Wavelet Regression Using a Lévy Prior Model

A thesis presented for the degree of

Doctor of Philosophy of the Diploma of Imperial College

by

Badr Missaoui

Department of Mathematics

Imperial College

180 Queen's Gate,

London SW7 2BZ

MAY 17, 2013

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:

Copyright

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author and lodged in the doctorate thesis archive of the college central library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in Imperial College, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement. Further information on the conditions under which disclosures and exploitation may take place is available from the Imperial College registry.

For my family FMS.

Abstract

This thesis is concerned with nonparametric regression and regularization. In particular, wavelet regression using a Lévy prior model is investigated. The use of this prior is motivated by the statistical properties, such as heavy-tails, common in many datasets of interest, such as those in financial time series.

The Lévy process we propose captures the heavy tails of the wavelet coefficients of an unknown function. We study the Besov regularity of the wavelet coefficients and establish the connection between the parameters of the Lévy wavelet prior model and Besov spaces. At first, we gave a necessary and sufficient condition such that the realizations of the prior model fall into a certain class of Besov spaces. We show that the tempered stable distribution preserves its functional form for different time scales. We prove that this scaling behaviour can model the exponential-decay-across-scale property of the wavelet coefficients without imposing any specified structure on the coefficients' energy.

We also introduce a Lévy wavelet mixture model to capture the sparseness of the wavelet coefficients. We show that this sparse model exhibits a thresholding rule.
We also study the Lévy tempered stable prior model under a Bayesian framework. For the prior specified, we gave a closed form to the posterior Lévy measure of the wavelet coefficients and estimate the hyperparameters of the prior model in both a simulation study and for the S&P 500 time series.

We focus on density estimation using a penalized likelihood approach. Primarily, we study the wavelet Tsallis entropy and Fisher information and give closed-form expressions for these measures when the wavelet coefficients are driven by a tempered stable process. Then, we develop an entropic regularization based on the wavelet

Tsallis entropy and show that the penalized maximum likelihood method improves the convergence of the estimates.

Acknowledgements

I would like to thank Dr Emma McCoy for supervising this thesis. Working in the area of wavelet and nonparametric estimation has been a true learning experience for me and I am immensely grateful for the help my supervisor have given me. I also want to record a debt of gratitude to Prof. Nick Bingham for his comments, insights, and discussions. Lastly, thanks to Chris Minas for his help, discussions and fun times.

Badr Missaoui

Contents

Abstract **5**

1 Introduction **15**
1.1 Chapter Outline 18

2 Data Analysis **20**
2.1 Introduction 20
2.2 Data Analysis 21
2.3 Conclusion 23

3 Wavelet Analysis **29**
3.1 Introduction 29
3.2 Fourier Transform 29
3.2.1 Fourier Transform and Regularity 31
3.2.2 Short-Time Fourier Transform 32
3.3 Wavelet Transform 34
3.4 Continuous Wavelet Transform 34
3.4.1 Regularity 38
3.4.2 Hardy Spaces 40
3.5 Multiresolution Analysis 42
3.5.1 Theoretical framework 43
3.6 Discrete Wavelet Transform 47
3.7 Besov spaces 50
3.8 Summary 54

4 Basics on Lévy processes **55**
4.1 Definitions 55
4.2 Subordinators 61
4.3 Stable processes 62
4.4 Stochastic Integral representation for Stable Processes 68
4.5 Tempered stable process 68
4.6 Asymptotic behaviour tempered stable distributions 71

4.7 Wiener Hopf Factorization . . . 73
4.8 Simulation of tempered stable processes . . . 76
4.9 Summary . . . 78

5 Wavelet Statistical Models **81**
5.1 The wavelet approach to nonparametric regression . . . 81
5.1.1 Linear Wavelet smoothing . . . 82
5.1.2 Nonlinear Wavelet smoothing . . . 83
5.1.3 Existing Wavelet Prior models . . . 89
5.1.4 Prior models . . . 90
5.2 Lévy wavelet prior and Besov spaces . . . 95
5.2.1 Tempered Stable prior and Besov spaces . . . 96
5.3 Summary . . . 100

6 Bayesian approach **101**
6.1 A priori Model . . . 101
6.2 Bayesian Thresholding . . . 104
6.3 Estimation of the hyperparameters of Tempered Stable Distributions 107
6.3.1 Method of Moments . . . 108
6.3.2 Estimation of the parameters for symmetric Tempered Stable distribution . . . 109
6.3.3 Maximum Likelihood estimators for symmetric Tempered Stable distributions. . . . 110
6.4 MLE Results for Parameter Estimation of Tempered Stable distributions 112
6.5 Summary . . . 115

7 Regularization **118**
7.1 Introduction . . . 118
7.2 Penalized Least Square . . . 119
7.3 Penalized Maximum Likelihood . . . 122
7.4 About Entropy . . . 123
7.5 Wavelet Entropy . . . 127
7.6 Wavelet Entropy for $1/f^{\alpha}-$processes . . . 129
7.7 Wavelet Entropy for tempered stable processes . . . 131
7.8 Wavelet Fisher's information measure . . . 132
7.9 Wavelet Entropy Regularization . . . 134
7.10 Summary . . . 136

8 Conclusions and Future Work **138**
8.1 Conclusions . . . 138
8.2 Future Work . . . 139
8.3 Appendix A . . . 141
8.4 Appendix B . . . 142

List of Tables

2.1 Descriptive Statistics of log-prices for S&P500 24
2.2 Descriptive Statistics of log-returns for S&P500 24
2.3 Descriptive Statistics of Wavelet Coefficients of log-returns for S&P500 24
2.4 Parameter estimates of the log-Stable for the wavelet coefficients of the log-returns for S&P500 25

3.1 Hölder regulatity of Daubechies scaling functions (Table 3.3, Vidakovic (1999)) 50

6.1 Estimation of the parameters. $\sigma_{\Delta t} = (\Delta t)^{1/\alpha}\sigma_{\Delta t=1}$ for different number of simulation (Nb. Sim) 113
6.2 Estimation of the parameters for $N = 2048$ 114
6.3 Estimation of the parameters for $N = 4096$ 114

7.1 Penalized maximum likelihood estimates 137

List of Figures

2.1 Empirical PDF and CDF Fitting of the wavelets coefficients of the daily log-returns S&P500 . . . 25
2.2 Left and right tails of the CDF of the wavelets coefficients of the daily log-returns S&P500 . . . 26
2.3 Daily Log-prices S&P500 . . . 27
2.4 Daily Log-returns S&P500 . . . 27
2.5 Q-Q plot of Log-prices S&P500 . . . 28

3.1 Morlet's wavelet for $w_0 = 5.336$ (real part). . . . 37
3.2 Mexican hat wavelet for $\sigma = 1$ (real part). . . . 37
3.3 The Cauchy wavelet for $\alpha = 3$ in time representation (real part). . . . 38
3.4 Signal discontinuity. . . . 41
3.5 Level-6 Multiresolution Analysis of EEG time series. . . . 47
3.6 Daubechies' scaling and wavelet functions with indices 2, 3, 4 and 8. . . 51

4.1 Probability density function of α-stable distributions for different tail indices;$\alpha = 0.5; 1; 1.5; 2$. . . . 67
4.2 Cumulative density function of symmetric tempered stable distributions for different tail indices. . . . 72
4.3 Sample paths of a TS(1.2,0.5,0.5), TS(1.5,0.5,0.5) and Brownian motion (solid line) . . . 79
4.4 Sample paths of a stable TS(1.7,1,0) (Dot) and a tempered stable TS(1.7,1,0.5) (solide) . . . 80

5.1 Left: soft thresholding at λ. Right: hard thresholding. . . . 84
5.2 SCAD Estimator . . . 85
5.3 Denoising using mutli-level decomposition . . . 87
5.4 Denoising using mutli-level decomposition . . . 89

6.1 Sample paths of a stable TS(1.7,1,0) (Dot line) and a tempered stable TS(1.7,1,0.5) (solid line) . . . 113
6.2 Probability Density fit of the S&P 500 data. . . . 116
6.3 Cumulative Density fit of the S&P 500 data . . . 117

7.1 Wavelet Tsallis Entropy of a tempered stable process for $N = 10$ and different $q = 1.01; 1.5; 3; 5; 10$. 132
7.2 Wavelet Fisher Information for tempered stable process for $N = 5; 10; 15; 20$. 134

List of Papers

The research presented in this thesis is being prepared for publication in the following working papers:

The following paper is submitted to *Journal of the Royal Statistical Society Series B*.
Missaoui B., McCoy E.J. Lévy Process Wavelet Prior Model.
Which has the abstract:
"In this paper, we introduce the Lévy wavelet prior model. We propose a Lévy process to capture the tail heaviness of the wavelet coefficients ubiquitous in many observed time series. We study the Besov regularity of the wavelet coefficients and establish the connection between the parameters of the Lévy wavelet prior model and Besov spaces. We gave a necessary and sufficient condition such that the realizations of the prior model fall into a certain class of Besov spaces. We prove that we can model the common phenomenon of an exponential decay across scale of the wavelet coefficients without imposing any specified structure on the variance of the wavelet coefficients. Moreover, we study a Bayesian approach under the Lévy prior model, and employ this to estimate the hyperparameters of the prior model in both a simulation study and for the S&P 500 time series."

The following paper is being prepared for submission to *Entropy*.
Missaoui B., McCoy E.J. The Wavelet Entropy of Tempered Stable Processes.
Which has the following abstract:
"In this paper, we extend the work of Zunino et al. (2006) to non-Gaussian processes. We compute the wavelet Tsallis entropy and the wavelet Fisher's information for tem-

pered stable processes. Also, we introduce a regularization approach for estimating nonparametric wavelet regression. We follow here the Bayesian approach where we set the prior distribution to be tempered stable and we focus on finding a solution to the wavelet-based regression through a regularization approach based on the Tsallis entropy functional."

Chapter 1

Introduction

In recent years, nonparametric regression has gained popularity as a method of estimating real functions and Bayesian techniques have been successfully proposed to capture the statistical properties of real functions in a univariate or multivariate framework.
Many nonparametric regression estimators have been proposed. The most popular of which include smoothing splines and kernel estimators (Green and Silverman, 1994). The models are of the form

$$y_i = g(t_i) + \varepsilon_i, \quad i = 1, .., n, \tag{1.1}$$

with various assumptions on ε_i and t_i. Nevertheless, these estimators may not be satisfactory if, for example, the underlying is characterized by singularities. Therefore, wavelet-based estimators have been introduced to recover the edge structure of signals. The approach consists of expanding the data as a wavelet series and extracting the wavelet coefficients. Donoho and Johnstone (1994) and Donoho et al. (1995) demonstrate that the wavelet estimators are optimal when the signal belongs to a certain class of Besov spaces. The connection between Besov spaces and wavelet statistical models was first studied by Abramovich et al. (1998).

Since then, it has been shown that the standard thresholding approaches are outperformed by Bayesian methods; see Abramovich et al. (1998) and Baraniuk and

Choi (1999). Mallat (1989) used the Generalized Gaussian Distribution (GGD) to model the wavelet coefficients for images as the wavelet coefficients show a heavy tailed behavior and the variance of their marginal distribution is exponentially decaying across scale. Achim et al. (2001) and Boubchir and Fadili (2006) model the heavy-tailed behavior using an α-stable distribution as the prior and give the form of the variances across scales which ensure an exponential decay across scales. Recently, Wolpert, Clyde and Tu (2011) considered a nonparametric regression model in which the estimation is done using the atomic decomposition of the function of interest. They propose the Lévy Adaptive Regression Kernel (LARK) model which gives a prior based on Lévy random fields to the parameters of the atomic decomposition. In Wolpert, Clyde and Tu (2011, section 4), they investigate the LARK model under symmetric α-stable random fields and provide conditions for the LARK model to remain in the same Sobolev space ($B^s_{2,2}$) as the generating kernels. To illustrate their methodology, they consider only the symmetric Gamma and Cauchy processes in their inference and simulation study. While a good elicitation of the hyperparameters of the LARK model with the symmetric Gamma prior is achieved, this is not the case with the Cauchy prior.

We similarly consider a nonparametric regression based on a Lévy wavelet prior model, but provide necessary and sufficient conditions on the hyperparameters such that the model remains in Besov spaces ($B^s_{p,q}$) with general p and q (this is our main Theorem 32). Our approach uses a Lévy tempered stable distribution as a prior for the wavelet coefficients, where the key ingredient is the scaling property of tempered stable distributions. We demonstrate that the scaling behaviour of the tempered stable distribution can model the desired form of the wavelet coefficients without imposing any specified structure on the variance. It can be shown that our Lévy prior model is a special case of Wolpert, Clyde and Tu's model. If the generation function is a wavelet and the wavelet coefficients are independently tempered stable distributed, the proposed Lévy prior model can be viewed as the LARK model.

The first motivation for using the tempered stable as a prior is that its parameters decrease faster than those of the stable distribution and make the second and higher moments finite. The second motivation is that the tempered stable density is

characterized by correlated variables and makes the variance of the underlying decay exponentially across scale.

In the same spirit of the sparse model of Abramovich, Sapatinas and Silverman (1998), we introduce the Lévy sparse prior model within a Bayesian framework. We postulate that the prior on the discrete wavelet coefficients is defined as a mixture of a tempered stable distribution and a point mass at zero. The usual approach is to extract the significant wavelet coefficients by thresholding, where the choice of the thresholding rule is then an important step in the estimation procedure. Abramovich, Benjamini, Donoho and Johnstone (2006) use the false discovery rate (FDR) method to set the threshold for the wavelet coefficients, which is based on the principle of controlling the FDR in multiple hypothesis testing. Johnstone and Silverman (2005) explore the empirical Bayes methods for level-dependent threshold selection, where they consider a mixture of a heavy-tailed density and a mass point at zero for each wavelet coefficient and show that the empirical Bayes method is stable compared to the FDR approach. In the Lévy wavelet sparse model, we consider the universal thresholding proposed in Donoho and Johnstone (1994,1995).

Assuming a tempered stable as a priori, we give a closed form to the *a posteriori* Lévy measure of the wavelet coefficients by applying some Bayes rule. We use this posterior Lévy measure to estimate the tempered stable distribution.

We also present a regularization approach for estimating nonparametric regression functions. We follow here the Bayesian approach where we set the prior distribution to be tempered stable and we focus on finding a solution to the wavelet-based regression through a Penalized Maximum Likelihood (PML) approach or Maximum a Posteriori (MAP) estimation. Here, we explore a PML based on generalized entropy functionals, namely the discrete formulation of the Tsallis entropy.

1.1 Chapter Outline

We give a brief description of the contents in each chapter.

Chapter 2

We study the characteristics of various datasets in order to motivate the models we propose. We consider data from the S&P 500 and show that Lévy type distributions are appropriate to model the wavelet coefficients.

Chapter 3

We give a brief introduction to the wavelet analysis and Besov spaces, the theory of which is necessary in later Chapters.

Chapter 4

We introduce the Lévy process which forms the structure of the wavelet prior model. We proved the key ingredient of our results, the scaling properties of tempered stable distributions. In this Chapter, we aim to provide basic definitions and establish a novel approach in using the wavelet theory to approximate the Wiener-Hopf factors of a Lévy process.

Chapter 5

We introduce the Lévy wavelet prior model and study the regularity and sparsity properties of the prior model in Besov spaces. In addition to this, we state the connection between the hyperparameters of the model and Besov spaces.

Chapter 6

We study the tempered stable prior model within the Bayesian approach in the wavelet domain and give a closed form of the Lévy density of the wavelet coefficients. We also propose a Lévy sparse mixture which is a mixture of tempered stable distribution and a point mass at zero. We prove that this model yields to a thresholding rule. Furthermore, we introduce an approach to estimate the hyperparameters of the prior model within a Bayesian framework. The results of Chapters 5 and 6 have been submitted (Missaoui and McCoy, 2011).

Chapter 7

We develop the wavelet entropic regularization approach. Inspired by the work of Zunino et al. (2006), we compute the wavelet Tsallis entropy and Fisher information

for a tempered stable process and proposed a penalized likelihood method for density estimation. The results of this part are ready to be submitted (Missaoui and McCoy, 2012)

Conclusion

In this concluding Chapter, we give an overview of the results obtained from the wavelet regression in the thesis. Furthermore, we discuss potentially extensions of the work presented in this.

Chapter 2

Data Analysis

2.1 Introduction

In financial time series analysis, the most common assumption is that asset returns are assumed to be Gaussian distributed. However recent empirical research shows that financial data exhibit a fat-tailed behaviour which the Gaussian distribution fails to model (Rachev, 2003). Many studies considered replacing the Gaussian distribution with fat-tailed distributions e.g. the hyperbolic distribution (Bingham and Kiesel, 2001) and Gaussian mixtures (McLachlan and Peel, 2000). But these distributions are not scale-invariant; that is, their characteristics change with different time intervals. Another possibility is the Lévy stable distributions, first studied by Mandelbrot (1963). The Lévy stable distribution has fatter tails and preserves the scaling property but does not make the second and higher moments finite. The Lévy stable process is also defined by i.i.d increments and it is not characterized by correlated stochastic variables.

To remedy to these deficiencies, we will develop in this report a model based on the tempered stable distribution. In addition of having all higher moments finite, this distribution exhibits fat tails and preserves the scaling properties.

2.2 Data Analysis

Here we consider data from the S&P 500 and show that the wavelet coefficient time series of the log-returns of the data can be described by a Lévy stable type distribution; refer to section (4.3) for more details on the Lévy stable distribution and Chapter 3 for details on the wavelet transform.

We concentrate the analysis on the daily S&P 500 ranging from January 1, 1980 to December 31, 2004 (www.yahoo.com). A plot of daily closing prices for S&P500 are given in Figure (2.3). We also present daily log-returns in Figure (2.4).

In Tables (2.1) and (2.2) we give the statistics for the log-prices and the log-returns of the daily, weekly and monthly S&P500 closing prices. It is notable that the daily log-return series show small skewness and high kurtosis which implies that the empirical distribution is almost symmetric and exhibits thicker tails than normality. The weekly and monthly log-returns time series tends to be normally distributed. That is, we will restrict our wavelet analysis to the daily log-returns.

In Table (2.3), we give the statistics of the wavelet coefficients for the daily log-returns. The coefficients were calculated using the Daubechies 3 wavelet; refer to Daubechies (1992, Chapters 6 and 7) for more details. We notice clearly that the fat-tailed behaviour of the log-returns is translated to the wavelet coefficients.

The Q-Q plots in Figure (2.5) show also that the empirical distribution of the daily log-prices and the wavelet coefficients follow a non-Gaussian distribution.

The estimates of the four parameters $(\alpha, \beta, \gamma, \delta)$ of the Lévy stable distribution are in Table (2.4), where $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma \geq 0$ and $\delta \in \mathbb{R}$. We will describe the full parameterization in Chapter 4, an important parameter is α, the tail index, which gives an indication whether the times series is normally distributed (which has $\alpha \approx 2$). A tail index $\alpha = 1.73$ shows that the daily wavelet time series present fat tails. For the weekly and monthly time series, the empirical density approaches a Gaussian distribution.

The parameter γ can be interpreted as the volatility of the underlying process and we can verify from Table (2.4) that γ is time-dependent and thus the underlying cannot be modeled with a stochastic process with i.i.d increments.

The statistical properties of the wavelet coefficients of the daily log-returns also show that as the sampling time interval Δt increases, the mean increases and the ratio of the mean and the standard deviation decreases. This is consistent with the characteristics of a power law. Recently, Wei and Billings (2009) have shown that foreign exchange rates obey power-laws and therefore belong to the class of selfsimilar processes. In Figures (2.1) and (2.2), we present a comparison between the Gaussian and Lévy stable densities and cumulative density functions in fitting the wavelet coefficient series. We can see that the Gaussian density is not an appropriate model and that the Lévy stable density provides a much better central and extremal fit.

We can also see that the leptokurtic behaviour of the daily log-returns and the wavelet coefficients series disappears for weekly and monthly time intervals.

We observe that the Lévy stable distribution describes the daily log-returns well at different time intervals. However with the assumption of i.i.d increments and having an infinite second moment, the Lévy stable process is not an appropriate model for an underlying having scaling properties.

An alternative is to model the empirical distribution with a truncated Lévy flight (TLF) distribution. The probability distribution of a TLF process is defined by:

$$f(x) = \begin{cases} 0, & \text{if } l > 0, \\ L(x), & \text{if } -l < x < l, \\ 0, & \text{if } x < -l. \end{cases}$$

where

$$L(x) = \frac{1}{\pi} \int_0^{+\infty} \exp(-\gamma w^{\alpha}) \cos(wx) dw$$

is the symmetric Lévy stable distribution with tail index α and γ the scale parameter. The parameter l is the cutoff length.

The TLF distribution has a finite second moment and thicker tails. Mantegna and Stanley (1995) showed that the scaling property of the TLF distribution is preserved for small time intervals and the TLF tends to the Gaussian distribution as the time interval increases. However, the TLF is not an infinitely divisible distribution and

therefore it does not correspond in a natural way to a Lévy process; see Chapter 4 for background on Lévy processes.
What we need is a model with the same characteristic of the Lévy stable distribution but with finite second- and higher-order moments. That is, a model which behaves in a short time like a stable process and in long time tends to Brownian motion; see section 4.6. Such a model is the tempered stable distribution, which we will introduced in section 4.5.

2.3 Conclusion

In this chapter we introduced the data sets which form the motivation for our study. We have discussed the log-prices and the log-returns of the S&P500. We also investigated the characterization of the wavelet coefficients time series.
In the next chapter we introduce the basics of wavelet analysis that we need in the development of our framework.

Table 2.1: Descriptive Statistics of log-prices for S&P500

log-price	Daily	Weekly	Monthly
Mean	6.045	6.046	6.050
St Dev	0.801	0.802	0.801
Skewness	-0.053	-0.054	-0.061
Kurtosis	-1.242	-1.243	-1.244
Minimum	4.587	4.611	4.626

Table 2.2: Descriptive Statistics of log-returns for S&P500

log-returns	Daily	Weekly	Monthly
Mean	0.000	0.002	0.008
St Dev	0.011	0.022	0.045
Skewness	-1.743	-0.495	-0.895
Kurtosis	38.822	3.043	3.551
Minimum	-0.229	-0.130	-0.245
Maximum	0.087	0.085	0.124

Table 2.3: Descriptive Statistics of Wavelet Coefficients of log-returns for S&P500

log-returns Wavelet	Daily	Weekly	Monthly
Mean	0.000	0.001	0.000
St Dev	0.011	0.022	0.044
Skewness	-1.971	-0.008	-0.366
Kurtosis	51.496	4.051	4.112
Minimum	-0.206	-0.009	-0.155
Maximum	0.070	0.083	0.110

Table 2.4: Parameter estimates of the log-Stable for the wavelet coefficients of the log-returns for S&P500

log-Stable Wavelet	α	β	γ	δ
Daily	1.7365	0.0643	0.0058	0.0011
Weekly	1.8654	0.0581	0.0144	0.0011
Monthly	1.8822	-0.7439	0.0292	0.0040

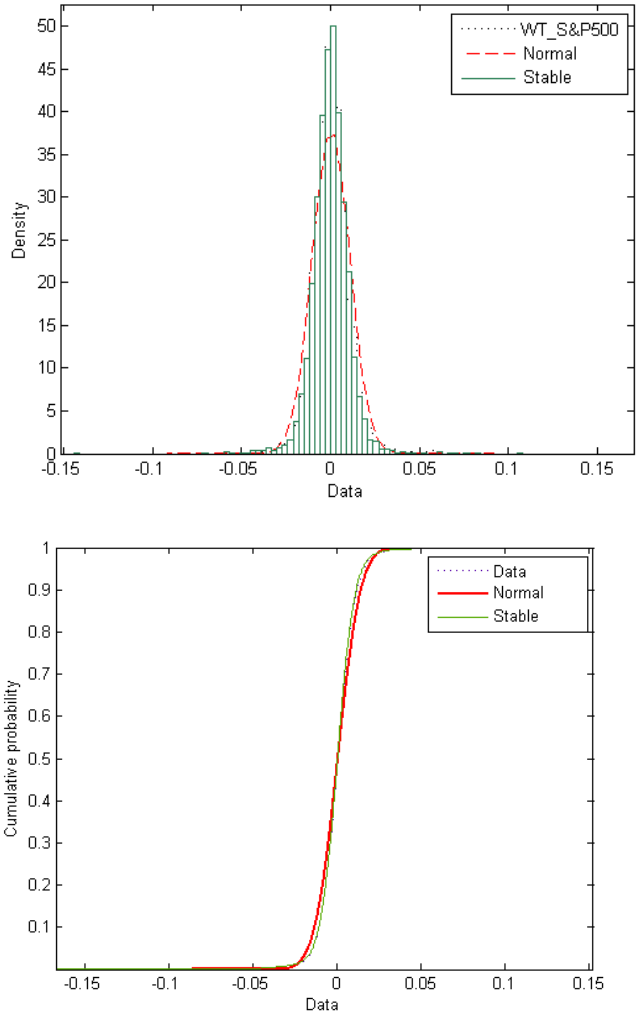


Figure 2.1: Empirical PDF and CDF Fitting of the wavelets coefficients of the daily log-returns S&P500

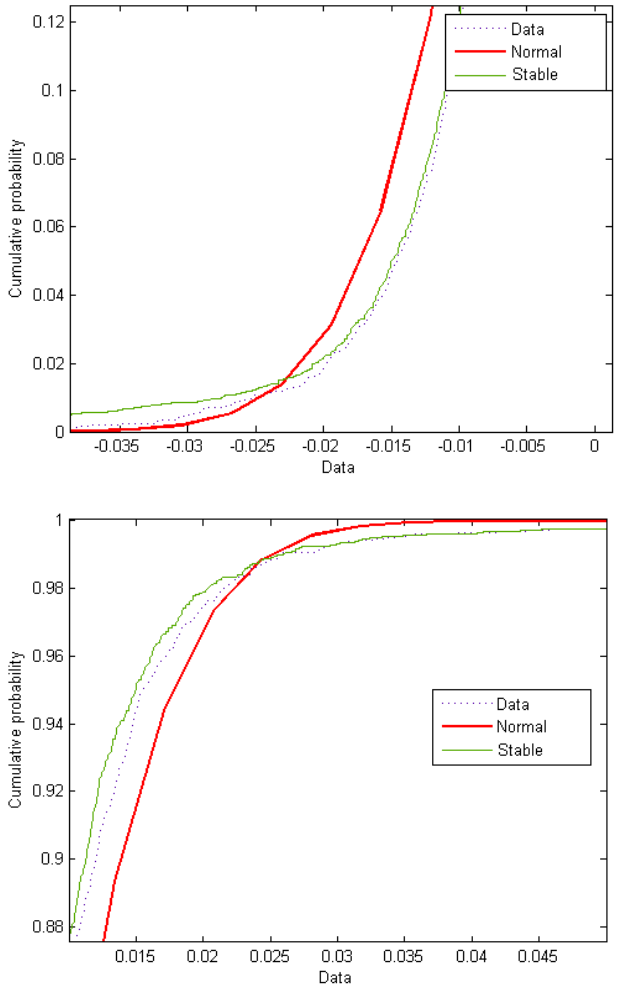


Figure 2.2: Left and right tails of the CDF of the wavelets coefficients of the daily log-returns S&P500

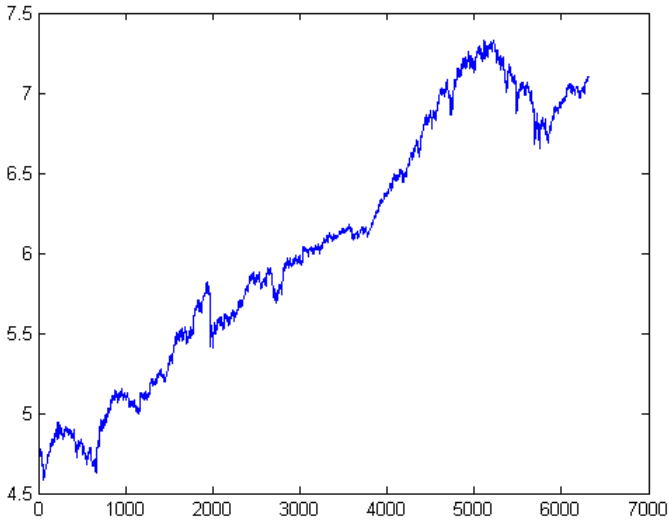


Figure 2.3: Daily Log-prices S&P500

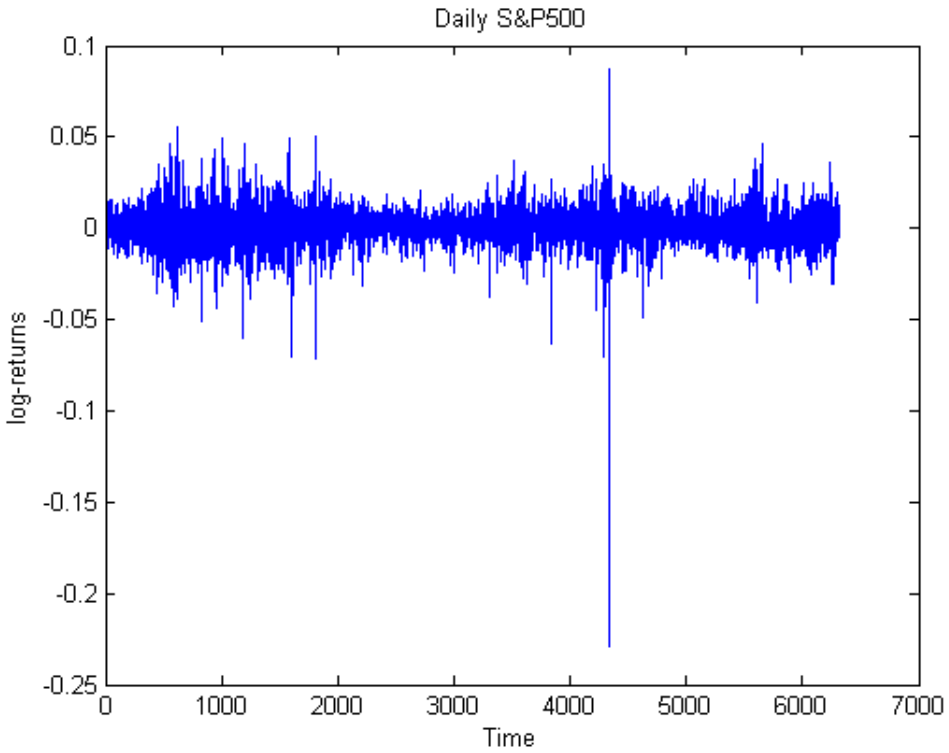


Figure 2.4: Daily Log-returns S&P500

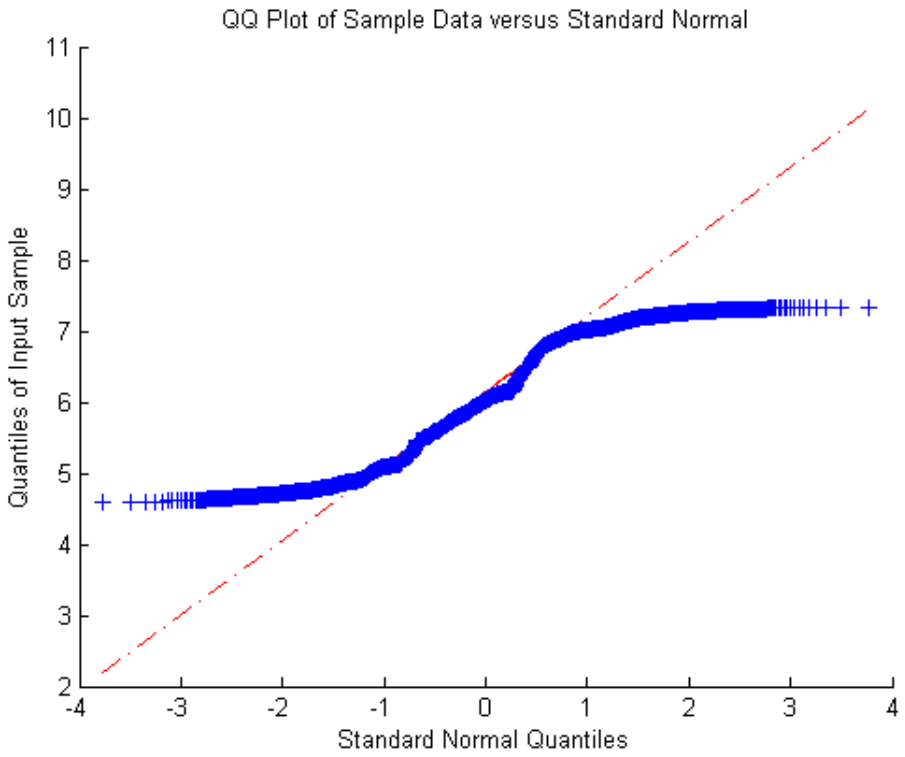


Figure 2.5: Q-Q plot of Log-prices S&P500

Chapter 3

Wavelet Analysis

3.1 Introduction

Until 1970, the predominant assumption was that signals were appropriately modeled as realizations of Gaussian processes, in consequence under the assumption of stationarity, linear algorithms based on the Fourier transform were considered optimal. The situation has been revolutionized with the development of novel image processing techniques in the 1980's. Images are poorly modeled by Gaussian processes, and the edge components of image such as contours are often more important. The use of nonlinear algorithms became necessary, which opened signal processing to modern mathematics and wavelet analysis. The inadequacy of the Fourier transform to describe non-stationary behavior of a function will be the basis of the motivation to introduce this concept.

3.2 Fourier Transform

For simplicity, we will introduce the Fourier analysis in one-dimension; refer to Papoulis (1987) and Dym and McKean (1972) for more details. We recall that a Hilbert space is a space which has an inner product $\langle x, y \rangle$ and which is complete with respect to the norm $||x|| = \langle x, x \rangle^{1/2}$. We also recall that a space E is called *complete*

if every Cauchy sequence in E converges in E. We denote $L^2(\mathbb{R})$ the collection of all complex-valued Lebesgue-measurable functions in $\mathbb{R}$ such that $\int_{\mathbb{R}} |f(t)|^2 \mathrm{d}t < \infty$. That is,

$$L^2(\mathbb{R}) = \left\{ f : \mathbb{R} \to \mathbb{C} \mid \int_{\mathbb{R}} |f(t)|^2 \mathrm{d}t < \infty \right\}.$$

$L^2(\mathbb{R})$ is a Hilbert space with the inner product

$$\langle f, g \rangle = \int_{\mathbb{R}} f(t)\overline{g(t)}dt,$$

and the norm $||f|| = \left(\int_{\mathbb{R}} |f(t)|^2 \mathrm{d}t\right)^{1/2} < \infty$, where $\overline{g}$ means the complex conjugate of g.

Definition 1. *Assume that $f \in L^2(\mathbb{R})$, the Fourier transform of f is the function $\mathcal{F}f$ defined by*

$$\mathcal{F}f(w) = \int_{\mathbb{R}} e^{-itw} f(t) \mathrm{d}t. \tag{3.1}$$

We usually write $\hat{f}$ for $\mathcal{F}f$ and it is a function of a parameter w which we call the angular frequency. Assuming $\hat{f} \in L^2(\mathbb{R})$, the following proposition gives the inverse Fourier transform.

Proposition 2. *If $\hat{f} \in L^2(\mathbb{R})$ and $f \in L^2(\mathbb{R})$, then*

$$f(t) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itw} \hat{f}(w) \mathrm{d}w. \tag{3.2}$$

The Fourier transform preserves the inner scalar product (Parseval formula):

$$\int_{\mathbb{R}} f(t)\bar{g}(t) \mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(w)\bar{\hat{g}}(w) \mathrm{d}w.$$

For $f = g$, we obtain the energy conservation known as the Plancherel formula:

$$\int_{\mathbb{R}} |f(t)|^2 \mathrm{d}t = \frac{1}{2\pi} \int_{\mathbb{R}} |\hat{f}(w)|^2 \mathrm{d}w.$$

We write $f * g$ for the convolution $f * g(x) = \int f(x-t)g(t)dt$.

Proposition 3. *If $f, g \in L^2(\mathbb{R})$, then almost everywhere*

$$\widehat{f * g}(w) = \hat{f}(w)\hat{g}(w).$$

Given a function f we define the translation operator T_b by $T_b f(x) = f(x-b)$ for $b \in \mathbb{R}$ and the dilation operator D_a by $D_a f(x) = |a|^{-1/2} f(x-a)$ for $a \in \mathbb{R}$-$\{0\}$. It is easy to prove that

$$\begin{aligned} \mathcal{F}T_b f(w) &= e^{-ibw}\hat{f}(w), \\ \mathcal{F}D_a f(w) &= |a|^{1/2}\hat{f}(aw). \end{aligned}$$

3.2.1 Fourier Transform and Regularity

The Fourier transform provides a way to characterize the overall regularity as well as the related concept of the frequency scale of a function. Roughly speaking, the regularity measures the smoothness of a function while the frequency scale measures how quickly the function oscillates; a natural way to describe the global regularity of a function and capture its frequency scale is to consider functional Sobolev spaces (see e.g Adams (1978)).
The Sobolev space $H^m(\mathbb{R})$ is defined to be the set of all functions $f \in L^2(\mathbb{R})$ such that for every α with $|\alpha| \leq m$, the partial derivative $\partial^\alpha f \in L^2(\mathbb{R})$, that is

$$H^m(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}) | \partial^\alpha f \in L^2(\mathbb{R}), \alpha \in \mathbb{R} \right\}.$$

$H^m(\mathbb{R})$ is a Banach space equipped with the norm

$$||f||_{H^m} = \sum_{|\alpha| \leq m} ||\partial^\alpha f||_{L^2(\mathbb{R})}.$$

Sometimes the properties of regularity of a function cannot be seen directly from the above definition of the Sobolev spaces. Therefore, it has been shown that it is interesting to investigate the regularity by means of the Fourier transform. The

Sobolev space $H^m(\mathbb{R})$ can be equivalently defined by

$$H^m(\mathbb{R}) = \left\{ f \in L^2(\mathbb{R}) | \int_{\mathbb{R}} (1 + |w|^2)^\alpha |\hat{f}(w)|^2 \mathrm{d}w < +\infty, \alpha \in \mathbb{R} \right\}.$$

The Fourier transform allows us to study the global regularity and to determine which Sobolev space the function belongs to. However, it does not describe and analyze the local regularity since a function can be locally regular without being globally regular; e.g., the Heaviside function is constant over $\mathbb{R}$-$\{0\}$ but does not belong to any Sobolev space with positive index.

Moreover, many signals in reality, such as speech, music and images, have changing frequency characteristics, i.e. frequencies vary with time, and the Fourier transform does not contain information on the time evolution of the frequencies. To achieve a time-dependent frequency analysis, we introduce the Gabor transform characterized by the short-time Fourier transform.

3.2.2 Short-Time Fourier Transform

To convey simultaneous time and frequency localization in a signal, Gabor (1940) introduced the short-time Fourier transform (STFT). In practice, this Gabor transform consists in multiplying the function by a window function h which is nonzero for only a short period of time. For $u, w \in \mathbb{R}$, the STFT for a function $f \in L^2(\mathbb{R})$ is defined by

$$TS(u, w) = \int_{\mathbb{R}} f(x) \overline{h(x-u)} e^{-iwx} \mathrm{d}x.$$

Putting $u = na$ and $w = mb$ where $n, m \in \mathbb{Z}$ and $a, b \in \mathbb{R}$, we can compute the Fourier coefficients of the product

$$d_{mn}(f) = \int_{\mathbb{R}} f(x) \overline{h(x-na)} e^{-imbx} \mathrm{d}x.$$

The coefficients d_{mn} represent in this formulation the intensity of the frequency mb at na. So, if we can find a good lattice $\{na, mb\}_{n,m \in \mathbb{Z}}$, then the function f can be

written using the Fourier series expansion

$$f(x) = \sum_{n,m} d_{mn}(f) h(x - na) e^{imbx}.$$

We will assume in the Gabor representation that the window function h satisfies

$$A \leq \sum_{n=1}^{N} |h(x - na)|^2 \leq B,$$

for $A, B \geq 0$. The Gabor transform is numerically stable when B is not too large. Morlet suggested to compute the inner product of f with an analyzing function $\frac{1}{\sqrt{a_0}} \psi\left(\frac{x-x_0}{a_0}\right)$ which we expect to be localized in frequency and in time. The function ψ is called *wavelet*.

If we denote by x_ψ and ν_ψ the mean values of x and ν with respect to the probability measure $|\psi(x)|^2 \mathrm{d}x$ and $|\hat{\psi}(\nu)|^2 \mathrm{d}\nu$, then we can write

$$\begin{aligned} x_\psi &= \int_{-\infty}^{+\infty} x |\psi(x)|^2 \mathrm{d}x, \\ \nu_\psi &= \int_{-\infty}^{+\infty} \nu |\hat{\psi}(\nu)|^2 \mathrm{d}\nu. \end{aligned}$$

We can also write the resolution of the analyzing wavelet ψ in terms of spreads around x_ψ and ν_ψ. The spreads are measured by the variance and defined by

$$\begin{aligned} \Delta x_\psi^2 &= \int_{-\infty}^{+\infty} |x - x_\psi|^2 |\psi(x)|^2 \mathrm{d}x, \\ \Delta \nu_\psi^2 &= \int_{-\infty}^{+\infty} |\nu - \nu_\psi|^2 |\hat{\psi}(\nu)|^2 \mathrm{d}\nu. \end{aligned}$$

The joint time-frequency resolution is defined by the product of the two resolutions $\Delta x_\psi \Delta \nu_\psi$. This joint resolution depends on a_0. This represents in itself a downfall of the Gabor transform and the wavelet analyzing function; that is, it is impossible to find a_0 which gives at the same time good frequency and good time resolution. This means that a window function h or a wavelet ψ with finite energy compactly supported in time and frequency does not exist. This is justified by the Heisenberg

uncertainty principle, which states

Theorem 4. *Let $\psi \in L^2(\mathbb{R})$ and for any $\nu_\psi, x_\psi \in \mathbb{R}$, we have*

$$\Delta x_\psi \Delta \nu_\psi \geq \frac{1}{2},$$

and the minimum value $\frac{1}{2}$ is reached for Gaussian functions.

3.3 Wavelet Transform

One of the reasons for the creation of the wavelet transform is to define a window function that is well-localized in time and frequency. The purpose of this section is to give a short introduction to wavelets. For more extensive treatments, see, for example, Mallat (1999), Meyer (1992) and Walden and Percival (2001). The term *wavelets* is used to refer to a dictionary of basis functions. The wavelet structure may be considered as a generation of an orthonormal wavelet basis for functions $g \in L^2(\mathbb{R})$.

3.4 Continuous Wavelet Transform

A wavelet ψ is a real function in $L^2(\mathbb{R})$ satisfying the *admissibility condition*

$$C_\psi = \int_0^{+\infty} \frac{|\widehat{\psi}(w)|^2}{w} \mathrm{d}w < \infty, \tag{3.3}$$

where $\widehat{\psi}$ is the Fourier transform of ψ and the integral of ψ has to vanish

$$\int \psi(x) \mathrm{d}x = 0. \tag{3.4}$$

Also, the square of the wavelet should integrate to unity,

$$\int_{-\infty}^{\infty} \psi^2(x) dx = 1.$$

The integral (3.4) is the value of the Fourier transform of ψ at 0. Setting $\widehat{\psi}(0) = 0$, we ensure that the *admissibility condition* is verified.

One way of selecting a wavelet function is given in the following theorem.

Theorem 5. *Let ϕ and $\phi^{(n)}$ be $L^2(\mathbb{R})$ functions for $n > 0$. Let $\phi^{(n)}(x) \neq 0$ and satisfies (3.4). Then, $\psi(x) = \phi^{(n)}(x)$ is a wavelet function.*

Proof. We need to prove that $C_\psi < \infty$.

$$\begin{aligned} C_\psi &= \int_0^{+\infty} \frac{|\widehat{\psi}(w)|^2}{w} dw \\ &= \int_{|w|\leq 1} |w|^{2n-1} |\hat{\phi}(w)|^2 dw + \int_{|w|>1} \frac{|w|^{2n} |\hat{\phi}(w)|^2}{|w|} dw \\ &\leq \int_{|w|\leq 1} |\hat{\phi}(w)|^2 dw + \int_{|w|>1} |w^n \hat{\phi}(w)|^2 dw \end{aligned}$$

Finally, $C_\psi \leq \|\phi\|_2^2 + \left\|\phi^{(n)}\right\|_2^2 < \infty$. □

The regularity of a wavelet is characterized by its moments. A wavelet has m vanishing moments if

$$\int_{-\infty}^{+\infty} x^p \psi(x) \mathrm{d}x = 0, \quad \text{for} \;\; p = 0, ..., m-1.$$

The wavelet function is then obtained by dilation and translation of ψ:

$$\psi_{t,s}(x) = T_t D_s \psi(x) = \frac{1}{\sqrt{s}} \psi\left(\frac{x-t}{s}\right).$$

The continuous wavelet transform Wf of a function $f \in L^2(\mathbb{R})$ is defined by

$$Wf(t,s) = \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{s}} \psi\left(\frac{x-t}{s}\right) dx = f * \overline{\psi_s}(t),$$

where $\overline{\psi_s}(t) = \frac{1}{\sqrt{s}} \psi\left(-\frac{t}{s}\right)$. Using the Pancherel formula, we might compute the wavelet coefficients in Fourier space via

$$Wf(t,s) = \sqrt{s} \int_{-\infty}^{\infty} \hat{f}(w) \bar{\hat{\psi}}(sw) e^{iwt} dw. \tag{3.5}$$

Like the Fourier transform, the wavelet transform has an inversion formula which allows us to reconstruct the function from its wavelet transform.

Theorem 6. *Let $\psi \in L^2(\mathbb{R})$ be an analysing wavelet, then for any $f \in L^2(\mathbb{R})$, we have*

$$f(t) = \frac{1}{C_\psi} \int_0^{+\infty} \frac{ds}{s^2} \int_{-\infty}^{+\infty} Wf(x,s)\psi_{t,s}(x)dx. \tag{3.6}$$

Proof. Using (3.5), we can write the right hand side of (3.6) as

$$\begin{aligned} g(t) &= \int_0^{+\infty} \int_{-\infty}^{+\infty} Wf(x,s)\frac{1}{\sqrt{s}}\psi\left(\frac{t-x}{s}\right)dx\frac{ds}{s^2}, \\ &= \int_0^{+\infty} \frac{ds}{s} \int_{-\infty}^{+\infty} dx \int_0^{+\infty} dw\hat{f}(w)\bar{\hat{\psi}}(sw)e^{iwt}\frac{1}{s}\psi\left(\frac{t-x}{s}\right), \\ &= \int_0^{+\infty} \int_0^{+\infty} \hat{f}(w)\bar{\hat{\psi}}(sw)\hat{\psi}(sw)e^{iwt}dx\frac{ds}{s}, \\ &= \int_0^{+\infty} dw\hat{f}(w)e^{iwt} \int_0^{+\infty} \frac{ds}{s}|\hat{\psi}(sw)|^2, \\ &= C_\psi f(t), \end{aligned}$$

where we used the inverse formula for the Fourier transform of f. □

Also, we have the Pancherel formula in the form

$$||f||^2_{L_2(\mathbb{R})} = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} |Wf(x,s)|^2 dx\frac{ds}{s^2}. \tag{3.7}$$

Here, we will give some examples of analyzing wavelet functions.

Example 1 (Morlet's Wavelet)

The Morlet wavelet was used first in seismic signals and geophysical explorations. It is complex-valued and obtained by shifting the Gaussian function in the Fourier domain:

$$\psi(t) = e^{iw_0t}e^{-t^2/2},$$

where w_0 is a constant. This wavelet is not admissible, since it is not of zero mean. Its Fourier transform does not vanishes at the origin: $\hat{\psi}(0) = \sqrt{2\pi}e^{-w_0^2/2}$. However, for large values of w_0, the value of $\hat{\psi}(0)$ is small and therefore the Morlet wavelet approximates a true wavelet. Figure (3.1) shows the real part of the Morlet wavelet

with $w_0 = 5.336$.

Example 2 (Mexican hats)

The Mexican hats are defined as the second derivative of the Gaussian function (see Figure (3.2)). The normalized Mexican hat wavelet is

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}} \left(\frac{t^2}{\sigma^2} - 1 \right) e^{-t^2/(2\sigma^2)}.$$

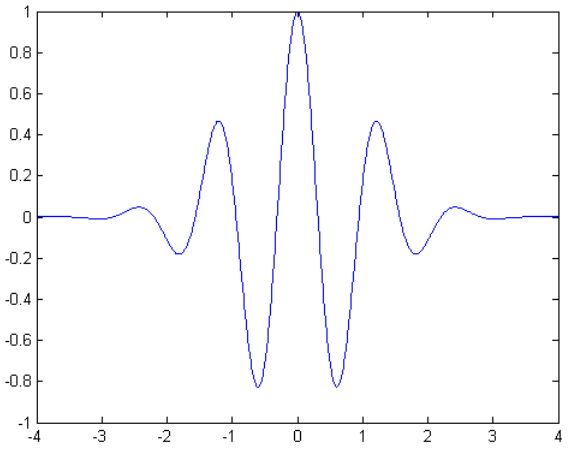


Figure 3.1: Morlet's wavelet for $w_0 = 5.336$ (real part).

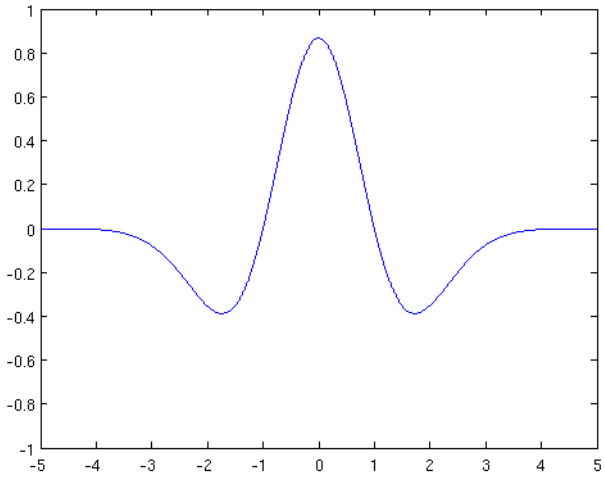


Figure 3.2: Mexican hat wavelet for $\sigma = 1$ (real part).

Example 3 (Cauchy wavelets)
The Cauchy wavelets are analytic functions and their general expression is (Figure 3.3)

$$\psi(t) = \frac{1}{2\pi}\frac{\Gamma(\alpha+1)}{(1-it)^{1+\alpha}}, \quad \alpha > 0, \tag{3.8}$$

where $\Gamma(t) = \int_0^\infty t^{z-1}e^{-t}dt$. The Fourier transform of Cauchy wavelet is

$$\widehat{\psi}(w) = \begin{cases} w^\alpha e^{-w} & \text{if } w > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, the Cauchy wavelet transform is closely related to the analysis of the analytic functions over half-plane; refer to section 3.4.2 for more details. In section 4.7, we present an application of Cauchy wavelets to compute the Wiener-Hopf factors.

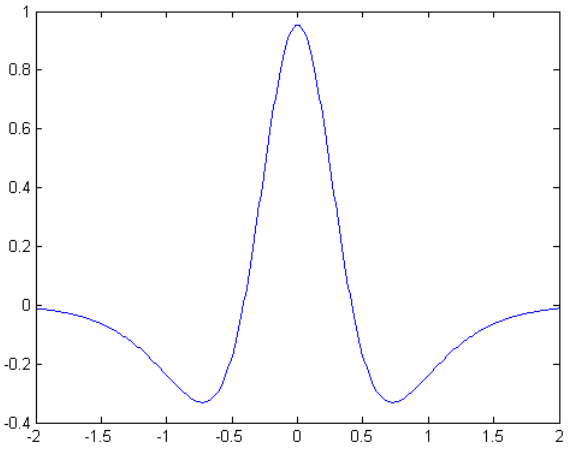


Figure 3.3: The Cauchy wavelet for $\alpha = 3$ in time representation (real part).

3.4.1 Regularity

One of the most important properties of wavelets is their high accuracy at fine scales. This feature is intrinsically embedded in their construction. Therefore, it is natural that wavelets may be used to characterize regularity properties of singular structures, and the decay of the wavelet coefficients across scales is related to the

Lipschitz (global) or pointwise Lipschitz (local) regularity.
Let $0 < \alpha < 1$. We introduce the α-Lipschitz space:

$$L^\alpha = \{f : \mathbb{R} \to \mathbb{C},\ \exists K < \infty : |f(x) - f(y)| \leq K|x - y|^\alpha,\ \forall x, y \in \mathbb{R}\}\,.$$

The conventional way to characterize α-Lipschitz functions is based on the Fourier transform. We can show that if f is bounded and satisfies

$$\int_{\mathbb{R}} |\hat{f}(w)|(1 + |w|^\alpha)dw < \infty, \tag{3.9}$$

then $f \in L^\alpha$. The converse is in general false (see e.g Adams (1978)).
For wavelet analysis, we will see that the results are similar. The following result follows the same philosophy as the analogue for Fourier transform of Lipschitz functions:

Theorem 7. *Let $f \in L^1(\mathbb{R})$ be α-Lipschitz over $[a, b]$ and ψ an analyzing wavelet such that $\psi \in L^1(\mathbb{R})$ and $x^\alpha\psi \in L^2(\mathbb{R})$. Then there exists $C > 0$ such that*

$$\forall (t, s) \in [a, b] \times \mathbb{R}^+ \quad , \quad |Wf(t, s)| \leq Cs^{\alpha+1/2}. \tag{3.10}$$

Conversely, if f is bounded and $Wf(t, s)$ satisfies (3.10), then f is α-Lipschitz on $[a + \epsilon, b - \epsilon]$ for any $\epsilon > 0$.

The inequality (3.10) is a similar to the Fourier condition (3.9) where s plays the role of the inverse of the frequency. It gives an asymptotic behaviour of the decay of the wavelet transform when s goes to zero.
The continuous wavelet transform can also characterize the regularity of f at a point t_0. Jaffard (1991) gave necessary and sufficient conditions on the wavelet transform for estimating the α-Lipschitz of f at t_0. We consider a compactly supported wavelet ψ having n vanishing moments.

Theorem 8. *Let $f \in L^2(\mathbb{R})$ be α-Lipschitz at t_0 and $\alpha \leq n$. Then, there exists $A > 0$*

such that

$$\forall (t,s) \in [a,b] \times \mathbb{R}^{*+} \quad , \quad |Wf(t,s)| \leq As^{\alpha+1/2}\left(1 + \left|\frac{t-t_0}{s}\right|^{\alpha}\right). \tag{3.11}$$

Conversely, if $\alpha < n$ and α is not an integer, and there exists A and $\alpha' < \alpha$ such that

$$\forall (t,s) \in [a,b] \times \mathbb{R}^{*+} \quad , \quad |Wf(t,s)| \leq As^{\alpha+1/2}\left(1 + \left|\frac{t-t_0}{s}\right|^{\alpha'}\right). \tag{3.12}$$

then f is α-Lipschitz at t_0.

We refer to Mallat (1998) and Jaffard (1991) for the proof of the results (3.10), (3.11) and (3.12).
This analysis was refined by Jaffard (1991), who has shown that wavelet analysis can be viewed as a simple version of a larger program, namely two-microlocalization theory introduced by Bony (1981) to study the solution of hyperbolic partial differential equations. We refer to Jaffard and Meyer (1993) for a background and references.
Figure (3.4) shows the discontinuity response of a locally perturbed signal in the first level detail coefficients and the wavelet coefficient at the discontinuity point.

3.4.2 Hardy Spaces

In this section, we give the basic definitions and results of Hardy spaces that we need in our development in section 4.7. Our aim will be to use Hardy spaces to approximate the Wiener-Hopf factors of a Lévy process. For a complete and rigorous exposition of Hardy spaces, see Duren (1970) or Koosis (1980).

Definition 9. *We say that an analytic function f belongs to $H^2_+(\mathbb{R})$ if*

$$\sup_{y>0}\left(\int |f(x+iy)|^2 dx\right)^{1/2} < \infty.$$

The analogue holds for $H^2_-(\mathbb{R})$ but the supremum is on the lower-half plane ($y < 0$). Another characterization of Hardy spaces was given by Paley and Wiener (1934).

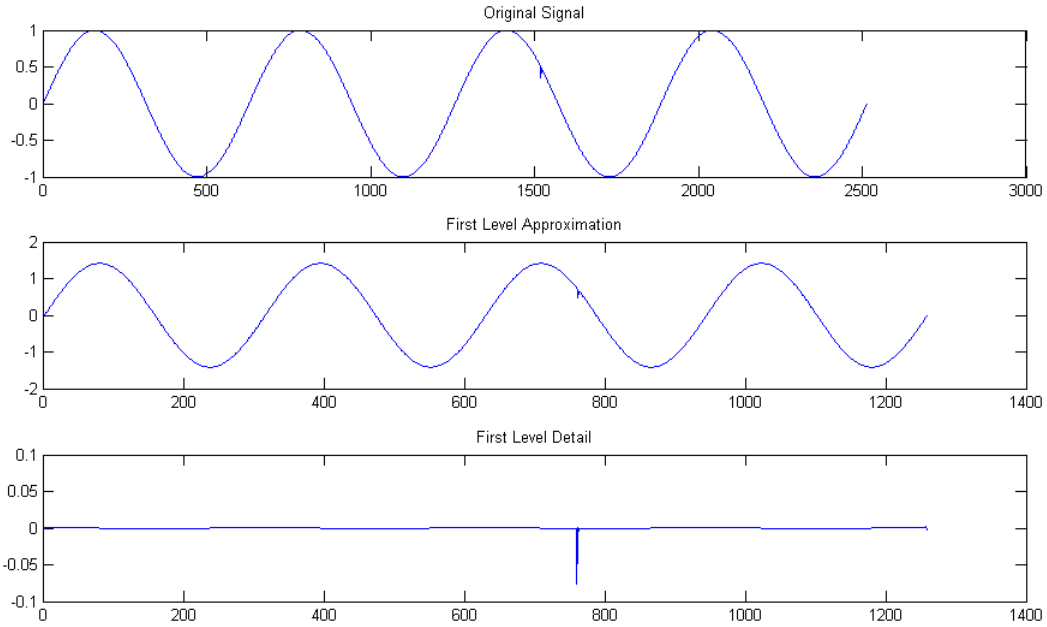


Figure 3.4: Signal discontinuity.

Definition 10.

$$\begin{aligned} H^2_+(\mathbb{R}) &= \{f \in L^2(\mathbb{R}) | \ \ \text{supp}(\hat{f}) \subset [0, +\infty)\}, \\ H^2_-(\mathbb{R}) &= \{f \in L^2(\mathbb{R}) | \ \ \text{supp}(\hat{f}) \subset (-\infty, 0]\}. \end{aligned}$$

The two definitions of Hardy spaces given in Defintions 9 and 10 are similar; their equivalence follows easily from the Parseval formula.
$H^2_+(\mathbb{R})$ (resp. $H^2_-(\mathbb{R})$) is a closed subspace of $L^2(\mathbb{R})$ with functions having only positive (resp. negative) frequencies. They can be identified with $H^2_\pm(\mathbb{R}) = \mathcal{F}_\pm(L^2(\mathbb{R}))$; see Ruth and Van Fleet (2009).
Using the Plancherel formula, the standard perpendicular decomposition

$$L^2(\mathbb{R}) = L^2(\mathbb{R}_-) \oplus L^2(\mathbb{R}_+).$$

becomes the non-trivial decomposition

$$L^2(\mathbb{R}) = H^2_+(\mathbb{R}) \oplus H^2_-(\mathbb{R}) \tag{3.13}$$

The functions that belong to $H^2_+(\mathbb{R})$ (resp. $H^2_-(\mathbb{R})$) are called progressive (resp. regressive). Also, the orthogonal projectors $P^{\pm} : L^2(\mathbb{R}) \to H^2_{\pm}(\mathbb{R})$ are expressed in terms of the Hilbert transfrom as follows:

$$P^{\pm} = \frac{1}{2}(1 \pm iH), \tag{3.14}$$

where $\hat{H}f(w) = -i\ \mathrm{sgn}(w)f(w)$ is the representation of the Hilbert transform in Fourier space. We recall here that the Hilbert transform is

$$Hf(t) = \frac{1}{\pi} \lim_{\epsilon \to 0} \int_{|t-x|>\epsilon} \frac{f(t)}{t-x} dt.$$

As an example, the Cauchy wavelet defined in (3.8) is progressive. Its Fourier transform is $\hat{g}(w) = w^{\alpha} e^{-w} 1_{\{w>0\}}$. That means, its Fourier coefficients for negative frequencies are zero.
The decomposition (3.13) gives an easy way of decomposing the analyzed function and the analyzing wavelet into a sum of functions belonging to $H^2_+(\mathbb{R})$ and $H^2_-(\mathbb{R})$. Let $g \in L^2(\mathbb{R})$. Using Theorem 6, we have

$$g(x) = \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} \psi_{a,b}(x) Wg(a,b) db,$$

where $\psi_{a,b}(x) = a^{-1/2}\psi(\frac{x-b}{a})$ with $a > 0$ and $Wg(a,b) = \langle g, \psi_{a,b} \rangle$.
The perpendicular splitting (3.13) provides us with a way of decomposing g into $g = g_1 + g_2$. It is enough to split the wavelet coefficients of g into a sum of Wg_+ and Wg_-, where $Wg_+ \in H^2_+(\mathbb{R})$ and $Wg_- \in H^2_-(\mathbb{R})$, and then reconstruct $g_1 = T(Wg_+)$ and $g_2 = T(Wg_-)$ where T is the reconstruction wavelet operator which takes Wg to g.

3.5 Multiresolution Analysis

Mallat (1999) and Meyer (1992) introduced the concept of multiresolution analysis (MRA) as a tool to construct a wavelet orthonormal basis. This notion comes

within the framework of multiscale approaches for which the work of Haar, Franklin and Littlewood-Paley are the precursors. In this section, we will introduce Mallat's representation of the MRA, which states that wavelets can be designed from discrete filters. For more details, see Mallat (1999).

3.5.1 Theoretical framework

Definition 11. *A multiresolution analysis is a sequence of subspaces of* $L_2(\mathbb{R})$ *denoted by* $(V_j)_{j\in\mathbb{Z}}$*, which has the following properties:*

$$\begin{aligned} V_j &\subset V_{j+1}, \\ \bigcap_{j\in\mathbb{Z}} V_j &= 0, \\ \overline{\bigcup_{j\in\mathbb{Z}} V_j} &= L_2(\mathbb{R}), \\ f(t) \in V_0 &\iff f(2^j t) \in V_j, \end{aligned}$$

and there exists a function $\phi(t) \in V_0$ *such that* $(\phi(\cdot - k))_{k\in\mathbb{Z}}$ *is an orthonormal basis for* V_0*.*

Roughly speaking, multiresolution analysis decomposes $L_2(\mathbb{R})$ into nested sequences of $(V_j)_{j\in\mathbb{Z}}$ named the *approximation spaces*. The space V_j provides a better approximation to a function f in $L_2(\mathbb{R})$ as j tends to infinity. The spaces $W_j \subset W_{j+1}$ such that W_j is a subset of V_{j+1} are called *detail spaces*. If we approximate $f_{j+1}(t) \in V_{j+1}$ by $f_j(t) \in V_j$, then $g_j(t) \in W_j$ holds the details we need to add to $f_j(t)$ to recover $f_{j+1}(t) = f_j(t) + g_j(t)$.
In the sequel, we will often write $\phi_n(t) = \phi(t - n)$. For H a Hilbert space and $(\phi_n)_{n\in\mathbb{Z}} \in H$, if there exist two constants $0 < A < B$ such that

$$A \sum_{n\in\mathbb{Z}} c_n^2 \leq \left\| \sum_{n\in\mathbb{Z}} c_n \phi_n(t) \right\|^2 \leq B \sum_{n\in\mathbb{Z}} c_n^2,$$

for any sequence $(c_n)_{n\in\mathbb{Z}}$, then the sequence $(\phi_n)_{n\in\mathbb{Z}}$ is a Riesz basis of H (Daubechies (1992) 0.0.10).

Theorem 12. *For $\{V_j\}_{j\in\mathbb{Z}}$ a multiresolution analysis for $L_2(\mathbb{R})$, there is a unique function $\phi \in L_2(\mathbb{R})$ such that*

$$\{\phi_{jk} : x \to 2^{j/2}\phi(2^j x - k)\}_{k\in\mathbb{Z}},$$

forms an orthonormal basis of the space V_j.

Since ϕ belongs to V_0 and V_0 belongs to V_1, we can write ϕ as a linear combination of $\{\phi_{1,k}\}_{k\in\mathbb{Z}}$. This means that there exists a square integrable sequence $\{h_k\}_{k\in\mathbb{Z}}$ such that

$$\phi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k\phi(2t - k), \tag{3.15}$$

and the Fourier transform of $\hat{\phi}$ of ϕ satisfies

$$\hat{\phi}(w) = m_0\left(\frac{w}{2}\right)\hat{\phi}\left(\frac{w}{2}\right), \tag{3.16}$$

where $m_0(w) = \frac{1}{\sqrt{2}}\sum_{k\in\mathbb{N}} h_k e^{-ikw}$. It is straightforward to see that m_0 is a 2π-periodic function. The function m_0 is called the *scaling filter*. One remarkable and useful property is that the orthonormality of the $(\phi_n)_{n\in\mathbb{Z}}$ can be restated in terms of $\hat{\phi}$ and m_0.

Theorem 13. *Assume $\phi \in L^2(\mathbb{R})$. The set $(\phi_n)_{n\in\mathbb{Z}}$ forms an orthonormal basis if and only if*

$$\sum_{k\in\mathbb{Z}} \left|\hat{\phi}(w + 2\pi k)\right|^2 = 1. \tag{3.17}$$

Moreover, if ϕ generates a multiresolution analysis $(V_j)_{j\in\mathbb{Z}}$, then

$$|m_0(w)|^2 + |m_0(w + \pi)|^2 = 1. \tag{3.18}$$

Proof. See Ruch and Van Fleet (2009) Theorem 5.5. □

The multiresolution approximation at resolution j of a function $f \in L_2(\mathbb{R})$ is obtained by projecting over the subspace V_j. We denote $P_{Vj}f$ the orthogonal projection

of f on V_j. These projections will construct Riesz bases of wavelets in the Hilbert space $L_2(\mathbb{R})$.

If $f \in L_2(\mathbb{R})$, the difference $P_{Vj+1}f - P_{Vj}f$ between resolutions j and $j+1$ is the orthogonal projection of f on W_j, the orthogonal complement of V_j in V_{j+1}. We write

$$V_j \oplus W_j = V_{j+1}, \tag{3.19}$$

and

$$P_{V_j}f + P_{W_j}f = P_{V_{j+1}}f. \tag{3.20}$$

The complement $P_{W_j}f$ provides the details that appear at scale $j+1$.

Hence, we construct a function ψ such that $\{x \mapsto \psi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of W_0. The function ψ is the *wavelet* and for any $j \in \mathbb{Z}$ the set $\{\psi_{jk} : x \to 2^{j/2}\psi(2^j x - k)\}_{k\in\mathbb{Z}}$ forms an orthonormal basis of W_j. Moreover, as W_0 belongs to V_1, there exists a sequence of real variables $\{g_k\}_{k\in\mathbb{Z}}$ such that

$$\psi(t) = \sqrt{2}\sum_{k\in\mathbb{Z}} g_k \psi(2t-k). \tag{3.21}$$

This defines the second wavelet filter m_1 such as $\hat{\psi}(w) = m_1(w/2)\hat{\psi}(w/2)$ and defined by

$$m_1(w) = \frac{1}{\sqrt{2}}\sum_{k\in F} g_k e^{-ikw}.$$

By mimicking m_0, the function m_1 also satisfies

$$\begin{aligned} |m_1(w)|^2 + |m_1(w+\pi)|^2 &= 1, \\ m_0(w)\overline{m_1(w)} + m_0(w+\pi)\overline{m_1(w+\pi)} &= 0. \end{aligned}$$

In practice, we normally choose m_1 to be expressed in terms of m_0 and we have $m_1(w) = -e^{-iw}\overline{m_0(w+\pi)}$.

Moreover, by iterating the equation (3.19), we obtain

$$V_j \oplus W_j \oplus ... \oplus W_{j'-1} = W_{j'}, \qquad if \;\; j < j'.$$

If $j' \to +\infty$ (and resp. also $j \to -\infty$), we obtain two decompositions:

$$\begin{aligned} L_2(\mathbb{R}) &= V_j \oplus \overline{\bigoplus_{j'=j}^{+\infty} W_{j'}}, \qquad \text{for all } j \in \mathbb{Z}, \\ L_2(\mathbb{R}) &= \overline{\bigoplus_{j'=-\infty}^{+\infty} W_{j'}}. \end{aligned}$$

The union of all Riesz bases in each of these direct sums provides several wavelet bases

$$\begin{aligned} \mathcal{B}_j &= \{\phi_{jk} : k \in \mathbb{Z}\} \cup \{\psi_{j'k} : j' \geq j, k \in \mathbb{Z}\}, \\ \mathcal{B} &= \{\psi_{jk} : j \in \mathbb{Z}, k \in \mathbb{Z}\}. \end{aligned}$$

Using equation (3.19), we can write

$$\begin{aligned} V_0 &= V_{-1} + W_{-1} \\ &= V_{-2} + W_{-2} + W_{-1} \\ &= V_{-J} + W_{-J} + ... + W_{-1}, \end{aligned}$$

This leads to the decomposition of a signal $x(t)$

$$\begin{aligned} x(t) &= A_1(t) + D_1(t) \\ &= A_2(t) + D_2(t) + D_1(t) \\ &= A_J(t) + D_J(t) + ... + D_1(t), \end{aligned}$$

where $D_i(t) \in W_{-i}$ is called the detail at level i, and $A_i(t) \in V_{-i}$ is called the approximation at level i.

Figure (3.5) shows a multilevel wavelet denoising of the EEG time series. We use Matlab function `wavedec` to perform a multilevel analysis to six levels with Daubechies-3 wavelets; refer to section 3.6 for background on Daubechies wavelet. Then we reconstruct the approximations at various levels using the Matlab function `wrcoef`. A

significant denoising occurs with level-6 approximation coefficients. The data corresponds to the recording on a different subject in the left occipital electrode (O1), with linked earlobes reference (http://www.vis.caltech.edu/ rodri/data.htm).

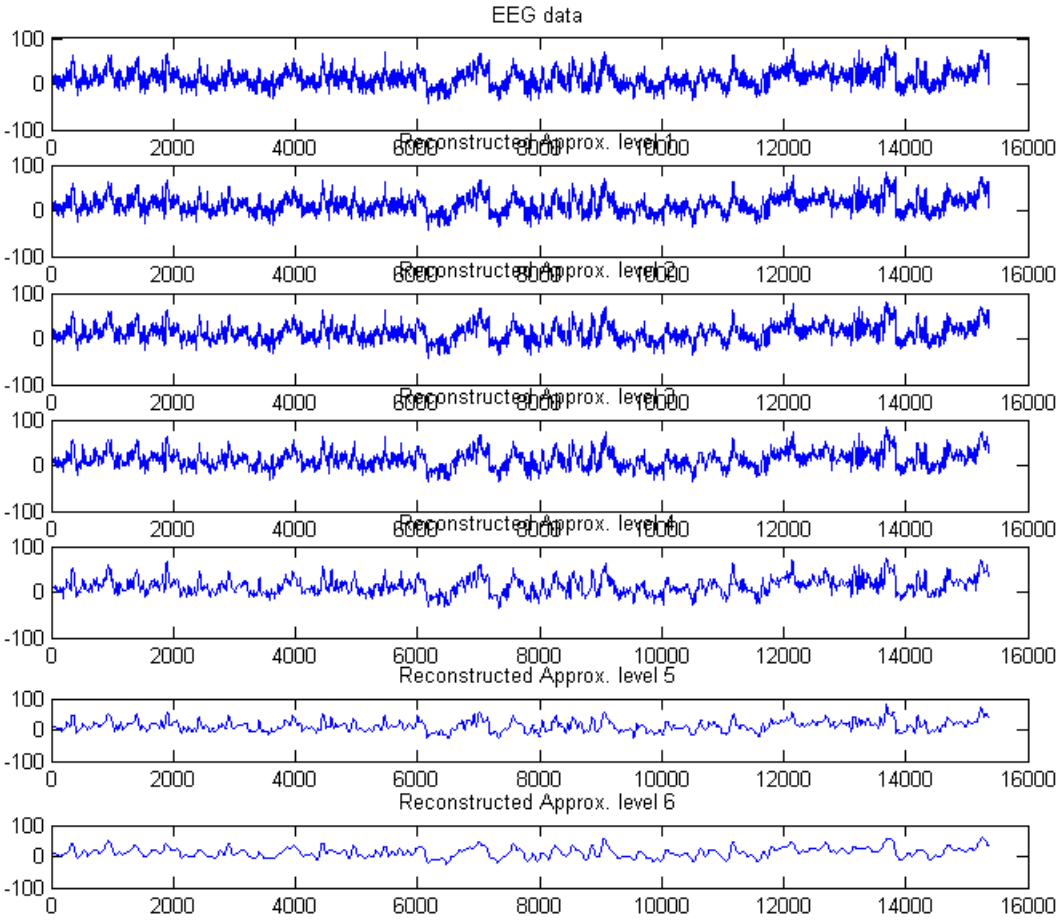


Figure 3.5: Level-6 Multiresolution Analysis of EEG time series.

3.6 Discrete Wavelet Transform

Like the Fourier transform, the wavelet transform has a discrete and a continuous version. Following the approach of Daubechies (1992), wavelet series are generated

by dilations and translations of a function ψ, called the mother wavelet:

$$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k), \qquad j, k \in \mathbb{Z},$$

The set of $\{\psi_{jk}|j, k \in \mathbb{Z}\}$ forms an orthonormal basis in $L^2(\mathbb{R})$. Given the wavelet basis, the wavelet series representation of a function $g \in L^2(\mathbb{R})$ is then

$$g(t) = \sum_{j,k\in\mathbb{Z}} d_{jk}\psi_{jk}(t),$$

where the wavelet coefficients d_{jk} are given by

$$d_{jk} = \int_{\mathbb{R}} g(t)\psi_{jk}(t)\mathrm{d}t.$$

The Discrete Wavelet Transform (DWT) can be computed with a convolution of discrete filters. Mallat (1987) suggested a method of constructing wavelet orthogonal bases using the Fast Fourier Transform (FFT).
Let the sequences $\{h_k\}_{k\in\mathbb{Z}}$ and $\{g_k\}_{k\in\mathbb{Z}}$ defined in (3.15) and (3.21) be a pair of response filters. We construct a scaling function ϕ and the corresponding wavelet function ψ in $L_2(\mathbb{R})$ such that

$$\begin{aligned} \widehat{\phi}(w) &= \prod_{k=1}^{+\infty} m_0\left(\frac{w}{2^k}\right), \\ \widehat{\psi}(w) &= m_1\left(\frac{w}{2}\right)\widehat{\phi}\left(\frac{w}{2}\right). \end{aligned}$$

These equations show that the infinite product converge uniformly on compact sets, and that ϕ has a compact support; see Daubechies (1999) Chapter 7. Using discrete filters reduces the complexity of the wavelet design. Instead of constructing an appropriate scaling function and the corresponding wavelet function, we simply choose the discrete set of coefficients of the filters.
Daubechies' idea was to construct a compactly supported orthonormal wavelet basis with compact support depending on an integer N which defines the support and the degree of regularity, namely *Daubechies wavelet DAUBN*. The goal was then to find

a trigonometric polynomial m_0 satisfying $m_0(0) = 1$ and

$$|m_0(w)|^2 + |m_0(w+\pi)|^2 = 1,$$

and $\prod_{j=1}^{N} m_0(\frac{w}{2^j}) 1_{[-\pi,\pi]}(\frac{w}{2^N})$ converges to $\prod_{j=1}^{\infty} m_0(\frac{w}{2^j})$ in L^2.
Daubechies studied this problem in terms of $|m_0(w)|^2$ instead of m_0 and proved that the trigonometric polynomial of the following form is a good choice:

$$m_1(w) = |m_0(w)|^2 = \sum_{k=0}^{N-1} \binom{2N-1}{k} \left(\frac{1+\cos w}{2}\right)^{2N-1-k} \left(\frac{1-\cos w}{2}\right)^{k} \tag{3.22}$$

Daubechies' fundamental result was that the choice of the polynomial $|m_0(w)|^2$ with degree $2N-1$ led to a compactly supported orthonormal scaling function ϕ_N with support on $[0, 2N-1]$ and degree of regularity $\alpha_N \sim (1 - \frac{\log 3}{2\log 2})N$ when $N \to \infty$. Moreover, since $m_1(w)$ has a root $w = \pi$ of multiplicity N, $m_1(w)$ can be written in terms of the Strang and Fix condition which states that any polynomial of degree $\leq N-1$ can be written of the form $P(x) = \sum_{k\in\mathbb{Z}} \langle P|\phi_N(x-k)\rangle \, \phi_N(x-k)$.
For example, When $N = 1$, we obtain *DAUB2* Daubechies and we find

$$|m_0(w)|^2 = \frac{1+\cos w}{2}.$$

The correct choice for m_0 would be $m_0(w) = \frac{1+e^{-iw}}{2}$. Therefore, we can calculate the scaling function ϕ

$$\widehat{\phi}(w) \;=\; \lim_{N\to\infty} \prod_{j=1}^{N} \frac{1}{2}\left(1 + \exp\left(-\frac{iw}{2^j}\right)\right).$$

We can verify that $\widehat{\phi}(w) = \frac{1-e^{-iw}}{iw}$ which implies that $\phi(x) = 1_{x\in\{[0,1]\}}$.
Let α_H be the supremum of β such that

$$\int (1+|w|^{\beta})|\hat{\phi}(w)|dw < \infty.$$

Let α_H be the exponent of the Hölder space $\mathbb{C}^{\alpha_H}$. We recall that a Hölder space $\mathbb{C}^s$

is defined by (Vidakovic, 1999 p.37)

1. $0 < s < 1$, $\mathbb{C}^s(\mathbb{R}) = \Big\{ f \in L_\infty(\mathbb{R}) : \sup \frac{|f(x+h)-f(x)|}{|h|^s} < \infty \Big\}$,
2. $s = n + s'$, $0 < s' < 1$, $\mathbb{C}^s(\mathbb{R}) = \big\{ f \in L_\infty(\mathbb{R}) \cap \mathbb{C}^n(\mathbb{R}) : \frac{d^n}{dx^n} f \in \mathbb{C}^{s'}(\mathbb{R}) \big\}$.

N	1	2	3	4	5	6	7	8	9	10
α_H		0.550	0.915	1.275	1.596	1.888	2.158	2.415	2.661	2.902

Table 3.1: Hölder regulatity of Daubechies scaling functions (Table 3.3, Vidakovic (1999))

From Table 3.1, we can see that *DAUB4* is the first differentiable wavelet with Hölder coefficient $\alpha_H = 1.275 > 1$. Figure (3.6) shows the plots of Daubechies' wavelets for different indices 2, 3, 4 and 8.

3.7 Besov spaces

Like Sobolev spaces, it has also been shown that in Besov spaces the smoothness of functions can be expressed in terms of approximation schemes and described through derivatives and differences; refer to Triebel (1992).

In this section, we introduce some relevant aspects of Besov spaces. Of especial importance is the connection between the wavelet theory and the Besov spaces. This connection is the key ingredient of the study of the smoothness of the prior model. The relationship between the hyperparameters of the prior model and the parameters of Besov spaces has been already investigated in Abramovich et al. (1998). For more details on functional spaces, refer to Triebel (1983).

We consider the family of Besov spaces $B^s_{p,q}(I)$, $I = [0, 1]$, $0 < s < \infty$, $0 < p \leq \infty$, and $0 < q \leq \infty$. These spaces have, roughly speaking, s derivatives in $L_p(I)$; The third parameter q allows one to make finer distinctions in smoothness. For $p = q = 2$, $B^s_{p,q}(I)$ is the Sobolev spaces $H^s(I)$.

For $p < 1$ or $q < 1$, these spaces are not quasi-Banach spaces, but rather quasi-normed linear spaces. This means that the triangular inequality may not hold and it

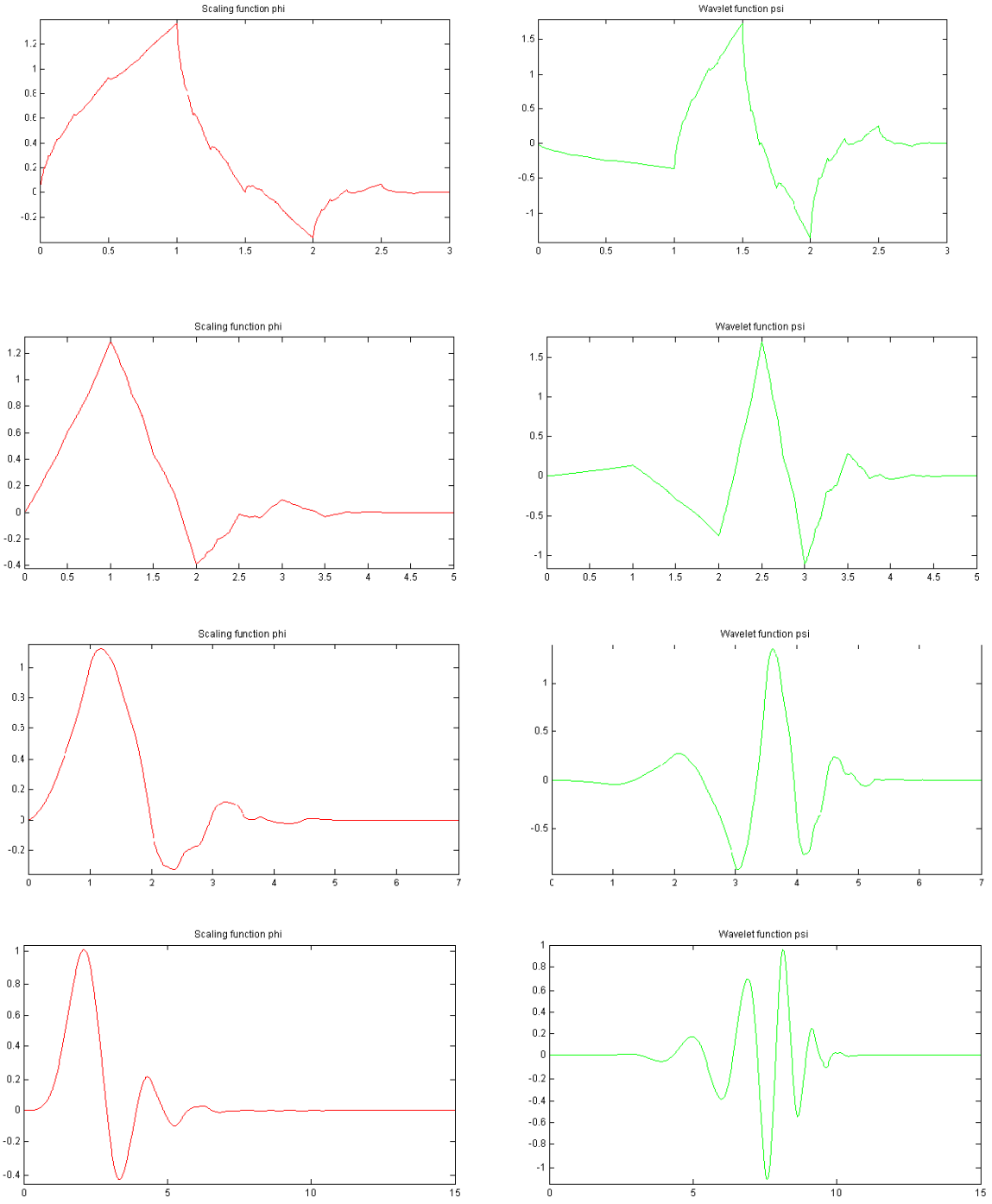


Figure 3.6: Daubechies' scaling and wavelet functions with indices 2, 3, 4 and 8.

is replaced by the inequality

$$||f+g||_{B^s_{p,q}(I)} \leq C(||f||_{B^s_{p,q}(I)} + ||g||_{B^s_{p,q}(I)}). \tag{3.23}$$

where C is a constant, f and g in $B^s_{p,q}(I)$. We recall here that a quasi-normed space is a quasi-Banach space if it is complete with respect to the quasi-norm in question. For simplicity, we shall continue to call these quasi-norms norms. We also wish to characterize Besov spaces $B^s_{p,q}$ in terms of means of differences.
For any $h \in \mathbb{R}$, we define $\Delta^0_h f = f(x)$ and let for $x \in I$, the mixed differences $\Delta^k_h f$ adapted to I of order k and step h is defined by

$$\Delta^k_h f(x) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} f(x+hj), \quad k \in \mathbb{N}, \quad h \in \mathbb{R}. \tag{3.24}$$

We recall that Δ^k_h can also be defined by

$$\Delta^{k+1}_h f(x) = \Delta^k_h f(x+h) - \Delta^k_h f(x), \quad k = 0, 1, ... \tag{3.25}$$

We define the $L_p(\mathbb{R})$-mixed modulus of continuity or means of differences, $0 < p \leq \infty$ as

$$w_r(f,t)_p = \sup_{|h| \leq t} \left(\int_{I_{rh}} |\Delta^k_h f(x)|^p dx \right)^{1/p}, \tag{3.26}$$

with the usual change to an essential supremum when $p = \infty$. Given $s > 0$, $0 < p \leq \infty$, and $0 < q \leq \infty$, choose $r \in \mathbb{Z}$ with $r - 1 \leq s \leq r$. Then the Besov space seminorm is defined as

$$|f|_{B^s_{p,q}(I)} = \left(\int_0^\infty (t^{-s} w_r(f,t)_p)^q \frac{dt}{t} \right)^{1/q}, \tag{3.27}$$

again with a supremum when $q = \infty$. The Besov space is defined by

$$B^s_{p,q} = \{ f \in L_p(I),\ ||f||_{B^s_{p,q}(I)} < \infty \}, \tag{3.28}$$

where

$$||f||_{B^s_{p,q}(I)} = |f|_{B^s_{p,q}(I)} + ||f||_{L_p(I)}, \tag{3.29}$$

and $L_p(I)$ is the class of functions f with $||f||_p = (\int_I |f(x)|^p dx)^{1/p} < \infty$ if $0 < p < \infty$ and $||f||_\infty = \inf(m : |f(x)| \leq m)$ if $p = \infty$. Since the Besov spaces can be characterized via moduli smoothness on I and admit wavelet decompositions, we can define a Besov sequence norm on the wavelet coefficients.

In the following we give the definition of the Besov norm as a function of the wavelet coefficients of a function which belongs to the Besov space.

To provide a better understanding of what is going on we will introduce the Littlewood-Paley decomposition. Calderón in 1960 introduced a well-known decomposition of the identity operator. Let $\psi \in L^2(\mathbb{R})$ such as $\hat{\psi} \in \mathcal{C}^\infty$, and compactly supported with $\text{Supp}(\hat{\psi}) \neq 0$. We define the convolution operator Q_t as $Q_t f = f * \psi_t(x)$ where $\psi_t(x) = \frac{1}{t}\psi\left(\frac{x}{t}\right)$. The Calderón formula is written as

$$\int_0^{+\infty} Q_t \circ Q_t^* \frac{dt}{t} = C_\psi I, \tag{3.30}$$

where Q_t^* is adjoint operator.

If we define $\Delta_j = Q_{1/2^j} \circ Q^*_{1/2^j}$, then the Calderón formula (3.30) is replaced by the Littlewood-Paley decomposition

$$f = \sum_{j \in \mathbb{Z}} \Delta_j f.$$

Therefore, the characterization of Besov space in terms of Littlewood-Paley is

$$B^s_{p,q} = \left\{ f \in \mathcal{S}' \mid \left(\sum_{j=0}^{\infty} 2^{jsq} ||\Delta_j f||^q_{L^p(\mathbb{R})} \right)^{1/q} < \infty \right\},$$

where $\mathcal{S}'$ is the set of all tempered distributions on $\mathbb{R}$.

For $f \in B^s_{p,q}$, suppose that f has a wavelet series expansion $f(x) = \sum_{j,k \in \mathbb{Z}} d_{jk}\psi_{jk}(t)$ such that $\sum |\hat{\psi}^2(2^{-j}\xi)| = 1$ *almost everywhere* for $\xi \in \mathbb{R}$.

Armed with Littlewood-Paley decomposition, the Besov sequence norm for f is defined by

$$||f||_{B^s_{p,q}} = |d_{00}| + \left\{ \sum_{j=0}^{\infty} 2^{js'q} \left(\sum_{k=0}^{2^j-1} |d_{jk}|^p \right)^{q/p} \right\}^{1/q}, \quad p > 0,\ 1 \leq q < \infty,$$

with $s' = s + \frac{1}{2} - \frac{1}{p}$. The norm when $q = \infty$ is given by

$$\| f \|_{B^s_{p,q}} = |d_{00}| + \sup_{j \geq 0} \left\{ 2^{js'} \left(\sum_{k=0}^{2^j-1} |d_{jk}|^p \right)^{1/p} \right\}.$$

This characterization of Besov norms goes back to Meyer (1992). In simple words, it means that one can control the regularity of a function f by controlling the amplitude of its wavelet coefficients.

3.8 Summary

In this chapter, we have introduced the basic definitions of wavelet theory and Besov spaces we need in the development of our model. In the following chapter we will introduce Lévy processes. We establish a novel approach in using the wavelet theory to approximate the Wiener-Hopf factors of a Lévy process using Hardy spaces corresponding to the wavelet domain. To our knowledge, these results have not been investigated before.

Chapter 4

Basics on Lévy processes

This chapter introduces the Lévy process which forms the structure of the model we develop in Chapter 5. We only aim to provide basic definitions and results. For an extensive review of Lévy processes, see Sato (1999) or Bertoin (1996).
We will first define infinitely divisible distributions and describe their connection to Lévy processes. Then, we will discuss on the construction of Lévy processes and give some examples of these processes.
We denote $\mu * \nu$ the convolution of the measure μ and ν, that is $\mu * \nu(A) = \int_{\mathbb{R}} \nu(A - x)\mu(\mathrm{dx})$, where A is a set and $A - z = \{x - z |\ x \in A\}$.

4.1 Definitions

Definition 14. *A probability measure μ is infinitely divisible if, for $n \in \mathbb{N}$, there is probability measure ν such that*

$$\mu = \underbrace{\nu * ... * \nu}_{n} = \nu^{*n}.$$

Theorem 15. *A probability measure μ is infinitely divisible if only if for all $n \in \mathbb{N}$, there exists a probability measure μ_n with characteristic function ϕ_n such that*

$$\phi_\mu(u) = (\phi_n(u))^n,$$

for all $u \in \mathbb{R}$.

Equivalently, the law ϕ_X of a random variable X is infinitely divisible if there is a sequence of i.i.d random variables $X_1, ..., X_n$ copies of X such that $X_1 + ... + X_n$ has distribution ϕ_X.

Proposition 16. *Let* $(\mu_n)_{n\in\mathbb{N}}$ *be a sequence of probability measures in* $\mathbb{R}$. *Denote* $\hat{\mu}$ *the characteristic function of* μ.
If $\hat{\mu}(z)$ *converges to a function* $\phi(z)$ *and if* $\phi(z)$ *is continuous at* $z = 0$, *then* $\phi(z)$ *is the characteristic function of a probability measure* μ *such that* $\mu_n \to \mu$.

Proof. This is Lévy's continuity theorem; see Williams (1991). □

Proposition 17.

(i) If μ *and* ν *are infinitely divisible measures, then* $\mu * \nu$ *is infinitely divisible.*

(ii) if μ *is an infinitely divisible measure, then* $\hat{\mu}(z) \neq 0$ *for all* $z \in \mathbb{R}$.

Proof. See Sato (1999) Lemmas 7.4 and 7.5. □

The Gaussian, Gamma, α- stable and Poisson distributions are examples of infinitely divisible distributions. A uniform distributed variable is not infinitely divisible.
The next theorem, the Lévy-Khintchine formula, gives a complete characterization of an infinitely divisible distribution via its characteristic function.

Theorem 18. *If the law* ϕ_X *of a random variable* X *is infinitely divisible, then its characteristic function satisfies*

$$\hat{\phi}_X(z) = \exp\left[i\gamma z - \frac{\sigma^2}{2}z^2 + \int_{-\infty}^{+\infty}(e^{izx} - 1 - izx1_{|z|\leq 1})\nu(\mathrm{d}x)\right], \tag{4.1}$$

where γ *and* σ *are real values and* ν *is a measure on* $\mathbb{R}$-$\{0\}$ *satisfying* $\int_{\mathbb{R}}(x^2 \wedge 1)\nu(\mathrm{d}x) < \infty$.

Proof. : See Theorem 8.1 Sato (1999). □

We give a general definition of a Lévy process.

Definition 19. *Let $\mathbb{P}$ be a probability measure on $(\Omega, \mathcal{F})$.*
A real-valued process $X = (X_t)_{t\geq 0}$ is called a Lévy process for $(\mathbb{P}, \Omega, \mathcal{F})$ if the following conditions are satisfied:

1. *The paths of X are right-continuous with left limits,*
2. *For $0 \leq s \leq t$, the increment $X_{t+s} - X_t$ is independent of $(X_u, 0 \leq u \leq t)$,*
3. *For $0 \leq s \leq t$, $X_{t+s} - X_t$ has the same distribution as X_s,*
4. *$\mathbb{P}(X_0 = 0) = 1$.*

If X verifies only the conditions 2 to 4, it is called a *Lévy process in law.*
The Poisson process is the simplest non-deterministic Lévy process. The compound Poisson process and the Brownian motion are also Lévy processes.

Definition 20. *Let X be a Lévy process and assume we observe X_t at a regular time interval Δ. Define*

$$S_n(\Delta) = \sum_{k=0}^{n-1} Y_k, \tag{4.2}$$

where $Y_k = X_{(k+1)\Delta} - X_{k\Delta}$ is a sequence of i.i.d random variables with the same distribution as X_Δ. Then $\{S_n\}_{n\in\mathbb{N}}$ is called a random walk.

For any $n \in \mathbb{N}$ and any $t > 0$, we have

$$X_t = X_{\frac{t}{n}} + (X_{2\frac{t}{n}} - X_{\frac{t}{n}}) + ... + (X_{n\frac{t}{n}} - X_{(n-1)\frac{t}{n}}).$$

We notice that we can define a Lévy process from a sum of i.i.d of random variables with the same distribution as $X_{\frac{t}{n}}$. In addition to this, we can conclude that:

Proposition 21. *If X is a Lévy process, then the law ϕ_{X_1} of the random variable X_1 is infinitely divisible. Conversely, if ϕ_{X_1} is infinitely divisible then there exists a Lévy process from which it arises in this way.*

Define the characteristic function of X_t

$$\Phi_t(u) = \mathbb{E}(\exp(iuX_t)), \quad t \geq 0, \quad u \in \mathbb{R}. \tag{4.3}$$

We have

$$\Phi_{t+s}(u) = \mathbb{E}\left(\mathrm{e}^{iuX_{t+s}}\right) = \mathbb{E}\left(\mathrm{e}^{iu(X_{t+s}-X_s)}\mathrm{e}^{iuX_s}\right), \tag{4.4a}$$

$$= \mathbb{E}\left(\mathrm{e}^{iuX_{t+s}}\right)\mathbb{E}\left(\mathrm{e}^{iuX_s}\right), \tag{4.4b}$$

$$= \Phi_t(u)\Phi_s(u), \tag{4.4c}$$

Equation (4.4c) together with the function $t \to \Phi_t(u)$ being continuous, implies that the unique solution of (4.4c) is exponential function

$$\Phi_t(u) = \exp(-t\Psi(u)). \tag{4.5}$$

where $\Psi : \mathbb{R} \to \mathbb{R}$ is a continuous function.
From Definition (19) and Proposition (21), it is not straightforward to see how a Lévy process is constructed. One way is to define a Lévy process via its characteristic function and the Lévy-Khintchin representation.
Using the Lévy-Khintchin representation, the characteristic function of X is giving by

$$\mathbb{E}(\exp(iuX_t)) = \exp(-t\Psi(u)), \tag{4.6}$$

where

$$\Psi(u) = -i\gamma u + \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty}(e^{iux} - 1 - iux1_{|x|\leq 1})\nu(\mathrm{d}x),$$

γ and σ are real values and ν is a measure on $\mathbb{R}$-$\{0\}$ satisfying $\int_{\mathbb{R}}(x^2 \wedge 1)\nu(\mathrm{d}x) < \infty$. The function Ψ is called the characteristic exponent.
From the Lévy-Khintchin representation, we see that a Lévy process consists of three components: a constant, a Brownian component and a pure jump component. The triple (γ, σ, ν) is called the *characteristic triple*.
• If $\int_{|x|<1}|x|\nu(dx) < \infty$, we can write the characteristic exponent in the form

$$\Psi(u) = -i\gamma' u + \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty}(e^{iux} - 1)\nu(\mathrm{d}x), \tag{4.7}$$

where $\gamma' = \gamma - \int_{|x|<1}|x|\nu(dx)$.

• If $\int_{|x|\geq 1} |x|\nu(dx) < \infty$, the characteristic exponent can take the form

$$\Psi(u) = -i\gamma^{''}u + \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{+\infty}(e^{iux} - 1 - iux)\nu(\mathrm{d}x), \tag{4.8}$$

where $\gamma^{''} = \gamma + \int_{|x|\geq 1} |x|\nu(dx)$. For Equations 4.7 and 4.8, see Sato (1999) Remark 8.4.

A fundamental concept and absolutely necessary for the study of Lévy processes is its Lévy measure, which we define by

Definition 22. *Let X be a Lévy process in $\mathbb{R}$ and A be a Borel subset of $\mathbb{R}$-$\{0\}$. The measure defined by*

$$\nu(A) = \mathbb{E}\left[\sum_{0\leq s\leq t} 1_A(\Delta X_s) \neq 0\right],$$

is called a Lévy measure of X.

The process $(\Delta X_t)_{t\geq 0}$ defined by $\Delta X_t = X_t - X_{t-}$, where $X_{t-} = \lim_{s\to t} X_s$, is called the jump process associated to the Lévy process $(X_t)_{t\geq 0}$.

The Lévy measure counts the size and the frequency of the jumps of the process of size in A up to t. It also contains useful information on the path properties and the variability of the Lévy process.

Proposition 23. *Let X be a Lévy process in $\mathbb{R}$ with triplet (γ, σ, ν).*

(i) If $\gamma = 0$ and $\nu(\mathbb{R}) < \infty$ or $\gamma = 0$, $\nu(\mathbb{R}) = \infty$ and $\int_{|x|\leq 1} |x|\nu(dx) < \infty$, then the sample paths $X_t(w)$ have a finite variation for each $t \in (0, \infty)$.

(ii) If $\gamma \neq 0$ or $\nu(\mathbb{R}) = \infty$, then the sample paths $X_t(w)$ have a infinite variation for each $t \in (0, \infty)$.

Proof. See Theorem 21.9 in Sato (1999). □

As an example, if $\nu = 0$, then X is a Brownian motion. For a Poisson process, X has $\nu = \lambda\delta_1$, where δ_1 is the Dirac measure at 1.

Let C_k be the k^{th} absolute moment of X_t, and define

$$C_k(t) = \mathbb{E}(|X_t|^k),\ t \geq 0.$$

It is straightforward to verify that for any $s, t \geq 0$

$$C_k(t+s) = \sum_{k=0}^{n} \binom{n}{k} C_k(t) C_{n-k}(s).$$

A very nice result of Sato (1999), Theorem 25.3, shows that for $k \geq 2$, the k^{th} moment is finite if the k^{th} moment of the Lévy measure exists, that is $\int_{|x|\geq 1} |x|^k \nu(dx) < \infty$. For the case where $k = 1$, C_1 is finite if and only if $\int_{|x|\geq 1} |x| \nu(dx) < \infty$.
The cumulants of a Lévy process with characteristic triple (σ, ν, γ) at time $t = 1$ are

$$\mathbb{E}(X_1) = \gamma + \int_{|x|\geq 1} x\nu(dx), \tag{4.9a}$$

$$C_2(X_1) = \sigma^2 + \int_{-\infty}^{+\infty} x^2 \nu(dx), \tag{4.9b}$$

$$C_n(X_1) = \int_{-\infty}^{+\infty} x^n \nu(dx), \ for\ n \geq 3. \tag{4.9c}$$

To conclude our basic introduction to Lévy processes, we give some examples:
• **Brownian Motion with drift.** The process X defined by $X_t = \mu t + \sigma W_t$ where W is a standard Brownian motion is a Lévy process. Its characteristic exponent is $-iu\mu + u^2\sigma^2/2$.
• **Poisson processes.** The probability distribution of Poisson process N with intensity λ concentrated on $n \in \mathbb{N}$ is $P_\lambda(N_t = n) = e^n \lambda^n / n!$. Its characteristic exponent is

$$\Psi(u) = \mathbb{E}(\mathrm{e}^{uN_t}) = \lambda t(1 - e^{iu}) = \int (1 - e^{iux}) \lambda \delta_1(dx).$$

Using the Lévy-Khinchine formula, the Poisson process is a Lévy process with Lévy measure $\lambda\delta_1$.
• **Compound Poisson processes.** Let N be a Poisson process and consider $X_t = \sum_{k=1}^{N_t} Y_k$ where $\{Y_i : i = 1, ...\}$ is a sequence of i.i.d random variables distributed with F. The characteristic exponent of X is

$$\Psi(u) = \lambda \int (1 - e^{iux}) F(dx).$$

The compound Poisson process is a Lévy process with Lévy measure $\lambda F(dx)$.

4.2 Subordinators

In this section, we consider the class of subordinators as possible choice of a Lévy process. Subordinators will be relevant to the developments in the simulation method in section 4.8.

A subordinator $Y = (Y_t)_{t\geq 0}$ is a Lévy process which takes values in $\mathbb{R}^+$. Such a process has clearly $\nu\{\mathbb{R}^-\backslash\{0\}\} = 0$ and its trajectories are almost surely non-decreasing. We can interpret a subordinator as a random time. We describe a subordinator using Laplace transforms rather than Fourier transforms. We define the Laplace exponent of Y by

$$\phi(u) = -\frac{1}{t}\log \mathbb{E}(\exp(-uY_t)) = \delta u + \int_0^{+\infty} (\exp(ux) - 1)\nu(\mathrm{d}x), \tag{4.10}$$

where $\int_0^{+\infty}(1 \wedge x)\nu(\mathrm{d}x) < \infty$.

Theorem 24. *Let X be a Lévy process on $\mathbb{R}$ with characteristic triple (γ, σ, ν). If T is a subordinator with Laplace exponent ϕ, independent of X, then the subordinated process $Y := X(T)$ is a Lévy process with Lévy characteristics*

$$\begin{aligned} \gamma_Y &= \gamma\delta + \int_{(0,+\infty)} \nu(\mathrm{ds}) \int_{|x|\leq 1} x p_X(\mathrm{dx}), \\ \nu_Y(\mathrm{dx}) &= \nu(\mathrm{dx})\delta + \int_{(0,+\infty)} p_X(\mathrm{dx})\nu(\mathrm{ds}), \\ \sigma_Y &= \sigma\delta. \end{aligned}$$

where p_X is the distribution function of X.

Proof. For a proof, refer to Sato (1999), Theorem 30.1. □

The transformation above from X to Y is called the subordination by T.
Since this process is non-decreasing, it is used to study and to time-change other Lévy processes.

Examples of Subordinators:

A simple example of a subordinated Lévy process is Compound Poisson process which is a random walk time changed by a Poisson process $X_t = S_{N_t}$ where $S_n = \sum_{i=1}^{n} Y_i$ and N is an independent Poisson process. $\{Y_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d random variables distributed with the same law f.

Consider a measure μ_α on $(0, +\infty)$ with density

$$\mu_\alpha(\mathrm{d}x) = \frac{c}{x^{1+\alpha}} 1_{\{x>0\}} \mathrm{d}x,$$

where $c > 0$. Then, μ_α is the Lévy measure of subordinator if and only if $0 < \alpha < 1$. Another example of a time changed Lévy process which is widely used is to subordinate a drifted Brownian motion with an independent subordinator Y. The resulting time-changed process X defined as $X_t = W(Y_t) + \mu Y_t$ is a Lévy process [Theorem 4.3, Cont and Tankov (2004)]. The Lévy measure ν of the time-changed Brownian motion with drift μ is given by

$$\nu_X(x) = \int_0^{+\infty} e^{-\frac{(x-\mu y)^2}{2y}} \frac{\nu_Y(\mathrm{d}x)}{\sqrt{2\pi y}},$$

where ν_Y is the Lévy measure of Y.

4.3 Stable processes

Stable processes are special cases of Lévy processes introduced by Paul Lévy in the 1920s. They generalize the Gaussian processes and their distributions are heavy tailed and skewed. Stable processes are selfsimilar processes for which a change of time scale is translated as a change of spatial scale. In this section, we give the definition and some properties of the stable processes and distributions. The stable laws were studied in great details in Zolotarev (1983), Samorodnitsky and Taqqu (1994).

A random variable has a stable distribution if it has the property that the sum of two independent copies of the variable have the same distribution up to scale and location parameters. Here, stability means that the shape of the distribution of the

random variable is preserved under sums. Let X_1 and X_2 be two copies of random variable X. The definition of stability can then be stated as follows: for $\{a_1, a_2\} \in \mathbb{R}$, $a_1X_1 + a_2X_2$ is of the same type as X, that is, $a_1X_1 + a_2X_2 \stackrel{d}{=} AX + B$ for $\{A, B\} \in \mathbb{R}$ with $A > 0$, where $\stackrel{d}{=}$ means equality in distribution (Nolan, 2013).
The familiar case of the normal distribution represents a stable distribution in its own right. The formal definition of a stable distribution is stated as follows in Feller (1971):

Definition 25. *A random variable X is said to be stable if*

$$X_1 + ... + X_n \stackrel{d}{=} c_nX + d_n, \tag{4.11}$$

where $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of n independent copies of X, and $c_n > 0$ and $d_n \in \mathbb{R}$.

Definition 26. *If the distribution of X_1 is stable, then a Lévy process X is said to be a stable process.*

It is shown that necessarily $c_n = n^{1/\alpha}$; see Feller (1971). The parameter α is called then the *index*. The distribution is said to be strictly stable if $d_n = 0$. To make the connection with the Lévy-Khintchin representation, we define the Lévy measure

$$\nu(x) = A\frac{dx}{|x|^{1+\alpha}}1_{\{x<0\}} + B\frac{dx}{x^{1+\alpha}}1_{\{x>0\}}, \tag{4.12}$$

where $0 < \alpha < 2$, $A, B \geq 0$. The parameters A and B describe the skewness of the distribution. If $A = B$, we said that the stable distribution is symmetric. A Gaussian variable is an α-stable variable with $\alpha = 2$. For $\alpha = 1$, we obtain a Cauchy random variable.
Therefore, under this formulation, the stable distributions represent a four-parameter family of distributions denoted by $S(\alpha, \beta, \sigma, \mu)$ and an alternative definition using the characteristic function is given by

Definition 27. *A random variable X is said to have a stable distribution if there are*

parameters $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, *and* $\mu \in \mathbb{R}$

$$\begin{aligned}\phi(u) &= \mathbb{E}(\exp(iuX)) \\ &= \begin{cases}\exp\{-\sigma^\alpha|u|^\alpha(1 - i\beta\text{sgn}(u)\tan(\frac{\pi\alpha}{2})) + i\mu u\}, & \text{if } \alpha \neq 1, \\ \exp\{-\sigma|u|(1 + i\frac{2\beta}{\pi}\text{sgn}(u)\ln(|u|) + i\mu u\}, & \text{if } \alpha = 1.\end{cases}\end{aligned}$$

For $\alpha < 2$,

$$\beta = \frac{A - B}{A + B}, \tag{4.13}$$

$$\mu = \gamma + \frac{B - A}{1 - \alpha},\ \alpha \neq 1 \tag{4.14}$$

$$\sigma = \begin{cases}(-(A + B)\Gamma(-\alpha)\cos(\pi\alpha/2))^{1/\alpha}, & \text{if } \alpha \neq 1 \\ \frac{\pi}{2}(A + B), & \text{if } \alpha = 1\end{cases} \tag{4.15}$$

The parameter β is the skew parameter and it represents the asymmetry of the distribution and it must satisfy $-1 \leq \beta \leq 1$. When $\beta = 0$ is said that the distribution is symmetric. The parameter σ represents the scale parameter. It can be seen as the volatility of the process. μ is the shift parameter.

The above formula is not the only form of the characteristic function. The literature on stable distributions contains many variations of parametrization which offer greater and lesser degrees of convenience for algebraic manipulation or numerical analysis purposes. The parametrization which we will use for the stable distribution is the parametrization of Zolotarev (1986):

$$\phi(u) = -\exp\{\nu^{-1/2}(\ln|u| + \tau - i\frac{\pi}{2}\theta\text{sgn}(t) + \gamma(\nu^{-1/2} - 1))\}, \tag{4.16}$$

where $\gamma = \lim_{n\to\infty}(\sum_{k=1}^{n}\frac{1}{k} - \ln(n)) \simeq 0.57721$ is Euler's constant. The parameters ν, θ and τ vary within the following bounds:

$$\nu \geq 1/4, \qquad |\theta| \leq \min(1, 2\sqrt{\nu} - 1), \qquad |\tau| < \infty,$$

and are connected with the parameter α, β and σ by the relations

$$\alpha = 1/\sqrt{\nu}, \tag{4.17a}$$

$$\beta = \theta \max(1, 1/(2\sqrt{(}\nu) - 1))\text{sgn}(1 - 1/\nu), \tag{4.17b}$$

$$\mu = \begin{cases} 0 & \text{if } \nu \neq 1, \\ \frac{\pi}{2}\tan(\pi\theta/2) & \text{if } \nu = 1, \end{cases} \tag{4.17c}$$

$$\sigma = \begin{cases} \exp(\tau/\sqrt{\nu} - \gamma(1 - 1/\sqrt{\nu})) & \text{if } \nu \neq 1, \\ \exp(\tau/\nu + \ln(\cos(\pi\theta/2)) - \ln(\pi/2)) & \text{if } \nu = 1. \end{cases} \tag{4.17d}$$

In general, α-stable laws don't have an explicit closed form for their probability density except for normal, Cauchy and Lévy distributions.

• *Normal distribution* $N(\mu, \sigma^2) = S(2, 0, \sigma/2, \mu)$ has a density

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

• *Cauchy distribution* $\text{Cauchy}(\mu, \nu) = S(1, 0, \nu, \mu)$ has a density

$$f(x) = \frac{1}{\pi} \frac{\nu}{\nu^2(x-\mu)^2}.$$

• *Lévy distribution* $\text{Lévy}(\mu, \nu) = S(1/2, 1, \nu, \nu + \mu)$ has a density

$$f(x) = \sqrt{\frac{\nu}{2\pi}} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\nu}{2(x-\mu)}\right) 1_{x>\mu}.$$

When $\alpha = 2$, the α-stable distribution coincides with the Gaussian distribution for which all moments exist. When $\alpha < 2$, the α-stable law has an infinite variance and heavy tails with asymptotic power law behaviour.

Proposition 28. *(Nikias and Shao, 1995)(Nolan, 1997)*

If $X \sim S(\alpha, \beta, \sigma, \mu)$ *for* $0 < \alpha < 2$, *then*

$$\begin{aligned} \lim_{x\to\infty} x^\alpha P(X > x) &= C_\alpha \frac{1+\beta}{2}\sigma^\alpha, \\ \lim_{x\to\infty} x^\alpha P(X < -x) &= C_\alpha \frac{1-\beta}{2}\sigma^\alpha, \end{aligned}$$

where (Zygmund, 1959)

$$C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x dx\right)^{-1} = \begin{cases} \frac{1-\alpha}{\Gamma(2-\alpha)\cos(\pi\alpha/2)}, & \text{if } \alpha \neq 1, \\ 2/\pi, & \text{if } \alpha = 1. \end{cases}$$

It is shown that when $\alpha < 1$, the mean of the α-stable distribution is equal to the shift parameter μ. Moreover, the p^{th} absolute moment $\mathbb{E}(|X|^p) = \int |x|^p f(x)dx$ exists if and only if $p < \alpha$.

In addition to this, the probability density p^α of a stable process satisfies the scaling property

$$p^\alpha(t, x) = t^{-1/\alpha} p^\alpha(1, t^{-1/\alpha}x)$$

Naturally, if we consider that $\{Y_1, ..., Y_n\}$ a sequence of i.i.d α-stable random variables distributed as $S(\alpha, 0, \gamma, 0)$, the generated random walk $S_n = Y_{n\Delta} = \sum_{k=1}^n (Y_{k\Delta} - Y_{(k-1)\Delta})$ as defined in (4.2) is an α-stable random variable and its characteristic function is given by

$$\psi_n(u) = [\exp(-\sigma|u|^\alpha)]^n = \exp(-n\sigma|u|^\alpha). \tag{4.18}$$

We can easily verify that the inverse Fourier transform of (4.18) gives the probability density function

$$p(S_n) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-n\sigma|u|^\alpha} e^{-iuS_n} du. \tag{4.19}$$

By changing the variable $\tilde{S_n} = S_n n^{1/\alpha}$, we verify easily that

$$p(\tilde{S_n}) = p(S_n)n^{1/\alpha}, \tag{4.20}$$

that is, α-stable distributions preserve the scaling properties.

These scaling properties are very important. It means that α-stable distributions for

different scale Δ, have similar functional forms. Figure (4.1) shows the pdfs of the α-stable densities for different tail indices.

For the Gaussian case (i.e $\alpha = 2$), we also obtain scaling properties for the variance $\sigma_n = \sqrt{n}\sigma$ (*finite variance*). This is also supported by the Central Limit theorem (CLT) which states that distribution of the sum of i.i.d random variables of finite variance converges to Gaussian distribution. However, this is not true for α-stable distributions (*infinite variance*). Nevertheless, there is a generalized CLT in the stable case, namely, Gnedenko and Kolmogorov CLT, which states that sum of n i.i.d random variables converges to α-stable distribution.

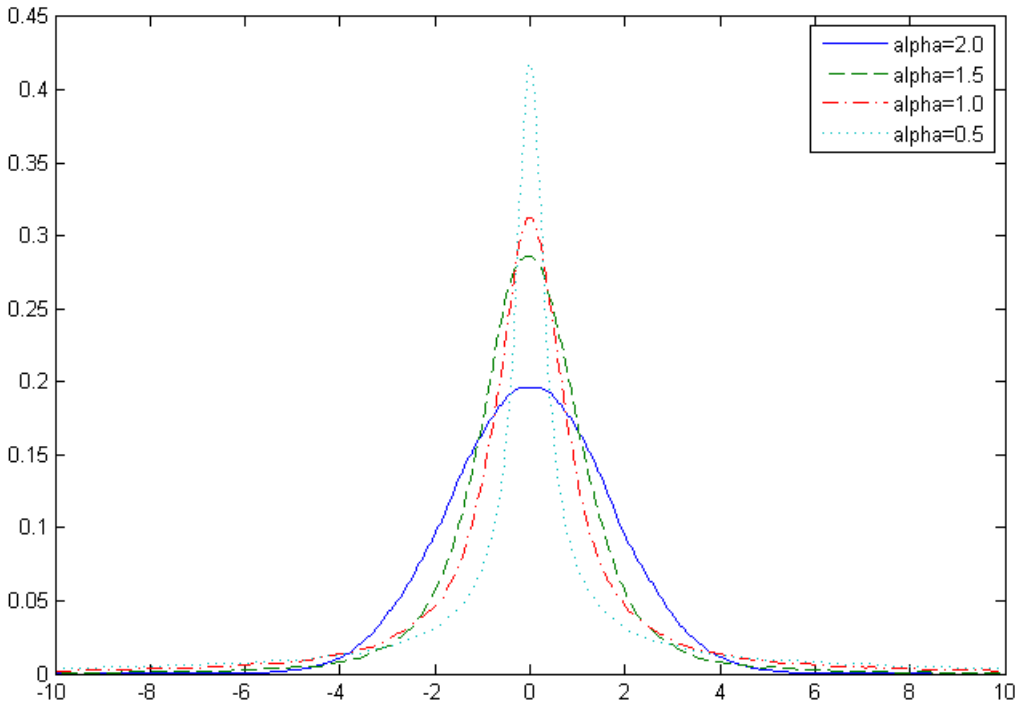


Figure 4.1: Probability density function of α-stable distributions for different tail indices;$\alpha = 0.5; 1; 1.5; 2$.

4.4 Stochastic Integral representation for Stable Processes

Stochastic integrals are useful and give us a convenient way to study the regularity of an α-stable process. Here, we will give only some definitions. See Samorodnitsky and Taqqu (1994), Chapter 10 for background and references.

Let $X = (X_t)_{t\in\mathbb{R}}$ be an $\alpha-$stable process and $(E, \mathcal{E}, m)$ be a metric space. We know from Samorodnitsky and Taqqu (1994) that X admits the integral representation

$$X_t = \int_E f(t,x)M(dx) + \eta(t), \tag{4.21}$$

where η is a deterministic function and M is an $\alpha-$stable random measure on $(E, \mathcal{E})$ with control measure m and skewness intensity $\beta : E \to [-1, 1]$. The function f belongs to the vector space F defined for $\alpha \neq 1$ by

$$F = \mathcal{L}^\alpha(E, \mathcal{E}, m) = \left\{ f : \ f \text{ measurable and } \int_E |f(x)|^\alpha m(dx) < \infty \right\},$$

and for $\alpha = 1$ by

$$\begin{aligned} F &= \mathcal{F}(m, \beta) \\ &= \left\{ f : \ f \in \mathcal{L}^1(E, \mathcal{E}, m) \text{ and } \int_E |f(x)\beta(x) \ln |f(x)||m(dx) < \infty \right\}. \end{aligned}$$

4.5 Tempered stable process

In this section, we introduce the tempered Lévy stable distribution. The motivation for using such a distribution is that its parameters decrease faster than those of the stable distribution and make the second and higher order moments finite.
The first attempt to make the variance finite has been made by Mantegna and Stanley (1995). They introduced the Truncated Lévy Flight distributions. Mantegna and Stanley's truncation results in a distribution which is not infinitely divisible. Novikov (1994) and Koponen (1995) developed a new structure of Truncated Lévy stable distributions which are infinitely divisible. The extension of the Koponen distributions,

namely the tempered stable distributions, have been studied in great detail in Rosiński (2007) and Boyarchenko and Levendorskii (2000).

The tempered stable Lévy measure is obtained by a exponential tilting transformation of the α-stable Lévy measure. That is, we multiply the α-stable Lévy measure by a pointwise in space decreasing exponential function. Such a transformation respects the integrability constraints of Lévy measures.
The tempered stable distribution of our interest takes the form of Carr, Geman, Madan and Yor (CGMY)(2005) and the tempered stable Lévy measure is given by

$$\nu(\mathrm{d}x) = Ap\frac{\exp(-G|x|)}{|x|^{1+\alpha}}1_{\{x<0\}}\mathrm{d}x + Aq\frac{\exp(-Mx)}{x^{1+\alpha}}1_{\{x>0\}}\mathrm{d}x, \tag{4.22}$$

where the parameters p, q describe the skewness of the distributions with $p+q=1$, $G, M \geq 0$ are the truncation parameters and $A > 0$, $0 < \alpha < 2$.

Using the Lévy-Khintchine representation, the characteristic function for a random variable X associated to a tempered stable Lévy measure is given by

$$\mathbb{E}(\exp(iuX)) = \exp\left(\int(\mathrm{e}^{iux}-1)\nu(\mathrm{d}x)\right) = \exp(-\Psi(u)),$$

where

$$\Psi(u) = -A\Gamma(-\alpha)[p(M-iu)^\alpha - pM^\alpha + q(G+iu)^\alpha - qG^\alpha].$$

The parameter A is the overall level of variability of the process. The parameters G and M control the skewness and the importance of large negative and positive moves. The parameter α controls the selfsimilarity of the process.
Yor and Madan (2005) showed that the CGMY process X can be represented as a time-changed Brownian motion

$$X_t = aZ_t + W_{Z_t},$$

where $(Z_t)_{t\geq 0}$ is a subordinator independent of the Brownian motion $(W_t)_{t\geq 0}$ and with Laplace transform

$$\mathbb{E}(e^{-uZ_t}) = \exp\left(tA\Gamma(-\alpha)\left\{2(2u+GM)^{\alpha/2}\cos(\xi(u)\alpha) - M^\alpha - G^\alpha\right\}\right),$$

where $u \in \mathbb{C}$, $\xi(u) = \arctan(\sqrt{2u-\lambda^2}/\tilde{\lambda})$, $\lambda = (G-M)/2$ and $\tilde{\lambda} = (G+M)/2$. Moreover, The subordinator Z_t has a Lévy density

$$\nu_z(t) = f(t)\nu_0(t), \tag{4.23}$$

where

$$f(t) = \frac{2^{\alpha/2}\Gamma(\frac{\alpha}{2}+\frac{1}{2})\mathrm{e}^{t\lambda^2/2 - t\tilde{\lambda}/4}}{\sqrt{\pi}} D_{-\alpha}(\tilde{\lambda}\sqrt{t}),$$

with D_α the parabolic cylinder function (Abramowitz and Stegun (1970)), and

$$\nu_0(t) = A\frac{2^{-\alpha/2}\sqrt{\pi}}{\Gamma(\frac{\alpha}{2}+\frac{1}{2})}\frac{1}{t^{1+\alpha/2}}1_{t>0} = \frac{K}{t^{1+\alpha/2}}1_{t>0}.$$

The case where $p = q = 1/2$ and $M = G = \lambda$ corresponds to the symmetric tempered stable distribution denoted $TS(\alpha, \sigma, \lambda)$, and for $0 < \alpha < 2$, $\alpha \neq 1$, the characteristic function has a closed form:

$$\begin{aligned}\Psi(u) &= A\lambda^\alpha\Gamma(-\alpha)\left[\frac{1}{2}(1-i\frac{u}{\lambda})^\alpha + \frac{1}{2}(1+i\frac{u}{\lambda})^\alpha - 1\right] &(4.24)\\ &= A\Gamma(-\alpha)\left[(\lambda^2+u^2)^{\alpha/2}\cos(\alpha\arctan(\frac{u}{\lambda})) - \lambda^\alpha\right]. &(4.25)\end{aligned}$$

Definition 29. *A Lévy process X is a tempered stable process if $X_1 \sim TS(\alpha, \sigma, \lambda)$.*

The cumulants of the tempered stable distribution are computed using the derivatives of the characteristic exponent, and they are of the form

$$C_n(X) = A\lambda^{\alpha-n}\Gamma(-\alpha-n)\left(p(-1)^n + q\right).$$

We assume we observe a discrete random variable with fixed sampling interval Δt. The probability density function $p_{\Delta t}$ at a time scale Δt of the tempered stable distri-

bution has a one-to-one correspondence with its characteristic function expressed via the inverse Fourier transform

$$p_{\Delta t}(x) = p(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \phi(k)\exp(ikx)\mathrm{d}x,$$

where $\phi(k) = e^{-\Psi(k)}$. Under the scaling transformations

$$\begin{cases} A_{\Delta t} = \Delta t A_1, \\ G_{\Delta t} = \Delta t^{-\frac{1}{\alpha}} G_1, \\ M_{\Delta t} = \Delta t^{-\frac{1}{\alpha}} M_1. \end{cases} \tag{4.26}$$

it is easy to prove that the tempered stable distribution exhibits also scaling properties

$$\begin{cases} x_{\Delta t} = \Delta t^{\frac{1}{\alpha}} x_1, \\ p_{\Delta t}(x) = \frac{p_1(x_1)}{\Delta t^{\frac{1}{\alpha}}}. \end{cases}$$

Therefore, the distribution at time scale Δt could be fully computed from the distribution at scale time $\Delta t = 1$. This result was confirmed in an empirical assessment of the analysis of the Standard and Poor's 500 index presented in Mantegna and Stanley (1995) in the case of truncated Lévy distribution.
Figure (4.2) shows the cdfs of the symmetric tempered stable distributions for different tail indices.

4.6 Asymptotic behaviour tempered stable distributions

In Chapter 2 we concluded that the tempered stable distribution is more suitable to model the data. Here, we will review the short and long time behaviour of the tempered stable process $X = (X_t)_{t\geq 0}$. This asymptotic behaviour has been studied in Rosiński (2007). Theorem 3.1 in Rosiński (2007) states that the tempered stable process behaves in short time like a stable process and in long time tends to Brownian motion.
Below, $\xrightarrow{d}$ denotes the weak convergence of a process to the space $D([0,+\infty),\mathbb{R})$ of

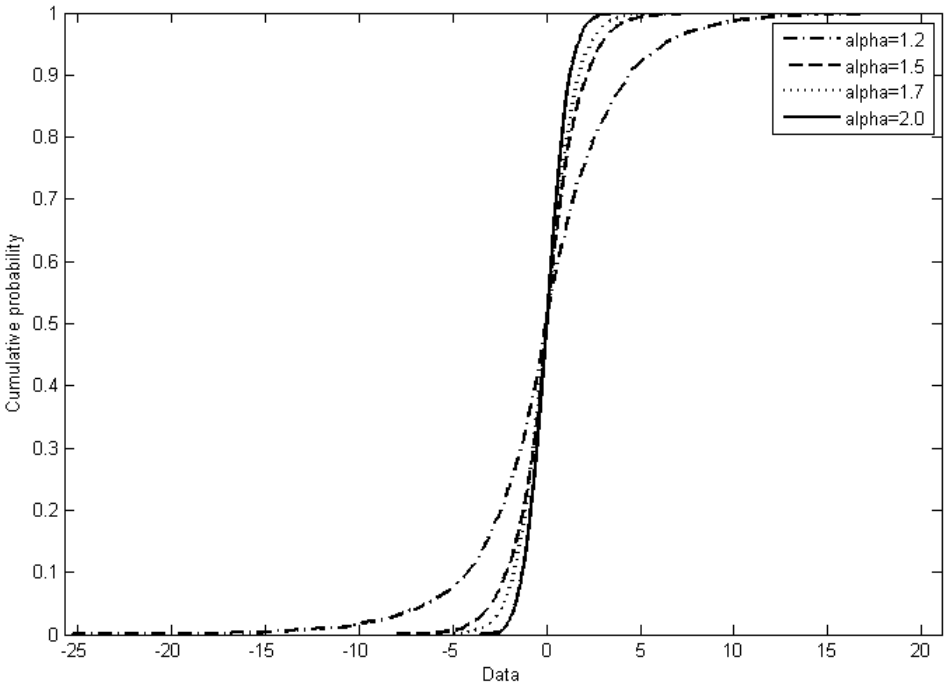


Figure 4.2: Cumulative density function of symmetric tempered stable distributions for different tail indices.

càdlàg functions equipped with the Skorohod topology.

Let X be a tempered stable Lévy process.

(i) *Short time behaviour:* If $\alpha \neq 1$ then

$$h^{-1/\alpha} X_{ht} \xrightarrow{d} X_t^{\alpha} \text{ if } h \to 0,$$

where $X^{\alpha} = (X_t^{\alpha})_{t \geq 0}$ is an α-stable process with $X_t^{\alpha} \sim S(\alpha, \beta, t\sigma, 0)$.

If $\alpha = 1$, then

$$h^{-1} X_{ht} - a_h t \xrightarrow{d} X_t^1 \text{ if } h \to 0,$$

where $a_h = \log h \int_{\mathbb{R}} x\nu(dx)$, and X^1 is a stable process with $X_t^1 \sim S(1, \beta, t\sigma, tb_0)$ where $b_0 = \int_{\mathbb{R}} x(1 - \log x^2)\nu(dx)$.

(ii) *Long time behaviour*: If $\alpha \in [1,2)$ then

$$h^{-1/2} X_{ht} \xrightarrow{d} cB_t \text{ if } h \to \infty,$$

where $B = (B_t)_{t\geq 0}$ is a centered standard Brownian motion and $c^2 = \Gamma(2-\alpha) \int_{\mathbb{R}} x^2 \nu(dx)$. If $\alpha \in (1,2)$ then

$$h^{-1/2} X_{ht} - k_h t \xrightarrow{d} cB_t \text{ if } h \to \infty,$$

where c is as above and $k_h = \alpha h \Gamma(1-\alpha) \int_{\mathbb{R}} x\nu(dx)$.
When $\alpha = 1$, the symmetry assumption of ν will be ensured by adding a_h and k_h and ν will be well defined in probability.

4.7 Wiener Hopf Factorization

We present in this section a fundamental path decomposition tool, known as the Wiener-Hopf factorization. Our main focus will be to establish a novel approach in using the wavelet theory to approximate the Wiener-Hopf factors of a Lévy process using Hardy spaces corresponding to the wavelet domain. See Sato (1999) section 9 and Bertoin (1996) section 6 for background and references on Wiener-Hopf factorization.
Let $X = (X_t)_{t\geq 0}$ be a Lévy process with Lévy exponent ψ. As starting point, we take the Wiener-Hopf factorization that is commonly used in probability theory:

$$\mathbb{E}(e^{iuX_T}) = \mathbb{E}(e^{iu\overline{X}_T})\mathbb{E}(e^{iu\underline{X}_T}), \quad \forall u \in \mathbb{R}. \tag{4.27}$$

where T denotes an exponential distributed random variables with mean q^{-1} that is independent of the process X, and $\overline{X}$ and $\underline{X}$ are its supremum and infinimum processes, respectively:

$$\overline{X}_t = \sup_{0\leq s\leq t} X_s, \qquad \underline{X}_t = \inf_{0\leq s\leq t} X_s, \qquad \forall t \geq 0.$$

If T is as above, the density of T equals $q\mathrm{e}^{-qt}dt$ implies that, for every $u \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}(\mathrm{e}^{iuX_T}) &= q\int_0^{+\infty} e^{-qt}\mathbb{E}\left(e^{iuX_t}\right), \\ &= \frac{q}{q+\psi(u)}.\end{aligned}$$

Define functions $\phi^{\pm}(u)$ by the formulas

$$\phi^{+}(u) = \mathbb{E}(e^{iu\overline{X}_T}),$$

$$\phi^{-}(u) = \mathbb{E}(e^{iu\underline{X}_T}).$$

Then immediately yields to the second form of the Wiener-Hopf factorization formula:

$$\frac{q}{q+\psi(u)} = \phi^{+}(u)\phi^{-}(u).$$

Since the trajectories of the supremum process $\overline{X}$ are non decreasing, $\phi^{+}(u)$ admits an analytic continuation into the upper half-plane and similarly, $\phi^{-}(u)$ into the lower half-plane.

In many examples of Lévy processes, no explicit formula for $\phi^{\pm}(u)$ is available, and one needs approximations, particularly for simulation purposes.

Boyarchenko and Levendorskii (2002) obtained integral formulas for the Wiener-Hopf factors. If $q > 0$ is large enough, $\Re(q + \Psi(u)) > 0$ for all u such that $\Im u \in (\lambda_-, \lambda_+)$. They proved that the Wiener-Hopf factor in the upper half-plane $\Im u > 0$ is

$$\phi^{+}(u) = \exp\left[\frac{1}{2\pi i}\int_{\Im\eta=w_-} \frac{u\ln(q+\Psi(\eta))}{\eta(u-\eta)}d\eta\right],$$

where $w_- \in (\lambda_-, 0)$. Similarly, in the lower half-plane $\Im u < 0$

$$\phi^{-}(u) = \exp\left[-\frac{1}{2\pi i}\int_{\Im\eta=w_+} \frac{u\ln(q+\Psi(\eta))}{\eta(u-\eta)}d\eta\right],$$

where $w_+ \in (0, \lambda_+)$. Boyarchenko and Levendorskii (2008) used the Fast Fourier Transform (FFT) and inverse FFT to compute the realizations of the Wiener-Hopf

factors.

Now, we will reformulate the Wiener-Hopf factorization in the language of wavelets. In the following theorem, we derive new formulas for the factors using wavelet transforms. We recall here that the Cauchy's differentiation formula (Krantz, 2008, section 3.1.1) is defined by

$$f^{(n)}(a) = \frac{n!}{2\pi i}\int_{\gamma}\frac{f(z)}{(z-a)^{n+1}}dz,$$

where $f : U \to \mathbb{C}$ is an analytic function, U is an open subset of $\mathbb{C}$, and γ is the circle forming the boundary of $D = \{z : |z - z_0| \leq r\} \subset U$. Let V an open subset of $\mathbb{C}$ containing U. If F an analytic function defined on V such that $F(z) = f(z)\ \forall z \in U$, then F is called an analytic continuation of f.

Theorem 30. *Let X be a Lévy process. $\phi^+(u)$ (resp., $\phi^+(u)$) admits an analytic continuation into the upper (resp., lower) half-plane given by*

$$\phi^{\pm}(u) = \exp\left[\frac{2i}{\pi}\int_0^{+\infty} ada\int_{-\infty}^{+\infty}\frac{\Psi^{'}(b+ia)}{q+\Psi(b+ia)}\frac{1}{(u-b\mp ia)^2}db\right],$$

Proof. Define the function $s(u) = \ln\left(q(q+\Psi(u))^{-1}\right)$. We consider the analyzing Cauchy wavelet $\psi(z) = \frac{1}{\pi}\frac{1}{(z+i)^2}$. It is an analytic function over the upper-half plane $\Im z > 0$ and the wavelet $\psi_{a,b}$ generates the Hardy space $\mathbb{H}^2_+(\mathbb{R})$ when $a > 0$.
Using the generalized Cauchy integral formulas, the wavelet coefficients of function s are then (see Meyer, 1993 p.26)

$$\begin{aligned} Ws(a,b) &= \langle s, \psi_{a,b}\rangle = \frac{1}{\pi}\int_{-\infty}^{\infty} s(x)\frac{a\sqrt{a}}{(x-b-ia)^2}dx, \\ &= 2ia\sqrt{a}s'(b+ia), \\ &= 2ia\sqrt{a}\frac{\Psi^{'}(b+ia)}{q+\Psi(b+ia)}. \end{aligned}$$

Therefore, the Wiener-Hopf factor $\phi^+(u)$ is continued analytically into the upper half-plane $\Im u > 0$ as follows:

$$\ln(\phi^+(u)) = \left[\frac{2i}{\pi}\int_0^{+\infty} ada\int_{-\infty}^{+\infty}\frac{\Psi^{'}(b+ia)}{q+\Psi(b+ia)}\frac{1}{(u-b-ia)^2}db\right].$$

For $\phi^{-}(u)$ into the lower half-plane $\Im u < 0$, the same results hold. This amounts to putting

$$\ln(\phi^{-}(u)) = \left[\frac{2i}{\pi}\int_{0}^{+\infty} ada \int_{-\infty}^{+\infty} \frac{\Psi'(b+ia)}{q+\Psi(b+ia)} \frac{1}{(u-b+ia)^2} db\right].$$

□

4.8 Simulation of tempered stable processes

A Lévy process can be generated from an infinitely divisible random variable. It is shown in Rosiński (2001) that Lévy processes $X = (X_t)_{t\geq 0}$ can be represented by a convergent series of the form

$$X_t = \sum_{i=1}^{\infty} \left[H(\Gamma_i, V_i)1_{U_i \leq t} - tc_i\right], \quad 0 \leq t \leq T,$$

where Γ_i are the arrival times of Poisson process with parameter 1, V_i are *i.i.d* random variables belonging to an Euclidean space S, U_i is *i.i.d* sequence of uniform random variables on $[0, T]$ and $c_i \in \mathbb{R}$. The function $H : (0, \infty) \to \mathbb{R}$ is a measurable function defined such that for each $v \in S$, $r \to |H(r, v)|$ is non-increasing.
To generate symmetric tempered stable random variables $TS(\alpha, a, \lambda)$, Rosiński (2001,2007) uses a series representation of the form

$$X_1 = \sum_{j=1}^{\infty} \min\left(V_j\left(\frac{\alpha\Gamma_j}{a}\right)^{-1/\alpha}, \left(\frac{E_j U_j^{1/\alpha}}{\lambda}\right)\right),$$

where U_j is a *i.i.d* sequence of uniform random variables in $[0, 1]$, E_j and $E_j^{'}$ *i.i.d* sequences of exponential variables of parameter 1 and $\Gamma_j = E_1^{'} + ... + E_j^{'}$, V_j a *i.i.d* Rademacher sequence of discrete random variables attaining values ± 1 with probability 1/2.
The Rosiński series representation is an extension of the LePage series of the sym-

metric α-stable random variables. LePage (1981) shows that if the series

$$S = \sum_{j=1}^{+\infty} V_j \Gamma_j^{-1/\alpha}, \tag{4.28}$$

converges almost surely, then the distribution of S is strictly stable. That is, the series $\sum_{j=1}^{+\infty} V_j \Gamma_j^{-1/\alpha}$ converges to a random variable $S \stackrel{d}{=} S(\alpha, (C_\alpha^{-1}\mathbb{E}(|V_1|^\alpha)^{1/\alpha}), 0, 0)$. Therefore, Rosiński series can viewed as LePage series for which the stable jumps $V_j \left(\alpha\Gamma_j/a\right)^{-1/\alpha}$ are cut at the level of $\frac{E_j U_j^{1/\alpha}}{\lambda}$.

The partial sums of the Rosiński series representation introduces an approximation bias due to the truncation procedure. This bias could be significant and at the same time, hard to quantify. To remedy this deficiency, Yor and Madan (2005) used the formulation of the tempered stable process as a time-changed Brownian motion. They suggested a simulation algorithm of the CGMY process based on a Rosiński's rejection approach; see Rosiński (2001). Poirot and Tankov (2006) suggested an alternative approach based on the construction of an equivalent measure under which the tempered stable process becomes a stable process whose exact simulation is well-established. They used the same simulation strategy used by Yor and Madan (2005) where the CGMY process is represented as a time-changed Brownian motion with drift, and the law of the time change is absolutely continuous with respect to that of the one sided stable $\alpha/2$ subordinator. They approximate the CGMY process by

$$\tilde{X}_t = A\tilde{Z}_t + \sqrt{\tilde{Z}_t} z,$$

where z is a standard normal random variable and $\tilde{Z}$ is the approximated subordinator process given by

$$\tilde{Z}_t = td + \sum_{i=1}^{\infty} y_i 1_{\Gamma_i \leq t} 1_{f(y_i) \geq u_i},$$

where

$$f(y) = e^{-B^2/2} \frac{\Gamma\left(\frac{\alpha}{2} + \frac{1}{2}\right)}{\Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{1}{2}\right)} 2^\alpha \left(\frac{B^2 y}{2}\right)^{\alpha/2} I(\alpha, B^2 y, B^2 y/2),$$

where

$$I(\nu, a, \lambda) = \int_0^\infty x^{\nu-1} e^{-ax-\lambda x^2} dx,$$

The parameter d defined by

$$d = \int_0^\epsilon y \frac{K}{y^{1+\alpha/2}} \mathrm{d}y = \frac{K\epsilon^{1-\alpha/2}}{1-\alpha/2},$$

represents the expected value of the truncated jumps below ϵ, the jump size

$$y_i = \frac{\epsilon}{(1-u_i)^{2/\alpha}},$$

and $\Gamma_i = \sum_{j=1}^{i} -\log(1-u_i)/\lambda$ are the arrival times of a Poisson process with parameter

$$\lambda = \int_\epsilon^\infty \frac{\mathrm{d}y}{y^{1+\alpha/2}} = \frac{2}{\alpha}\frac{1}{\epsilon^{\alpha/2}},$$

where u_i is a sequence of i.i.d uniform random variables.
Figure (4.3) shows the simulated path of tempered stable processes for different tail index parameters. Also, Figure (4.4) shows clearly the impact of the tempering on the variability of the stable sample trajectory.

4.9 Summary

In the next chapter, we will introduce the wavelet Lévy prior model and state the connection between the hyperparameters of the Lévy-based probabilistic model and Besov spaces. To this end, we review the existent prior models. Most of these models place a Gaussian prior on the wavelet coefficients. This motivates the extension to a non-Gaussian prior, where in the next chapter, we present our novel approach.

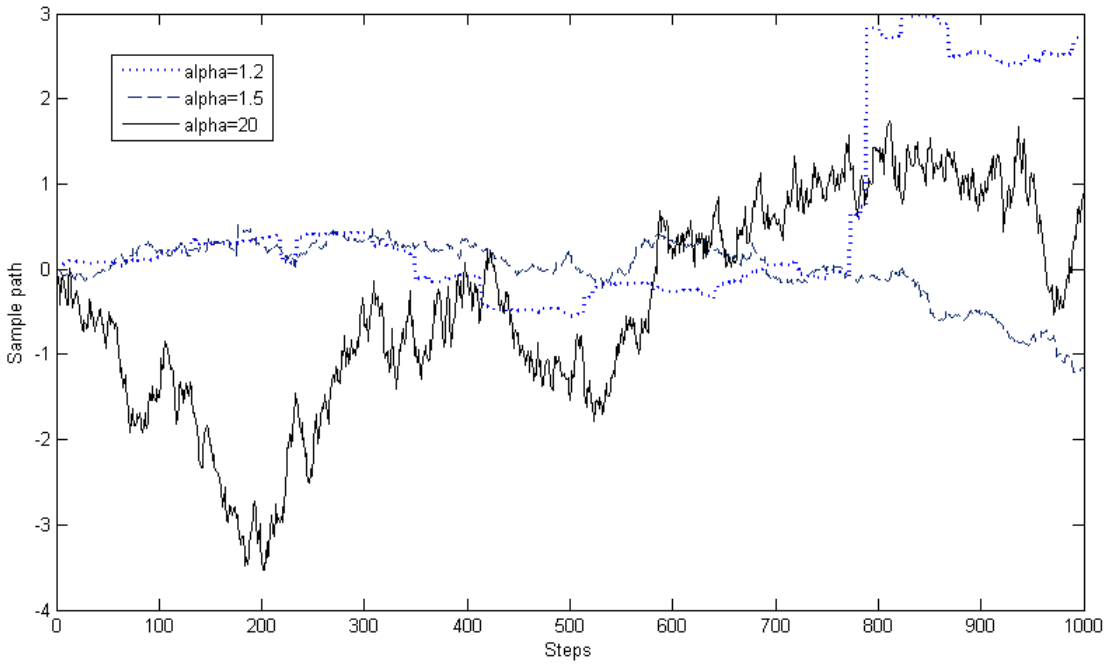


Figure 4.3: Sample paths of a TS(1.2,0.5,0.5), TS(1.5,0.5,0.5) and Brownian motion (solid line)

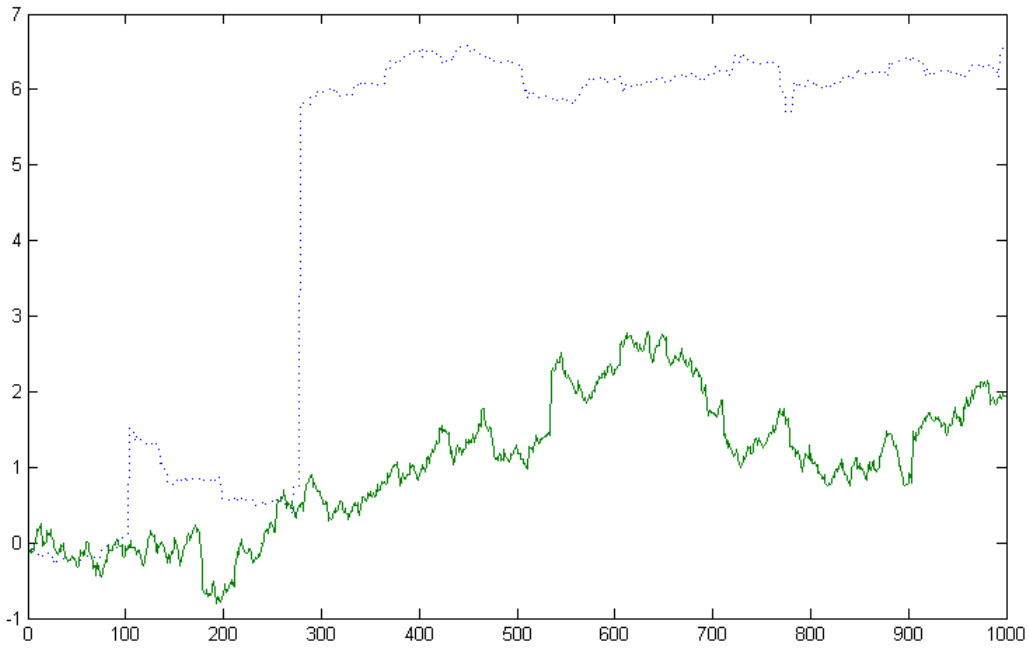


Figure 4.4: Sample paths of a stable TS(1.7,1,0) (Dot) and a tempered stable TS(1.7,1,0.5) (solide)

Chapter 5

Wavelet Statistical Models

5.1 The wavelet approach to nonparametric regression

We consider the standard univariate regression at regular points $\{t_i\}_{i=1,...,n}$:

$$y_i = g(t_i) + \epsilon_i, \qquad i = 1, ..., n,$$

where ϵ_i are independent $N(0, \sigma_\epsilon^2)$ random variables. The goal is to recover the underlying function g from the noisy data y_i without assuming any particular parametric structure for g.

We assume that the values $t_i \in [0, 1]$ are deterministic and are equally spaced; e.g., $t_i = i/n$.

A commonly used approach in nonparametric estimation is to expand the unknown function g in terms of an orthogonal set of basis functions such as Fourier series and wavelet series. It is often appropriate to expand the function g in wavelet series. The wavelet basis allow a localization of g in both frequency and time domains. Usually, the choice of the basis is governed by the property that we wish to be revealed by the decomposition.

The wavelet estimator can be linear or nonlinear. The linear estimator (see section 5.1.1) consists of using a kernel in the regression. In the nonlinear wavelet estimator, we decompose the empirical wavelet coefficients and keep only the components that

are significant. The negligible coefficients are put to zero.

5.1.1 Linear Wavelet smoothing

Linear wavelet smoothing consists of using a kernel for the nonparametric regression. The regression is linear in the sense that the kernel estimator is a linear function of the original data.

Antoniadis, Grégoire and McKeague (1995) suggest a wavelet orthogonal series estimator of g given by

$$\hat{g}(x) = \sum_{i=1}^{n} y_i \int_{A_i} K_J(x, y)dy,$$

where $\{A_i\}_{i=1}^n$ is a set of disjoint intervals such that $X_i \in A_i$ and

$$\bigcup_{i=1}^{n} A_i = [0, 1].$$

The parameter level J can be seen as the smoothing parameter. A very large value of J may lead to oversmoothing the unknown function g. The wavelet reproducing kernel K_j is defined by

$$K_j(x, y) = 2^j K(2^j x, 2^j y),$$

where

$$K(x, y) = \sum_k \phi(x - k)\phi(y - k).$$

The projection of g on the space V_j is

$$P_j g(x) = \int 2^j K(2^j x, 2^j y)g(y)dy.$$

Antoniadis (1996) chooses the value of J minimizing the cross validation score

$$CV(J) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{g}_{(i)}(x_i))^2,$$

where $\hat{g}_{(i)}$ is the leave-one-out estimator obtained with the i^{th} observation removed from the original data set. Another approach to wavelet linear smoothing is to regularize the residual sum of the squares with a measure of roughness, for instance, the second derivative of the unknown function g. The regression is to find g such that it minimizes the functional problem

$$S_\lambda(g) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))^2 + \lambda \int_0^1 g''(x)dx, \tag{5.1}$$

where λ is the smoothing parameter. In the case where we replace g'' by $g^{(m)}$, the solution is explicitly given in Kimeldorf and Wahba (1970a,b); see Wahba (1990) for background and references. If $g \in B^s_{2,2}([0,1])$, Antoniadis (1996) shows that the solution of the problem (5.1) is

$$g_\lambda(x) = \sum_{k=0}^{2^{j_0}-1} c_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{n}\sum_{k=0}^{2^j-1}\hat{d}_{jk}\psi_{jk}(x),$$

where $\hat{d}_{jk} = \frac{d_{jk}}{1+\lambda 2^{2sj}}$. To choose the smoothing parameter λ, Amato and Vuza (1997) use generalized cross validation, while the selection rule of Antoniadis (1996) is based on risk minimization and the knowledge of the noise variance σ^2_ε.

5.1.2 Nonlinear Wavelet smoothing

It has been shown by many researchers that linear wavelet smoothing cannot handle functions with varying degrees of smoothness. To remedy this deficiency, Donoho and Johnstone (1994,1995,1998) and Donoho et al. (1995) introduced a nonlinear wavelet estimator based on thresholding rules. They suggest to include only the largest detail coefficients in the wavelet estimator.
The most common thresholding methods are the *hard* threshold

$$\delta^H_\lambda(d_{ij}) = \begin{cases} 0, & \text{if } |d_{ij}| \le \lambda, \\ d_{ij}, & \text{if } |d_{ij}| > \lambda. \end{cases}$$

and the soft threshold

$$\delta_\lambda^S(d_{ij}) = \begin{cases} 0, & \text{if } |d_{ij}| \leq \lambda, \\ d_{ij} + \lambda, & \text{if } d_{ij} < -\lambda, \\ d_{ij} - \lambda, & \text{if } d_{ij} > \lambda. \end{cases}$$

Bruce and Gao (1996) demonstrate that due to its discontinuity, hard thresholding

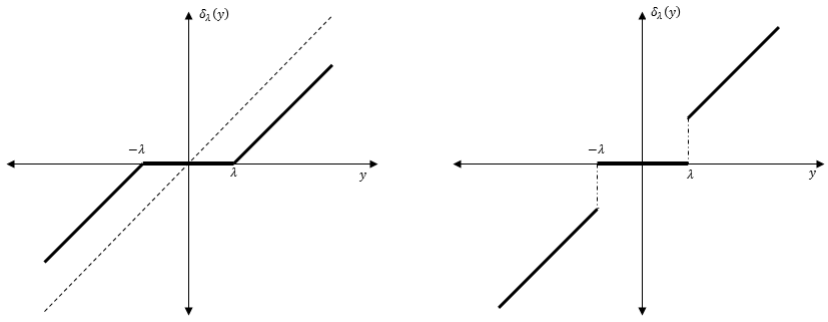


Figure 5.1: Left: soft thresholding at λ. Right: hard thresholding.

produces larger variance and it is unstable and sensitive to changes to the data while soft thresholding shifts the estimated coefficients by λ, creating large bias when the coefficients are large. To remedy these drawbacks, Bruce and Gao (1996) proposed semisoft or firm thresholding which inherits the properties of both hard and soft thresholding.

$$\delta_{\lambda_1,\lambda_2}^F(d_{ij}) = \begin{cases} 0, & \text{if } |d_{ij}| \leq \lambda_1, \\ \operatorname{sgn}(d_{ij}) \frac{\lambda_2(|d_{ij}| - \lambda_1)}{\lambda_2 - \lambda_1}, & \text{if } \lambda_1 < |d_{ij}| < \lambda_2, \\ d_{ij} - \lambda, & \text{if } |d_{ij}| > \lambda_2. \end{cases}$$

From the definition above, we notice that firm thresholding requires two threshold values, hence potentially making the estimation more difficult. Gao (1998) proposed the *nonnegative garrote* with one threshold and comparable to the *firm* thresholding rules

$$\delta_\lambda^G(d_{ij}) = \begin{cases} 0, & \text{if } |d_{ij}| \leq \lambda, \\ d_{ij} - \frac{\lambda^2}{d_{ij}}, & \text{if } |d_{ij}| > \lambda. \end{cases}$$

Fan and Li (2001) proposed the SCAD (Smoothly Clipped Absolute Deviation) thresholding given by

$$\delta_\lambda^{SCAD}(d_{ij}) = \begin{cases} \text{sgn}(d_{ij})\max(0, |d_{ij}| - \lambda), & \text{if } |d_{ij}| \leq 2\lambda, \\ \frac{(\alpha-1)d_{ij} - \alpha\lambda\text{sgn}(d_{ij})}{\alpha-2}, & \text{if } 2\lambda < |d_{ij}| \leq \alpha\lambda, \\ d_{ij}, & \text{if } |d_{ij}| > \alpha\lambda. \end{cases} \tag{5.2}$$

The SCAD estimator can be seen as a combination of hard thresholding for large

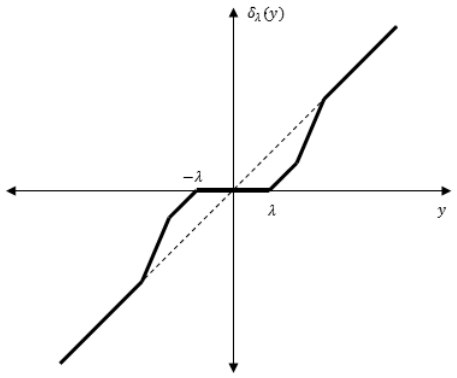


Figure 5.2: SCAD Estimator

values of d and soft thresholding for small d, with a piecewise linear interpolation in between. Antoniadis and Fan (2001) have recommended to use the value of $\alpha = 3.75$ based on Bayesian argument.

The choice of the thresholding parameter λ is of great importance in the estimation procedure. There are various methods for choosing the threshold value λ. There are two types of thresholds: the global threshold, which is applied to all wavelet coefficients, and the level-dependent threshold, where different values of λ are chosen for each resolution level j. The choice of the threshold type depends on the regression problem. Johnstone and Silverman (1997) demonstrate that the level dependent ap-

proach outperforms the global approach when the noise in the data is stationary and correlated.

- **The minimax threshold**

Donoho and Johnstone (1994) proposed a threshold based on the minimax approach given by

$$\lambda^M = \sigma_\epsilon \lambda_N^*,$$

where N is the sample size, and λ_N^* is the value satisfying

$$\lambda_*^N = \inf_\lambda \sup_d \left\{ \frac{R_\lambda(d)}{N^{-1} + R_{oracle}(d)} \right\},$$

with $R_\lambda(d) = \mathbb{E}(\delta_\lambda(d) - d)^2$ and $R_{oracle}(d)$ is the optimal risk obtained by an oracle. Donoho and Johnstone (1994) proposed two oracles : *diagonal linear projection* (DLP), which involves a linear operator and decides when to "keep" or "kill" the wavelet coefficients, and *diagonal linear shrinker* (DLS), defining how much to shrink the coefficients.

- **The universal threshold**

Donoho and Johnstone (1994) suggested another thresholding strategy based on the minimax approach using the universal threshold

$$\lambda^U = \sigma_d \sqrt{2 \ln N}, \tag{5.3}$$

where the parameter σ_d represents the standard deviation of the empirical wavelet coefficients. Only the detail coefficients at fine scales are used to estimate this parameter, since the fine scales contain mainly pure noise. Donoho and Johnstone (1995) suggested using the median absolute deviation from the median (MAD) of d_{J-1}, the wavelet coefficients associated to the resolution level $J-1$,

$$\sigma_d = \frac{\text{median}(|d_{J-1} - \text{median}(d_{J-1})|)}{0.6745},$$

The universal threshold ensures with high probability that the data does not contain the noise after the thresholding. This is because, if we consider the random variables $X_1, ..., X_n$ are independent and independent distributed $N(0, 1)$, then

$$P\left((\max_{1\leq i\leq n}|X_i| > \sqrt{c\log n}\right) \sim \frac{\sqrt{2}}{n^{c/2-1}\sqrt{c\pi\log n}}, \quad \text{as } n \to \infty.$$

For $c = 2$, we have

$$P\left((\max_{1\leq i\leq n}|X_i| > \sqrt{2\log n}\right) \sim \frac{1}{\sqrt{\pi\log n}}, \quad \text{as } n \to \infty.$$

Therefore, one can define the universal threshold (5.3). The universal threshold sets

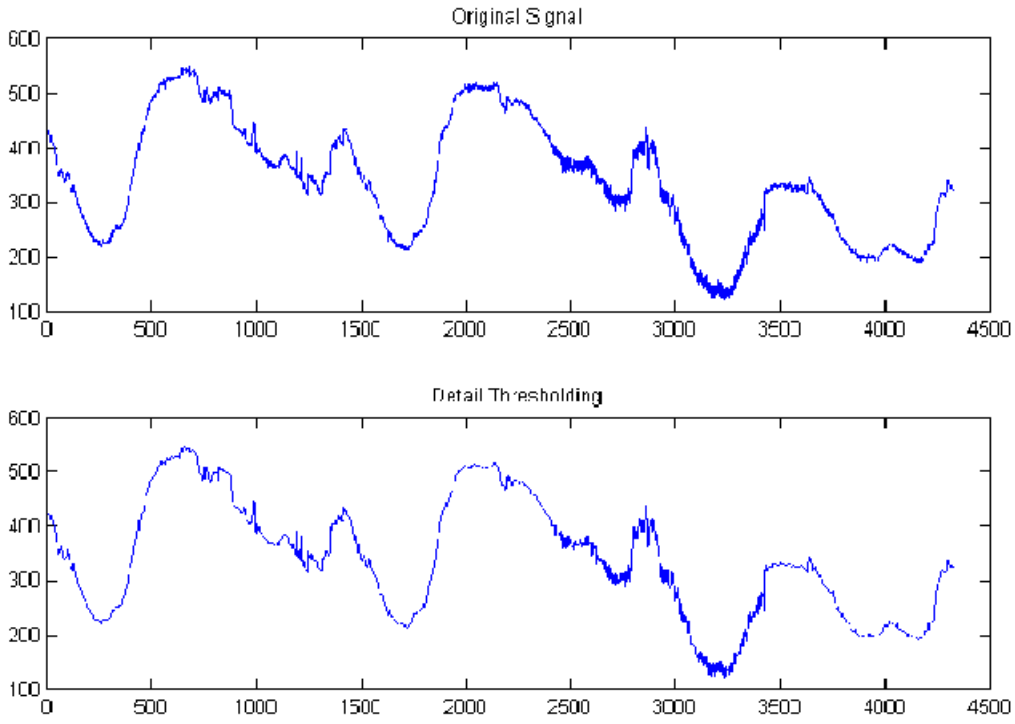


Figure 5.3: Denoising using mutli-level decomposition

most of the coefficients and leaves the coefficients that allow to estimate the signal intact.

Figure (5.3) shows that the hard thresholding 83.40% of the wavelet coefficients were killed to zero, but the 99.99% of the energy in the signal was retained.

- **The SureShrink threshold**

Donoho and Johnstone (1995) proposed the SureShrink, which consists in choosing a threshold for the wavelet coefficients for each resolution level. This threshold is based on minimizing Stein's unbiased estimator of l_2-risk (Stein (1981)). SureShrink threshold λ_j^S is defined by

$$\lambda_d^S = \text{argmin}_{0\leq\lambda\leq\lambda^u} SURE(d_{jk}, \lambda_j),$$

where

$$SURE(d_{jk}, \lambda_j) = 2^j - 2\sum_{k=0}^{2^j-1} 1_{\{|d_{jk}|\leq\lambda_j\}} + \sum_{k=0}^{2^j-1}(|d_{jk}| \wedge \lambda_j)^2.$$

Other thresholding and shrinkage procedures have been proposed in the literature. Antoniadis and Fan (2001) developed a threshold based on a penalized least-squares approach. They proved that different penalty functions correspond to different type of thresholding.

The wavelet coefficients obtained by the above different thresholding rules are used to the reconstruct the function g. The reconstructed function is then

$$\hat{g}_\lambda(x) = \sum_{k=0}^{2^{j_0}-1} c_{j_0k}\phi_{j_0k}(x) + \sum_{j=j_0}^{J}\sum_{k=0}^{2^j-1} \delta\hat{d}_{jk}\psi_{jk}(x).$$

In practice, the resulting function is obtained by simply performing the inverse discrete wavelet transform to $(c_{j_0k}, \delta\hat{d}_{jk})$. Therefore, the reconstruction procedure can be summarized as follows

$$y \overset{\text{DWT}}{\longrightarrow} (c_{j_0k}, \hat{d}_{jk}) \overset{\text{thresholding}}{\longrightarrow} (c_{j_0k}, \delta\hat{d}_{jk}) \overset{\text{IDWT}}{\longrightarrow} \hat{g}_\lambda.$$

Figure (5.4) shows that a significant denoising occurs with the level-4 decomposition coefficient using Daubechies 4 wavelets.

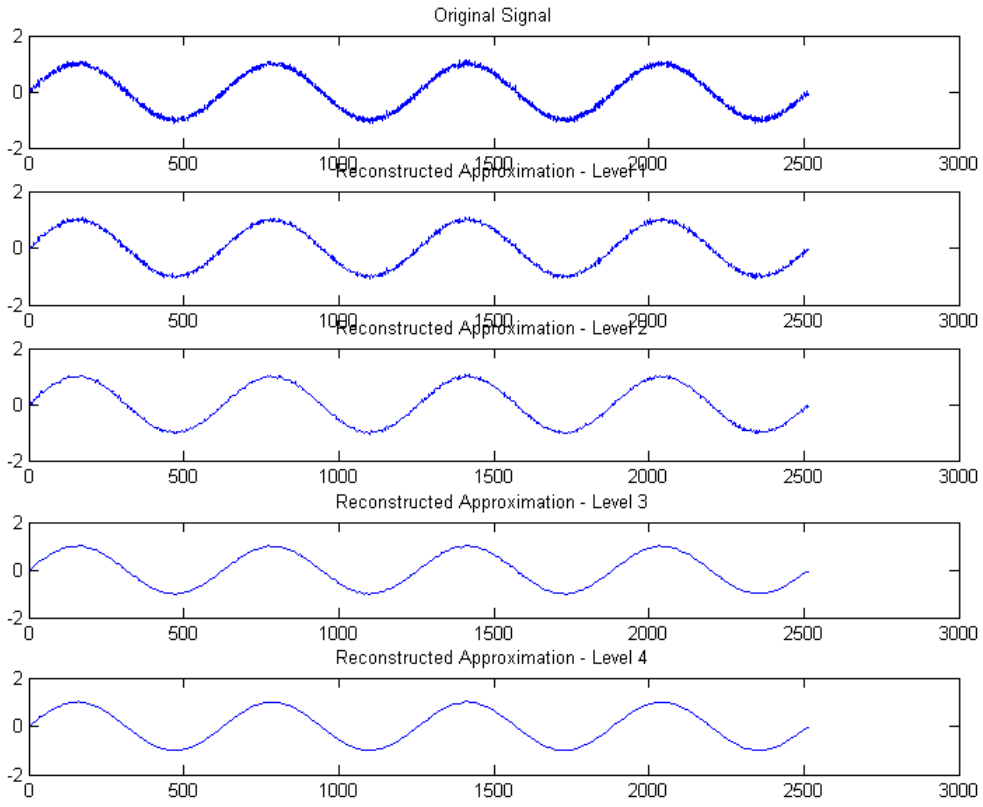


Figure 5.4: Denoising using mutli-level decomposition

5.1.3 Existing Wavelet Prior models

In the Bayesian framework, several wavelet probabilistic models have been studied. Since Mallat (1989) suggested the use of the Generalized Gaussian Distribution (GGD) for image compression, there have been several attempts to model the statistical properties of wavelet coefficients using GGD priors; see Moulin and Liu (1998) and Sendur and Selesnick (2002).

In the real world, the variance of the wavelet coefficients tends to decay exponentially across scale. This exponential decay property is related to the signal regularity, thus to the theory of Besov spaces. The connection between wavelet probabilistic models and Besov spaces was first studied in the paper of Abramovich et al. (1998) in their Gaussian mixture model.

Before we introduce the wavelet tempered stable prior model and its connection to the

Besov spaces, we give an overview on the existing prior models for wavelet coefficients.

5.1.4 Prior models

The Bayesian wavelet approach proceeds by assigning a prior distribution to the wavelet coefficients to capture their sparsity. A first approach based on the Gaussian prior was proposed by Chipman et al.(1997). They assume that the distribution of d_{jk} given s_{jk} is $N(s_{jk}, \sigma^2)$ and the coefficients s_{jk} are mutually independent. Each s_{jk} is a mixture of two Gaussians,

$$s_{jk}|\gamma_j \sim \gamma_j N(0, c_j^2\tau_j^2) + (1-\gamma_j)N(0, \tau_j^2),$$

where γ_j are Bernoulli random variables with $P(\gamma_j = 1) = p_j$. The parameters τ_j, c_j, p_j have to be chosen and are the same for all coefficients for a given level j. The small wavelet coefficients are then described by $N(0, \tau_j^2)$, and the large coefficients are modeled by $N(0, c_j^2\tau_j^2)$. The parameter p_j can be seen as the proportion of the non-negligible coefficients at each level.
Under squared error loss, the posterior mean of s_{jk} is

$$\hat{s}_{jk} = d_{jk}\left(\frac{c_j^2\tau_j^2}{\sigma^2 + c_j^2\tau_j^2}P(\gamma_j = 1|d_{jk}) + \frac{\tau_j^2}{\sigma^2+\tau_j^2}P(\gamma_j = 0|d_{jk})\right), \tag{5.4}$$

where

$$P(\gamma_j = 1|d_{jk}) = \frac{p_j\pi(d_{jk}|\gamma_j = 1)}{p_j\pi(d_{jk}|\gamma_j = 1) + (1-p_j)\pi(d_{jk}|\gamma_j = 0)},$$

and $\pi(d_{jk}|\gamma_j = 1) \sim N(0, \sigma^2 + c_j^2\tau_j^2)$, and $\pi(d_{jk}|\gamma_j = 0) \sim N(0, \sigma^2 + \tau_j^2)$.
This Bayesian approach is an adaptive shrinkage rule. The shrinkage function can be seen as an interpolation between two linear functions with slopes $\frac{c_j^2\tau_j^2}{\sigma^2+c_j^2\tau_j^2}$ and $\frac{\tau_j^2}{\sigma^2+\tau_j^2}$.
Clyde, Parmigiani and Vidakovic (1998) proposed the prior on the s coefficients to

be a mixture of a Gaussian distribution and a point mass at 0. That is,

$$\begin{aligned} s_{jk}|\gamma_j, \sigma^2 &\sim \gamma_j N(0, c_j\tau_j^2) + (1-\gamma_j)\delta(0), \\ \gamma_j &\sim Bernoulli(p_j), \\ \eta\nu/\sigma^2 &\sim \chi^2_\nu. \end{aligned}$$

The posteriori mean of $s|\gamma$ is

$$\mathbb{E}(s|d,\gamma) = \Gamma(I_n + C^{-1})^{-1}d,$$

$\Gamma = (\text{diag}\{\gamma_j\})_j$ and $C = (\text{diag}\{c_j\})_j$ Under squared error loss, the posterior mean is

$$\mathbb{E}(s|d) = \sum_\gamma \pi(\gamma|d)\Gamma(I_n + C^{-1})^{-1}d,$$

where $\pi(\gamma|d) = q_\gamma / \sum_{\gamma'} q_{\gamma'}$ and

$$q_\gamma = \prod_{i=1}^{N} \left(\frac{p_i}{1+p_i}(1+c_i)^{-1/2}\right)^{\gamma_i} \left(\lambda\nu + y'y - \sum_{i=1}^{N} \gamma_i d_i^2/(1+c_i^{-1})\right)^{(N+\nu)/2},$$

which corresponds to a nonlinear shrinkage of the coefficients towards 0.
Abramovich, Sapatinas and Silverman (1998) suggested a mixture of a Gaussian distribution and a Dirac distribution with mass at zero:

$$s_{jk} \sim \pi_j N(0, \tau_j^2) + (1-\pi_j)\delta(0),$$

where $0 \leq \pi_j \leq 1$, $\delta(0)$ is the point mass at zero and the d_{ij}'s are independent. Abramovich et al. (1998) show that the realizations of their model can be shown to belong to certain classes of Besov spaces and choose the parameters τ_j^2 and π_j to decrease exponentially:

$$\tau_j^2 = c_1 2^{-aj}, \quad \pi_j = \min(1, c_2 2^{-bj}),$$

with c_1, c_2,$a > 0$ and $b \geq 0$. They stated that a function g belongs to $B^s_{p,q}$ if the parameters a, b are chosen to satisfy

$$s + 1/2 - b/p - a/2 \leq 0,$$

with equality when $q = \infty$ and $1 \leq p < \infty$.
Abramovich et al. (1998) showed that given σ, the posterior median is

$$\text{median}(s_{jk}) = \text{sgn}(d_{jk})\max(0, \zeta_{jk}),$$

where

$$\zeta_{jk} = \frac{\tau_j^2}{\sigma^2 + \tau_j^2}|d_{jk}| - \frac{\tau_j \sigma}{\sqrt{\sigma^2 + \tau_j^2}}\Phi^{-1}\left(\frac{1 + \min(w_{jk}, 1)}{2}\right),$$

$$w_{jk} = \frac{1 - \pi_j}{\pi_j}\frac{\sqrt{\sigma^2 + \tau_j^2}}{\sigma}\exp\left(-\frac{\tau_j^2 d_{jk}^2}{2\sigma^2\sigma^2 + \tau_j^2}\right),$$

and Φ^{-1} is the inverse of the standard normal distribution function.
While a and b are well defined and connected to the Besov space parameters, the hyperparameters c_1 and c_2 have no good interpretation and their estimation is difficult. Rivoirard (2004) proposed a priori of the wavelet coefficients based on Pareto distributions. He suggested that s_{ij} are independent and modeled as follows:

$$s_{jk} \sim F_j,$$

where

- $F_j = (F_j^+ + F_j^-)/2$,
- F_j^- is the reflection of F_j^+ about 0,
- F_j^+ is the distribution of $\min(\alpha_j X_j - \alpha_j, \mu_j)$, and X_j is a Pareto random variable with parameter p,

- α_j and μ_j are nonnegative real values.

Moreover, Rivoivard (2004) suppose that the coefficient α_j takes the form

$$\alpha_j = C2^{-j\delta},$$

where C and δ_j are positive constants, and $\mu_j = \sqrt{\max(M^2 - 2\log(\alpha_j^p))}$ for a large positive constant M.

As already noted, Mallat (1989) was the first to propose the GGD wavelet model in compression. Subsequently, several estimators adopted the GGD prior to model the statistics of the wavelet coefficients; see, for example, Simoncelli and Adelson (1996), Moulin and Liu (1999) and Chang et al.(2000a, 2000b).

The GGD is defined by

$$f(x) = \frac{\nu\eta(\nu)}{2\sigma\Gamma(1/\nu)} \exp\left(-[\eta(\nu)|x|/\sigma]^\nu\right),$$

where $\eta(\nu) = \Gamma(3/\nu)/\Gamma(1/\nu)$, ν is the shape parameter and σ is the scale parameter. For $\nu = 2$, we obtain the Gaussian distribution and for $\nu = 1$ we have the Laplacian distribution.

The probability density function of the GGD has rapid exponential decay to infinity for decreasing values of ν. Because of this behaviour, the GGD model is unable to model the heavy tailed behaviour of the wavelet coefficients.

The GGD model with shape parameter $0.3 < \nu < 1$ provides a reasonable model for the distribution of wavelet coefficients of an image (Mallat, 1989). It can be applied to each orientation and each scale except at the coarsest scale. Moulin and Liu (1999) have shown that the MAP estimator by a priori GGD is a thresholding by contraction with a threshold equal to λ, where

$$\lambda = \frac{2-\nu}{2(1-\nu)} \left(2(1-\nu)^{1/(2-\nu)}\right) \eta(\nu)^{\nu/(2-\nu)} \sigma_\epsilon^{2/(2-\nu)} C^{-\nu/(2-\nu)},$$

Note that in the case of the Laplacian distribution, the MAP estimator is identical to a soft thresholding type estimator (Donoho and Johnstone, 1994), with threshold

$\lambda = \sqrt{2}\sigma_\epsilon^2$ where $c = \sigma_\epsilon(\log N)^{-1/2}$. When ν is small tending to 0, the estimator is a hard thresholding type estimator (Donoho and Johnstone, 1994), with threshold $\lambda = \sigma_\epsilon\sqrt{6\sqrt{3}/e}\nu^{-1/2}$ where $\nu = \frac{3\sqrt{3}/e}{\log N}$. This shows that the classical wavelet thresholding estimators are a special case of this Bayesian MAP estimator. Assuming all wavelet coefficients are independent and identically distributed at each scale, we have

$$d_{jk} \sim GGD_{\nu_j}(0, \sigma_j^2),$$

where ν_j measures the tail-heaviness of the wavelet coefficients.

Baraniuk and Choi (1999) generalized the results of Abramovich et al. (1998) to the more general case of the GGD wavelet model. Assuming exponential decay of the wavelet coefficients, the GGD model they suggest takes the form

$$d_{jk} \sim GGD_\nu(0, \sigma_j^2) \quad \text{with} \quad \sigma_j = 2^{-j\beta}\sigma_0. \tag{5.5}$$

Baraniuk and Choi (1999) gave a necessary and sufficient condition for realizations of the GGD model in (5.5) to belong to a certain class of Besov spaces. We have the following theorem (Theorem 1, Baraniuk and Choi (1999)):

Theorem 31. *Suppose each wavelet coefficient d_{jk} is distributed as in (5.5) with $\beta > 0$ and $\sigma_0 > 0$. Then, for $0 < p, q < \infty$, the realizations of the GGD model are almost surely in $B^s_{p,q}(I)$ if only if $\beta > s + 1$.*

Achim et al. (2001) proposed the a priori α-stable distribution to capture the heavy tailed behaviour of the wavelet coefficients' heavy tail. In Achim et al. (2001), the authors proposed a Bayesian estimator of type CEP (Conditional Expectation Posterior). However, no analytical form was provided for this Bayesian estimator except for some cases, which requires a numerical implementation of an integral, making the calculation unstable and time-consuming. The estimation of the hyperparameters of the a priori model remains a very crucial task. If different estimation procedures are available in the absence of the noise, it becomes much more difficult in its presence.

Portilla et al. (2003) proposed a Bayesian statistical model based on a priori scales

mixture of Gaussian, namely, the Gaussian scale mixture. The wavelet coefficients are modeled as follows:

$$d_{jk} \sim \sqrt{z_j} u,$$

where z_j is an independent scalar random variable, and u is a zero-mean Gaussian vector.

The statistical model proposed by Portilla et al. (2003) is effective for capturing leptokurtic behavior of the wavelet coefficients and taking into account the correlation between adjacent coefficients. The Gaussian scale mixture includes a variety of distributions, such as the α-stable distribution, the GGD and the symmetric Gamma distribution family.

Also, the use of Lévy processes and non-Gaussian distributions in Bayesian nonparametric regression has been gaining popularity; see, for instance, Johnstone and Silverman (2005), Abramovich, Benjamini, Donoho and Johnstone (2006), Caron and Doucet (2008), Griffin and Brown (2011), Wolpert, Clyde and Tu (2011) and Polson and Scott (2012).

5.2 Lévy wavelet prior and Besov spaces

Choosing a Hilbert space $L^2(\mathbb{R})$ with a frame or basis and projecting the underlying process into a subspace with lower dimension where only few components are significant facilitates obtaining a prior; see, for example, Mallat (1989), Simoncelli and Portilla (1998), Grenander and Srivastava (2001). Indeed, the probability density function of these components in such representations are far from being Gaussian. They are leptokurtic with heavy tails. In this section, we introduce the symmetric tempered stable distribution as a novel Bayesian statistical prior to model the wavelet coefficients i.e.,

$$d_{jk} \sim TS_{\sigma_j}(\alpha, c, \lambda). \tag{5.6}$$

where σ_j and (α, c, λ) are respectively the implied variance and the hyperparameters of the tempered stable distribution.

5.2.1 Tempered Stable prior and Besov spaces

Besov spaces form rich class of function spaces. The advantage of Besov spaces is that they are a much more general tool in describing the smoothness and the regularity properties of functions. In this section we demonstrate the relationship between the Besov space parameters and the hyperparameters of the tempered stable prior model. Our argument will be based on the equivalence between the Besov norm of a function and its wavelet coefficients.

First, we will show that the tempered stable distribution is characterized by correlated variables and an exponential decay of the variance across scale.

In order to determine the relationship between the Besov space parameters and the hyperparameters of the Lévy wavelet prior model, we will first show that the scaling behavior of the tempered stable distribution is translated to the second moment.

To do so, we compute the variance at time scale Δt:

$$\sigma^2_{\Delta t} = \int_{-\infty}^{+\infty} x^2 \nu(\mathrm{d}x), \tag{5.7a}$$

$$= \int_0^{+\infty} x^2 Ap \frac{e^{-G_{\Delta t}x}}{x^{1+\alpha}} \mathrm{d}x + \int_0^{+\infty} x^2 Aq \frac{e^{-M_{\Delta t}x}}{x^{1+\alpha}} \mathrm{d}x, \tag{5.7b}$$

$$= A_{\Delta t}\Gamma(2-\alpha)(pG^{\alpha-2}_{\Delta t} - qM^{\alpha-2}_{\Delta t}). \tag{5.7c}$$

Inserting the scaling transformations (4.26) in the second term of equation (5.7c), we obtain

$$\sigma^2_{\Delta t} = \Delta t^{\frac{2}{\alpha}} \sigma^2_1, \tag{5.8}$$

where $\sigma^2_1 = A_1\Gamma(2-\alpha)(pG^{\alpha-2}_1 - qM^{\alpha-2}_1)$ refers to the variance at time $\Delta t = 1$.

Using the transformations (4.26), the variance at time scale Δt is completely defined by the variance at $\Delta t = 1$.

As a consequence, if the distribution of the wavelet coefficients is a tempered stable distribution, then the wavelet coefficients decay exponentially across scale.

In the following theorem, we demonstrate the connection between the Besov spaces and the TS model (5.6):

Theorem 32. *Let the function f have a wavelet series expansion where each wavelet coefficient d_{jk} is TS distributed. Then, for $0 < p, q \leq \infty$, $f \in B^s_{p,q}(I)$ almost surely if only if $\frac{1}{\alpha} > s + \frac{1}{2}$.*

Proof. By setting the time scale to a dyadic time scale $\Delta t = 2^{-j}$ and σ_j to the variance of the signal at scale j (i.e. $\sigma_j = \sqrt{\text{var}(d_{jk})}$), we have

$$\mathbb{E}(||d_j||_p^p) = \mathbb{E}\left(\sum_{k=0}^{2^j-1} |d_{jk}|^p\right) = \sum_{k=0}^{2^j-1} \mathbb{E}(|d_{jk}|^p) = 2^{j(1-p/\alpha)}\nu_p\sigma_0^p,$$

here ν_p is the p^{th} absolute moment of the TS distribution.

We now show that $2^{-j(1-p/\alpha)}||d_j||_p^p$ converge to $\nu_p\sigma_0^p$ almost surely as $j \to \infty$. Given $\epsilon > 0$, Chebychev's inequality implies that

$$\begin{aligned}
\sum_{j=0}^{\infty} \mathbb{P}\left(|2^{-j(1-p/\alpha)}||d_j||_p^p - \nu_p\sigma_0^p|\right) > \epsilon \quad &\leq \quad \epsilon^{-2}\sum_{j=0}^{\infty} 2^{-j(1-p/\alpha)}\text{var}(||d_j||_p^p), \\
&\leq \quad \epsilon^{-2}\sum_{j=0}^{\infty} 2^{-j(1-p/\alpha)}\sum_{k=0}^{2^j-1}\mathbb{E}(|d_{jk}|^{2p}), \\
&\leq \quad \epsilon^{-2}\nu_{2p}\sigma_0^{2p}\sum_{j=0}^{\infty} 2^{-j} < \infty,
\end{aligned}$$

which is equivalent to

$$2^{-j(1-p/\alpha)}||d_j||_p^p \to \nu_p\sigma_0^p, \quad \textit{almost surely as } j \to \infty, \tag{5.9}$$

by the first Borel-Cantelli lemma.

Let d_{jk} be the wavelet coefficients of a function g in the tempered stable prior model. For a finite q, we have $\nu_p < \infty$. So, the function g belongs to $B^s_{p,q}(I)$ if and only if

$$\sum_{j=0}^{\infty} 2^{js'q}||d_j||_p^q < \infty,$$

where $s' = s + 1/2 - 1/p$.

Using the result (5.9), the series converges if and only if $1/\alpha > s + 1/2$.

For $q = \infty$, we still have $\nu_p < \infty$. Hence, the function g belongs to $B^s_{p,q}(I)$ if and only if

$$\sup_{j\geq 0}(\sum_{j=0}^{\infty} 2^{js'}||d_j||_p^p) < \infty,$$

which is equivalent to

$$\sup_{j\geq 0} \sum_{j=0}^{\infty} 2^{js'} 2^{j(1/p-1/\alpha)} < \infty,$$

which holds if and only if $1/\alpha > s + 1/2$. □

The result is close to the results of Abramovich et al.(1998). It simply shows that the Lévy wavelet prior model can model the dependence structure between the wavelet coefficients and the exponential decay across scale properties without imposing any specified structure on the variance of the wavelet coefficients. Similar results apply for a large class of wavelet coefficients models such as the Lévy sparse mixture model we define in section 6.2.

Abramovich and Silverman (2000) also develop stochastic expansions for overcomplete wavelet dictionaries with arbitrary scales and locations. They use Gaussian priors to model the wavelet coefficients and restrict the variance to decrease exponentially across scales to ensure that the stochastic expansions fall into a certain class of Besov spaces. Their model can be viewed as a special case of the LARK model. Wolpert, Clyde and Tu (2011) investigate the regularity of the random series generated by the LARK model with Gaussian priors. They prove the same regularity conditions as in Abramovich and Silverman (2000).

Our model can also be viewed as a Lévy Adaptive Regression Model model where the generator kernel in LARK is a wavelet function and the coefficients β_j have independent tempered stable distributions. To place the work of Wolpert, Clyde and Tu in a proper setting, we consider the standard regression model (1.1). LARK

assumes that the regression function has an atomic decomposition

$$g(x_i) = \sum_{0 \leq J} f(x_i, w_j)\beta_j = \int f(x_i, w_j)\mathcal{L}(dw),$$

where $f(x_i, w_j)$ are atoms at x_i and $\mathcal{L}(dw) = \sum \beta_j \delta_{w_j}(dw)$ is a random measure. Wolpert, Clyde and Tu (2011) also investigate the regularity conditions of the LARK model with symmetric α-stable priors. They give regularity conditions such that the LARK model falls into a certain class of Sobolev spaces ($B^s_{2,2}$), but for general p and q the problem remains open. Following a different approach, in Theorem 32, we give regularity conditions of the Lévy wavelet model to remain in a certain class of Besov spaces with general p and q.

One possible way to define the distribution of the wavelet coefficients $\{d_{jk}\}_{j,k\in\mathbb{Z}}$ of a stochastic process $(X_t)_{t\in\mathbb{R}}$ is to use the stochastic-integral representation of X. Pesquet-Popescu (1999) has already investigated the statistical properties of the wavelet coefficients of an α-stable process and proved the following proposition.

Proposition 33. *If X is an α-stable process, then its wavelet coefficients also form an α-stable process.*

Proof. Sketch of proof.
The wavelet coefficients of X are given by $d_{jk} = \int_{\mathbb{R}} X_t \psi_{jk}(t)dt$. Using the stochastic representation of an α-stable process defined in (4.21), we have by Fubini's theorem

$$d_{jk} = \int_E \left(\int_{\mathbb{R}} f(t, x)\psi_{jk}(t)dt \right) M(dx) = \int_E g_j(k, x)M(dx),$$

where $g_j(k, x) = \int_{\mathbb{R}} f(t, x)\psi_{jk}(t)dt$. It has been shown in Pesquet-Popescu (1999) that the function $g_j(k, .)$ belong to $\mathcal{L}^\alpha(E, \mathcal{E}, m)$. Therefore, the wavelet coefficients d_{jk} are stable integrals and form an α-stable process. □

The question is then whether we can investigate the distribution of the wavelet coefficients for a tempered stable process. To do so, we need to extend the work of Rajput and Rosinski (1989) to define the stochastic integral driven by a tempered stable random measure. The idea is to investigate whether we can prove that the

parameters of tempered stable processes can be written as function of a tempered stable random measure on $(E, \mathcal{E})$ as stated in section 4.4 for stable processes.

5.3 Summary

In this chapter we have proposed a novel approach for modeling the wavelet coefficients. This approach assumes the prior belief that the wavelet coefficient follow a Lévy wavelet prior model and Besov spaces and gave a necessary and sufficient condition such that the realizations of the prior model fall into certain class of Besov spaces.

Chapter 6

Bayesian approach

In the previous Chapter, we conducted a classification of the state of the art on the problem of denoising using wavelet methods. These methods try to find a solution that has a certain regularity. This can be imposed in different ways, and carried out using multiscale transformations in the Bayesian framework. In the Bayesian context, in order to reconstruct a signal free of noise while maintaining the details, it is necessary to impose constraints on the reconstructed solution, which amounts to choosing a prior model.
In Chapter 5, we introduced an adequate Lévy prior distribution to model the statistical properties of the sparse wavelet coefficients. In this Chapter, we will introduce various techniques to estimate the hyperparameters of the model.

6.1 A priori Model

In the Bayesian approach, a prior model is applied to the wavelet coefficients. The appropriate model should capture all the characteristics and the asymptotic behavior of the wavelet coefficients. Our Bayesian approach assumes that in the prior model the coefficients are mutually independent. We start by modeling the Lévy *a posteriori* measure of the wavelet coefficients and then estimating the hyperparameters of the tempered stable prior distribution. Like in Abramovich, Sapatinas and Silverman (1998), we prove that our Lévy sparse mixture of a tempered stable distribution and

a point mass at zero is equivalent to a type of thresholding rule.
We consider the standard univariate regression at regular points $\{t_i\}_i$:

$$y_i = g(t_i) + \epsilon_i, \qquad i = 1, ..., n, \tag{6.1}$$

with ϵ_i independent $N(0, \sigma_\epsilon^2)$ random variables. The goal is to recover the underlying function g from the noisy data y_i without assuming any particular parametric structure for g.

Wavelet Bayesian methods are applied after the data has been projected into a wavelet space. The regression problem becomes

$$d = \theta + \eta,$$

where $d = Wy$, $\theta = Wg$ and $\eta = W\epsilon$, which is also a white noise process since the wavelet transform W is orthogonal.
The a posteriori probability density of the wavelet coefficients is giving by

$$p(d) = \int \phi(d - \theta) \mathrm{d}P(\theta),$$

where $\mathrm{P}(\theta)$ is the a priori distribution of the wavelet coefficients and ϕ is the likelihood function.
In our Bayesian framework, we assume that the a priori distribution $\mathrm{P}(\theta)$ is a tempered stable distribution

$$\theta \sim TS_\sigma(\alpha, c, \lambda),$$

and

$$\phi(d - \theta) = \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \exp\left(-\frac{1}{2}(d - \theta)^2 / \sigma_\epsilon^2\right).$$

The Lévy measure of the a posteriori distribution can be computed when the a priori Lévy measure is tempered stable.

Proposition 34. *Assuming that θ has a tempered stable Lévy measure as in (4.22) and $\alpha \neq 1$. Then the measured set of the wavelet coefficients d has a Lévy measure*

ν_d given by

$$\begin{aligned}\nu_d(t) &= \frac{\exp(-at^2/2)}{\sqrt{2\pi\sigma_\epsilon^2}}(\sigma_\epsilon)^{-\alpha/2}\Gamma(-\alpha)\{Ap\exp\left((-2at+\lambda)^2\sigma_\epsilon^2/4\right)D_\alpha((-2at+\lambda)\sigma_\epsilon) \\ &+ Aq\exp\left((2at+\lambda)^2\sigma_\epsilon^2/4\right)D_\alpha((2at+\lambda)\sigma_\epsilon)\},\end{aligned}$$

where $a=\frac{1}{\sigma_\epsilon^2}$ and D_ν the parabolic cylinder function.

To establish this result, we use the integral in Gradshteyn and Ryzhik, (1980) *p.337*:

$$\int_0^{+\infty} x^{\nu-1}\exp(-\beta x^2-\gamma x)\mathrm{d}x = (2\beta)^{-\frac{\nu}{2}}\Gamma(\alpha)\exp\left(\frac{\gamma^2}{8\beta}\right)D_{-\nu}\left(\frac{\gamma}{\sqrt{(2\beta)}}\right)$$
$$(Re(\nu)>0,\ Re(\beta)>0),$$

where D_ν is the parabolic cylinder function defined by the formula 19.5.3 in Abramowitz and Stegun (1970):

$$D_\nu(x)=\frac{\exp(-\frac{x^2}{4})}{\Gamma(-\nu)}\int_0^{+\infty}\exp(-xt-t^2/2)t^{-\nu-1}\mathrm{d}t,\ for\ Re(\nu)<0. \tag{6.2}$$

Proof. Let ν_θ be the Lévy measure of θ where

$$\nu_\theta(s)=Ap\frac{e^{-\lambda|s|}}{|s|^{1+\alpha}}1_{s<0}+Aq\frac{e^{-\lambda s}}{s^{1+\alpha}}1_{s>0}.$$

The Lévy measure of d is given by

$$\nu_d(t)=\int_{-\infty}^{+\infty}\phi(t-s)\nu_\theta(s)ds$$

where

$$\phi(t-s)=\frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left(-\frac{(t-s)^2}{2\sigma_\varepsilon^2}\right).$$

The calculation of this integral is given in detail in Appendix A. □

Remark: We also verified numerically the validity of Equation (6.2) for different values of $\alpha\in(0,2)$ and $\alpha\neq 1$. For this, we used the hypergeometric representation of the parabolic cylinder function; see Abramowitz and Stegun (1970) p.695.

The C++ code implementation of the parabolic cylinder function is available at http://www.proba.jussieu.fr/pageperso/tankov/code/cgmy.cpp.

6.2 Bayesian Thresholding

Donoho and Johnstone (1994) proposed a nonlinear estimator of the underlying function based on an appropriate selection of the wavelet coefficients and demonstrated the efficiency of shrinkage methods in the wavelet domain. They suggested comparing the wavelet coefficients to an appropriately chosen parameter λ, keeping only the significant coefficients.

In a similar spirit to the prior mixture model (5.5), we add a point mass at zero to ensure a thresholding rule in the Bayesian fashion; see also Denison, Holmes, Mallick and Smith (2002). Our proposed Bayesian thresholding method postulates that the prior is a tempered stable mixture model

$$\theta_j \sim \pi_j TS_{\sigma_j}(\alpha, c, \lambda) + (1 - \pi_j)\delta(0), \tag{6.3}$$

where $0 \leq \pi_j \leq 1$, $\delta(0)$ is point mass at zero and the $\theta_{j,k}$'s are independent. π_j represents the proportion of the thresholded wavelet coefficients containing 'signal'. Abramovich and Sapatinas and Silverman (1998) assumes that π_j is of the form $\pi_j = \min(1, 2^{-\beta j}C)$ where $\beta, C \geq 0$. In the case of self-similar processes, the overall regularity depends on the self-similar parameter, which corresponds to the parameter α in the Lévy wavelet prior model. That is, the probability π does not depend on the resolution level j, which means that the proportion of non-zero wavelet coefficients is the same at each level. That said, the tempered stable wavelet prior mixture model will then take the form

$$\theta_j \sim \pi TS_{\sigma_j}(\alpha, c, \lambda) + (1 - \pi)\delta(0), \tag{6.4}$$

where $0 \leq \pi \leq 1$.

Each d_{jk} in the model (6.4) is either 0 with probability $1 - \pi$ or is tempered stable distributed with zero mean and variance σ_j with probability π.

For a given set of coefficients d, we construct a set of $(n-k)$ nonzero values $\overline{d}$ using the hard thresholding rule $\overline{d} = d1_{|d|>\lambda}$ with $\lambda = \sigma_\varepsilon\sqrt{2\ln N}$. Throughout our discussion, we have assumed that σ_ϵ is known. In practice, σ_ε could be estimated using $\sigma_\epsilon = \text{MAD}(d)/0.6745$, where MAD is the median absolute deviation.

The maximum likelihood approach can be used to estimate the parameters of the mixture prior model. The likelihood function is given by

$$\text{L}(\overline{d}, \psi; \pi) \sim (1-\pi)^k \pi^{n-k} \prod_{i=1}^{n-k} TS(\overline{d}_i; \psi),$$

where ψ is the vector of the hyperparameters of the tempered stable distribution.
It is straightforward to verify that the function $h(\pi) = (1-\pi)^k\pi^{n-k}$ is maximized at $\pi_{max} = (n-k)/n$. So we can write the optimal cumulative distribution function of θ as

$$F(x) = \frac{n-k}{n} TS(x; \psi_{max}) + \frac{k}{n}\epsilon(x), \tag{6.5}$$

$\epsilon(x) = 1$ if $x > 0$ and $\epsilon(x) = 0$ if $x \leq 0$ and ψ_{max} is the vector of the hyperparameters estimated with the nonzero coefficients.
Given a prior distribution ν and observed data d, the posterior risk is

$$\mathbb{E}_d L(a, \theta) = \int L(a, \theta)\nu(d\theta|d),$$

where $L(a, \theta)$ is the *loss function* in which $a \in \mathbb{R}^n$ denotes an estimate and $\theta \in \mathbb{R}^n$ denotes the true parameter value, and the Bayes estimator is defined by

$$\hat{\theta}_\nu(d) = \text{argmin}_a \mathbb{E}_d L(a, \theta).$$

If we consider that L is a q-power function of the form $L_q(a, \theta) = |a-\theta|^q$, the Bayesian estimator takes the form

$$\hat{\theta}_\nu(d) = \text{argmin}_a \int |a-\theta|^q \nu(d\theta|d). \tag{6.6}$$

The sparse mixture prior model (6.4) can be written as

$$\nu(d\mu) = \pi\gamma(\mu)d\mu + (1-\pi)\delta_0(d\mu), \tag{6.7}$$

where $\gamma(\mu)$ is tempered stable symmetric and unimodal. Here, the unimodality means that $\gamma(\mu)$ is decreasing in μ for $\mu \geq 0$. This model can be viewed as a special case of the model considered by Johnstone and Silverman (2005).
The following proposition shows that the Lévy sparse mixture prior (6.7) is a thresholding rule.

Proposition 35. *Suppose that the prior has mixture form (6.7) for $\pi > 0$. The posterior median $\hat{\theta}_\nu(d)$ ($q = 2$ in (6.6)) is a threshold rule: there exists $t(\pi) > 0$ such that $\hat{\theta}_\nu(d) = 0$ if and only if $|d| \leq t(\pi)$.*

Proof. The proof mimics the proof of the Proposition 2.1 section 2.3 in Johnstone (2011). The marginal density is

$$p(d) \;=\; \int \phi(d-s)\nu(ds) = (1-\pi)\phi(d) + \phi g(d),$$

where the

$$g(d) = \phi * \gamma(d) = \int \phi(d-s)\gamma(s)ds.$$

The prior and the posterior densities are given respectively by

$$\nu(s) = \begin{cases} 1-\pi, & \text{if } s = 0 \\ \pi\nu(\mu), & \text{if } s \neq 0, \end{cases}$$

and

$$\nu(s|d) = \nu(s)\phi(d-s)/p(d) \tag{6.8a}$$

$$= \begin{cases} (1-\pi)\phi(d)/p(d), & \text{if } s = 0 \\ \pi\nu(s)\phi(d-s)/p(d), & \text{if } s \neq 0, \end{cases} \tag{6.8b}$$

From the posterior density (6.8) we have that $\mathrm{P}(s = 0|d) > 0$ if the non-zero weight $\pi < 1$. In other words, the existence of an atom at 0 in the prior implies that there is a corresponding atom at 0 in the posterior. Consequently, by symmetry

$$\mathrm{P}(s < 0|d = 0) = \mathrm{P}(s > 0|d = 0) < 1/2,$$

and

$$\mathrm{P}(s < 0|d = 0) < \frac{1}{2} < \mathrm{P}(s \leq 0|d = 0),$$

so we have $\hat{s}_\nu(0) = 0$.
From (6.8), we have $x \to \mathrm{P}(s < 0|d)$ and $x \to \mathrm{P}(s \leq 0|d)$ are continuous and strictly decreasing functions, and so there exists $-t(\pi) > 0$ such that

$$-t(\pi) < d < t(\pi) \Rightarrow \hat{s}_\nu(d) = 0.$$

□

In our Bayesian thresholding approach, rather than using the universal threshold to estimate the wavelet coefficients, we can consider a threshold value based on the properties of the stable distribution. This would lead to the construction of an appropriate thresholding rule and we can then measure its performance against the other thresholding procedures introduced in section 5.1.

6.3 Estimation of the hyperparameters of Tempered Stable Distributions

In the Bayesian approach, it is necessary to estimate the hyperparameters of the *a priori* distribution. In the wavelet domain, the hyperparameters are estimated from the observable wavelet coefficients. There are several estimation methods: maximun likelihood, EM-Algorithm, method of moments and cumulants, quantile methods, etc. The α-stable distributions do not have finite moments of order greater or equal to α, but for raw data we can calculate empirical moments of any order. The tempered stable distributions are a reasonable model for fitting real data having a finite second

and higher moments and they show faster decay than predicted by α-stable distributions.

This section is devoted to the estimation of the hyperparameters of the tempered stable distributions. We will state how to estimate the parameters of the symmetric tempered distribution using the method of moments of Zolotarev (1986). Then, we provide an algorithm for the Maximum Likelihood Estimation (MLE) for the symmetric TS distributions using the Midpoint Rule and the Discrete Fourier transform, a method introduced by Menn and Rachev (2004). The method of moments is used to find the initial seed of the hyperparameters in the maximization algorithm.

6.3.1 Method of Moments

In this section, we shall outline the method of moments to estimate the hyperparameters of the stable distribution which has been introduced by Zolotarev (1986). Let $X = (X_1, ..., X_n)$ be a set of independent and identically distributed stable random variables defined by their characteristic function given in (4.16), we transform X to a new set of random variables such that all its moments are well-defined and can be computed. We define

$$U_i = \text{sign}(X_i), \quad V_i = \log|X_1|, \ i = 1, ..., n. \tag{6.9}$$

Using the equalities established in Zolotarev (1986),

$$\nu = \frac{6}{\pi}\text{Var}(V) - \frac{3}{2}\text{Var}(U) + 1, \quad \theta = \mathbb{E}(U), \quad \tau = \mathbb{E}(V), \quad \nu = \frac{1}{\alpha^2}, \tag{6.10}$$

we define an unbiased estimator of ν:

$$\hat{\nu} = \frac{6}{\pi}S_V^2 - \frac{3}{2}S_U^2 + 1, \tag{6.11}$$

where

$$S_U^2 = \frac{\sum_{i=1}^{n}(U_i - \overline{U})}{n-1} \to \mathbb{E}(U^2) - \mathbb{E}(U)^2 = \text{Var}(U), \tag{6.12}$$

and

$$S_V^2 = \frac{\sum_{i=1}^{n}(V_i - \overline{V})}{n-1} \to \mathbb{E}(V^2) - \mathbb{E}(V)^2 = \mathrm{Var}(V). \tag{6.13}$$

Using the equality (6.10), we have

$$\widehat{\alpha} = \frac{1}{\sqrt{\widehat{\nu}}}. \tag{6.14}$$

We can now give an estimator of the scale parameter σ using (6.10) and (4.17d)

$$\widehat{\sigma} = \frac{1}{2}\exp(\widehat{\tau}/\sqrt{(}\widehat{\nu} - \gamma(1 - 1/\sqrt{(}\widehat{\nu})))), \quad if\ \nu \neq 1, \tag{6.15}$$

where

$$\widehat{\nu} = \frac{1}{n}\sum_{i=1}^{n} V_i. \tag{6.16}$$

6.3.2 Estimation of the parameters for symmetric Tempered Stable distribution

We now consider the symmetric tempered stable distributions. Using Zolotarev's representation, a biased estimator is given by

$$\widehat{a} = \frac{\exp[\widehat{\alpha}(\overline{V} + \gamma) - \gamma]}{-\cos(\alpha\pi/2)\Gamma(-\widehat{\alpha})}, \tag{6.17}$$

where $\overline{V} = \frac{\sum_{i=1}^{n}\log(|x_i|)}{n}$.
Once $\widehat{a}$ is known, we can give a biased estimator of λ

$$\widehat{\lambda} = \left(\frac{\mathrm{Var}(X)}{\widehat{a}\Gamma(2-\widehat{\alpha})}\right)^{1/(\widehat{\alpha}-2)}. \tag{6.18}$$

6.3.3 Maximum Likelihood estimators for symmetric Tempered Stable distributions.

To estimate the hyperparameters of the tempered stable distribution, we use maximum likelihood estimation. We need to compute the likelihood function for *i.i.d* tempered stable random variables

$$f(x;\alpha,a,\lambda)=\prod_{i=1}^{n}f(x_i;\alpha,a,\lambda). \tag{6.19}$$

where f is the density function of the tempered stable distribution. Tempered stable random variables are defined by their characteristic function $\varphi(u)=e^{\Psi(u)}$, namely

$$\Psi(u) \;=\; a\Gamma(-\alpha)[(\lambda^2+u^2)^{\alpha/2}\cos(\alpha\arctan(u/\lambda))-\lambda^{\alpha}], \tag{6.20}$$

and we can compute the density function by applying a Fourier transform to the characteristic function

$$f(x)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{-iux}\varphi(u)du\approx\frac{1}{2\pi}\int_{-c}^{c}e^{-iux}\varphi(u)du \tag{6.21}$$

where c is unknown. In the case of fixed c, the later integral can be computed using a quadrature formula. The simplest integral approximations are the rectangular rules, and since the "midpoint rule" is computationally efficient, Menn and Rachev (2004) derived a density approximation using this method. Here, we will follow the same approach. We provide the connection between the integral by the Midpoint Rule and the Discrete Fourier Transform (DFT).

Recall that the Midpoint Rule calls for a partition of the interval $[-c,c]$ by the grid $-c=c_0<c_1<...<c_n=c$ and the following calculation:

$$\int_{-c}^{c}g(x)dx=\sum_{k=1}^{n}\int_{c_{k-1}}^{c_k}g(x)dx\approx\sum_{k=1}^{n}g(x_k)(c_k-c_{k-1}), \tag{6.22}$$

where $x_k=(c_k-c_{k-1})/2$.

The DFT is a mapping of a vector $y=(y_0,...,y_{N-1})\in\mathbb{C}^N$ onto a vector $z=$

$(z_0, ..., z_{N-1}) \in \mathbb{C}^N$ such that

$$z_k = \mathrm{DFT}(y)_k = \sum_{j=0}^{N-1} y_j \mathrm{e}^{-ik2\pi j/N}$$

Let $N = 2^m$, $c \in \mathbb{R}_+$, and $h = 2c/N$. Furthermore, define an equidistant grid over the interval $[-c, c]$ via $t_j = -c + jh$, $j = 0, ..., N$. The midpoints are given by $t_j^* = (t_j + t_{j+1})$, $j = 0, ..., N-1$.
Define the vectors $y \in \mathbb{C}^N$ and $C \in \mathbb{C}^N$ by

$$\begin{aligned} x_k &= -\frac{N\pi}{2c} + \frac{\pi}{a}k, \qquad k = 0, ..., N-1, && (6.23)\\ y_j &= (-1)^j \varphi(t_j^*), \qquad j = 0, ..., N-1, && (6.24)\\ C_k &= (-1)^k i \frac{2c}{N} e^{-i\frac{\pi k}{N}}, \qquad k = 0, ..., N-1. && (6.25) \end{aligned}$$

We have

$$\sum_{k=1}^{N-1} \varphi(t_j^*) e^{-it_j^* x_k}(t_{j+1} - t_j) = C_k . DFT(y)_k, \tag{6.26}$$

So application of the "Midpoint Rule" in the evaluation of the integral can be realized by using the Fast Fourier Transform algorithm.
One way to select c is to estimate the parameters $\widehat{\alpha}$, $\widehat{a}$ and $\widehat{\lambda}$ using (6.16), (6.17) and (6.18), simulate N tempered stable random variables and compute the density from these simulated values. Then, we compute the values of the density on the grid $(x_k)_k$ defined in (6.23) using (6.21). We choose c such that the L_1 distance between these values and the true density from the simulated random variables does not exceed 10^{-6}.

6.4 MLE Results for Parameter Estimation of Tempered Stable distributions

As already discussed, the tempered stable distributions provide us with a suitable model distribution for fitting real data with a finite second and higher moments and they show faster decay than predicted by α-stable distributions. This section describes a simulation study to verify the estimation method we introduced in section 6.3. For simplicity, we will consider the symmetric tempered stable distribution $TS(\alpha, c, \lambda)$ (i.e. $p = 1/2$, $q = 1/2$) and we estimate σ_ϵ using the universal estimator $\sigma_\epsilon = \frac{\text{MAD}(d)}{0.6745}$. We estimate the initial seed of the parameters using the method of moments introduced by Zolotarev (1986). This can be done since Rosiński (2007) proved that the short-term increments of the tempered stable process behave like a stable process; refer to section 4.6. Also, we observe that the characteristic function of d is given by the product of the characteristic functions of θ and η

$$\Phi_d(w) = \Phi_\theta(w)\Phi_\eta(w)$$

where

$$\Phi_\eta(w) = \exp(-\sigma_\varepsilon^2 w^2/2)$$

and

$$\Phi_\theta(w) = \exp(-\Psi(w)),$$

where Ψ is defined in 4.24. Then, we perform a Maximum Likelihood Estimation (MLE) using the Midpoint Rule and the Discrete Fourier Transform introduced by Menn and Rachef (2004) introduced in section 6.3.3.

We simulated the tempered stable process with different time intervals $\Delta t = 1; 0.5; 0.25$. Using the transformations (5.8), we notice that the scaling properties on the standard deviation are also verified empirically; see Table (6.1). The standard deviation computed from the realizations of the simulated process approximates the standard deviation calculate out of the formula (4.26). This confirms that the distributions of returns for various choices of Δt have similar functionals forms.

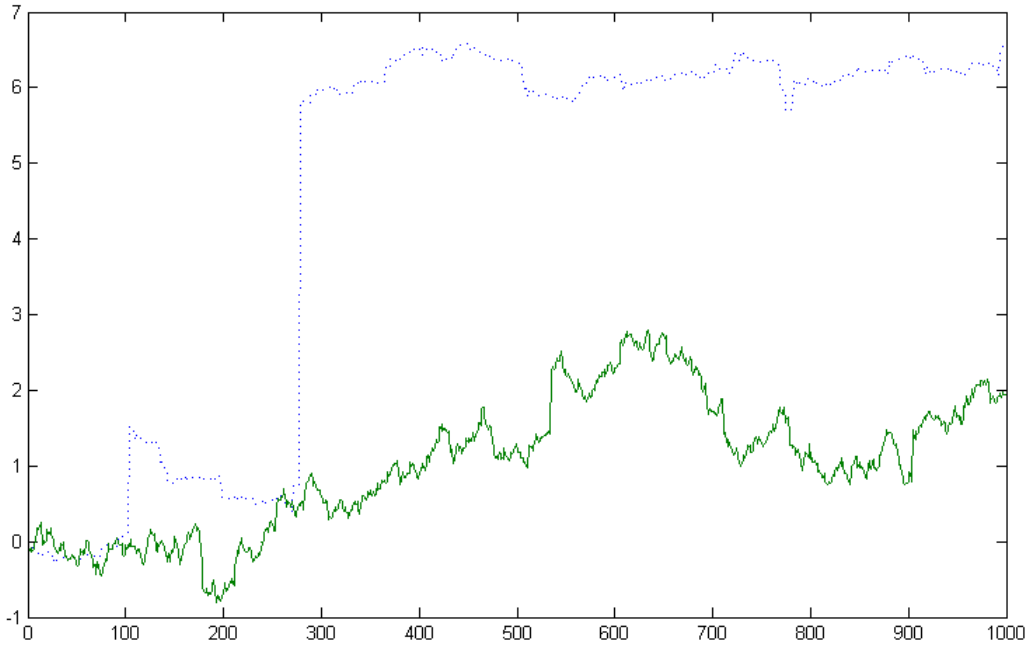


Figure 6.1: Sample paths of a stable TS(1.7,1,0) (Dot line) and a tempered stable TS(1.7,1,0.5) (solid line)

Nb. Sim	Δt	α	λ	c	σ	$\sigma_{\Delta t}$
1024	1	1.5	0.1000	1	**0.6722**	**0.6722**
512	0.5	1.5	0.1587	0.5	**0.4211**	**0.4235**
256	0.25	1.5	0.2520	0.25	**0.2190**	**0.2660**

Table 6.1: Estimation of the parameters. $\sigma_{\Delta t} = (\Delta t)^{1/\alpha}\sigma_{\Delta t=1}$ for different number of simulation (Nb. Sim)

We simulated 100 simulation runs of $TS(0.6; 1; 0)$, $TS(1.6; 0.0001; 3)$, $TS(1.3; 0.0001; 1)$ and $TS(1.6; 1; 0)$ random variables with sample size $N = 2048$ and $N = 4096$ using the algorithm of Tankov and Poirot (2005).
The result of the estimates are presented in Tables (6.2) and (6.3).
We can see that the estimation procedure produces good results. As expected, performance improves with sample size. Note, however, that the estimation of the parameters works less well in these simulation when the tail parameter $\alpha << 2$ and the truncation parameter λ is large (> 1). This is clear example of an extreme stress scenario which does not reflect real financial data.

	$\alpha = 1.3$	$c = 0.0001$	$\lambda = 1.0$
Mean	1.3107	0.0001	1.03
Std. err	0.037	3.10E-5	0.80
	$\alpha = 1.6$	$c = 1.0$	$\lambda = 0$
Mean	1.6077	1.04	2.5E-4
Std. err	0.065	0.037	0.0045
	$\alpha = 1.6$	$c = 0.0001$	$\lambda = 3$
Mean	1.616	0.0001	2.80
Std. err	0.0398	5.14E-5	0.57
	$\alpha = 0.6$	$c = 1$	$\lambda = 0$
Mean	0.6227	1.076	1.4E-5
Std. err	0.0173	0.050	3.20E-6

Table 6.2: Estimation of the parameters for $N = 2048$

	$\alpha = 1.3$	$c = 0.0001$	$\lambda = 1.0$
Mean	1.3042	0.0001	1.01
Std. err	0.011	6.550E-6	0.63
	$\alpha = 1.6$	$c = 1.0$	$\lambda = 0$
Mean	1.6025	1.02	6.1E-6
Std. err	0.032	0.019	0.0032
	$\alpha = 1.6$	$c = 0.0001$	$\lambda = 3$
Mean	1.611	0.0001	2.87
Std. err	0.0168	4E-5	0.37
	$\alpha = 0.6$	$c = 1$	$\lambda = 0$
Mean	0.6100	1.025	1.33E-5
Std. err	0.0121	0.030	2.2E-6

Table 6.3: Estimation of the parameters for $N = 4096$

To illustrate the application of this methodology to financial time series, we have studied the S&P 500 index. We estimate the hyperparameters from the time series on the index closing prices over the period January 1, 1980 to December 31, 2004. We formed the time series of the daily returns and we calculated the wavelet coefficients. The coefficients were calculated using the Daubechies-3 wavelet. The skewness of the returns of the wavelet coefficients is -1.97 and the kurtosis is 51.49. High kurtosis shows that the distribution of the returns of the wavelet coefficients exhibits a heavy tailed behavior.

We estimate the hyperparameters using the symmetric tempered distribution. Figures (6.2) and (6.3) shows the tempered stable fit to the daily wavelet coefficients. The fit is based on 3072 observations. It is evident from Figure (6.2) that the fit is good and the estimates of the hyperparameters of the wavelet coefficients are $\hat{\alpha} = 1.66$, $\hat{c} = 2.11\text{E-}5$ and $\hat{\lambda} = 6.91$.

6.5 Summary

In this chapter we developed a Lévy wavelet sparse model where the prior density of the wavelet coefficients is a mixture of a tempered stable density and a mass point at zero. A maximum likelihood estimation was performed using the Midpoint Rule and the discrete Fourier transform to infer the model parameters. This approach performs well except for some extreme cases.

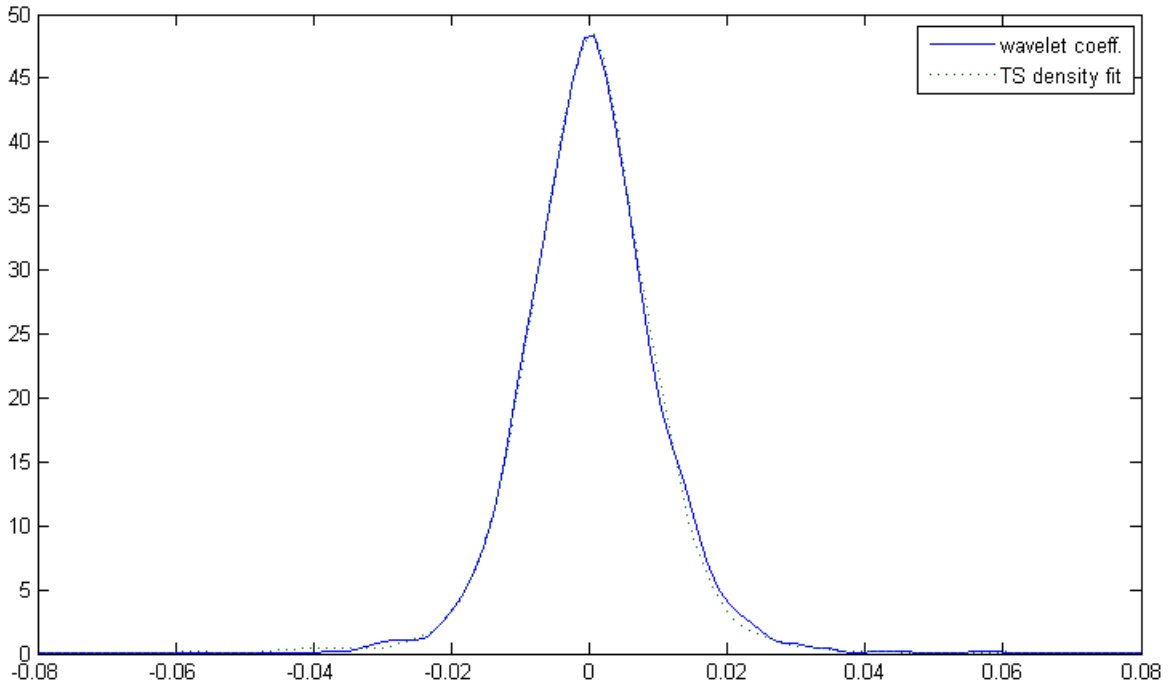


Figure 6.2: Probability Density fit of the S&P 500 data.

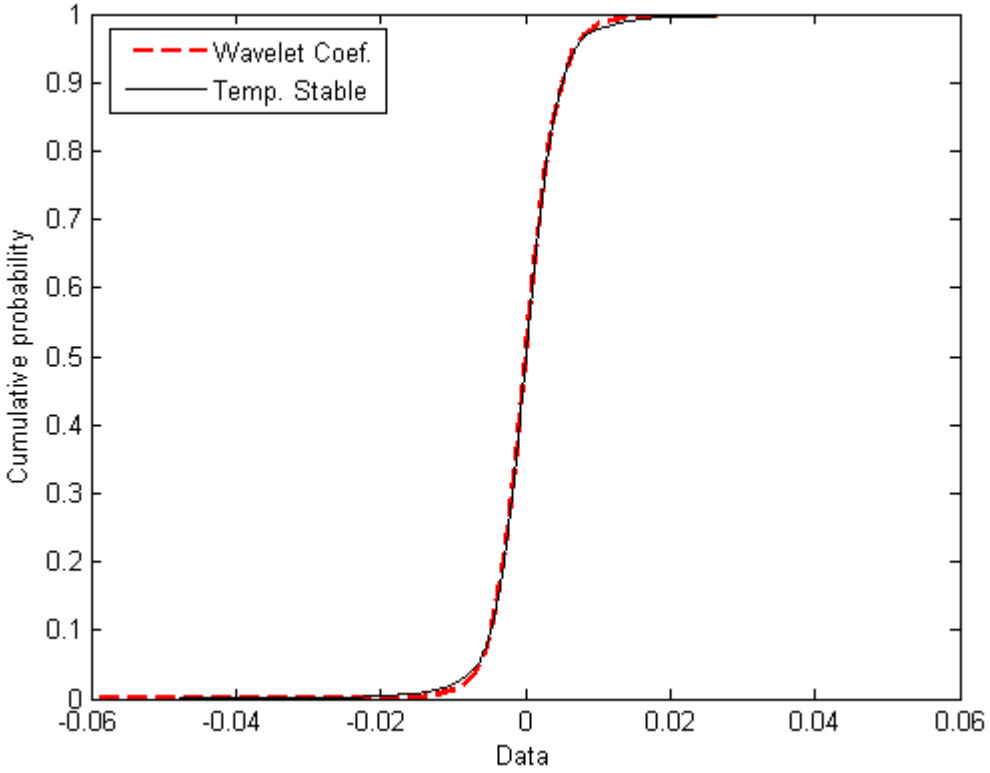


Figure 6.3: Cumulative Density fit of the S&P 500 data

Chapter 7

Regularization

7.1 Introduction

In this chapter, we present a wavelet regularization approach for estimating non-parametric regression functions. We follow here the Bayesian approach, where we set the prior distribution to a tempered stable and we focus on finding a solution to the wavelet-based regression through a Penalized maximum Likelihood (PML) approach. Density estimation methods based on penalized MLE offer a flexible and data-analytically appealing methodological tool. We will explore PMLE based on wavelet entropy functionals. Our approach considers a generalized form of the entropy functional based on the discrete formulation of the Tsallis entropy.
We also give explicit closed form expressions of the wavelet Tsallis entropy and the wavelet Fisher's information for tempered stable distributions. We will use the functional of the wavelet Tsallis entropy as a penalty function for the regularization problem. Here, not only do we introduce the wavelet regularization approach but we give a new methodology based on wavelet entropy functionals to improve the estimation of the hyperparameters of the tempered stable distributions.

7.2 Penalized Least Square

Let consider the standard regression model (1.1) and let W be the wavelet transform and $s = Wg$. We can also write $g = W^T s$ because of the orthogonality of the transform matrix.

If we denote the sampled data by Y_n, then the regression problem can be expressed as

$$Y_i = W^T s_i + \varepsilon_i, \quad i = 1, ..., n \tag{7.1}$$

where $\varepsilon \sim N(\sigma_\varepsilon^2)$.

An unbiased and least squares estimate of the regression problem (7.1) is obtained by minimizing the residual square error:

$$\min_s ||Y - W^T s||^2, \tag{7.2}$$

In high dimensional data, the matrix WW^T becomes sparse and close to singular making inverting the matrix impossible. Therefore, we may need to regularize the problem by adding a constraint, and minimizing instead the penalized least squares problem

$$||Y_n - W^T s||^2, \quad \text{subject to } \sum_{i=1}^{n} p_\lambda(|s_i|) < t, \tag{7.3}$$

where p_λ is the penalty function. The problem (7.3) can also be written

$$||Y_n - W^T s||^2 + \sum_{i=1}^{n} p_\lambda(|s_i|), \tag{7.4}$$

for a given regularization parameter $\lambda > 0$. The penalized least square regression corresponds to a thresholding or shrinking approach by choosing a particular penalty function. Fan and Li (2001) introduce the main properties of a good estimator. They are

1. Unbiasedness: The resulting estimator is approximately unbiased.

2. Sparsity: The resulting estimator is a thresholding rule, which automatically sets small estimated coefficients to zero to reduce model complexity.

3. Continuity: The resulting estimator is continuous to avoid instability in model prediction.

These conditions are translated into the penalty function to produce a good estimator with a good model selection. We have

1. Unbiasedness: if $p_\lambda^{'}(|s|) \to 0$ as $|s| \to \infty$.

2. Sparsity: if $\text{argmin}_{s\geq 0}(|s| - p_\lambda^{'}(|s|) = 0$.

3. Continuity: *if and only if* $\min_{s\geq 0}(|s| - p_\lambda^{'}(|s|) > 0$.

If the sampling points are equally spaced and $n = 2^J$, the formulation (7.4) of the minimization problem becomes

$$\frac{1}{2}||d - s||^2 + \sum_{i=1}^{n} p_\lambda(|s_i|), \tag{7.5}$$

where $d = WY$ are the wavelet coefficients. To obtain a good estimator that has the above properties, the penalty function should satisfy the necessary and sufficient conditions given above. Antoniadis and Fan (2001) gave the implications of these conditions on the solution of the problem (7.5)

Theorem 36. *Let p_λ be a nonnegative, nondecreasing and differentiable function in $(0, \infty)$. Assume that the function $s \mapsto -s - p_\lambda^{'}(s)$ is strictly unimodal on $(0, \infty)$. Then*

1. *The problem (7.4) has a unique solution satisfying $\hat{s}(-d) = -\hat{s}(d)$*

2. *The solution satisfies $\hat{s}(d) = (d - \text{sgn}(d)p_\lambda^{'}(|\hat{s}(d)|)1_{|d|>p_0}$, where $p_0 = \min_{s\geq 0} s - p_\lambda^{'}(s)$.*

3. If $p'_\lambda(s)$ is nonincreasing, then we have

$$|d| - p_0 \leq |\hat{s}(d)| \leq |d| - p'_\lambda(|d|).$$

4. If $p'_\lambda(|d|) \to 0$ as $|d| \to \infty$, then

$$|\hat{s}(d)| = d - p'_\lambda(|d|) + o(p'_\lambda(|d|)).$$

Proof. See Theorem 1 in Antoniadis and Fan (2001). □

The penalty function characterizes the behaviour and the structure of the estimator. The choice of the penalty function of the form $p_\lambda(s) = \lambda s^2$ assumes that the a priori distribution of s is Gaussian $N(0, \sigma^2/\lambda)$. The conditional probability of d_i given s_i is $N(s_i, \sigma^2)$, then the posterior density of s is proportional to

$$\exp\left(-\frac{1}{2}\left[\sum_i (d_i - s_i)^2 + \lambda \sum_i s_i^2\right]\right).$$

Such regularization is known in the literature as ridge regression, and it can be written as

$$\frac{1}{2}\sum_i (d_i - s_i)^2 + \lambda \sum_i s_i^2,$$

In ridge regression, no variable is excluded from the model; the coefficients are a function of λ and tend to zero as λ increases. Therefore, ridge regression does not perform model selection since it does not shrink coefficients to zero.
To remedy this perceived deficiency, Tibshiriani (1996) studied the case of an l_1 penalty function where the minimization of the normal likelihood is subject to a constraint $\sum_i |s_i| \leq t$. This is equivalent to choosing $p_\lambda(s) = \lambda|s|$. This yields to the Lasso (Least Absolute Shrinkage Selection Operator) (Tibshiriani (1996), Griffin and Brown (2010)) defined by

$$\frac{1}{2}\sum_i (d_i - s_i)^2 + \lambda \sum_i |s_i|,$$

The solution to the Lasso problem yields the soft-thresholding rule

$$\hat{s}_j = \text{sgn}(d_j)(|d_j| - \lambda)^+$$

Like ridge regression, the Lasso can be seen as a Bayesian method but assuming the detail coefficients s to be double exponentially distributed with density $\frac{\lambda}{2}\exp(-\lambda|s|)$; see Griffin and Brown (2011). However, the Lasso penalty function satisfies the sparsity and continuity conditions but not the unbiasedness property.
Therefore, Fan and Li (2001) have introduced the SCAD (Smoothly Clipped Absolute Deviation) penalty function that has all properties, and defined by

$$p_\lambda(|s|) = \begin{cases} \lambda|s|, & \text{if } |s| \leq \lambda, \\ -\frac{|s|^2 - 2a\lambda|s| + \lambda^2}{2(a-1)}, & \text{if } \lambda < |s| \leq a\lambda, \\ \frac{(a+1)\lambda^2}{2}, & \text{if } |s| > a\lambda. \end{cases}$$

The resulting solution of the SCAD penalization is given by (5.2). The pair (λ, a) could be estimated using cross-validation or generalized cross-validation; see Craven and Wahba (1977). Although, The SCAD penalty function satisfies all conditions to produce a good estimator, it is nonconvex and its implementation is computationally expensive.

7.3 Penalized Maximum Likelihood

The penalized maximum likelihood estimate is another way to find the estimators of regression coefficients, without fitting the noise. We can also consider the penalized maximum likelihood estimate which is defined by

$$\min_\theta -\frac{1}{n}\sum_{i=1}^{n} \log \mathcal{L}(y_i;\theta) + \lambda \sum_{i=1}^{n} p(|s_i|) \tag{7.6}$$

The penalized likelihood estimate performs (simultaneously) variable selection, parameter of the prior distribution estimation and facilitates the study of the asymptotic

properties of the estimators.

There have been many studies dealing with the regularization problem with Bayesian modeling in the wavelet domain. Antoniadis and Fan (2001) presented a general characterization of the wavelet-based PML estimates for a class of additive smooth penalty functions.

We propose an entropic regularization approach for estimating the wavelet-based non-parametric regression. The difficulty in this approach is the choice of the appropriate entropy. The entropy functional should reflect the statistical properties of the signal. Most of the existing entropic regularization methods use the Boltzmann entropy as a penalty function. This choice is too restrictive since, the Boltzmann entropy can model only positive additive distributions and does not take into account correlations between adjacent constituents of the signal. To remedy this, Gull and Skilling (1999) use the intrinsic correlation function method to decorrelate the reconstructed image. Unfortunately, the correlation structure in most of the cases is unknown, therefore it has to be constructed from the historical data.

Tsallis et al. (1988) proposed a generalization of the Boltzmann entropy to treat non-additive systems. He proved that the maximum Tsallis entropy can lead to solutions in the Lévy-type distributions. We will develop a wavelet-based penalized maximum likelihood using the Tsallis entropy functional.

7.4 About Entropy

Entropy was first introduced in physics and it has enjoyed a special status. It is associated with the second law of thermodynamics: the entropy of an isolated system is increasing over time. Later, Boltzmann connected the entropy, defined as a macroscopic description of the state of a system, to the microscopic description of the same system.

The Boltzmann-Gibbs (BG) entropy of a sequence of random variables X is given by

$$S(X) = -k \sum_{i=1}^{N} p_i \ln p_i,$$

where k is the Boltzmann-Gibbs constant in thermodynamics or taken to be 1 in information theory and the probability $p_i = P(X = x_i)$. If $p_i = 1/N$, we obtain the well-known entropy

$$S_B = k \ln N.$$

Let $\mathrm{P}(X, Y)$ be the joint probability of X and Y, and $\mathrm{P}(X|Y)$ the probability of X conditional to Y. We denote $\mathrm{P}(x_i)$ the probability $\mathrm{P}(X = x_i)$, $\mathrm{P}(y_i)$ the probability $\mathrm{P}(Y = y_i)$, $\mathrm{P}(x_i, y_i)$ the probability $\mathrm{P}(X = x_i, Y = y_i)$ and $\mathrm{P}(x_i|y_i)$ and $\mathrm{P}(x_i|y_i)$ are the conditional probabilities $\mathrm{P}(X = x_i|Y = y_i)$ and $\mathrm{P}(Y = y_i|X = x_i)$.
The joint entropy $S(X, Y)$ is defined by

$$S(X, Y) = -\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{P}(x_i, y_j) \ln \mathrm{P}(x_i, y_j).$$

The conditional entropy is

$$\begin{aligned} S(X|Y) &= -\sum_{j=1}^{N} \mathrm{P}(y_j) S(X|Y = y_i), \\ &= -\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{P}(x_i|y_j)\mathrm{P}(y_j) \ln \mathrm{P}(x_i|y_j). \end{aligned}$$

Using the Bayes rule

$$\mathrm{P}(X, Y) = \mathrm{P}(X)\mathrm{P}(Y|X) = \mathrm{P}(Y)\mathrm{P}(X|Y),$$

we can write

$$\begin{aligned} S(X, Y) &= S(X) + S(Y|X), \\ &= S(Y) + S(X|Y). \end{aligned}$$

This means that the total entropy behave as the sum of its parts and we refer to the Boltzmann entropy as an additive measure.
The construction of the BG entropy relies on a series of axioms which we will study

here.

There are several set of axioms that allow one to define and obtain the BG entropy functional. We will follow here the structural definition of Fadeev (1957):

Definition 37. *If $S(p_1, ..., p_n)$ is the functional entropy of a discrete probability distribution $p_1, ..., p_n$ where $\sum p_i = 1$, and if the four following axioms are satisfied*

1. *$S(p_1, ..., p_n)$ is symmetric,*
2. *$S(p, 1-p)$ is a continuous function of p for $0 \leq p \leq 1$,*
3. *$S(p_1, ..., p_n) = S(p_1 + p_2, p_3, ..., p_n) + (p_1 + p_2)S(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$,*
4. *S(1/2,1/2)=1,*

then the structural entropy S is

$$S(p_1, ..., p_n) = -\sum_{i=1}^{n} p_i \ln p_i.$$

These axioms actually seem reasonable, given what we expect from a measure of information. In particular, axiom (3) means that when we change the scale, i.e. when we take a finer mesh of the space of events, the global information increases by the information in the subdivided mesh.

Rényi proposed replacing axiom (3) by:

Let $P = (p_1, ..., p_n)$ and $Q = (q_1, ..., q_n)$ be two probability distributions. Denote $P * Q = p_i.q_j$. Then the total entropy is

$$S(P * Q) = S(P) + S(Q).$$

However, this axiom is weaker than (3), and it is necessary to introduce an additional axiom to characterize the functional. It is this axiom that can lead to other entropies. To do so, Rényi extended the notion of probability distribution by defining what he called "incomplete" or generalized probability distributions. A finite discrete generalized probability distribution P is a sequence of nonnegative numbers $p_1, ..., p_n$ such

that $0 < \sum p_i = W(P) \leq 1$.

So we extend the notion of the functional entropy for generalized probability distributions by means of the four axioms mentioned previously and a last axiom, called the *mean-value condition* on the entropy, is stated as follows:

(5) Let $P = (p_1, ..., p_n)$ and $Q = (q_1, ..., q_m)$ two generalized probability distributions such that $W(P) + W(Q) \leq 1$, and $P \cup Q = (p_1, ..., p_n, q_1, ..., q_m)$. We have

$$S(P \cup Q) = f^{-1}\left(\frac{W(P)f(S(P)) + W(Q)f(S(Q))}{W(P) + W(Q)}\right),$$

where f is an arbitrary monotonic and continuous function. Rényi proved that the choice of the function f can only be one of the two types:

1. f is a linear function $f(x) = ax + b$ with $a \neq 0$ which leads to the functional entropy $S = -\frac{\sum_{i=1}^{n} p_i \ln p_i}{\sum_{i=1}^{n} p_i}$.

2. f belongs to the exponential family of type $f(x) = \mathrm{e}^{-(1-\alpha)x}$ with $\alpha \geq 0$ and $\alpha \neq 1$, when the functional entropy is

$$S_\alpha(P) = \frac{1}{1-\alpha} \ln\left(\frac{\sum_{i=1}^{n} p_i^\alpha}{\sum_{i=1}^{n} p_i}\right),$$

known as the Rényi entropy of order α.

We can easily see that the Boltzmann entropy is a special case of the Rényi entropy $\lim_{\alpha \to 1} S_\alpha(P) = S(P)$ for a ordinary probability distribution.

Tsallis (1988) postulated a generalized form of the functional entropy, namely the Tsallis entropy

$$S_q(P) = \frac{1}{1-q}\left(1 - \sum_{i=1}^{N} p_i^q\right), \quad q \in \mathbb{R}. \tag{7.7}$$

It is straightforward to verify that $\lim_{q \to 1} S_q(P) = S(P)$. From definition (7.7), the Tsallis entropy does not satisfy axiom (3). Tsallis proposed replacing axiom (3) with (3): $S_q(P * Q) = S_q(P) + S_q(Q) + (1-q)S_q(P)S_q(Q), \quad q \in \mathbb{R}$

which makes S_q nonadditive except for the case $q = 1$.

We have worked so far with discrete distributions. In a continuous framework, the

question is how to set a maximum entropy formalism. Prato and Tsallis (1999) considered the generalized entropy form

$$S_q(p(x)) = \frac{\int_{-\infty}^{+\infty} \frac{\mathrm{d}x}{\sigma}(\sigma p(x))^q}{q-1}, \tag{7.8}$$

subject to the constraints

$$\int_{-\infty}^{+\infty} \mathrm{d}x\ p(x) = 1,$$

on the mean and the second moment condition

$$\frac{\int_{-\infty}^{+\infty} \mathrm{d}x\ x^2(p(x))^q}{\int_{-\infty}^{+\infty} \mathrm{d}x\ (p(x))^q} = \sigma^2.$$

The maximum Tsallis entropy yields

$$p(x) = \frac{1}{\sigma}\left(\frac{q-1}{\pi(3-q)}\right)^{1/2} \frac{\Gamma\left(\frac{1}{q-1}\right)}{\Gamma\left(\left(\frac{3-q}{2(q-1)}\right)\right)} \frac{1}{\left(1+\frac{q-1}{3-q}\frac{x^2}{\sigma^2}\right)^{1/(q-1)}},$$

for $q > 1$. This distribution has infinite second moment and the asymptotic form

$$p(x) \sim x^{-2/(q-1)}.$$

Since the Lévy α-stable density f exhibits asymptotically a power law of the form

$$f(x) \sim x^{-1-\alpha}$$

where α is the tail index, then, q is related to α as follows:

$$q = \frac{3+\alpha}{1+\alpha}.$$

7.5 Wavelet Entropy

In 1979, Powell and Percival (1979) introduced the spectral entropy defined from the Fourier power spectrum. The spectral entropy is used to measure the regularity

of a signal and to quantify the distribution of spectral power. However, the Fourier transform requires stationary signals/processes and does not characterize the time-evolution of the frequency patterns. Therefore, the spectral entropy is not a time dependent functional.

To remedy such disadvantages, the entropy has been combined with the wavelet transform. This makes no assumption on the stationarity of the signal. The combined entropy functional is called *wavelet entropy*.

The wavelet Boltzmann-Gibbs entropy which amounts to computing the Boltzmann-Gibbs entropy on a probability density derived from the wavelet variance, has found applications in studying neurobiological data (Quian Quiroga et al., 2001), electroencephalographic (EEG) activity (Al-Nashash et al., 2003) and the complexity of stochastic processes (Zunino et al., 2007).

Here, the objective is to use the wavelet entropy functional to characterize the degrees of complexity of the underlying process. We present the framework of the wavelet entropy and compute its functional when the distribution of the wavelet coefficients is tempered stable.

Let ψ_{ij} be a family of wavelet orthonormal basis in $L^2(\mathbb{R})$. The wavelet coefficients for a given function $g \in \mathbb{R}$ are given by $d_{ij} = \int_{\mathbb{R}} g(t)\psi_{ij}(t)\mathrm{d}t$. The energy at resolution j is defined by

$$E_j = \mathbb{E}(||d_j||_2^2) = \sum_{k=0}^{2^j-1} \mathbb{E}(|d_{jk}|^2).$$

The corresponding total energy is obtained by summing the energy at all levels:

$$E_{tot} = \sum_{j=1}^{N} E_j.$$

Therefore, we define the wavelet energy at scale j as a normalized value

$$p_j = E_j/E_{tot}. \tag{7.9}$$

It is straightforward to see that $\sum_{j=1}^{N} p_j = 1$. This leads us to consider p_j as a time-scale density. This density can be used to compute information functionals,

such entropies and Fisher's information. Also, the wavelet energy is then a suitable tool for detecting and characterizing the edges and singularities in both time and frequency.

Consequently, we can define the wavelet entropy as

- For the Boltzmann wavelet entropy $S_B = -\sum_{i=1}^{N} p_i \ln p_i$,
- For the Tsallis wavelet entropy $S_T = \frac{1}{1-q}\left(1 - \sum_{i=1}^{N} p_i^q\right)$,

The wavelet entropy could be computed either using a critically sampled or oversampled filter bank; refer to Vetterli and Kovacević (1995) and Cvetković and Vetterli (1998) for detailed study on the critically sampled and oversampled filter banks respectively.

The wavelet entropy can be seen as a degree of "order" or "disorder" of the signal. If the signal is a random process, its wavelet decomposition will have significant contributions to the total wavelet energy from all frequency bands. Consequently, the relative wavelet energy will be almost equal for all resolutions and the wavelet entropy will attain its maximum. For an ordered process, the relative energy is resolved in one wavelet resolution level. The wavelet entropy in this case will take a very low value.

7.6 Wavelet Entropy for $1/f^{\alpha}-$processes

The concept of entropy has been used to measure the amount of information contained in stochastic processes. Entropic functionals, such as Boltzmann-Gibbs, Rényi and Tsallis were employed to quantify the complexity of self-similar processes.

We say that a stationary process X is an $1/f^{\alpha}-$process with $\alpha \in (0,1)$ if there exists a constant $c_f > 0$ such that

$$\lim_{f \to 0} S_X(f) / \left(c_f |f|^{-\alpha}\right) = 1,$$

where S_X is the spectral density function of X. The parameter α represents the long-range correlation of the underlying process X; see Beran (1994) and Percival

and Walden (2000) for background and references.
Some well-known and simple examples of $1/f^{\alpha}$-processes are the fractional Brownian motion and the fractional Gaussian noise.

Definition 38. *The fractional Brownian motion (fBm) is a mean-zero Gaussian process B_H with covariance*

$$\mathbb{E}(B_H(t)B_H(s)) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right)\mathbb{E}(B_H^2(1)),$$

where $0 \leq H < 1$

For $H = 1/2$, $B_{1/2}$ is the standard Brownian motion. The increment process $Y_H = \{t > 0 | Y_H(t) = B_H(t+1) - B_H(t)\}$ is known as the fractional Gaussian noise (fGn). We assume that X has a wavelet series expansion $X_t = \sum_{j,k} d_{jk}\psi_{jk}(t)$. The wavelet spectrum of X is defined (see Flandrin (1992)) by

$$\mathbb{E}(|d_{jk}^2|) = \int_{-\infty}^{+\infty} S_X(2^{-j}f)|\Psi(f)|^2 \mathrm{d}f,$$

where Ψ is the Fourier integral of ψ and f is the Fourier frequency.
The wavelet spectrum takes the form (Stoev et al., 2005)

$$\mathbb{E}(|d_{jk}^2|) \sim 2^{j\alpha}C(\psi, \alpha), \tag{7.10}$$

where $C(\psi, \alpha) = c_\gamma \int |f|^{-\alpha}|\Psi(f)|^2 \mathrm{d}f$ and c_γ is a constant. By direct application of Equation (7.10) into Equation (7.9), we obtain

$$p_j = 2^{(j-1)\alpha}\frac{1-2^{\alpha}}{1-2^{\alpha N}}, \quad j = 1, ..., N. \tag{7.11}$$

The Boltzmann wavelet entropy was calculated for the fBm and fGn (Zunino et al., 2007). They gave the Boltzmann wavelet entropy as

$$S_B(\alpha, N) = \frac{\alpha - 1}{\log_2 N}\left\{\frac{1}{1-2^{1-\alpha}} - \frac{N}{1-2^{(1-\alpha)N}}\right\} - \frac{1}{\log_2 N}\log_2\left[\frac{1-2^{\alpha-1}}{1-2^{(\alpha-1)N}}\right].$$

Ramírez-Pacheco and Roman (2011) extended this by evaluating the wavelet Tsallis

entropy for $1/f^{\alpha}$-processes, with

$$S_q(\alpha, N) = c_{N,q}\left\{\left(\frac{1-2^{\alpha}}{1-2^{\alpha N}}\right)^{q}\left(\frac{1-2^{\alpha q N}}{1-2^{\alpha q}}\right)\right\},$$

where $c_{N,q} = 1/(1-N^{1-q})$ is the normalizing factor.

7.7 Wavelet Entropy for tempered stable processes

In this section, we extend the results established so far for the wavelet Tsallis entropy where the underlying process is tempered stable process; see section 4.5 for background and references.

We assume the underlying process X has a wavelet series expansion $X_t = \sum_{j,k} d_{jk}\psi_{jk}(t)$, where each wavelet coefficient d_{jk} is tempered stable distributed, i.e $d_{jk} \sim TS_{\sigma_j}(\alpha, c, \lambda)$. Applying Theorem (5.2), the energy at resolution at level j is given by

$$E_j = \mathbb{E}(||d_j||_2^2) = 2^{j(1-2/\alpha)}\nu_2\sigma_1^2, \tag{7.12}$$

where ν and σ_1 are respectively the second absolute moment and the variance at scale $\Delta t = 1$ of the tempered stable distribution.

In order to evaluate the wavelet entropy, we compute the relative wavelet energy as in (7.9)

$$p_j = 2^{(j-1)(1-2/\alpha)}\frac{1-2^{1-2/\alpha}}{1-2^{N(1-2/\alpha)}}, \quad j = 1, ..., N. \tag{7.13}$$

Consequently, the wavelet Tsallis entropy is obtained from (7.13),

$$S_q(\alpha, N) = \frac{1}{q-1} - \frac{1}{q-1}\left(\frac{1-2^{(1-2/\alpha)}}{1-2^{N(1-2/\alpha)}}\right)^{q}\frac{1-2^{Nq(1-2/\alpha)}}{1-2^{q(1-2/\alpha)}}. \tag{7.14}$$

where $0 < \alpha < 2$.

We define the normalized wavelet Tsallis entropy in the form $S_q/S_{\max}$ where $S_{\max}$ is the maximum entropy value. Since N is the number of wavelet resolution levels, the normalized Boltzmann-Gibbs entropy $S_{\max} = \ln N$. For the Tsallis entropy, $S_{\max} = \frac{1-N^{1-q}}{q-1}$; refer to Tsallis et al. (1995) and Prato and Tsallis (1999) for back-

ground and references for Tsallis entropy

The parameter q in (7.14) gives further flexibility to analyse and to emphasize the

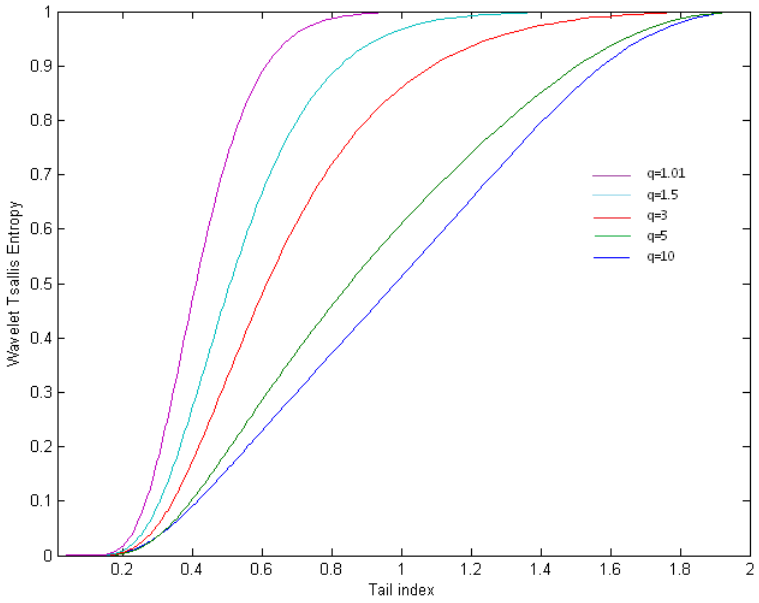


Figure 7.1: Wavelet Tsallis Entropy of a tempered stable process for $N = 10$ and different $q = 1.01; 1.5; 3; 5; 10$.

characteristics of the data. The question now is how the parameter q affects the entropy planes. Since $q \in \mathbb{R}$, we study entropy features in some key ranges of q. Figure (7.1) plots the wavelet Tsallis entropy of tempered stable processes. We observe that the parameter q stretches the form of the entropy as q is increasing. The maximum entropy is attained for the tail parameter $\alpha = 2$. This is in concordance with the fact that the entropy value is maximal for highly disordered signal (white noise signal $\alpha = 2$.)

7.8 Wavelet Fisher's information measure

Ramírez-Pacheco et al. (2011) extended the Fisher information concept to the wavelet domain. They studied the properties of the wavelet Fisher information for $1/f^{\alpha}$ processes.

The Fisher information can be used to detect non-stationary behaviour of a signal when the entropy fails to capture the full characterization of the signal.
In this work, we extend the Fisher's information measure to the wavelet domain and then give a closed form expression of this information for tempered stable processes. Let f_X be the probability density function of the random variable X. The Fisher information is defined by

$$\begin{aligned} I_X &= \int \left(\frac{\partial \log f_X(x)}{\partial x} \right)^2 f_X(x) \mathrm{d}x, \\ &= \int \left(\frac{\partial f_X(x)}{\partial x} \right)^2 \frac{\mathrm{d}x}{f_X(x)}. \end{aligned}$$

A discretized version of the Fisher's information is defined by

$$I_X = \sum_{k=1}^{N} \frac{(p_{k+1} - p_k)^2}{p_k}. \tag{7.15}$$

where p_k is the wavelet energy at scale k defined in (7.13).
Using the relative wavelet energy of Equation (7.13), we can extend the Fisher information measure to the wavelet domain. It is given by

$$I_X = \left(2^{1-2/\alpha} - 1\right)^2 \frac{1 - 2^{(N-1)(1-2/\alpha)}}{1 - 2^{N(1-2/\alpha)}}. \tag{7.16}$$

The wavelet Fisher information is decreasing in α and attains its minimum for $\alpha = 2$, *i.e.*, for a Gaussian process. For $\alpha \to 0$, I_X tends to infinity. Figure (7.2) gives the behaviour of the wavelet Fisher's information given different scaling parameters. The wavelet Fisher's information for tempered stable processes does not vary with N. The Fisher information is shown to describe the complexity associated with a random process or signal. The complexity here means the amount of irregularity in the underlying process. Therefore, it is able to detect the structural changes in the mean. Ramírez-Pacheco et al. (2011) studied the detection of the structural breaks in the mean of $1/f^\alpha$ processes using the wavelet Fisher's information.

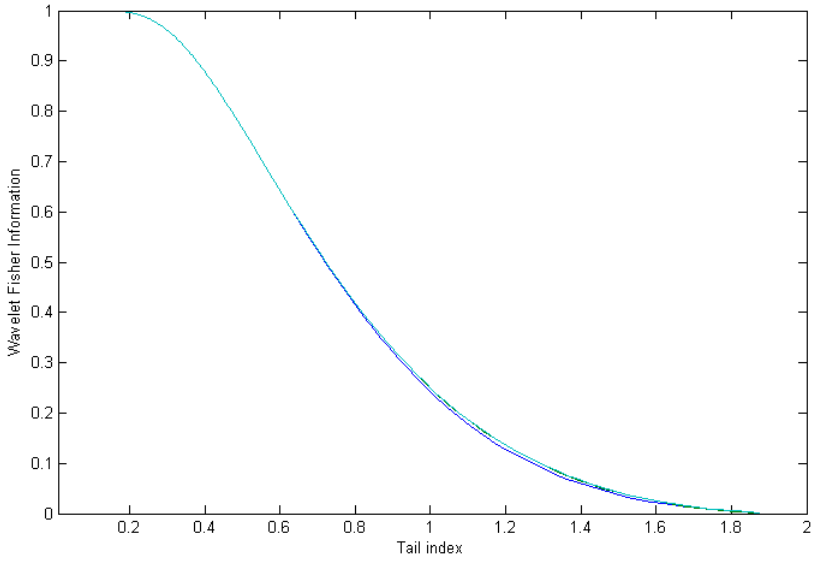


Figure 7.2: Wavelet Fisher Information for tempered stable process for $N = 5; 10; 15; 20$.

7.9 Wavelet Entropy Regularization

In practice, the noise present in the data leads to biased estimates of the parameters when the least squares method is used as a minimization criterion. This might cause an instability in density estimation. In order to get around this issue, regularized least squares approaches have been used.

Inspired by the work of Good and Gaskins (1971), we suggest an alternative penalty function that employs the wavelet Tsallis entropy in the penalized likelihood regularization. The PML method used in Good and Gaskins uses the maximization of the objective function to estimate the hyperparameters of the model

$$\omega(f) = L(f) - \Phi(f),$$

where L is the log-likelihood and Φ the roughness penalty defined by

$$\Phi(f) = 4\alpha \int \gamma'^2 dx + \beta \int \gamma''^2 dx$$

where $\gamma(x) = \sqrt{f(x)}$, $\alpha \geq 0$, $\beta \geq 0$ and $\alpha + \beta > 0$. They used the Rayleigh-Ritz method to minimize ω which makes use of the expansion

$$\gamma(x) = \sum_{i=0}^{\infty} \omega_i \phi_i(x),$$

where $\phi_0, \phi_1, ...$ are orthonormal functions such as Hermite or Fourier systems; see Courant and Hilbert (1953) for more background on the Rayleigh-Ritz method.

Like Good's roughness, we propose a penalty in the context of density estimation based on the wavelet Tsallis functional entropy. We consider that the PML estimate is given by

$$L(\alpha, c, \lambda; \mu) = -\sum_{i=1}^{N} \log f(x_i; \theta) + \mu S_q(\alpha, M), \tag{7.17}$$

where $\theta = (\alpha, c, \lambda)$ and $S_q(\alpha, M)$ is the wavelet Tsallis entropy defined in (7.14). The first term is the same goodness-of-fit measure as in the likelihood function (6.19) and the second term represents the functional entropic measure for the roughness. The parameter μ is the Lagrangian or regularization parameter and $M = \log_2 N$, where N is the number of observations.

The parameter μ determines the degree of regularization and is interpreted in the Bayesian framework as a hyperparameter. Maximum likelihood estimation depends sensitively on the noise in the data and oscillates largely so as to fit the noise. That is, the parameter μ can be seen as a parameter to stabilize the solution of the penalized maximum likelihood. Therefore, the choice of μ is crucial. There are several methods to choose μ; we refer to Härdle (1990) section 5 for an overview on different methods for assigning μ.

In maximum entropy applications, μ is chosen such that $\chi^2 = \frac{1}{2}\sum_{i=1}^{M}(d_i - s_i)^2 \approx N$. Titterington (1995) shows that this choice leads to underfitting of the data. However, inferring μ from the data using the Bayesian framework would assign an optimal value to μ:

$$p(s|d) = \int p(s|d, \mu) p(d|\mu) d\mu.$$

In our approach, we propose to choose the parameter μ as a global constant value instead of setting it dependent on the resolution level.
Our penalized approach is tested on simulated data. We compare the results to the ones obtained from a standard maximum likelihood method in section 6.3.3.
We considered data sets of 512, 1024 and 2048 samples generated from $TS(1.7; 0.5; 1)$ and for illustration, we fix the parameter $q = 2$.
Table (7.1) shows the results of the penalized maximum likelihood estimation of the hyperparameters (7.17). It illustrates the dependence of the estimates on the regularization parameter μ. The proposed regularization approach improves the estimation of the parameter λ and c compared to the maximum likelihood method, the results of which are depicted in Tables (6.2) and (6.3).
For each data set, we estimate the hyperparameters when μ has been fixed to a given value. We also report the sum of squared errors (SSE), i.e. the Euclidean L_2 distance, between the probability density functions with the true and estimated parameters. By looking at the values of SSE, it is evident that setting $\mu \approx N$ does not always give the best results, i.e. it does not always correspond to the minimum SSE. Therefore, we believe that treating μ as an additional model parameter can improve the estimation. The optimal regularization parameter μ will be investigated in future work.

7.10 Summary

In this chapter, we combine the wavelet transform with the Tsallis entropy functional into a new wavelet Tsallis entropy. We use this new effective wavelet entropy to develop a wavelet entropic regularization approach. Inspired by the work of Zunino et al. (2006), we compute the wavelet Tsallis entropy and Fisher information for a tempered stable process and propose a penalized likelihood method for the estimation of the hyperparameters of the tempered stable distribution. The results are promising, therefore a wavelet Bayesian context can be interpreted as a better model for density estimation.

N	μ	$\alpha = 1.7$	$c = 0.5$	$\lambda = 1$	SSE
512	128	1.6539	0.4332	0.8158	0.0525
	512	1.6810	0.4270	0.8250	0.0368
	1024	1.7025	0.4216	0.8370	0.0261
	2048	1.6539	0.4332	0.8158	0.0525
	4096	1.6980	0.4222	0.8319	0.0174
1024	128	1.7015	0.4756	0.8478	0.0069
	512	1.6996	0.4744	0.8448	0.0080
	1024	1.7115	0.4542	0.8493	0.0157
	2048	1.6473	0.4594	0.8580	0.0378
	4096	1.6386	0.4642	0.8550	0.0449
2048	128	1.6763	0.5232	0.9263	0.0047
	512	1.6693	0.5182	0.9099	0.0087
	1024	1.6693	0.6564	0.8113	0.0031
	2048	1.6993	0.5008	0.9510	0.0016
	4096	1.6659	0.5038	0.9601	0.0141

Table 7.1: Penalized maximum likelihood estimates

Chapter 8

Conclusions and Future Work

This Chapter summarizes the work of the thesis, discusses the contributions, and emphasizes areas for future research.

8.1 Conclusions

In this work, our aim is to estimate an unknown function or the underlying signal in the framework of a white noise nonparametric regression. In our study, the signal is approximated by a dictionary of wavelet functions which offers a parsimonious decomposition of the underlying.
The main issue here is to find an appropriate prior model that captures the sparsity of the wavelet coefficients of the signal. Given this *a priori* information, we can estimate the hyperparameters of the *a priori* distribution using a Bayesian approach.
In our framework, we introduce a novel approach to model the statistical properties of a nonparametric function via its wavelet coefficients. In Chapter 5, we establish a new prior model, namely the Lévy wavelet model to capture the heavy tails of the wavelet coefficients. We give necessary and sufficient conditions on the model parameters of the Lévy wavelet prior model to belong to a certain class of Besov spaces. In Chapter 6, we develop a Bayesian framework without imposing any specified structure on the variance of the wavelet coefficients. The outlined method produces promising results. We introduce a Lévy wavelet mixture model to capture the sparseness of the wavelet

coefficients, and show that this sparse model exhibits a thresholding rule. We also give a closed form expression of the posterior Lévy measure of the wavelet coefficients and estimate the parameters of the prior model in both a simulation study and for the S&P 500 time series. In Chapter 7, we propose an entropic regularization approach to stabilize the estimate of the hyperparameters of the wavelet Lévy model.

8.2 Future Work

Despite the promising results of the wavelet entropic regularization, there are some aspects in this approach that need to be improved. A key focus of immediate future work is to develop an efficient estimation of the smoothing parameter. We have restricted ourselves to the universal threshold by Donoho and Johnstone (1994). So, rather than using the universal thresholding in our Bayesian approach, we will investigate and construct a threshold value based on the properties of stable distributions, and compare it to the existent Bayesian thresholding algorithms. We hinted that wavelet regularization approaches coincide with the MAP rules. The MAP maximizes the posterior distribution and is a shrinkage rule. Therefore, we will investigate whether we can derive an analytical form of the MAP estimator of our Bayesian thresholding based on the Lévy sparse mixture model. If the mode of the posterior distribution cannot be given by a closed form, we will compute the MAP using Monte Carlo methods using simulated annealing. Andrieu and Doucet (2000) presented an original simulated annealing algorithm which is based on stochastic simulation of the Markov chain, which could be employed in this context.

Another consideration is to study the rates of convergence of the posteriori distribution for the Bayesian density estimation with Lévy tempered stable prior. Van Der Vaart and Van Zanten (2008) study the rates of convergence where the priors are Gaussian processes. Bochkina and Sapatinas (2009) have investigated the theoretical performance of Bayes factor estimators with independent and identically distributed errors that are not necessarily normally distributed. Other distributions were considered in Scricciolo (2007) and Castillo (2008) based on Dirichlet mixture priors.

By using a wavelet analysis, Bardet and Bertrand (2010) derive a nonparametric

estimator of the spectral density of a continuous time Gaussian process observed at random times. They prove that this estimator satisfies a central limit theorem with a convergence rate depending on the roughness of the process. Their results could be extended to non-Gaussian processes (e.g tempered stable processes). The expression of the spectral density estimator will be also non-Gaussian. The key tool to investigate this would the use of the stochastic integral representation for stable processes; see section 4.4.

8.3 Appendix A

Proof of Proposition 35

$$
\begin{aligned}
\nu_d(t) &= \int \phi(t-s)\nu_\theta(s)ds, \\
&= \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \int \exp\left(-a(t-s)^2\right) \nu_\theta(s)ds, \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} \int \exp(-as^2-bs)\nu_\theta(s)ds, \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} \int \exp(-as^2-bs) \left(Ap\frac{e^{-\lambda|s|}}{|s|^{1+\alpha}}1_{s<0} + Aq\frac{e^{-\lambda s}}{s^{1+\alpha}}1_{s>0} \right) ds, \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} (Ap\int_0^{+\infty} \exp(-as^2+bs)e^{-\lambda s}s^{-1-\alpha}ds \\
&+ Aq\int_0^{+\infty} \exp(-as^2-bs)e^{-\lambda s}s^{-1-\alpha}ds), \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} (Ap\int_0^{+\infty} \exp(-as^2-(-b+\lambda)s)s^{-1-\alpha}ds \\
&+ Aq\int_0^{+\infty} \exp(-as^2-(b+\lambda)s)s^{-1-\alpha}ds), \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} (Ap\int_0^{+\infty} \exp(-as^2-(-b+\lambda)s)s^{-1-\alpha}ds \\
&+ Aq\int_0^{+\infty} \exp(-as^2-(b+\lambda)s)s^{-1-\alpha}ds), \\
&= \frac{\exp(-at^2)}{\sqrt{2\pi\sigma_\epsilon^2}} (2a)^{\alpha/2}\Gamma(-\alpha)\{Ape^{\frac{(b+\lambda)^2\sigma_\epsilon^2}{4}}D_\alpha((b+\lambda)\sigma_\epsilon) \\
&+ Aqe^{\frac{(-b+\lambda)^2\sigma_\epsilon^2}{4}}D_\alpha((-b+\lambda)\sigma_\epsilon)\},
\end{aligned}
$$

where $a = \frac{1}{2\sigma_\epsilon^2}$ and $b = -\frac{t}{\sigma_\epsilon^2}$.

8.4 Appendix B

Assume that we aim to estimate an unobserved random variable X knowing the observations Y. A Bayesian estimator is defined as map

$$\mathcal{D}: \begin{array}{l} \mathcal{Y} \to \mathcal{X} \\ y \mapsto \widehat{x} = \mathcal{D}(y) \end{array}$$

where $\widehat{x}$ is the nearest neighbor of x. "Nearest" depends on the definition of a metric or distance between x and its estimate $\widehat{x}$. Define a nonnegative cost function C satisfying $C(x, \widehat{x}) = 0$.
The Bayes' risk is defined by

$$\begin{aligned} R(x, \mathcal{D}(y)) &= (C(x, \mathcal{D}(y))) \\ &= \int_{\mathcal{X}\times\mathcal{Y}} C(x, \mathcal{D}(y)) P_{X,Y}(x, y) dxdy \\ &= \int_{\mathcal{Y}} \int_{\mathcal{X}} C(x, \mathcal{D}(y)) P_{Y|X}(y|x) P_X(x) dxdy. \end{aligned}$$

The Bayesian estimator minimizes the Bayes' risk and its robustness depends on the choice of the cost function.
The MAP (Maximum a Posteriori) is a Bayesian estimator with cost function

$$C(x, \widehat{x}) = \begin{cases} 0 & \text{if } x = \widehat{x} \\ 1 & \text{otherwise} \end{cases}$$

Therefore, the maximum a posteriori is

$$\widehat{x}_{MAP} = \text{argmin} P_{X|Y}(x|y).$$

Also, we define the CEP (Conditional Expectation a Posteriori) as a Bayesian estimator with cost function

$$C(x, \mathcal{D}(y)) = (x - \mathcal{D}(y))^2$$

. Then, we can write

$$\widehat{x}_{CEP} = \mathbb{E}(X|Y) = \frac{\int_{\mathcal{X}} x P_{Y|X}(y|x) P_X(x)}{\int_{\mathcal{X}} P_{Y|X}(y|x) P_X(x)}.$$

Bibliography

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, 1978.

[2] F. Abramovich, Y. Benjamini, D. L. Donoho, and I. M. Johnstone, Adapting to unknown sparsity by controlling the false discovery rate, *Ann. Statist.*, vol.34(2):584-653, 2006.

[3] F. Abramovich, T. Sapatinas and B. Silverman, Wavelet hresholding via a Bayesian approach, *J.R. Statistics.Soc.B*, vol.60:725-749, 1998.

[4] A. Achim et al., Novel Bayesian multiscale method For speckle removal in medical ultrasound images. *IEEE Trans. Med. Imag*, vol.20:772-783, 2001.

[5] H. A. Al-Nashash, J. S. Paul, W. C. Ziai, D. F. Hanley and N. V. Thakor, Wavelet entropy for subband segmentation of EEG during injury and recovery. *Ann. Biomed. Eng.*, vol.31(6):653-658, 2003.

[6] U. Amato and D. T Vuza, Wavelet regularization for smoothly data, *Technical Report, Instituto per Applicazioni della Mathematica*, Napoli, 1994.

[7] C. Andrieu and A. Doucet, Simulated annealing for maximum a posteriori parameter estimation of hidden Markov models. *IEEE Trans. on Infor. Theo,*, vol.46(3):994-1004, 2000.

[8] A. Antoniadis, Smoothing noisy data with tapered coiflet series. *Scand. J. Statist*, vol.23:313-330, 1996.

[9] A. Antoniadis and J. Fan, Regularization of wavelets approximations. *Journal of the American Statistical Association.*, vol.96:939-967, 2001.

[10] A. Antoniadis, G. Gégoire and I. McKean, Wavelet methods for curve estimation, *J. Amer. Statist. Assoc*, vol.89:1340-1353, 1994.

[11] J. Bertoin, *Lévy Processes.* Cambridge University Press, 1996.

[12] R. Baraniuk and H. Choi, Wavelet statistical models and Besov spaces, *SPIE Technical Conference on Wavelet Applications in Signal Processing*, Denver 1999.

[13] J. M. Bardet and P. Bertrand, A nonparametric estimator of the spectral density of a continuous-time Gaussian process observed at random times. *Scand. J. Stat.*, vol.38: 458-476, 2010.

[14] J. Beran, *Statistics for Long-Memory Processes,* New York: Chapman & Hall, 1994.

[15] N. H. Bingham and R. Kiesel, Modelling asset return with hyperbolic distributions. *Knight, J., Satchell, S. (eds.) Asset return distributions. Butterworth-Heinemann.* p.1-20, 2001.

[16] N. Bochkina and T. Sapatinas, Minimax rates of convergence and optimality of Bayes factor wavelet regression estimators under pointwise risks. *Statistica Sinica*, Vol.19: 1389-1406, 2009.

[17] J-M. Bony, Calcul symbolique et propagation des singularités pour les EDP non linéaires. *Ann. Sci. Éc. Norm. Sup*, vol.14:209-249, 1981.

[18] L. Boubchir and J.M Fadili, A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior. *Pattern Recognition Letters*, vol.27:1370-1382, 2006.

[19] Boyarchenko, S. I., and S. Z. Levendorskii, *Non-Gaussian Merton-Black-Scholes Theory*, World Scientific, New Jercey, London, Singapore, Hong Kong, , 2002.

[20] M. Boyarchenko and S. Levendorskii, Refined and enhanced fast Fourier transform techniques, with an Application to the pricing of barrier options, 2008.

[21] P. Brown and J. Griffin, Inference with normal-gamma prior distributions in regression problems, *Bayesian Analysis*, vol.5(1): 171-188, 2010.

[22] A. G. Bruce and H-Y. Gao, Understanding WaveShrink: variance and bias estimation. *Biometrika*, vol.83(4):727-745, 1996.

[23] P. Carr, H. Geman, D. Madan, and M. Yor, The fine structure of asset returns: An empirical investigation, *Journal of Business*, vol.75:305-332, 2002.

[24] F. Caron and A. Doucet, Sparse Bayesian nonparametric regression. *ICML 08: Proceedings of the 25th international conference on Machine learning*, p.88-95, 2008.

[25] I. Castillo, Lower bounds for posterior rates with Gaussian process priors. *Electronic Journal of Statistics*, vol.2:1281-1299, 2008.

[26] S. Chang, B. Yu and M. Vetterli, Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image. Process*, vol.9(9):1522-1531, 2000a.

[27] S. Chang, B. Yu and M. Vetterli, Spatially adaptive wavelet thresholding with context modeling for image denoising. *IEEE Trans. Image. Process*, vol.9(9):1532-1546, 2000b.

[28] H. A. Chipman, E. D. Kolaczyk and R. E. McCulloch, Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc*, vol.92(440):1413-1421, 1997.

[29] Z. Ciesielki, G. Kerkyacharian and B. Roynette, Quelques espaces focntionnels associes a des processus gaussiens, *Studia Mathematica*. vol.107(2):171-204, 1993.

[30] M. A. Clyde, G. Parmigiani and B. Vidakovic, Multiple shrinkage and subset selection in wavelets. *Biometrika.*, vol.85:391-402, 1998.

[31] R. Cont and P. Tankov, *Financial modelling with Jump Processes*. Chapman and Hall, 2003.

[32] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol 1. New York and London: Internoscience, 1953.

[33] P. Craven and G. Wahba, Smoothing noisy data with spline functions. *Numer. Math.* vol.31(4):377-403, 1978.

[34] Z. Cvetković and M. Vetterli, Oversampled filter banks, *IEEE Trans. on Sig. Proc.*, vol.46(5):1245-1255, 1998.

[35] I. Daubechies, *Ten Lectures on Wavelets*, volume 61 of *CBMS-NSF Regional Conference Series in Applied Mathematics.* Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[36] D. G. T. Denison, C. C. Holmes, B. K. Mallick, A. F. M. Smith, *Bayesian Methods for Nonlinear Classification and Regression*, Wiley Series in Probability and Statistics, 2002.

[37] D. L. Donoho and I. M. Johnstone, Ideal adaptation by wavelet shrinkage. *Biometrica.*, vol.81:425-455, 1994.

[38] D. L. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, vol.90:1200-1224, 1995.

[39] D. L. Donoho and I.M. Johnstone, Minimax estimation via wavelet shrinkage. *Ann. Statist.*, vol.26(3):879-921, 1998.

[40] D. L. Donoho, I. M. Johnstone, G, Kerkyacharian and D. Picard. Wavelet shrinkage: Asymtoptia? (with discussion). *J. Roy. Statist. Soc. Ser. B*, vol.57(2):301-369, 1995.

[41] P. L Duren, *Theory of H^p Spaces*, Pure and Applied Math., 38, Academic Press, New York-London, 1970.

[42] H. Dym and H.P. McKean, *Fourier Series and Integrals*, Academic Press, NewYork 1972.

[43] D. K. Fadeev, Zum Begriff der Entropie einer endlichen Wahrscheinlichkeitsschermas, *Arbeiten zur Informationstheorie I, Berlin, Deutscher Verlag der Wissenschaften,*, p.85-90, 1957.

[44] J. Fadili and L. Boubchir, Analytical form for a Bayesian wavelet estimator of images using the Bessel K forms densities, *IEEE Trans. Image Process*, vol.14(2):231-240, 2005.

[45] J. Fan and R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Amer. Statist. Assoc.*, vol.96:1348-1360, 2001.

[46] W. Feller, *An Introduction to Probability Theory and its Applications*, vol 1 and 2, Wiley, New York, 1966.

[47] P. Flandrin, Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inf. Theory*, vol.38:910-917, 1992.

[48] D. Gabor, Theory of Communication. *J. Inst. Electr. Engrg.*, vol.93(3):429-457, 1946.

[49] H. Y. Gao, Wavelet Shrinkage denoising using the non-negative Garrote, *J. Comput. Graph. Statist*, vol.7(4):469-488, 1998.

[50] I. J Good and R. A. Gaskins, Nonparametric roughness penalties for probability densities, *Biometrika*, vol.58(2):255-277, 1971.

[51] P. J. Green and B.W. Silverman, *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Chapman and Hall, 1994.

[52] U. Grenander, X. Liu and A. Srivastava, Universal analytical forms for modeling image probabilities, *IEEE PAMI*, vol.24(9):1200-1214, 2001.

[53] J. E. Griffin and P. J. Brown, Bayesian hyper-lassos with non-convex penalization. *Australian and New Zealand Journal of Statistics*, vol.53(4):423-442, 2010.

[54] A. Grossmann and J. Morlet, Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. of Math. Anal.,* vol.15(4):723-736, 1984.

[55] S. F. Gull and J. Skilling, Quantified maximum entropy, The MEMSYS5 User's Manual. Maximum Entropy Data Consultants Ltd.

[56] W. Härdle, *Applied Nonparametric Regression*. Econometric Society Monograph Series 19, Cambridge University Press, 1990.

[57] S. Jaffard, Pointwise Smoothness, Two-microlocalization and wavelets coefficients. *Publications Mathematiques*, vol.35:155-168, 1991.

[58] S. Jaffard and Y. Meyer, Pointwise behaviour of functions. *Preprint, CERMA, ENPC, La Courtine*, 1993.

[59] I. M. Johnstone, Gaussian estimation: sequence and wavelet models. Draft version, December 27, 2011.

[60] I. M. Johnstone and B. W. Silverman, Wavelet threshold estimators for data with correlated noise, *J. Roy. Statist. Soc ser B*, vol.59(2):319-351, 1997.

[61] I. M. Johnstone and B. W. Silverman, Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences. *Ann. Statist.*, vol.32:1594-1649, 2004.

[62] I. M. Johnstone and B. W. Silverman, Empirical Bayes selection of wavelet thresholds, *Ann. Statist.*, vol.33(4):1700-1752, 2005 .

[63] G. Kimeldorf and G. Wahba, Some results on Tchebycheffian spline functions, *Journal of Mathematical Analysis and Applications*, vol.33(1):82-85, 1970.

[64] G. Kimeldorf and G. Wahba, A Correspondence between Bayesian estimation on stochastic processes and smoothing by splines, *Ann. Math. Statist.*, vol41(2):495-502, 1970.

[65] P. Koosis, *Introduction to H^p Spaces*, London Mathematical Society, Lecture Notes Series, 40, Cambridge University Press, Cambridge 1980.

[66] I. Koponen, Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process., *Physical Review E*, vol.52:1197-1199, 1995.

[67] S. G Krantz, *A Guide to Complex Variables*, Cambridge University Press, 2008.

[68] R. LePage, M. Woodroofe and J. Zinn, Convergence to a stable distribution via order statistics. *Ann. of Proba.*, vol.9(4):624-632, 1981.

[69] D. Leporini and J.C. Pesquet, Bayesian wavelet denoising: Besov priors and non-Gaussian noises, *Signal Processing* vol.81:55-67, 2001.

[70] S. G. Mallat, Multiresolution approximations and wavelet orthonormal basis of $L^2(\mathbb{R})$. *Trans. Amer. Math. Soc*, vol.315(1):69-87, 1989.

[71] S. G. Mallat, *A Wavelet Tour of Signal Processing.* Academic Press, 1999.

[72] B. Mandelbrot, The variation of certain speculative prices. *The Journal of Business.* vol.36(4):394-419, 1963.

[73] R. Mantegna and H. E. Stanley, Scaling behavior in the dynamics of an economic index. *Letters to Nature*, vol.376:46-49, 2002.

[74] G. McLachlan and D. Peel, *Finite Mixture Models.* Wiley Series in Probability and Statistics, 2000.

[75] C. Menn and S. T. Rachev , Calibrated FFT-based density approximation for α-stable aistributions. *Compt. Statist. and Data Analysis*, vol.50:1891-1904, 2006.

[76] Y. Meyer, *Wavelets and Operators.* Cambridge University Press, English Edition, 1992.

[77] Y. Meyer, *Wavelets: Algorithms and Applications.* Society for Industrial and Applied Mathematics, 1993.

[78] B. Missaoui and E. McCoy, Lévy wavelet prior model. *Submitted*, 2011.

[79] B. Missaoui and E. McCoy, Wavelet entropy of tempered stable processes. *In Preparation*, 2012.

[80] P. Moulin and J. Liu, Analysis of multiresolution image denoising schemes using generalized-Gaussian priors. in *Information Theory, IEEE Trans.*, vol.45(3):909-919, 1998.

[81] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, New York, 1995.

[82] J. P. Nolan, Numerical calculation of stable densities and distributions functions, *Comm. Statist. Stochastic Models*, vol.13(4):759-774, 1997.

[83] J. P. Nolan, Stable , *Stable Distributions - Models for Heavy Tailed Data*, Birkhauser, Boston, In progress 2013.

[84] E. A. Novikov, Infinitely divisible distributions in turbulence, *Physical Review E*, November, 1994.

[85] R. E.A.C. Paley and N. Wiener, Fourier Transforms in Complex Domains. *Colloquium Publ. Amer. Math. Soc.*, New York, vol.9, 1934.

[86] A. Papoulis, *The Fourier Integral and its Applications*, McGraw-Hill. New York, second edition, 1987.

[87] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis.* Cambridge University Press, 2000.

[88] B. Pesquet-Popescu, Statistical properties of the wavelet decomposition of certain no-Gaussian self-similar processes, *Signal Processing.* vol.75:303-322, 1999.

[89] J. Poirot and P. Tankov, Monte Carlo option pricing for tempered stable (CGMY) processes, *Asia Pacific Financial Markets*, vol.13(4), 2006.

[90] N. G. Polson and J. G. Scott. Local shrinkage rules, Lévy processes, and regularized regression.*J. Roy. Statist. Soc. B*, vol.74(2):287-311, 2012.

[91] J. Portilla, V. Strela, M.J. Wainwright, E.P. Simoncelli. Image denoising using scale mixtures of Gaussians in the wavelet domain, *IEEE Trans. Image Process.* vol.12(11):1338-1351, 2003.

[92] C. E. Powell and I.C. Percival, A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems. *J Phys A: Math Gen*, vol.12:2053-2071, 1979.

[93] D. Prato and C. Tsallis, Nonextensive foundation of Lévy distributions. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics.*, vol.60(2B):2398-2401, 1999.

[94] R. Quian Quiroga, O. A. Rosso, E. Basar, and M. Schürmann, Wavelet entropy in event-related potentials: a new method shows ordering of EEG oscillations, *Biological Cybernetics*, vol.84(4):291-299, 2001.

[95] S. T. Rachev, *Handbook of Heavy Tailed Distributions in Finance.* Elsevier B.V. 2003.

[96] B. S. Rajput and J. Rosiński, Spectral representations of infinitely divisible processes, *Probab. Th. Rel. Fields*, vol.82(3):451-487, 1989.

[97] J. Ramírez-Pacheco, D. Torres-Román, J. Argaez-Xool, L. Rizo-Dominguez, J. Trejo-Sanchez and F. Manzano-Pinzón, Wavelet q-Fisher information for scaling signal analysis. *Entropy*, vol.14(8):1478-1500, 2012.

[98] J. Ramírez-Pacheco and D. Torres-Román, Cosh-window behaviour of wavelet Tsallis q-entropies in $1/f^{\alpha}$ signals. *Electron Lett.* vol.47(3):86-87, 2011.

[99] V. Rivoirard, Thresholding procedure with priors based on Pareto distributions.*Soc. Est. Invest. Op.*, vol.13(1), 213-246, 2004.

[100] J. Rosiński, Tempering stable processes, *Stochastic Process Appl.*, vol.117(6):677-707, 2007.

[101] J. Rosiński,, Series representations of Lévy processes from the perspective of point processes, In *Lévy processes-Theory and Application.* O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick Eds., Birkhauser, Boston, p.401-415, 2001.

[102] O. A. Rosso, L. Zunino, D. G. Pérez, A Figliola, H. A. Larrondo, M. Garavaglia, M. T. Martín and A. Plastino, Extracting features of Gaussian self-similar stochastic processes via the Bandt-Pompe approach, *Physical Review E*, vol.76(6), 6 pages , 2007.

[103] D. Ruch and P. Van Fleet, *Wavelet Theory: An Elementary Approach with Applications*, John Wiley Sons, 2009.

[104] G. Samorodnitsky and M.S Taqqu, *Stable non-Gaussian Random Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, New York, 1994.

[105] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. vol.1:379-382, 1998.

[106] C. Scricciolo, On rates of convergence for Bayesian density estimation. *The Scandinavian Journal of Statistics*, vol.34:626-642, 2007.

[107] L. Sendur and I. W. Selesnick, Bivariate shrinkage functions for wavelet-based denoising exploiting interscale dependency, *Signal Processing, IEEE Trans.*, vol.50(11):2744-2756, 2002.

[108] E. P. Simoncelli and E. H. Adelson, Noise removal via Bayesian wavelet coring, in *Proc. IEEE Int. Conf. on Image Proc. - ICIP 1996*, vol.1:379-382, 1996.

[109] E. P. Simoncelli and J. Portilla, Texture characterization via joint statistics of wavelet coefficient magnitudes, 5^{th} *IEEE Int'l Conf on Image Processing*, vol.1:62-66, 1998.

[110] C. M. Stein, Estimation of the mean of a multivariate normal distribution, *Ann. Statist*, vol.:9:1135-1151, 1981.

[111] S. Stoev, M.S. Taqqu, C. Park and J.S. Marron, On the wavelet spectrum diagnostic for Hurst parameter estimation in the analysis of internet traffic, *Computer Networks*, vol.48(3):423-445, 2005.

[112] R. Tibshiriani, Regression shrinkage and selection via the Lasso, *J. Roy. Stat. Soc. B*, vol.58:267-288, 1996.

[113] D. M. Titterington, General structure of regularization procedures in image reconstruction, *Astronomy and Astrophysics*, vol.144:381-387, 1985.

[114] H. Triebel, *Theory of Function Spaces II.* Birkhäuser, Monographs in Mathematics, vol.84, 1992.

[115] C. Tsallis, S. V. F. Levy, A. M. C. Souza and R. Maynard, Statistical-Mechanical foundation of the ubiquity of Lévy distributions in nature, *Physical Review Letters*, vol.75(20):3589-3593, 1995.

[116] A. W. Van Der Vaart and H. Van Zanten, Rates of contraction of posterior distributions based on Gaussian process priors. *Annals of Statistics.* vol.36(3):1435-1463, 2008.

[117] M. Vetterli and J. Kovacević, *Wavelets and Subband Coding*, Prentice Hall, New Jersey, 1995.

[118] B. Vidakovic, *Statistical Modeling by Wavelets*, Wiley Series in Probability and Statistics, 1999.

[119] G. Wahba, *Spline Models for Observation Data*, Society For Industrial Mathematics, vol.59, 1990.

[120] H. L. Wei and A. Billings, Power-Law behaviour evaluation from foreign exchange market data using a wavelet transform method. *Physics Letters A*, vol.73(37):3324-3329, 2009.

[121] D. Williams, *Probability with Martingales*, Cambridge University Library, 1991.

[122] R. Wolpert, M. Merlise and C. Tu, Stochastic expansions using continuous dictionaries: Lévy adaptive regression kernels, *Ann. Stat.*, vol39(4):1916-1962, 2011.

[123] M. Yor and D. B. Madan, CGMY and Meixner subordinators are absolutely continuous with respect to one sided stable subordinators, Prépublication du Laboratoire de Probabilités et Modéles Aléatoires, 2005.

[124] V. M. Zolotarev, *One-dimensional Stable Distributions.* American Mathematical Society, 283, 1986.

[125] L. Zunino, D. G. Pérez,M. T. Martýn, A. Plastino,M. Garavaglia, and O. A. Rosso, Characterization of Gaussian self-similar stochastic processes using wavelet-based informational tools, *Physical Review E*, vol.75(2), 10 pages, 2007.

[126] A. Zygmund, *Trigonometric Series*, Cambridge Mathematical Library, 1959.