Journal of Empirical Finance*, 4(4):341–372.

[149] Rao, C. R. (2009). *Linear statistical inference and its applications*, volume 22. John Wiley & Sons, New York.

[150] Renó, R. (2003). A closer look at the Epps effect. *International Journal of Theoretical and Applied Finance*, 06(01):87–102.

[151] Rilling, G., Flandrin, P., and Goncalves, P. (2003). On empirical mode decomposition and its algorithms. *Proceedings of IEEE EURASIP Workshop on Nonlinear Signal and Image Processing NSIP03*.

[152] Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schürmann, M., and Başar, E. (2001). Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105(1):65–75.

[153] Rua, A. and Nunes, L. C. (2009). International comovement of stock market returns: A wavelet analysis. *Journal of Empirical Finance*, 16(4):632–639.

[154] Saadi, S., Gandhi, D., and Elmawazini, K. (2006). On the validity of conventional statistical tests given evidence of non-synchronous trading and non-linear dynamics in returns generating process. *Applied Economics Letters*, 13(5):301–305.

[155] Sahalia, Y. A., Fan, J., and Li, Y. (2013). The leverage effect puzzle: disentangling sources of bias at high frequency. *Journal of Financial Economics*, 109(1):224–249.

[156] Samoradnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian random processes: stochastic models with infinite variance*. Stochastic Modeling Series. Taylor & Francis.

[157] Selçuk, F. and Gençay, R. (2006). Intraday dynamics of stock market returns and volatility. *Physica A: Statistical Mechanics and its Applications*, 367:375–387.

[158] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 623–656.

[159] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. *Monographs on Statistics and Applied Probability*, 65:1–68.

[160] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222.

[161] Stărică, C. and Granger, C. (2005). Nonstationarities in stock returns. *Review of Economics and Statistics*, 87(3):503–522.

[162] Stenger, F. (2012). *Numerical methods based on sinc and analytic functions*, volume 20. Springer Science & Business Media, New York.

[163] Stoev, S., Pipiras, V., and Taqqu, M. S. (2002). Estimation of the self-similarity parameter in linear fractional stable motion. *Signal Processing*, 82(12):1873–1901.

[164] Stoev, S. and Taqqu, M. S. (2004). Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform. *Fractals*, 12(01):95–121.

[165] Suykens, J., Brabanter, J. D., Lukas, L., and Vandewalle, J. (2002). Weighted least squares support vector machines: robustness and sparse approximation. *Neurocomputing*, 48(1):85–105.

[166] Taqqu, M. S. and Teverovsky, V. (1998). A practical guide to heavy tails. chapter On Estimating the Intensity of Long-range Dependence in Finite and Infinite Variance Time Series, pages 177–217. Birkhauser Boston Inc., Cambridge, MA, USA.

[167] Tay, F. E. and Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4):309–317.

[168] Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79:61–78.

[169] Torrence, C. and Webster, P. J. (1999). Interdecadal changes in the ENSO-Monsoon system. *Journal of Climate*, 12(8):2679–2690.

[170] Vácha, L. and Baruník, J. (2012). Co-movement of energy commodities revisited: Evidence from wavelet coherence analysis. *Energy Economics*, 34(1):241–247.

[171] Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag, New York.

[172] Vidakovic, B. (2009). *Statistical modeling by wavelets*, volume 503. John Wiley & Sons, New York.

[173] Wang, J., Zhang, W., Li, Y., Wang, J., and Dang, Z. (2014). Forecasting wind speed using empirical mode decomposition and Elman neural network. *Applied Soft Computing*, 23:452–459.

[174] Weron, A. and Weron, R. (2000). Fractal market hypothesis and two power laws. *Chaos, Solitons and Fractals*, 11(1–3):289–296.

[175] Whaley, R. E. (2000). The investor fear gauge. *The Journal of Portfolio Management*, 26(3):12–17.

[176] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.

[177] Wild, P., Foster, J., and Hinich, M. J. (2014). Testing for non-linear and time irreversible probabilistic structure in high frequency financial time series data. *Journal of the Royal Statistical Society: Series A*, 177(3):643–659.

[178] Willinger, W., Taqqu, S. M., and Teverovsky, V. (1999). Stock market prices and long-range dependence. *Finance and Stochastics*, 3(1):1–13.

[179] Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30:79–82.

[180] Wood, R. A., McInish, T. H., and Ord, K. (1985). An investigation of transactions data for NYSE stocks. *Journal of Finance*, 40(3):723–39.

[181] Wu, M.-C. (2007). Phase correlation of foreign exchange time series. *Physica A: Statistical Mechanics and its Applications*, 375(2):633–642.

[182] Wu, Z. and Huang, N. E. (2004). A study of the characteristics of white noise using the empirical mode decomposition method. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 460(2046):1597–1611.

[183] Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1–41.

[184] Yu, L., Wang, S., and Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5):2623–2635.

[185] Zeng, Q. and Qu, C. (2014). An approach for Baltic Dry Index analysis based on empirical mode decomposition. *Maritime Policy & Management*, 41(3):224–240.

[186] Zhang, B.-L., Coggins, R., Jabri, M., Dersch, D., and Flower, B. (2001). Multiresolution forecasting for futures trading using wavelet decompositions. *Neural Networks, IEEE Transactions on*, 12(4):765–775.

[187] Zhang, X., Lai, K., and Wang, S.-Y. (2008). A new approach for crude oil price analysis based on empirical mode decomposition. *Energy Economics*, 30(3):905–918.

[188] Zhou, B. (1996). High-frequency data and volatility in foreign-exchange rates. *Journal of Business & Economic Statistics*, 14(1):45–52.

[189] Zhou, W.-X. and Sornette, D. (2003). Nonparametric analyses of log-periodic precursors to financial crashes. *International Journal of Modern Physics C*, 14(08):1107–1125.

[190] Zunino, L., Tabak, B. M., Pérez, D. G., Garavaglia, M., and Rosso, O. A. (2007). Inefficiency in Latin-American market indices. *The European Physical Journal B*, 60(1):111–121.

[191] Zunino, L., Zanin, M., Tabak, B. M., Pérez, D. G., and Rosso, O. A. (2010). Complexity-entropy causality plane: A useful approach to quantify the stock market inefficiency. *Physica A: Statistical Mechanics and its Applications*, 389(9):1891–1901.

Appendix A

ARIMA and SVR model parameters

In this appendix, we report the parameters used in the implementation of the forecasting models. The analysed data set consists of 128 time series of intraday observations of the S&P 500 index which were separately used to trained the EMD-SVR models. We report a summary of the obtained parameters using histograms.

A.1 ARIMA model parameters

One of the benchmark models was the ARIMA model. We fitted this model to each S&P 500 intraday time series and to its respective IMFs. The obtained parameters are worth to be mentioned since the input vector of the EMD-SVR forecasting models is based on these parameters. In Figure A.1, we report the histogram for the number of the autoregressive terms, p. Each Subfigure corresponds to the model fitted to the time series specified in the plot title.

Figures A.2 and A.3 report the order of the moving average term, q, and the number of differentiations needed to achieve stationarity, d, respectively.

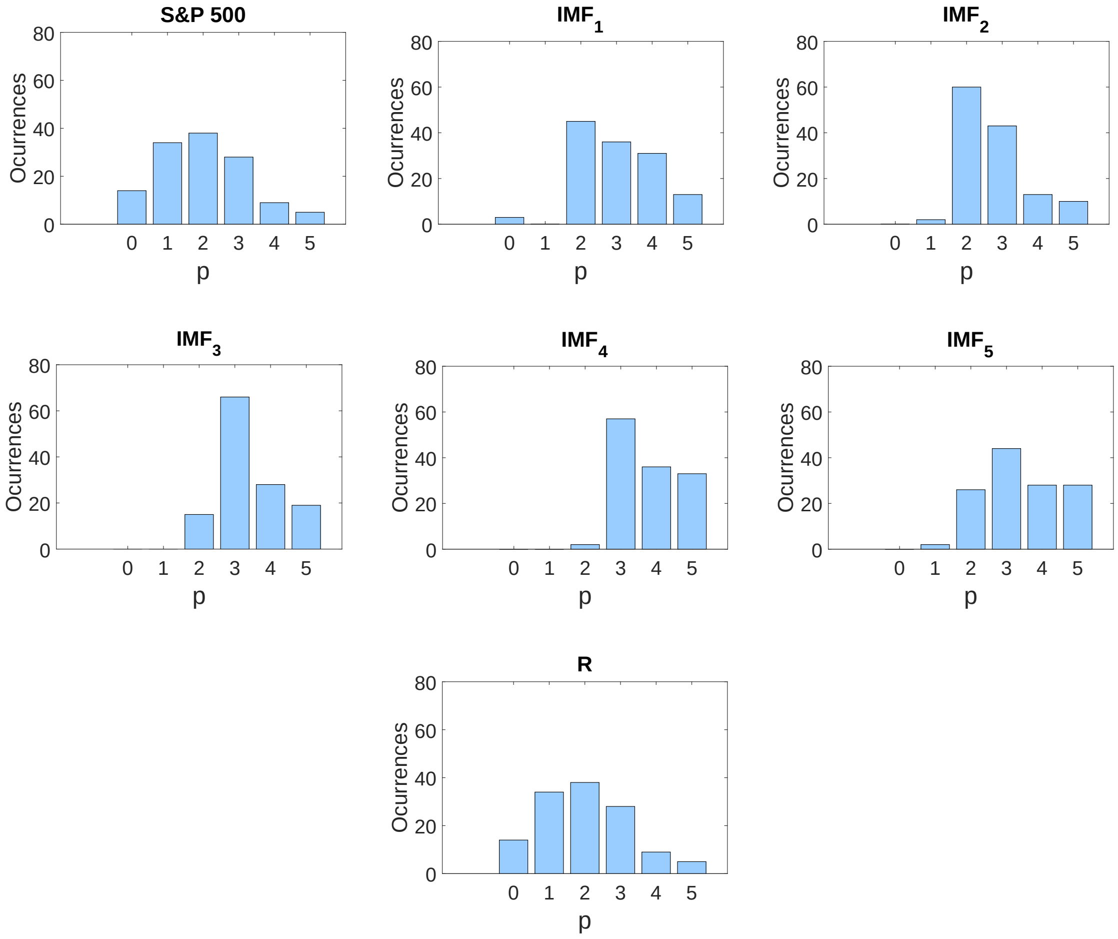


Fig. A.1 Number of autoregressive terms p for the ARIMA model fitted to the indicated time series.

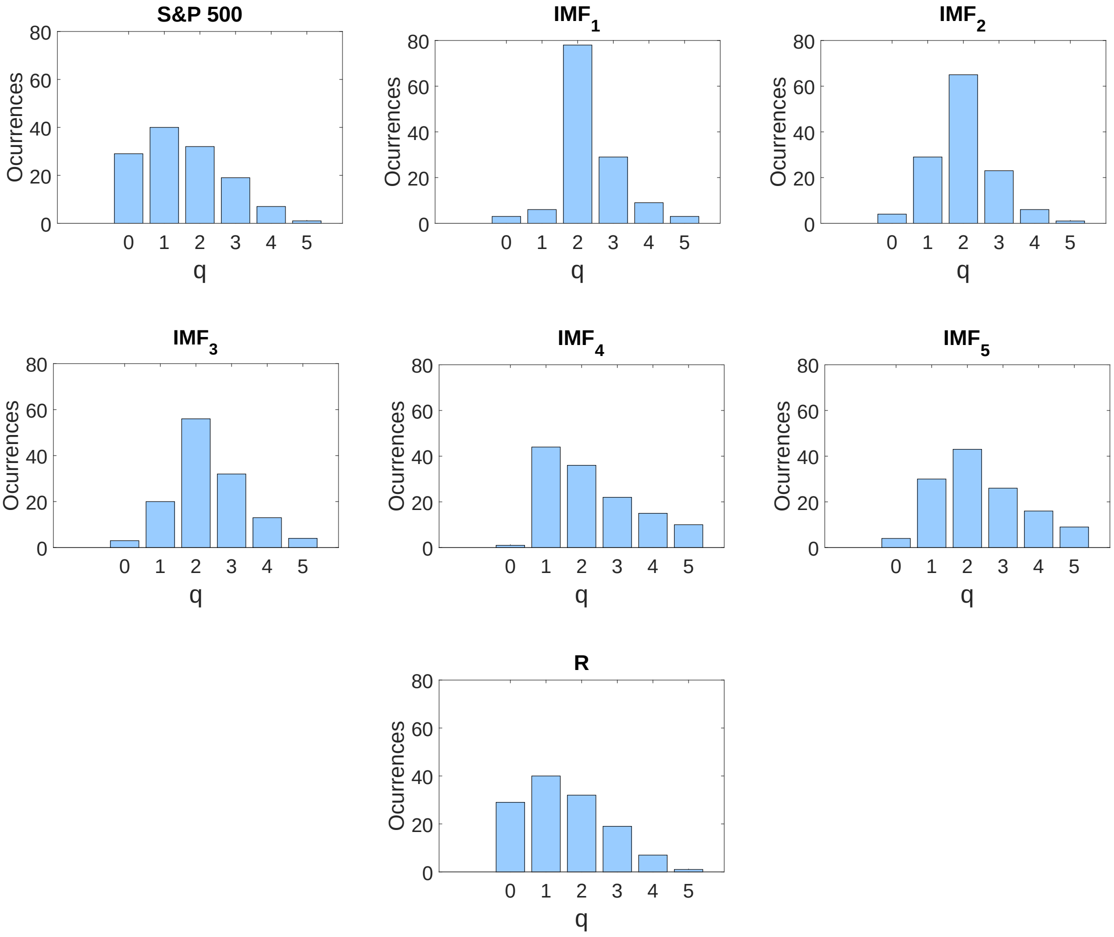


Fig. A.2 Order of the moving average term, q, for the ARIMA model fitted to the indicated time series.

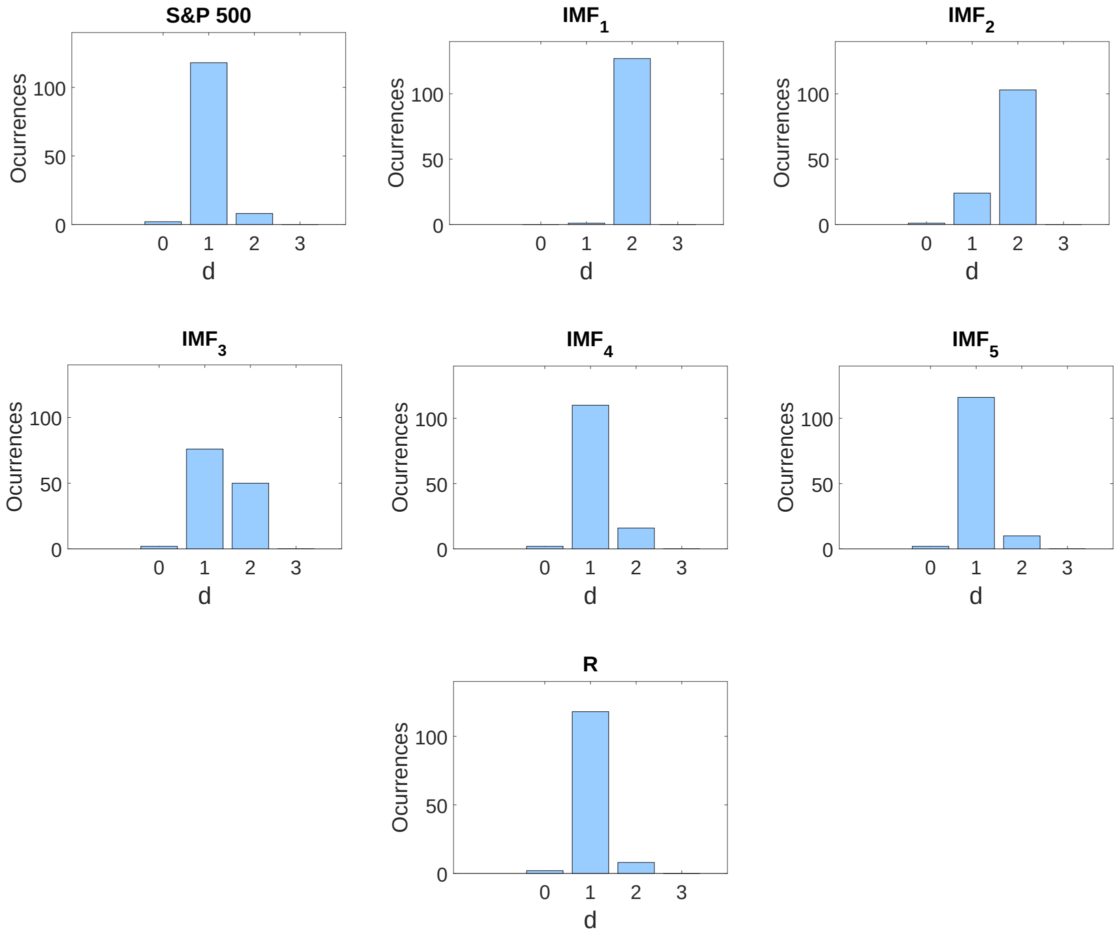


Fig. A.3 Number of differencing d, for the ARIMA model fitted to the indicated time series.

A.2 Support vector regression parameters

For the SVR models, we only report parameters for the 1-step ahead forecast and for models with input vector of length $m = p + d$. The SVR models were trained on the S&P 500 time series and their respective IMFs.

Figure A.4 reports the values obtained for the penalization constant C. A greater value of C corresponds to a greater error penalty, indicating that the objective is only to minimize the empirical risk creating a more complex model. A smaller value may cause the errors too be excessively tolerated, resulting in a poor approximation. If the data are noisy, then smaller values of C may be preferred.

Figure A.5 reports the values of ε, the insensitive tube radius. This parameter affects the smoothness or complexity of the approximation function, controls the accuracy of the approximation function and determines the number of support vectors. Smaller values of ε may lead to more support vectors and result in a complex model. Larger values of ε may cause the ε-insensitive tube to include too many data that are unseen by the model.

Figure A.6 reports the values of γ, the kernel width, which determines the flexibility of the resulting SVR model. Over-fitting occurs when this parameter is too large.

Figure A.7 reports the number of support vectors for each model. These vectors “support” the definition of the approximation function. Cherkassky and Ma [50] reported that a model performance is optimized when the percentage of support vectors is about 50% of the training data.

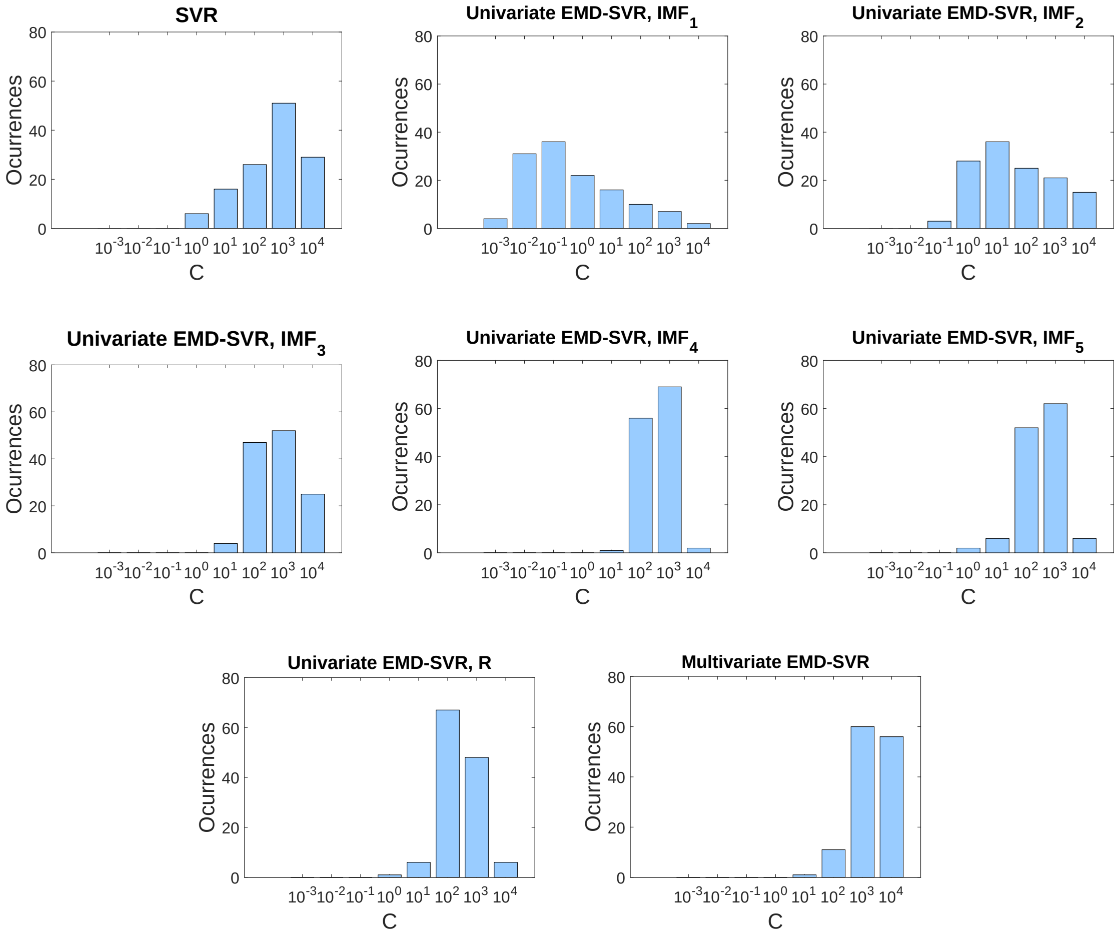


Fig. A.4 Regularization constant C for the SVR model fitted to the indicated time series.

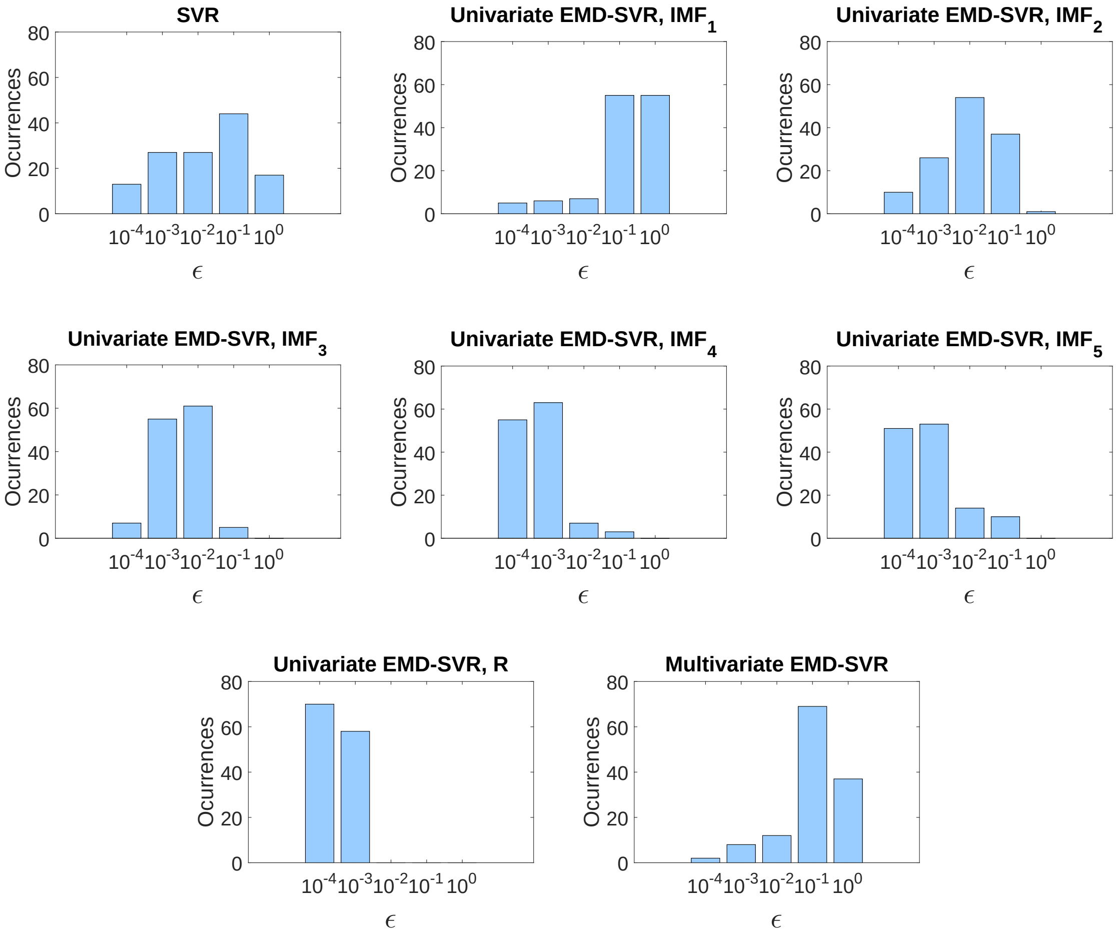


Fig. A.5 Insensitive coefficient ε for the SVR model fitted to the indicated time series.

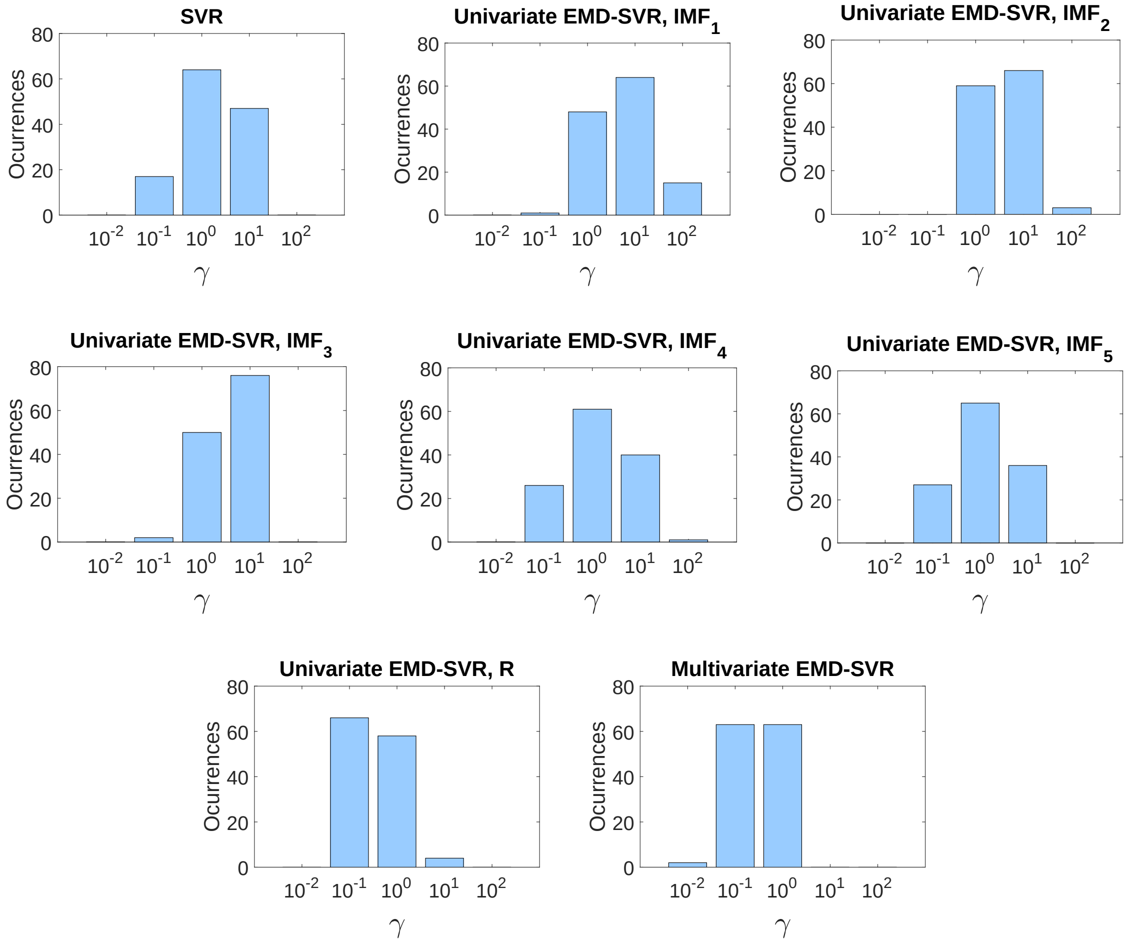


Fig. A.6 Kernel width parameter γ for the SVR model fitted to the indicated time series.

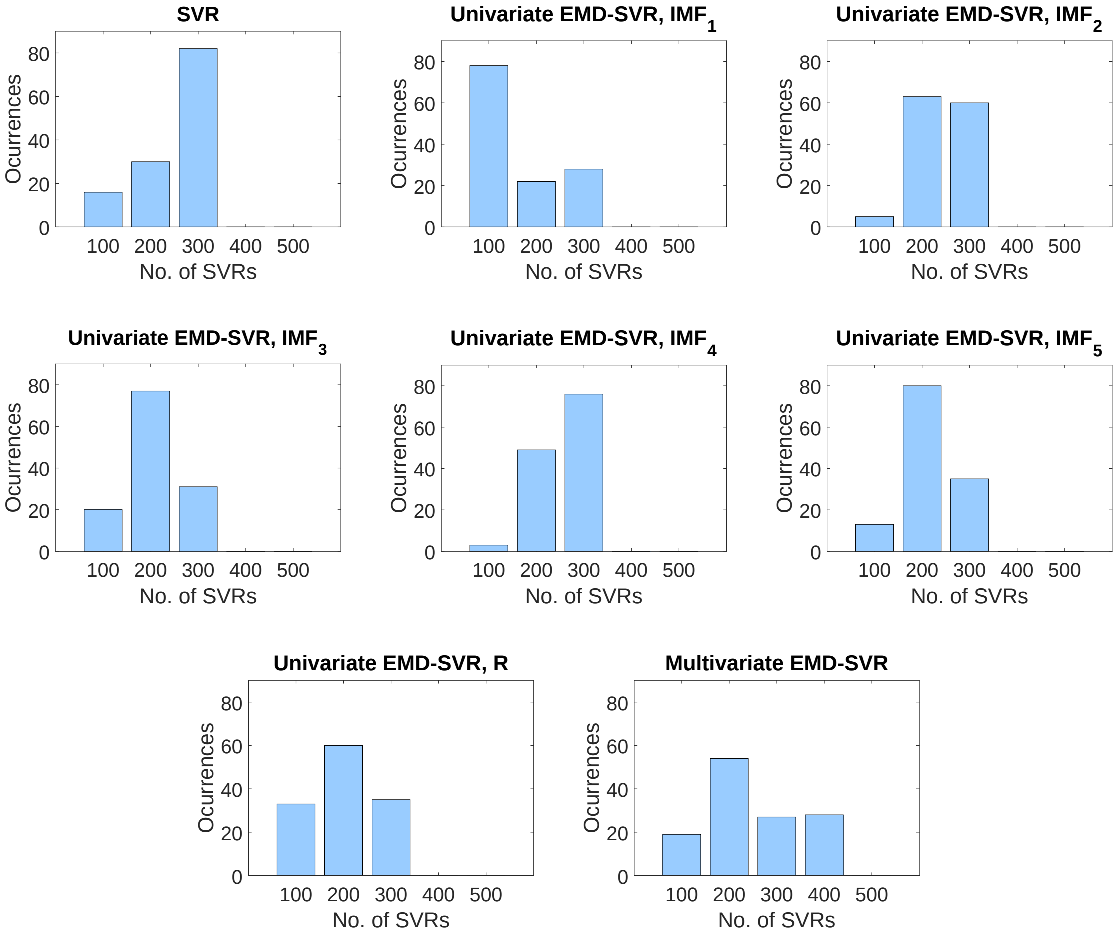


Fig. A.7 Number of support vectors for the SVR model fitted to the indicated time series.

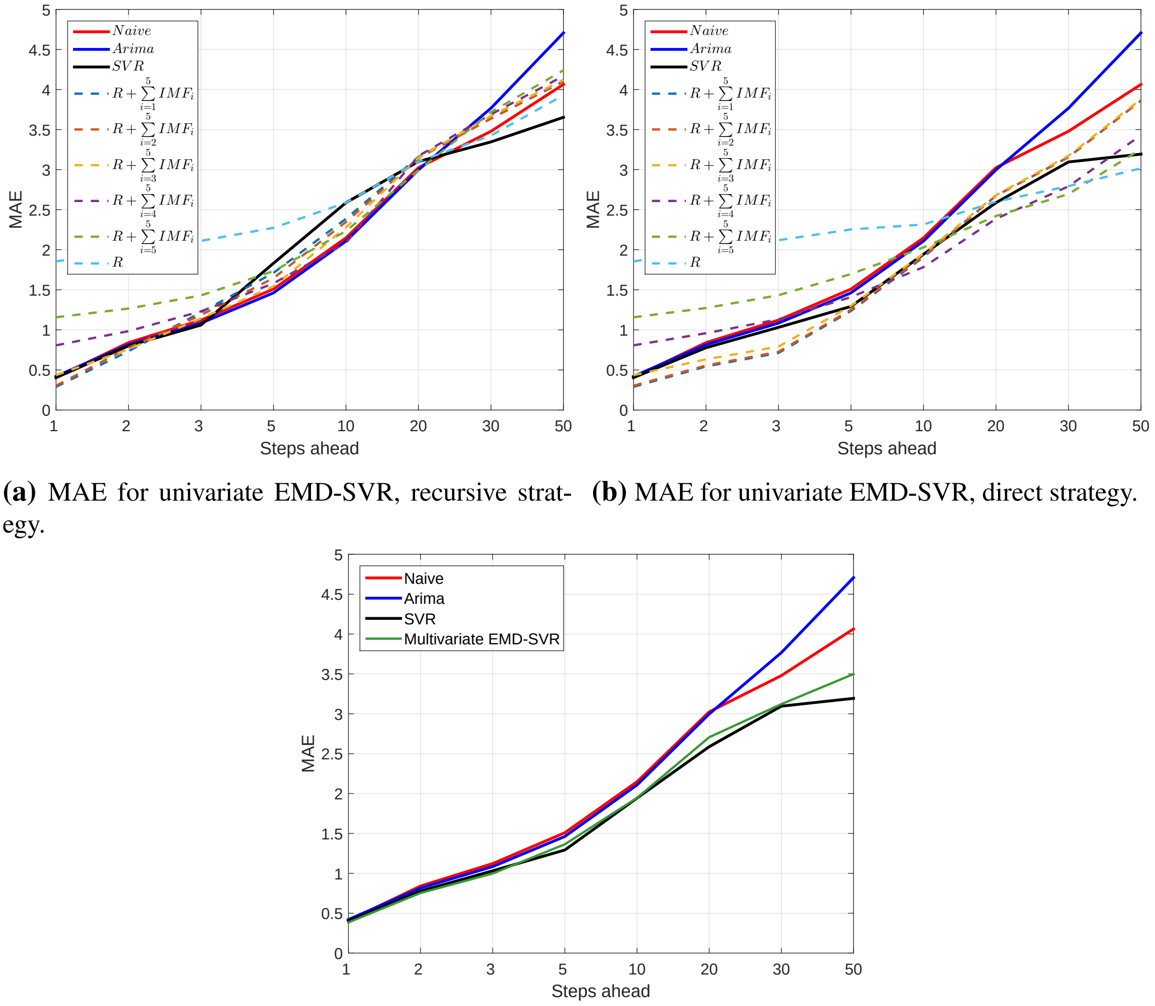
Appendix B

Forecasting results for the FTSE 100 index

The data set for the FTSE 100 index consists of 128 days of intraday data sample at 30-second intervals from the period May 2015 to September 2015. The London stock exchange opens at 8:00 am and closes at 4:30 pm EST, thus, a full trading day consists of 1020 prices. We excluded weekends and public holidays and we avoided market opening effects by eliminating the first 5 minutes in every day. We used a training set consisting of $N = 500$ prices and forecasted up to $h = 50$ steps ahead (25 minutes ahead). We tested three input vector with different lengths, $m = 1$, $m = 5$, and $m = p+d$. The following table reports the results for the input vector with the best results, that is, $m = p+d$ (p and d obtained from the ARIMA fitted model). In Table B.1, we report the MAE and its standard deviation (parentheses). In Figure B.1, we compare the MAE with respect to the forecasting horizon. Finally, in Table B.2, we report the results for the Wilcoxon test.

	Steps ahead	1		2		3		5		10		20		30		50	
	Model	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
	Naive	0.406	(0.489)	0.841	(0.738)	1.122	(1.054)	1.509	(1.498)	2.149	(1.898)	3.025	(2.731)	3.479	(2.795)	4.064	(3.831)
	ARIMA	0.420	(0.467)	0.816	(0.722)	1.085	(1.029)	1.462	(1.407)	2.108	(1.884)	2.997	(2.865)	3.767	(3.399)	4.707	(5.018)
Direct	SVR	0.404	(0.441)	0.778	(0.666)	1.031	(0.953)	1.293	(1.380)	1.943	(1.722)	2.587	(2.506)	3.096	(2.547)	3.193	(3.396)
	Multivariate	0.384	(0.415)	0.754	(0.644)	0.998	(0.876)	1.364	(1.298)	1.946	(1.782)	2.706	(2.488)	3.121	(2.702)	3.498	(2.867)
	$R+\sum_{i=1}^{5} IMF_i$	**0.289**	(0.322)	**0.543**	(0.463)	**0.712**	(0.680)	**1.237**	(1.237)	1.908	(1.679)	2.674	(2.312)	3.157	(2.604)	3.861	(3.642)
	$R+\sum_{i=2}^{5} IMF_i$	0.303	(0.352)	0.556	(0.472)	0.727	(0.686)	1.251	(1.253)	1.921	(1.690)	2.675	(2.308)	3.154	(2.606)	3.864	(3.643)
	$R+\sum_{i=3}^{5} IMF_i$	0.435	(0.461)	0.632	(0.592)	0.792	(0.752)	1.300	(1.357)	1.951	(1.728)	2.682	(2.309)	3.172	(2.596)	3.881	(3.653)
	$R+\sum_{i=4}^{5} IMF_i$	0.807	(0.813)	0.958	(0.905)	1.132	(1.026)	1.409	(1.398)	**1.783**	(1.625)	**2.387**	(2.159)	2.791	(2.424)	3.429	(3.298)
	$R+\sum_{i=5}^{5} IMF_i$	1.157	(1.211)	1.273	(1.345)	1.433	(1.481)	1.697	(1.764)	2.029	(1.846)	2.421	(2.281)	**2.695**	(2.410)	3.256	(3.175)
	R	1.855	(1.789)	1.986	(1.928)	2.119	(2.078)	2.254	(2.289)	2.315	(2.252)	2.600	(2.597)	2.794	(2.499)	**3.017**	(2.858)
Recursive	SVR	0.404	(0.415)	0.801	(1.130)	1.060	(1.552)	1.832	(3.437)	2.587	(4.212)	3.102	(3.756)	3.346	(3.744)	3.653	(4.150)
	Multivariate	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} IMF_i$	0.289	(0.322)	0.734	(0.789)	1.219	(1.421)	1.711	(1.782)	2.387	(2.286)	3.159	(2.598)	3.636	(3.298)	4.101	(3.461)
	$R+\sum_{i=2}^{5} IMF_i$	0.303	(0.352)	0.777	(0.922)	1.185	(1.362)	1.647	(1.653)	2.347	(2.279)	3.142	(2.576)	3.637	(3.317)	4.104	(3.525)
	$R+\sum_{i=3}^{5} IMF_i$	0.435	(0.461)	0.758	(0.769)	1.144	(1.223)	1.543	(1.569)	2.279	(2.277)	3.149	(2.532)	3.671	(3.295)	4.121	(3.508)
	$R+\sum_{i=4}^{5} IMF_i$	0.807	(0.813)	0.984	(0.894)	1.231	(1.076)	1.580	(1.596)	2.086	(1.993)	3.171	(2.556)	3.690	(3.183)	4.169	(3.371)
	$R+\sum_{i=5}^{5} IMF_i$	1.157	(1.211)	1.266	(1.343)	1.430	(1.488)	1.733	(1.927)	2.232	(2.212)	2.992	(2.730)	3.718	(3.230)	4.240	(3.595)
	R	1.855	(1.789)	1.983	(1.922)	2.111	(2.081)	2.273	(2.335)	2.588	(2.512)	3.101	(2.944)	3.425	(2.952)	3.929	(3.384)

Table B.1 MAE and std for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = p+d$ lagged values, the same input vector as the ARIMA model. The smallest MAE of each forecast horizon is set in boldface.



(a) MAE for univariate EMD-SVR, recursive strategy.

(b) MAE for univariate EMD-SVR, direct strategy.

(c) MAE for multivariate EMD-SVR.

Fig. B.1 MAE as a function of the forecast horizon for the considered forecasting models: naive, ARIMA, univariate and multivariate EMD-SVR with input vector $m = p + d$ lagged values.

	Model	1	2	3	5	10	20	30	50
	ARIMA	-1.023	2.173*	2.026*	1.461	0.614	1.311	0.313	-0.602
Direct	SVR	0.636	1.432	1.511	2.656**	1.724	2.140*	2.228*	2.798**
	Multivariate	3.023	6.157	6.014	6.890	6.647	8.335	7.272	8.521
	$R+\sum_{i=1}^{5} IMF_i$	3.815**	7.363**	7.515**	5.356**	3.303**	2.628**	2.730**	0.977
	$R+\sum_{i=2}^{5} IMF_i$	3.403**	6.599**	7.068**	5.073**	3.165**	2.654**	2.742**	0.989
	$R+\sum_{i=3}^{5} IMF_i$	-1.329	4.276**	5.510**	4.169**	2.737**	2.554*	2.604**	0.920
	$R+\sum_{i=4}^{5} IMF_i$	-6.124**	-1.201	-0.128	1.339	3.610**	3.976**	4.404**	3.153**
	$R+\sum_{i=5}^{5} IMF_i$	-7.239**	-3.408**	-2.086*	-0.913	1.296	3.377**	4.352**	3.893**
	R	-8.752**	-6.737**	-5.680**	-4.134**	-0.882	1.969*	3.620**	4.554**
Recursive	SVR	0.636	3.671**	2.843**	1.848	1.253	2.171*	2.835**	3.467**
	Multivariate	–	–	–	–	–	–	–	–
	$R+\sum_{i=1}^{5} IMF_i$	3.815**	3.413**	1.046	-0.357	-0.654	-0.925	-0.069	-0.017
	$R+\sum_{i=2}^{5} IMF_i$	3.403**	3.655**	1.118	0.195	-0.347	-0.718	-0.052	0.131
	$R+\sum_{i=3}^{5} IMF_i$	-1.329	2.621**	1.403	1.398	0.005	-0.839	-0.140	-0.050
	$R+\sum_{i=4}^{5} IMF_i$	-6.124**	-1.903	-1.394	0.297	1.161	-0.697	-0.366	-0.162
	$R+\sum_{i=5}^{5} IMF_i$	-7.239**	-3.344**	-2.024*	-0.854	0.666	0.583	-0.335	-0.785
	R	-8.752**	-6.794**	-5.650**	-4.209**	-2.018*	0.014	0.680	0.511

Table B.2 Z-statistic for the Wilcoxon signed-rank test for the null hypothesis that the naive model is as accurate as the studied models: ARIMA model, univariate and multivariate EMD-SVR models with input vector $m = p + d$. Top, direct strategy, bottom, recursive strategy.
* Statistically significant at the 5% confidence level
** Statistically significant at the 1% confidence level.