UNIVERSITÉ PARIS I PANTHÉON-SORBONNE
ÉCOLE DOCTORALE ED-465
CENTRE D'ÉCONOMIE DE LA SORBONNE

Thèse

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS I PANTHÉON-SORBONNE

Spécialité : Mathématiques Appliquées

présenté et soutenue par

Clément Goulet

Mardi 5 Décembre 2017

Signal Extractions with Applications in Finance

Thèse dirigée par Philippe de Peretti

JURY :

Jean-François Aujol	Professeur des Universités, Université de Bordeaux	Rapporteur
Christophe Chorro	Maitre de Conférences HDR, Université Paris I	Examinateur
Hayette Gatfaoui	Associate Professor HDR, IESEG School of Management	Examinateur
Olivier Guéant	Professeur des Universités, Université Paris I	Examinateur
Chafic Merhy	Docteur, Natixis Asset Management	Examinateur
Philippe de Peretti	Maitre de Conférences HDR, Université Paris I	Directeur de Thèse
Giulia Rotundo	Professor, University La Sapienza, Rome	Rapporteur

"Un jeune homme qui entasse pêle-mêle ses idées, ses inventions, ses études, ses lectures, doit produire le chaos; mais aussi dans ce chaos il y a une certaine fécondité qui tient à la puissance de l'âge".

Chateaubriand, *Mémoires d'outre-tombe*, XVIII,9.

Contents

Acknowledgements

I would like to express my sincere gratitude to my advisor, Philippe de Peretti, for his continuous support, his patience, and his immense knowledge that has directly contributed to the improvement of this PhD dissertation. He is never short of good advice regarding my future career. He also gave me the opportunity to teach a very interesting course and thanks to him, I know that I will try to keep teaching at University. Finally, we discovered a common passion for submarines, and perhaps one day we could transform this common passion into a joint research paper.

In addition to my advisor, I would like to thank the committee: Hayette Gatfaoui, for her very constructive comments on my work; Giulia Rotundo, for agreeing to serve as a reviewer in my jury. Her knowledge in applied mathematics and particularly in physics are a real asset to my work. I thank Jean-François Aujol for his review. Because most of my research is based on a article he wrote with Gilles Aubert, his presence is particularly important to me. I thank Christophe Chorro and Olivier Guéant, they always have been of good advice. Last but not least, I am grateful to Chafic Merhy for agreeing to be a member of my jury. The attendance of a highly skilled representative of the financial sector is most valuable.

I thank Matthieu Garcin who led me to discover exciting tools such as Wavelets or Total Variation techniques. I could not have met a better co-author than him. I am grateful for the time he spent making me learning how to properly write a research paper as well as the numerous very stimulating talks we had. His good mood as well as his particular sense of humour brought lightness and delicacy to these three years.

I thank the great academic fellows I met during these three years of PhD, the dynamic team of young researchers: ”le thé des jeunes économètres”; Marco Mazzola, for his advice and his proofreading. I thank Romain Poirson for his practitioner perspective. Lastly, I am also grateful to Mathieu Blanchard, for his advice regarding web-scraping, parallel computing and other informatics concepts I use in my dissertation

A friendly greeting to all my labmates! I know some of them since the Magistère d'économie, and I am glad we spent these six last years all together.

I would like to thank my friends and my family for supporting me spiritually throughout writing this thesis and my life in general.

I express my sincere gratitude to Mathilde, with whom I shared my life since the very beginning of our Phds. Our first meeting, at LabEx, was characterized by a long gale of laughter underlying her sunny disposition. Her brilliant mind as well as her good mood were a significant help all along these three years. Also I thank her for all her proofreading, and for all the english words she taught me, and all the good advice she had, either on my work or on my career.

I cannot write these acknowledgements without having a moving thought to the young child I was, when my grandfather, François Goure, taught me multiplication tables. At this time, I had concentration problems and it was difficult. But I succeed and today, I am proud of the path I accomplished. I am particularly proud as my PhD is somehow connected with Signal Theory and my grandfather graduated from Telecom Paris in 1944.

Chapter 1

Introduction

The concept of signal processing covers a wide range of issues such as signal compression, enhancement, pattern recognition and last, but not least, noise removal [1]. Noise removal is particularly challenging because it is generally considered as a pre-processing step providing the basis for other signal processing fields. Therefore, noise removal is of major concern as a large amount of available data is contaminated by noise. Analysing these data without removing the noise strongly affects the results achieved.

Noise can be defined as unwanted elements that modifies a signal and we introduce hereafter some concrete examples of noise removal.

Let's assume a sound engineer who is in charge of broadcasting a music concert at the radio. The engineer's microphone records all the sounds, including the roar from the crowd and other interferences. His or her work is therefore to distinguish between background noise and actual music. The main challenge of noise removal methods is that the original signal is unknown and one can only rely on beliefs to achieve his or her goal.

As in the previous example, in finance the original signal, the fair value of stocks, is often unobservable. Analysts only access to stock prices that often have an erratic behaviour unexplained by the theory. Black (1986) provides the following definition of noise in finance: "Noise in the sense of a large number of small events is often a cause factor much more powerful than a small number of large events can be." Indeed, according to him, the price of a stock reflects both the fundamental information available on the

[1]This work was achieved through the Laboratory of Excellence on Financial Regulation (Labex ReFi) supported by PRES heSam under the reference ANR10LABX0095. It benefited from a French government support managed by the National Research Agency (ANR) within the project Investissements d'Avenir Paris Nouveaux Mondes (investments for the future Paris New Worlds) under the reference ANR11IDEX000602.

specific firm as well as the noise generated by trades without motivations linked with the fundamentals. Noise in financial markets can have various causes depending on the time scale studied. Intra-day prices are contaminated by a micro-structure noise, which causes the stock price to deviate from its fundamental value **?**. Micro-structure noise results from intra-day phenomena, such as the execution latency, the bid-ask bounce or the discreteness of price changes. At a daily frequency the noise is originated by noise traders that are traders issuing buy or sell orders without connection with the firm fundamentals **?**. At larger frequencies, the noise is generated by phenomena such as market manipulation or financial bubbles, moving the stock price away from its fundamental.

Hence, in a financial context, the challenge is to recover the fundamental value of the stock. Noise removal have a direct implications in term of regulation, risk management, option pricing or stock price forecasting.

This PhD dissertation contributes to the applied mathematics literature in two different ways. First, by studying volatility spillovers around specific market events, Earnings Announcements, it directly contributes to the literature on applied econometrics to finance. Second, it contributes to the growing literature on signal processing and aims to offer improvements in noise removal methods. This PhD dissertation is based on three papers. The first paper provides an analysis of a signal occurring around Earnings Announcements. The latter two propose novel noise removal algorithms. For these two techniques, financial application are provided.

We assess below the contribution of the three articles.

The first field of interest deals with firm Earnings Announcements and historical volatility. Firms publish several times a year news regarding their financial situations or upcoming projects. The content of these publications is very informative and has direct consequences on investor behaviors and as a result, earning announcements is a widely studied topic. Following the seminal work of Ball and Brown (1968), a particular focus has been given on the jump in the announcer stock returns after the announcement. Various theoretical explanations have been proposed such as the asymmetry of information between investors, the existence of a liquidity premium or the media coverage... However, the publication content is directly linked to the uncertainty on the announcing firm. For scheduled announcements, Patell and Wolfson (1979) find a pre-announcement increase in the implied volatility of the announcer: investors are highly uncertain regarding the outcome of the announcement, which is consequently reflected by higher volatility. Donder and Vorst (1996) confirm this finding on a panel of 23 firms from the Amsterdam Stock Exchange, between June 1991 and December 1992. Moreover, Perignon and Isakov (2001) carry evidences of strong asymmetry in volatility following the announcement: the day following the announcement, the implied volatility increases (decreases) when bad (good) news have been published. This

finding is in line with the leverage effect of volatility[2].
Following on from these results, we want to test whether the uncertainty, carried by the announcer historical volatility, affects the volatility of peers. If such contagion is identified, it could have direct consequences in terms of regulation or portfolio in portfolio management[3]. In a signal theory perspective, this work aims at extracting a signal, that is a volatility spillover, and then to analyse it using a set of potential explanatory variables.

A major challenge is therefore to propose an efficient manner of testing for volatility contagion around EA.

The recent developments in econometric theory permit to measure contagion in volatility among financial markets. Kang and Yoon (2013) study the volatility spillovers between spot and future indices. Belgacem, Creti et al (2015) measure the effect of macroeconomic announcements on the volatility spillover between oil markets and the S&P. There results found evidences on oil market reaction to macro-economic news and therefore a spillover effect from oil market reaction to the US stock market (S&P 500). Lastly, Ben Rejeb and Arfaoui (2016) propose an analysis of volatility spillovers between emerging and developed countries. To quantify volatility spillovers, the authors use quantile regression. As a result, they observe that: interdependence increases during bullish markets while decreases during bearish markets.

Thus, the last contribution of this PhD is to propose an empirical analysis of historical volatility spillovers around earning announcements. On a panel of 406 S&P 500 firms and between April, 1st 2017 and August, 22nd 2017, we study the interdependence in volatility between the publisher and its co-sector. To be able to capture the contagion, we work on intra-day volatility (5min). Modelling of the volatility is performed using a Garman-Klass approach. Following Ben Rejeb and Arfaoui (2016), we model the volatility spillover with quantile regression. This permits to study the interdependence across all quantiles of the volatility conditional distributions. To test for a jump in volatility contagion, we use a panel regression method. Moreover, the spillover effect is analysed through variables such as: the outcome of the news, the surprise of the market, the announcer capitalization, the market sentiment toward the announcer...

Second, few denoising procedures consider a noise with a non-constant variance. In the concert example, we suppose now that for the last concert measure, people start to applaud slowly and, as soon as the music is getting weaker, applauds are getting louder and louder. In such case, the noise intensity increases through time. The sound

[2]The leverage effect is characterized by higher level of volatility when a firm reports bad news **?**

[3]For instance, the volatility of an equity portfolio can be decreased by removing firms generating volatility spillovers around there EA.

engineer has therefore to employ a methods being able to deal with the noise non-stationarity. Noises with non-constant variance, such as signal dependent noise, time dependent noise or others are observed in various fields. In Finance, the distribution of stock returns is known to be time dependent. Notably, at intra-day, daily and monthly frequencies, the probability distribution of stock returns is assumed to be Gaussian with zero expectation (expected return is null) and a time dependent variance. Indeed, the standard-deviation of returns is a common measure of uncertainty: periods with low (high) uncertainty are characterized by low (high) return standard-deviations. In data transformation, the transformation of irregularly spaced observations into regularly spaced ones, introduces an heteroscedastic noise. To our knowledge, few noise removal methods consider a noise with a non-constant variance. In terms of wavelet analysis, Gao (1997) extends the Donoho and Johnstone results to the case of an heteroscedastic noise. However, to estimate the non-constant variance, they implement a naive rolling window estimate. In Total Variation denoising, the *ROF* model can be adjusted to an heteroscedastic Gaussian noise. Indeed, the objective energy is composed by a data fidelity and a data regularity term. To give the priority to one of these two components, a regularization parameter is used. Thanks to the *discrepancy principle*, the regularization parameter is selected to match the noise variance. However, the discrepancy principle is known to over smooth the restored signal and an alternative selection procedure is given by Chambolle (2004). Hence, if one has a robust estimate of the non-constant noise, the original signal can be recovered. The challenge being to derive a robust estimate for the noise standard-deviation. In econometrics, Engle (1982) and Bollerslev respectively developped ARCH and GARCH models. Originally built for time series modelling, they are often used as filters. But, most of these models rely on the stationarity assumption of the underlying signal.

Hence, the second contribution of this Phd is to propose a novel denoising algorithm for a Gaussian non-constant noise. The technique we propose combines wavelet theory and total variation. Indeed, denoising a signal disrupted by a non-constant noise is divided in two stages. First, one has to derive a robust estimate of the noise standard-deviation. This is achieved by considering the non-constant noise part of the problem as a signal (the standard-deviation) multiplied by a zero mean and unit variance Gaussian noise. Estimating the standard-deviation is therefore similar to solve multiplicative noise problem. Thanks to the work of Aubert and Aujol (2008), such problem can easily be solved. As soon as a robust estimate of the noise standard-deviation has been obtained, the original signal can be recovered thanks to a standard wavelet filtering.

The last observation is that few denoising techniques consider the statistical properties of noises. For instance, let suppose that part of the concert is attended by a rather young audience. Their high voices is likely to distort the noise probability distribution and this specificity needs to be taken into account in the noise removal procedure of the engineer. In imagery, darker surfaces tend to be more positively skewed than lighter surfaces. For example, the shadow of an image element, such as a mountain, is a positively asymmetric noise. Besides, asymmetry is largely observed in Finance: the probability distribution of stock returns is known to be asymmetric and leptokurtic. However, a vast majority of the literature assumes the noise follows a Gaussian distribution. In wavelet analysis, Donoho and Johnstone (1994) set up a *universal threshold* for a Gaussian noise. Wavelet analysis consists in projecting a signal on a particular basis of functions. Gross structure are in *scale coefficients* while finer details are kept in *wavelet coefficients*. Once wavelets coefficients are derived, a threshold distinguishes the signal from the noise. The *universal threshold* is based on an estimate of the noise standard-deviation, and permits to distinguish among wavelet coefficients, those carrying signal from those carrying noise. An other popular approach in noise removal relies on *bounded variation functions* and even if this technique has been developed in the context of imagery, there exists a full range of possible applications. *Rudin-Osher-Fatemi* (ROF) (1992) are the first to propose such model, built for a Gaussian noise. This technique consists in finding the signal minimizing an energy based on both data fidelity and on data regularity. The data regularity term ensures the preservation of local singularities while keeping the signal as regular as it is. In their formulation, the data fidelity term is set to the *Mean-Squared-Error* (MSE), and according to Chambolle (2004), when the noise is Gaussian then MSE can be replaced by a quantity depending on the noise variance. A last advantage of this methods is that it enables the original signal to have local singularities, which is of great interest. However, the original formulation of the ROF model is known to have certain limitations. First, when the data fidelity term is set to the MSE, it leads to loss of contrast in the reconstructed image. As a result, Osher, Sole, and Vese (2003) propose a new formulation of the original model with a weaker norm for the data fidelity term. Chan and Esedoglu (2005) provide a theoretical justification for the use of a L1 norm as a data fidelity term. Second, it relies on the Gaussian assumption of the noise. Even if the use of Gaussian distributions is justified by the *central limit theorem* and by simplicity of implementation, Gaussian probability distributions suffer from strong properties such as symmetry and mesokurticity. Overall the years, many formulations of the ROF model have been introduced for removing more complex noise. Thanks to the Bayesian formulation of the ROF model[4] Le, Chartran, and Asaki (2007) develop a new version, for use with Poisson noise. As a matter of fact, Poisson noise occurs in imagery when the number of elements carrying energy (photon), is too little and so, make their fluctuations, detectable. Recently, Sciacchitano, Dong and Zeng (2015) adapt the ROF model to the case of a Cauchy noise. Like Poisson noise, Cauchy noise results from impulsive degradation and appears in various domains such as, radar and sonar applications, biomedical images or synthetic aperture radar.

[4] Bayesian formulation is known as the Maximum A Posteriori approach.

The first contribution of this Phd is to apply the *Maximum A Posteriori* approach to asymmetric and leptokurtic noise probability distributions such as the *Student*, the *skewed Gaussian* and the skewed Student. We believe that these formulations can be applied to a lot of issues apart finance.

This PhD thesis is based on a collection of submitted or to be submitted in peer-reviewed scientific journals.

We first establish backgrounds and fundamentals of the noise removal field and describe basic definitions in Chapter 2. We give a particular focus on Wavelets, Total Variation techniques and on Quantile regressions. To illustrate those techniques, some numerical simulations are provided. In Chapter 3, we introduce our analysis of volatility spillovers around Earning Announcements. Chapter 4 presents the algorithm we designed for removing an heteroscedastic noise. Chapter 5 introduces the *A Maximum A Posteriori approach to remove asymmetric and leptokurtic noise*. Finally, Chapter 6 gathers conclusions and some open questions.

Chapter 2

Mathematical Tools.

In this chapter, we recall some definitions related to *Wavelet theory* and *Calculus of Variations*. We briefly present how they can be applied to denoising problems and we provide numerical examples.

1 Wavelet theory

At the origin of wavelet theory is an empirical experiment: Jean Morlet was working within the oil company Elf Aquitaine. The signal he was interested in, was composed of high frequency parts with small duration, and low frequency parts with long duration. As the properties of windows Fourier transform, was not suited to its application, as it was smoothing all the relevant information, he built a new transform that preserves singularities and called it : the wavelet transform.

1.1 Theory

Every polynomial functions and more generally, every square integrable function can be decompose into a basis of functions. Well known basis of functions are polynomial and Fourier basis. However, these basis have some limitations as they are not accurate enough to take into account local singularities. When a signal is characterized by different range of frequencies, a common approach to analyse it, is to divide the signal into slices, to assume that each slice is stationary and then to apply a Fourier transform on it.

Definition 1. *The short time Fourier transform, F at the location $t \in \mathbb{R}$, associated to a given signal y, where $y \in L^2(\mathbb{R})$, with respect to a given window function g that is* *compactly supported, is :*

$$F(t,\omega) = \int_{\mathbb{R}} y(s)g(s-t)\exp^{-2\pi i\omega s} ds.$$

However, Fourier analysis has some drawbacks. The time and frequency information cannot be seen at the same time and so a resolution dilemma appears on the window's length: narrow window provides a poor frequency resolution and from a wide window results a poor time resolution. In other terms, we can have either a high resolution in the time domain or in the Fourier domain, but not at the same moment.

A natural improvement is provided by wavelet analysis.

Definition 2. *Let ψ be a complex valued function. ψ is a wavelet if its Fourier transform, $\hat{\psi}$, satisfies:*

$$\int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega = \int_{-\infty}^{0} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega = C_\psi < \infty.$$

From which results: $\int_{-\infty}^{+\infty} \psi(x)dx = 0.$

In the present section we go deeper in the definition.

A wavelet family is a collection of functions obtained by shifting and dilating the graph of mother wavelet.

Definition 3. *Let ψ be a mother wavelet, a wavelet family is a collection of functions $\psi_{a,b}$ such that:*

$$\forall x \in \mathbb{R}, \quad \psi_{a,b}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right).$$

Where $a \in \mathbb{R}$ is the scale parameter, and $b \in \mathbb{R}$, the location parameter.

Functions $\{\psi_{a,b}, \quad a,b \in \mathbb{Z}\}$ form an orthonormal basis of $L^2(\mathbb{R})$.

A strength of wavelet analysis is the ability to capture only the relevant details. For instance, one of the most famous wavelet has been design by the mathematician Alfred Haar in 1909. It serves as an orthonormal basis of functions for $L^2([0,1])$ signals. The *Haar* wavelet mother takes the form:

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Different wavelet families exist associated with different basis of function:

- ▷ *Haar Wavelets*, piecewise base of functions,
- ▷ *Battle-Lemarié Wavelets*, polynomial spline base of functions,
- ▷ *Daubechies Wavelets*, compactly supported base of functions.

For a complete review on wavelet families, see **?**, **?**.

The real wavelet integral transform of signal $y \in L^2(\mathbb{R})$, for a given wavelet mother ψ, is defined by:

$$W_\psi(a,b)y = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \psi\left(\frac{x-b}{a}\right) y(x)dx.$$

We precise that wavelet transform can be adapted to the notion of distributions, introduced later. Notably, if y is a tempered distribution of order m, and if the wavelet ψ is m times differentiable, then the above wavelet transform is well-defined. The original signal f can be recover with the inverse wavelet transform **?**.

Hence, an important property of the wavelet transform, is the ability to preserve local singularities, *ita est* parts where the signal $y \in L^2(\mathbb{R})$ is not differentiable. Various observed signals are characterized by local singularities. We can enumerate: images with rough edges, sunspot time series or extreme financial returns **?**. Local singularities contain substantial information that we want to preserve for analysis purposes.

First, singularities can be related to local polynomial approximations thanks to the Taylor formula and to Lipschitz continuity. In the following, let m be an integer and $m < \alpha < m+1$.

Definition 4. *A function $f : \mathbb{R} \to \mathbb{R}$, is pointwise Lipschitz, $\alpha > 0$ at $v \in \mathbb{R}$ if there exists $K > 0$, and a Taylor polynomial p_v of degree $m = \lfloor \alpha \rfloor$ such that:*

$$\forall x \in \mathbb{R},\ |f(x) - p_v(x)| \leq K|x-v|^\alpha.$$

A function is uniformly Lipschitz $\alpha > m$ on $[a,b]$ if it is pointwise Lipschitz for all $v \in [a,b]$ and if K is independent from v. We precise that the Lipschitz exponent, α may vary from abscissa. The signal f is singular at a point x, if it is not Lipschitz 1 at this point. If the signal f is uniformly Lipschitz in the neighbourhood of v, then one can verify that f is necessary m times continuously differentiable in this neighbourhood. More precisely, if y is one time continuously differentiable at a point x, then it is Lipschitz 1 at this point. If its derivative is discontinuous but bounded, f is still Lipschitz 1 at this point, and so, it is not singular.

Hence, if $0 \leq \alpha \leq 1$, then $p_v(x) = f(v)$ and Definition 4 becomes:

$$\forall x \in \mathbb{R},\ |f(x) - f(v)| \leq K|x-v|^\alpha.$$

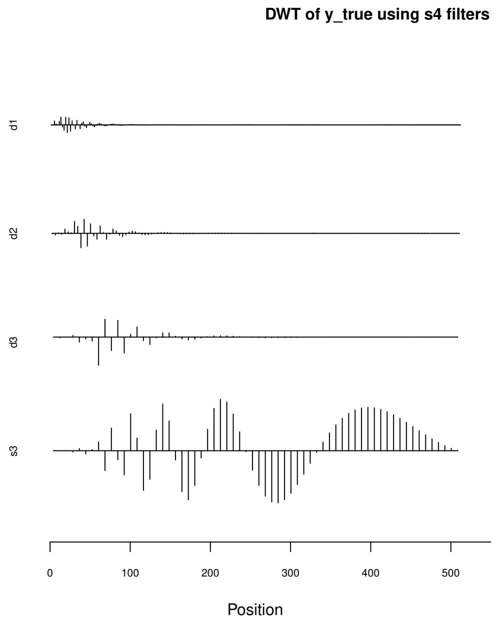


Figure 2.1: Representation of the decomposition of the Doppler signal on a wavelet basis, with a resolution dimension of three, and Daubechies wavelets with four null moments. s stands for the scale function whereas d for wavelet coefficients.

Thanks to the Lipschitz property, the signal can be decomposed as:

$$f(x) = p_v(x) + \varepsilon(x), \text{ with } |\varepsilon(x)| \leq K|x - v|^{\alpha},$$

and the wavelet transform estimates directly the exponent α by ignoring the $p_v(x)$ term. This is done with wavelets having $m > \alpha$ *vanishing moments*, for m a fixed integer.

Definition 5. *A wavelet ψ has m vanishing moments if :*

$$\int_{-\infty}^{+\infty} x^k \psi(x) dx = 0, \text{ for } 0 \leq k < m$$

.

The intuition is that a wavelet with *m vanishing moments* is orthogonal to polynomials of degree $m - 1$ **?**. Moreover, wavelets are useful to detect singularities, notably for Fractal Analysis. This is achieved thanks to the *Modulus Maxima Wavelet Transform.*

When we want a signal to be reconstruct from wavelet coefficients, that are only available at a scale $a < a_0$, then we can still obtained it with the *scaling function*. This function aggregates all wavelets coefficients at scales greater than one. More precisely it is defined by:

$$\phi_a(x) = \frac{1}{\sqrt{a}}\phi\left(\frac{x}{a}\right).$$

Therefore, the low-frequency representation of a signal y at a resolution a is:

$$Ly(a,b) = \langle y(x), \frac{1}{\sqrt{a}}\phi\left(\frac{x}{a}\right)\rangle.$$

The scaling function ϕ associated to *Haar* Wavelets is :

$$\phi(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

Wavelet bases

Multi-Resolution Analysis (MRA) sets the basis for the construction of orthogonal wavelets. The idea behind MRA is to approximate a function y at a resolution level 2^{-j} by a discrete grid of samples that provide local averages of y. The scale is sampled from a dyadic sequence $\{2^j\}_{j\in\mathbb{Z}}$

Definition 6. *A sequence $\{V_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is a multi-resolution approximate if the following six properties are satisfied.*

- ▷ $\forall(j,k) \in \mathbb{Z}^2,\ f(x) \in V_j \equiv f(x - 2^j b) \in V_j$.
- ▷ $\forall j \in \mathbb{Z},\ V_{j+1} \subset V_j$.
- ▷ $\forall j \in \mathbb{Z},\ f(x) \in V_j \equiv f(\frac{x}{2}) \in V_{j+1}$.
- ▷ $\underset{j\to\infty}{lim} V_j = \cap_{j=-\infty}^{\infty} V_j = \{0\}$.
- ▷ $\underset{j\to-\infty}{lim} V_j = closure(\cup_{j=-\infty}^{\infty} V_j) = L^2(\mathbb{R})$.

Various MRA exists : piecewise MRA, spline MRA ect. In practise, wavelet and scale coefficients can be approximated thanks to the *Cascade Algorithm* **?**.

1.2 Wavelet Shrinkage: Diagonal attenuation

As we said, wavelets permit to decompose a signal into rough and fine details. Henceforth they are well-designed for noise removal. We present in this section some results regarding noise removal techniques. We focus on thresholding estimators, that are used in *Chapter 4*. We introduce notions of *Oracle estimator*, *Soft and hard-thresholding*, *Stein Unbiased Risk Estimate* (SURE).

We consider the following problem: we want to recover the original unidimensional signal $y \in L^2(\mathbb{R})$ from noisy observations s.

$$s(x) = y(x) + \varepsilon(x),$$

where for all x in $\mathbb{R}$, $\varepsilon(x) \sim \mathcal{N}(0, \sigma^2)$.

For the next part, $\{x_1, ..., x_N\}$ is an $\mathbb{R}$ discretization.

Oracle Estimators

Oracle estimators permit to define a lower bound of the estimated error introduced by the noise removal procedure. Suppose that the operator D exists such that $\hat{y} = Ds$:

$$Ds = \sum_{n=1}^{N} a_n \left(s(x_n) \right).$$

It comes that the operator D is linear when a_n is independent from s.

For a given signal, we want to find the coefficient a that minimizes the estimation error $r(D, y)$, where

$$r(D, y) = \mathbb{E}\left[|y - \hat{y}|^2 \right] = \sum_{n=1}^{N} \mathbb{E}\left[|y(x_n) - a_n(s(x_n))|^2 \right]. \tag{2.1}$$

By developing Equation 2.1, it comes that $\mathbb{E}\left[|y(x_n) - a_n(s(x_n))|^2\right] = y(x_n)^2(1 - a_n)^2 + a_n^2\sigma^2$. The risk is minimum for $a_n = \frac{y(x_n)^2}{y(x_n)^2 + \sigma^2}$. This Oracle estimor generates a denoising risk of:

$$r_{or}(y) = \sum_{n=1}^{N} \frac{y(x_n)^2}{y(x_n)^2 + \sigma^2}$$

.

We remark that this risk depends directly on the unobserved true signal y and so, it is a theoretical result serving as a lower bound for the risk estimation. Selecting the following non-linear projector:

$$a_n = \begin{cases} 1 \text{ if } |y(x_n)| > T \\ 0 \text{ else} \end{cases}$$

leads to a risk of the oracle projection of :

$$r_{or}(y) = \sum_{n=1}^{N} min(y(x_n)^2, \sigma^2).$$

Threshold Estimators

Two threshold functions exist: the *hard-threshold* and the *soft-threshold*. These two functions recover the original signal by either, setting to zero all coefficients below a given threshold, and therefore assessing that all large wavelet coefficients represent the singularities of the original signal; or by diminishing the amplitude of large wavelet coefficients by the threshold.

Hard-threshold

The hard-threshold function $\rho_{hard} : \mathbb{R} \to \mathbb{R}$ is defined by:

$$\text{for } u \in \mathbb{R}, \quad \rho_{hard}(u) = \begin{cases} u \text{ if } |u| > T \\ 0 \text{ else.} \end{cases}$$

The risk associated to the hard-threshold can be calculated:

$$\mathbb{E}\left[(y(x_n) - \rho_{hard}(s(x_n)))^2\right] = \begin{cases} |\varepsilon_n|^2 \text{ if } |s(x_n)| > T \\ |y(x_n)|^2 \text{ else} \end{cases}$$

Hence, we observe that the risk generated by an hard-thresholding is greater than the Oracle risk:

$$r_{hard}(y) \geq r_{or}(y).$$

Soft-threshold

Similarly, the risk associated to a soft-thresholding is bounded by below by the Oracle risk. Indeed, as we previously said, soft-thresholding consists in reducing the amplitude of wavelet coefficients $\langle \psi_{j,k}, s \rangle$ by the threshold value. Hence, it ensures that estimated coefficients $\langle \psi_{j,k}, \hat{y} \rangle$ are just below the true coefficients. This permits to cancel sharp variations of s induced by the noise.

The soft-thresholding function is defined by :

$$\text{for } u \in \mathbb{R}, \quad \rho_{soft}(u) = \begin{cases} u - T \text{ if } u > T \\ u + T \text{ if } u < -T \\ 0 \text{ else.} \end{cases}$$

With $|\rho_{soft}(s(x_n))| \leq |y(x_n)|$.

Upper-Bound Error

The risk associated with thresholding estimator is close the one of an oracle projector.

Theorem 1. *Let* $T = \sigma\sqrt{2\log_e N}$*, the risk* r *of a soft or hard-thresholding estimator satisfies,* $\forall\ N > 4$*,*

$$r(y) \leq (2\log_e + 1)(\sigma^2 + r_{or}(y))$$

A proof of Theorem 1 is available in **?**.

The presence of N in the universal threshold expression is linked to the Gaussian assumption on the additive noise: the larger the number of observations, the larger the noise amplitude. The challenge in the choice of the threshold is to not delete signal. Therefore, a rule of thumb is to set T just above the largest amplitude of the noise. It has been proved that the maximum noise amplitude has an high probability of being just below $T = \sigma\sqrt{2\log_e N}$.

Threshold improvement

A natural improvement consists in finding the threshold minimizing the reconstruction error. However, we discussed previously that the lower bound of the reconstruction error, the Oracle risk, depends directly on the signal energy. Henceforth, a challenge consists in estimating the Oracle risk.

In the framework of soft-thresholding, if $|s(x_n)| < T$, then the risk equals $|y(x_n)|$. Knowing that $\mathbb{E}\left[|s|^2\right] = |y|^2 + \sigma^2$, the signal energy, $|y|^2$ is given by:

$$|y|^2 = \mathbb{E}\left[|s|^2\right] - \sigma^2$$

A robust estimator of the Oracle risk is given by the *Stein Unbiased Risk Estimator* (SURE) **?**:

$$SURE(s, T) = \sum_{n=1}^{N} C(s(x_n)), \tag{2.2}$$

with, in case of a soft-threshold:

$$\text{for } u \in \mathbb{R}, \quad C(u) = \begin{cases} u^2 - \sigma^2 \text{ if } u \leq T \\ u^2 + T^2 \text{ if } u > T \end{cases}$$

Donoho and Johnstone (1994) prove that the SURE estimator is an unbiased estimate of the Oracle risk. Hence, a natural manner to improve the threshold procedure, is to select the threshold $\hat{T}$, minimizing the SURE estimate **?**:

$$\hat{T} = \underset{T \in \mathbb{R}^+}{\operatorname{argmin}} \; SURE(s, T).$$

But, even if the SURE is an unbiased estimate of the Risk, its variance remains unknown. As a result, when the signal energy is lower than the noise variance, that is:

$$||y||^2 < N\sigma^2,$$

the optimal threshold may be to small. In this particular case, Mallat (2000) suggests to use an hard-threshold **?**.

We provide a denoising example on the *HeaviSine*, introduced by Donoho and Johnstone (1994) and defined as follows **?**:

$$y(x_n) = 4\sin(4\pi x_n) + \operatorname{sign}(x_n - 0.3) - \operatorname{sign}(0.72 - x_n),$$

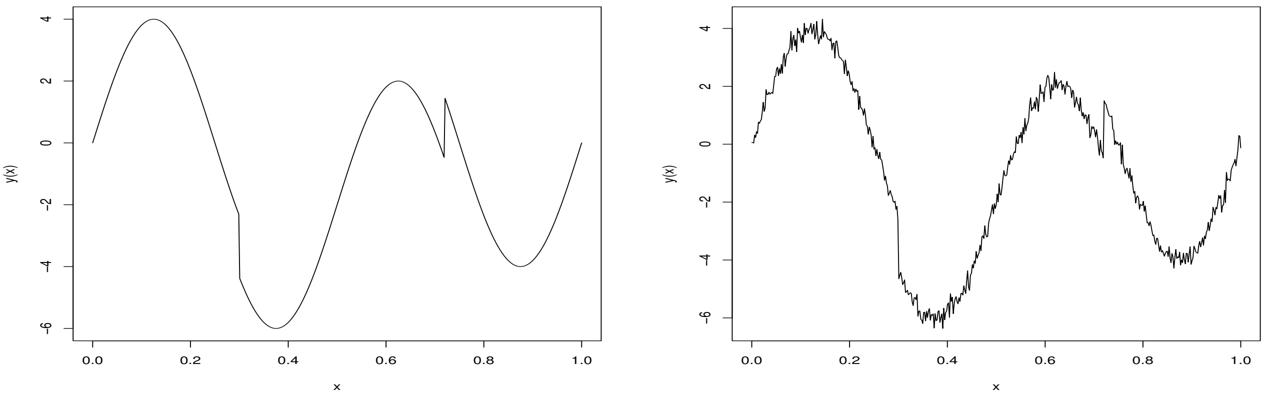
for x_n forming a $[0, 1]$ discretization.

This signal has two jumps around $x_n = 0.3$ and $x_n = 0.7$, and is depicted in Figure 4.1. Therefore, we compute the estimate of the reconstructed error for standards filters and thresholds. They are depicted in Figure 2.3. As a mother wavelet, we select the *Daubechies* wavelet, with four null moments.

Regarding the obtained risk estimates, we observe that the *Universal Threshold*, for both *Hard filtering* and *Soft filtering* lead to higher MSE. In this brief example *SureShrink* using a *Hard Filter* lead to the lowest reconstruction error. Moreover, Figure 2.4 depicts well the distinction between the attenuation of wavelet coefficients, induced by a soft- filtering and the cancellation of wavelet coefficients that are below the threshold. Indeed, by reducing the amplitude of wavelet coefficients, *Soft-filtering* smooths the two singularities of the HeaviSine function. On the contrary, the *Hard filter* by keeping all wavelets above the threshold, the singularities of the signal are preserved.

2 Total Variation techniques

Total Variation techniques consider signals of bounded variation and allow original signals to have discontinuities **?**, **?**, **?**.



(a) Original Signal (b) Noisy Signal

Figure 2.2: Original and Noisy representations of the Heavisine function, for a signal-to-noise ratio of 19. The HeavySine is characterized by singularities at $x_n = 0.3$ and $x_n = 0.7$.

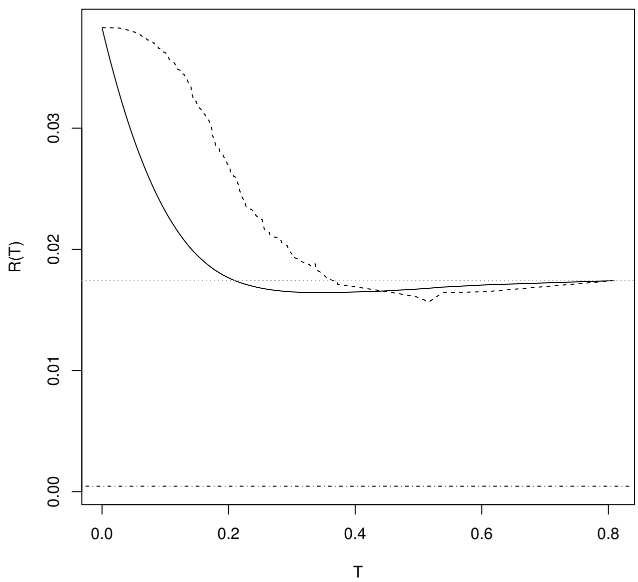
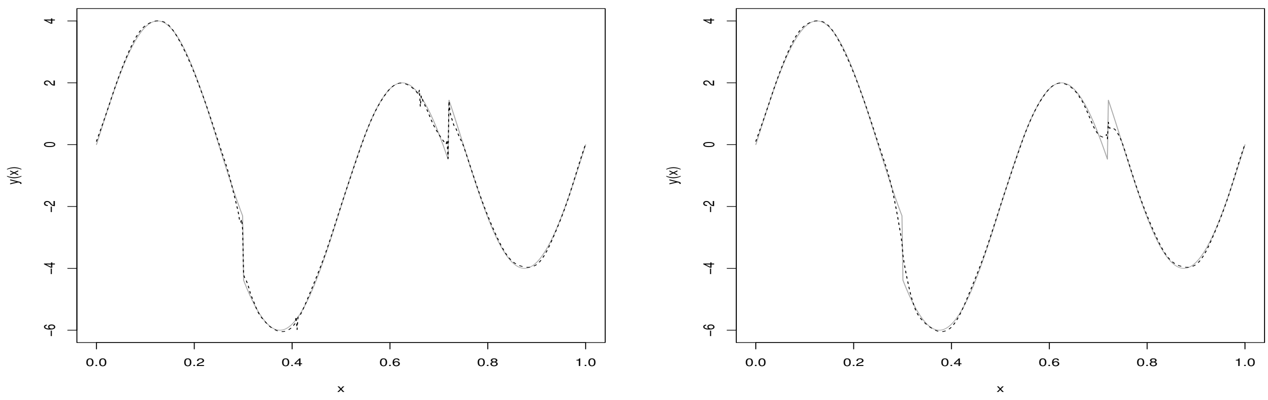


Figure 2.3: Representation of the *Mean Squared Error*, in function of the threshold T, for the *Universal Threshold*, dotted lines, *hard-threshold*, dashed lines, *soft-threshold*, solid line. We also graph the *Mean Squared Error* associated to the *Oracular Threshold*.



(a) Hard Filter (b) Soft Filter

Figure 2.4: Original and Estimates of the true signal using *Universal Threshold*, for *Hard* and *Soft* filters. Dotted black lines are estimated signals and grey lines are the original ones.

Total variation denoising were developed for image denoising but are used in various other contexts such as image deblurring, super resolution, inpainting or scale detection **?**. One explanation for this success is the simplicity of the Total Variation operator and its ability to penalize highly oscillating patterns (noise) while preserving the signal regularity. Besides, Total Variation methods are simple enough to produce few artefacts and can be applied to signals of various dimensions. The standard denoising problem consists in recovering a signal y disrupted by an additive noise: ε[1].

$$s(x) = y(x) + \varepsilon(x),$$

where $\forall\ x$, $\varepsilon(x)$ follows a known probability distribution, $s \in L^2(\mathbb{R})$ and $y \in L^1(\mathbb{R})$. In this part, we introduce notions of *distributions*, *Radon measure*, *bounded variation space*, and the *bounded variation semi norm.*

Distribution theory was introduced to generalize the notions of function and derivative to the case of discontinuous functions. The idea behind the notion of distribution, is that, to know the behaviour of a function f it is enough to know the value of the following integral: $\int_{\mathbb{R}} f(x)\phi(x)dx$, for a set of well-defined of functions ϕ. In the noise removal context, distribution theory allows the signals to be recovered, to have some discontinuities. They will be assume to be in $L^1(\Omega)$, and no more in $C^1(\Omega)$.

Functions ϕ are called *test functions* and are C^∞ with compact support. The set of

[1]Again, we consider only the case of a unidimensional signal.

such functions is denoted $\mathcal{D}$.

Definition 7. *The set of applications T linear, continuous and definite for all test functions and taking value in $\mathbb{C}$, is called the set of distributions, denoted $\mathcal{D}'$.*
The distribution T is such that, $\forall\ \phi$ in $\mathcal{D}$, $\phi \rightarrow < T, \phi >$, with $< T, \phi >$ in $\mathbb{C}$.

A common distribution is the Dirac. The Dirac distribution in a is defined by:

$$< T, \phi >= \phi(a).$$

T is denoted δ_a for :

$$\delta_a(x) = \begin{cases} +\infty \text{ if } x = a \\ 0 \text{ otherwise} \end{cases}$$

Moreover, a distribution is said to be *regular*, if it exists a function f locally summable ($f \in L^1_{loc}(\mathbb{R})$) on a bounded interval, such that:

$$\forall \phi \in \mathcal{D} \quad < T, \phi >= \textstyle\int_{\mathbb{R}} f(x)\phi(x)dx.$$

The set of distributions forms a vectorial space.

The notion of derivability can be extended to distributions, and thanks to them, non-differentiable functions can be differentiated. We are interested in the following quantity:

$$\forall \phi \in \mathcal{D} \quad < T'_f, \phi >= \textstyle\int_{\mathbb{R}} f'(x)\phi(x)dx.$$

By integrating by part, we obtain:

$$< T'_f, \phi >= [f(x)\phi(x)]_{-\infty}^{+\infty} - \int_{\mathbb{R}} f(x)\phi'(x)dx.$$

As ϕ are with compact support, then $[f(x)\phi(x)]_{-\infty}^{+\infty} = 0$ and we come up with the *Distributional derivative* notion.

Definition 8. *Every distribution T is derivable, and its derivative is the distribution T', defined by:*

$$\forall \phi \in \mathcal{D} \quad < T'_f, \phi >= -\textstyle\int_{\mathbb{R}} f(x)\phi'(x)dx.$$

Distributions generalize Radon measures. Let $\mathcal{K}$ be the vectorial space of functions $\phi \in \mathbb{D}$, with compact support: they are null outside an interval $[a, b]$. Contrary to $\mathcal{D}$, $\mathcal{K}$ is a metrisable space, hence sequential. $\mathcal{K}$ has the following notion of convergence: A sequence (ϕ_n) of elements in $\mathcal{K}$, converges toward ϕ if:

- ▷ It exists an interval $[a, b]$ such that, $\forall n, \forall x \notin [a, b] : \phi_n(x) = 0$.
- ▷ The sequence (ϕ_n) converges uniformly toward ϕ.

The notion of bounded variation is, in this particular setting, linked to the *Radon measure.*

Definition 9. *A Radon measure is an application, linear and sequentially continuous:* $\mu : \phi \in \mathcal{K} \to \mu(\phi) =< \mu, \phi >\in \mathbb{C}$*, verifying:*

- ▷ $< \mu\lambda\phi + \Psi >= \lambda < \mu, \phi > + < \mu, \Psi >$.
- ▷ *For a sequence* (ϕ_n) *of elements in* $\mathcal{K}$ *converging toward* ϕ *in* $\mathcal{K}$*,* $\underset{n\to\infty}{lim} < \mu, \phi_n >=< \mu, \phi_n >$.

Radon measures are locally finite, and inner regular, which permits to introduce the notion of function in a space of bounded variation (BV).

Definition 10. *A function* f *is in* $BV(\Omega)$*, where* Ω *is an open subset of* $\mathbb{R}$*, if it is integrable, and there exists a Radon measure* Df *such that:*

$$< Df, \phi >= -\int_\Omega f(x)\phi'(x)dx, \;\; with\ \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}).$$

When f is smooth, $Df(x) = \nabla f(x)$ and the *total variation semi norm* of f is:

$$||f||_{TV(\Omega)} = \int_\Omega |Df| := \sup\left\{\int_\Omega f(x)\phi'(x)dx, \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}), \forall x \in \Omega, |\phi(x)| \leq 1\right\}$$

When f is smooth, TV is equivalently the integral of its first derivative magnitude:

$$||f||_{TV(\Omega)} = \int_\Omega |f'(x)|dx.$$

In the sequel, the quantity $|Df|$ will be denoted the total variation of f.

Moreover, the two following properties provide some intuition on the material used for the demonstration of the existence and the uniqueness of a solution to the energy minimization problem defined above[2]. They permit to apply the Direct Method in the calculus of variation **?**.

Property 1. *(Sequential Lower SemiContinuity) Let f_n be any sequence in $BV(\Omega)$ such that $f_n \to f$ in $L^1(\Omega)$ then:*

$$\int_\Omega |f'(x)|dx \leq \lim\inf_{n\to\infty} \int_\Omega |f_n'(x)|dx.$$

Property 2. *(Approximation by smooth functions) If f in $BV(\Omega)$ then there exists a sequence $(f_n)_{n\in\mathbb{N}}$ of $C^\infty(\Omega)$ functions such that:*

$$f_n \to f, \ in \ L^1(\Omega),$$

and

$$\int_\Omega f_n'(x)dx \to \int_\Omega f'(x)dx.$$

Before introducing the main denoising model based on TV, we present the following variational model where the energy to minimize is only composed of the noise log-likelihood **?**.

Log-Likelihood Estimator

Suppose the law of the noise is known, then an approach to recover the original signal y is simply to maximize the noise log-likelihood.

We recall that the likelihood L associated with the probability distribution P is :

$$L(x_1, ...x_N, \theta) = \prod_{n=1}^{N} P(X = x_n|\theta),$$

Where $P : \mathbb{R} \times \Theta \to [0, 1]$ is the probability density function, $\theta \in \Theta$ is a parameter vector and $\{x_1, ...x_N\}$ is a discretization of $[a, b]$. We look for a minimizer $y^\star$ of :

$$-\sum_{n=1}^{N} log\left(P\left(X = s(x_n) - y(x_n)|\theta\right)\right),$$

that stands for the discrete version of the functional:

$$-\int_a^b H_\theta(x, y, \dot{y})dx. \tag{2.3}$$

[2]For an interested reader, we refer to **?** and to **?**.

with $H_\theta\left(x,y,\dot{y}\right) = log\left(P(X = s(x) - y(x)|\theta)\right).$

For the particular case where $\forall\ x$, $\varepsilon(x) \sim \mathcal{N}(0,\sigma^2)$, Equation 2.3 becomes:

$$J[y] = \frac{1}{2}log(2\pi\sigma^2) + \frac{1}{2\sigma^2}\int_\Omega \left(s(x) - y(x)\right)^2 dx \tag{2.4}$$

According to the *Direct Method of Variational Calculus*, a solution to Equation 2.4 must satisfies Euler-Lagrange equations:

$$\frac{\partial H}{\partial y} - \frac{d}{dx}\frac{\partial H}{\partial \dot{y}} = 0,$$

The log-likelihood approach only focuses on the *a priori* knowledge on the noise distribution, and thus, does not take into account the knowledge we have on the behaviour of the true signal y. A natural improvement is therefore to incorporate any *a priori* we have on y.

Tichonov regularization

A first popular approach to recover an original signal perturbed by an additive noise was introduced by Tichonov in 1963 **?**. This approach consists in penalizing large oscillations, led by the noise, through the L2 norm of the gradient. More precisely, the restored signal $\hat{y}$ is solution of the following convex problem:

$$J[y] = \frac{1}{2}||y - x||_2^2 + \frac{\lambda}{2}\int_\Omega |\dot{y}(x)|^2 dx, \tag{2.5}$$

where $\lambda > 0$ is a regularization parameter.

This problem has the advantage to be easily solve, with for instance, Euler-Lagrange equations:

$$(y - x) + \lambda\dot{y}(x) = 0.$$

However, the use of the L2 norm prevents the original signal y to have discontinuities, and so, non-smooths approaches are more appropriate.

Rudin, Osher, Fatemi model

Rudin, Osher and Fatemi (1992) are the first to develop a noise removal algorithm based on variational calculus including Total Variation. Even if the model they propose is built for denoising problems, a generalized version of it permits to deal with other signal analysis problematics. A total variation approach consists in minimizing an energy composed by a data fidelity term and by a data regularity constraint. The theoretical demonstration of ROF has been done in Lions and Chambolle (1997) **?**.

The objective of the ROF approach is to find a function y in $BV(\Omega)$, minimizing the following energy:

$$J[y] = \frac{\lambda}{2}||s-y||_2^2 + \int_\Omega |\dot{y}(x)|dx. \tag{2.6}$$

The data fidelity term ($||s-y||_2$) ensures the closeness with the true signal, while the data regularity term ($\int_\Omega |\dot{y}(x)|dx$) ensures the solution to be in $BV(\Omega)$. In the case of impulse noise, it can be replace by a L1 norm. The parameter λ controls for the signal regularity.

The ROF model can easily be transposed into a Bayesian framework. In this context, the formulation belongs to the class of *Maximum A Posteriori* problems, and the aim is therefore to select among a set of signals, the most probable under a posterior probability distribution, $P(Y|S)$. From Bayes's rule, we know that:

$$P(Y|S) = \frac{P(S|Y)P(Y)}{P(S)},$$

and as maximizing $P(Y|S)$ is equivalent to minimize $-P(Y|S)$, our problem consists in solving:

$$-\log\left(P(Y|S)\right) = -\log\left(P(S|Y)\right) - \log\left(P(Y)\right) + \log(P(S)).$$

As $\log(P(S))$ does not depend on Y, we focus on $-\log\left(P(S|Y)\right) - \log\left(P(Y)\right)$. The term $P(Y)$ contains any *a priori* we have on y behaviour, and as in a total variation context, $y \in BV(\Omega)$, the Gibbs's prior is selected:

$$P(Y) = \exp\left(-\mu \int_\Omega |\dot{y}(x)|dx\right),$$

where $\mu > 0$ stands for a regularization parameter.

For in the MAP formulation, the ROF model becomes:

$$J[y] = \frac{1}{2\sigma^2}||s-y||_2^2 + \mu \int_\Omega \mu|\dot{y}(x)|dx. \tag{2.7}$$

These two formulations are equal for $\frac{1}{\lambda} = \mu\sigma^2$.

Problem Solving

To derive a solution to equation 2.7, we rewrite the problem as follows:

$$J[y] = \int_{\Omega} H(x, y(x), \dot{y}(x))dx.$$

For $H(x, y, \dot{y}) = \frac{\lambda}{2}(s(x) - y(x))^2 + |\dot{y}(x)|$.
A minimizer $y^\star$ of the functional J must satisfies the *Euler-Lagrange* equation:

$$\frac{\partial H}{\partial y} - \frac{\partial}{\partial x}\frac{\partial H}{\partial \dot{y}} = 0 \tag{2.8}$$

However, the ROF model has two limitations.

Staircasing effect

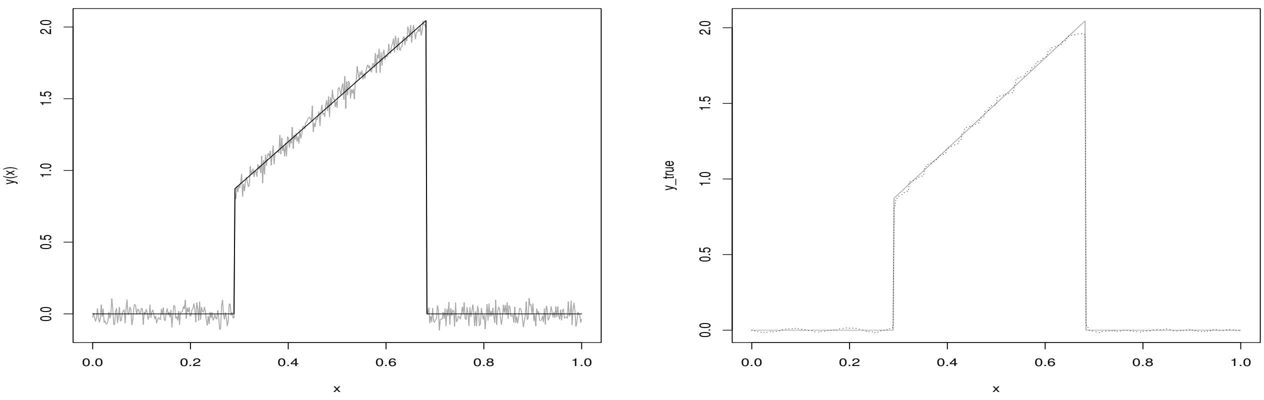
First, TV methods are characterized by the *staircasing effect*. This can be illustrated through an affine one dimensional signal perturbed by a Gaussian noise, illustrated in Figure 2.5. The estimate obtained will not be smooth but piecewise constant, and so, the methodology fails in recovering the homogeneity of the original signal, this comes from the non-differentiability of the L1 norm in 0. Hence, when the original signal is constant, the ROF methodology leads to poor approximation results. A natural improvement consists in replacing the L1 norm in the Total Variation part by a more general function called *potential functions* (PFs). Nikolova (2005) is the first to set up a generalization of regularization functions **?**. $|\dot{y}(x)|$ is generalized to the potential function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$. These functions should be chosen such as edges in images or breaking points are preserved. In our dissertation, we consider this particular PF: $\phi_\beta(t) = \sqrt{t^2 + \beta^2}$[3] The parameter β is strictly positive.

Regularization parameter selection

A main challenge is to select the regularization parameter λ. When the noise probability distribution is Gaussian, then a natural may to select λ is through a SURE approach. A first range of methods consist

In the particular case, where the noise is Gaussian, a natural way of choosing λ is through the *Stein Unbiased Risk Estimate* (SURE) **?**. Indeed it provides a robust manner of estimating the risk $\mathbb{E}\left[(y - \hat{y})^2\right]$. When the noise is Gaussian distributed, an estimate of the risk, that does not depend on the unknown signal y, can be derived.

[3]Other forms of potential functions ϕ, can be found in **?**.



(a) Original and Noisy signals

(b) Original and Denoised signals

Figure 2.5: Illustration of the staircasing effect. On the right, the original triangle-shaped signal disrupted by an additive Gaussian noise of variance 0.05. In the ROF result (right figure), the noise has been removed but a staircasing effect is visible on the slope. This example is taken from **?**.

Indeed, let $\hat{y}$ be an estimate of y, then:

$$\begin{aligned}
\mathbb{E}||y-\hat{y}||_2^2 &= \mathbb{E}||y+s-s-\hat{y}||_2^2 \\
&= -|\Omega|\sigma^2 + ||s-\hat{y}||_2^2 + 2Cov(s,\hat{y}) \\
&= -|\Omega|\sigma^2 + ||s-\hat{y}||_2^2 + 2\sigma^2 \sum_{x_n} \frac{\partial \hat{y}(x_n)}{\partial s(x_n)}
\end{aligned}$$

The term σ^2 is the true or estimated noise variance, and $|\Omega|$ is the number of elements in the discretization of Ω.

Aside from giving the risk of an estimator, the SURE estimate is useful for selecting the regularization parameter, $\lambda \in \Lambda$, associated to any model.

In this framework, selecting λ amounts to solve the following minimization problem:

$$\lambda^\star = \underset{\lambda\in\Lambda}{\operatorname{argmin}} \; -|\Omega|\sigma^2 + ||s-\hat{y}_\lambda||_2^2 + 2\sigma^2 \sum_{x_n} \frac{\partial \hat{y}_\mu(x_n)}{\partial s(x_n)}$$

Variations of the SURE have been proposed such as Stein Unbiased Gradient Estimator of the Risk (SUGAR) or Generalized Stein Unbiased Risk Estimator (GSURE) **?**, **?**.

An other automatic selection procedure is the *discrepancy principle*. This approach consists in selecting λ so that the data fidelity term: $\int_\Omega (s(x)-y(x))^2\, dx$ matches the noise variance: σ^2 **?**, **?**. The ROF procedures is expressed as follows.

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \, ||y||_{TV} \text{ subject to } \int_\Omega (s(x) - y(x))^2 \, dx = \sigma^2 |\Omega|.$$

But, as underlined by Getreuer (2012), the discrepancy principle is known to slightly overestimate the Mean Squared Error **?**. As a result, Chambolle (2004) propose an algorithm to select λ so that it equals approximately σ^2 **?**.

However, these two methods depend directly on the noise probability distribution, which limit their applications to signal disrupted by noise following other probability distributions.

Recently, Deledalle et al. propose an innovative manner of selecting the regularization parameter without any a priori on the noise probability distribution. However, they consider a MAP formulation of the ROF model. Using the Karush-Khun-Tucker condition, they propose a closed form formula to get μ **?**. In this setting, the computed value corresponds to the minimum value of the regularization parameter above which the solution remains constant. See **?** for more details on the procedure.

ROF extensions

Lastly, various extensions of the ROF model exist. The inclusion of higher derivatives in the ROF model permits to better recover the curvature of the original signal. To our knowledge, various formulations exists and for a complete review, we refer the reader to **?**. These methods are therefore well-designed to treat the staircasing effect.

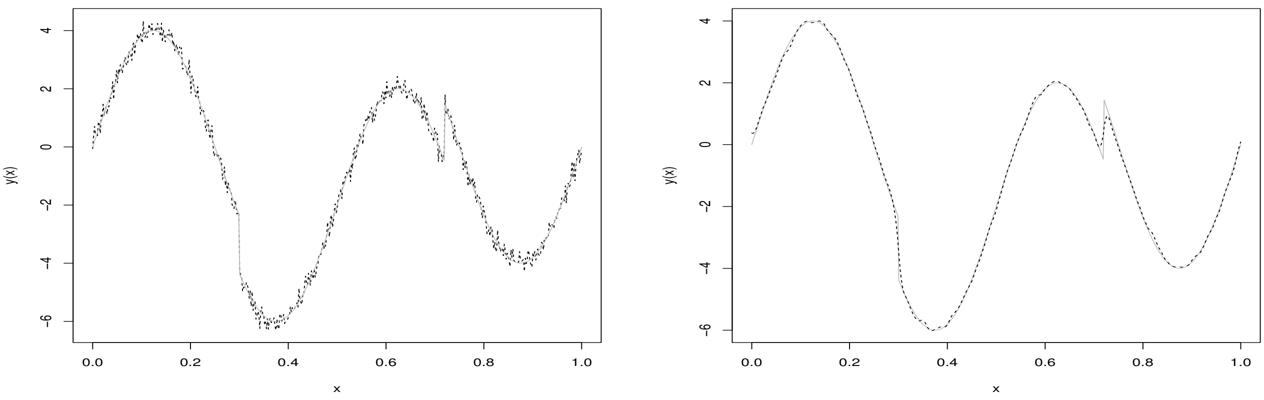
Also, non-local ROF extensions have drawn attentions. These methods are based on a simple observation: a signal can have similar regions, but far from each others. The neighborhood filter is therefore any filter which restores a element of a signal by taking an average of the values of neighboring elements with a similar pattern **?**.

Numerical approximation

To solve numerically *Euler-Lagrange* equations, we embed Equation 2.8 into an dynamical equation.

$$\begin{aligned} & \tfrac{d}{dj} y = -\tfrac{\partial}{\partial y} H + \tfrac{d}{dx} \tfrac{\partial}{\partial \dot{y}} H \\ \Leftrightarrow \quad & y_{j+1} = y_j + \delta \left(-\tfrac{\partial}{\partial y_j} H + \tfrac{d}{dx} \tfrac{\partial}{\partial \dot{y}_j} H \right). \end{aligned}$$

As a result, at each iteration of the algorithm, an estimate of y, denoted y_j is obtained: The parameter $\delta > 0$ controls for the speed of convergence. The initial condition can be set arbitrarily to s or to $\frac{1}{|\Omega|} \int_\Omega s(x) dx$. Iterations can be stopped when a given



(a) Log-likelihood (b) MAP

Figure 2.6: Original and Estimates of the true signal using *Log-likelihood* and MAP models. Dotted black lines are estimated signals and grey lines are the original ones. The number of iterations is set to 100, and the starting point of the numerical approach is set to $\dfrac{1}{card([a,b])}\displaystyle\int_{\Omega} s(x)dx$.

convergence criterion has been reached [4].

In most cases, observed signals are discrete, and so gradients must be approximated. We recall here some finite difference methods.

Let $\dot{y}$ be the first derivative of y with respect to x. The purpose of finite difference method is to approximate $\dot{y}$. From now on $\{x_1, ..., x_n\}$ forms a discretization of Ω, and $\dot{y}(x_n)$ denotes the derivative at point x_n.

▷ central

$$\dot{y}(x_n) = \frac{y(x_{n+1}) - y(x_{n-1})}{2}.$$

▷ upward

$$\dot{y}(x_n) = y(x_{n+1}) - y(x_n).$$

▷ backward

$$\dot{y}(x_n) = y(x_n) - y(x_{n-1}).$$

[4]For instance, iterations can be stopped when $J[y]$ is stable enough.

3 Quantile Regressions

Even if quantile regressions have been theorized in the late seventies by Roger Kœncker, the underlying concept exists for centuries. At the end of the 18th century, a Croatian priest proposed a general algorithm to minimize absolute deviances in order to compute a median regression, which appears as an alternative to the so-called ordinary least squares. While ordinary least squares focus on a move in a variable given a move in expectation of an explanatory variable, quantile regressions study the move in a variable given a move in a chosen part of the whole conditional probability distribution. It has the advantage to discriminate the relation between two variables in the whole conditional distribution, and is employed is various domains. For instance, in the french *Carnet de Santé*, the weight-age curve is built using quantile regression. Indeed, conditionally to each age is given the possible weights at quantiles 25, 50 and 75. Moreover, quantile regressions are used in several fields of finance to measure relations between returns and volumes traded or to quantify volatility spillovers **?**, **?**.

3.1 Basics

Quantile regressions directly depend on the concept of *quantiles*.

Definition 11. *Let X be a random variable with cumulative density function $F_X(x) = P(X \leq x)$, the $\tau - st$ quantile of X is defined by:*

$$Q_X(\tau) = \inf x : F_X(x) > \tau,$$

where $0 < \tau < 1$ is the quantile level.

Known cases are: the median $Q_X(0.5)$, the first quartile $Q_X(.25)$, the third quartile $Q_X(.75)$. For example, in finance, a particular attention is given to extreme quantiles of the financial return distribution, as they corresponds to extreme events. It is easy to observe that $Q_X : (0,1) \to \mathbb{R}$, is a non-decreasing function of τ.

Quantile regressions use the notion of conditional quantiles.

Definition 12. *Let X be a random variable, explained by the random variable Y and let $F_{X|Y}(x|y) = P(X \leq x|Y = y)$ be the conditional cumulative density function. Then the $\tau - th$ conditional quantile is defined by:*

$$Q_{X|Y}(\tau) = \inf \left\{ x : F_{X|Y}(x|y) > \tau \right\},$$

where $0 < \tau < 1$ is the quantile level.

The standard ordinary least squares (OLS) model is defined as follows:

$$X = Y'\beta + \Sigma,$$

where $X = (x_1, ..., x_n)$, $Y = (y_1, ..., y_n)$, $\beta = (\beta_0, \beta_1)$, $\Sigma = (\varepsilon_1, ..., \varepsilon_n)$ such as $\mathbb{E}[\Sigma] = 0$.

In this context, $\mathbb{E}[X|Y] = Y'\beta$, and so the OLS, by paying attention on the conditional relation in mean, hides relations occurring in other parts of the conditional distribution. The aim of quantile regression is therefore to depict the relation between two random variables in the whole conditional distribution.

$$X = Y'\beta(\tau) + \Sigma(\tau), \tag{2.9}$$

where $Q_{X|Y}(\tau) = Y'\beta(\tau)$.

As in the OLS model, $\beta_0(\tau)$ corresponds to the intercept. $\beta_1(\tau)$ models a marginal change in the $\tau - th$ quantile of X resulting from a marginal change in Y.

Quantile regressions have two major advantages.

First, it allows to study the impact of a predictor on different quantiles of the conditional distribution, and thus, provide a complete picture on the relation between two variables. For example, if one is interested in sailing, he/she can studies if large waves occurs more in the Atlantic ocean during winter.

Second, by using the absolute deviation rather than the squared deviation, quantile regressions are robust to outliers, which is not the case for the OLS regression. Indeed, a great attention is given to the sensitivity of an estimator to the presence of outliers. An estimator is robust, when it is not influenced by the presence or absence of outliers. Particularly, the sensitivity of median, like other quantiles, to the presence of outliers, is bounded, which is of great interest. We precise that this bound is independent from the distance of the outlier **?**.

Lastly, the inference results concerning quantile regressions are independent from the noise distribution.

3.2 Estimation

The estimation of quantile regressions differs from the estimation of the OLS regression, in the sense that absolute deviations are considered instead of squared deviations.

For instance, to estimate the median ($Q_X(0.50)$) of the vector of realizations X, one has to solve the following optimization problem:

$$Q_X(0.50) = \underset{b\in\mathbb{R}}{\text{argmin}}\ \mathbb{E}|X-b|,$$

and the sample median solves $\min_{b\in\mathbb{R}} \sum_n |x_n - b|$. Equivalently, if we are interested in the estimation of $Q_{X|Y}(0.50)$ such as:

$$Q_{X|Y}(0.50) = Y'\beta(0.50),$$

then an estimate $\hat{\beta}(0.50)$ of $\beta(0.50)$ verifies:

$$\hat{\beta}(0.50) = \min_{\beta(0.50)\in\mathbb{R}} \sum_n |x_n - y_n'\beta(0.50)|.$$

This can easily be generalized to all quantiles $\tau \in (0,1)$:

$$Q_X(\tau) = \underset{b\in\mathbb{R}}{\text{argmin}}\ \mathbb{E}|X-b|,$$

however, as the L1 norm is not differentiable in 0, the following form of the above equation is implemented instead:

$$Q_X(\tau) = \underset{b\in\mathbb{R}}{\text{argmin}}\ \mathbb{E}\phi_\tau(X-b),$$

with : $\phi_\tau : \mathbb{R} \to \mathbb{R}+\ 0$, $u \to \phi_\tau(u) = u(\tau - \mathbf{1}_{u<0})$. ϕ_τ is referred as the *quantile loss function.*

A conditional quantile estimate $\hat{Q}_{X|Y}(\tau) = Y'\hat{\beta}(\tau)$, solves:

$$\hat{Q}_{X|Y}(\tau) = \min_{\beta(\tau)\in\mathbb{R}^2} \mathbb{E}\phi_\tau(X - Y'\beta(\tau)), \tag{2.10}$$

To solve Equation 2.10, we transform it into a linear programming problem **?**:

$$x_n = \hat{Q}_{X=x_n|Y=y_n}(\tau) + \varepsilon_n = y_n'\beta(\tau) + (u_n - v_n),$$

with $u_n = \varepsilon_n \mathbf{1}_{\varepsilon_n>0}$ and $v_n = |\varepsilon_n|\mathbf{1}_{\varepsilon_n<0}$.

Therefore, we come up with:

$$\min_{\beta(\tau)\in\mathbb{R}^2} \sum_n \phi_\tau\left(x_n - y_n'\beta(\tau)\right) \leftrightarrow \begin{cases} \min\limits_{\beta(\tau)\in\mathbb{R}^2, U\in\mathbb{R}_+^N, V\in\mathbb{R}_+^N} \tau U + (\tau-1)V \\ \text{such that } X - Y'\beta(\tau) = U - V, \end{cases} \tag{2.11}$$

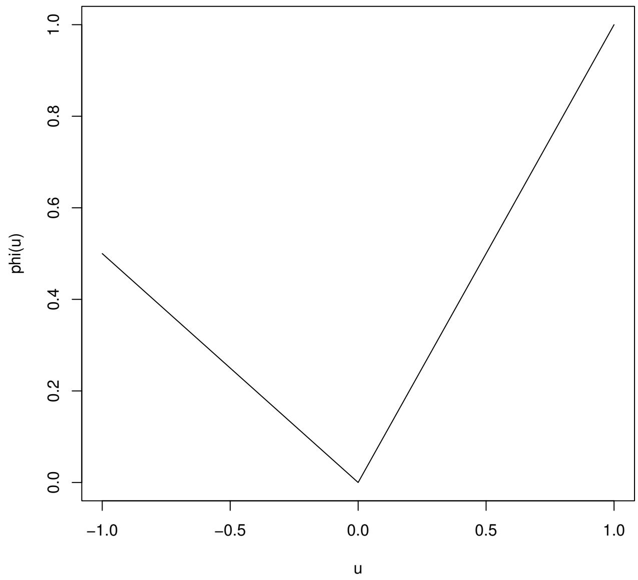


Figure 2.7: Representation of the ϕ_τ for $\tau = 0.5$

where $U = (u_1, ...u_N)$ and $V = (v_1, ...v_N)$.

Various algorithms exist to solve the above problem: the Simplex Method, Frisch-Newton Interior Point Method or others. For the interested reader, we refer he/she to **?**.

3.3 Statistical Properties

The coefficient estimate, $\hat{\beta}$ is characterized by two properties: consistency and asymptotic normality. Following He and Portnoy (1997), basis of the asymptotic theory for quantile regressions are the following assumptions:

- ▷ **H.1** The distribution functions of X given Y, $F()$, is absolutely continuous with continuous densities functions $f()$, that is uniformly bounded away from 0 and ∞ at $\xi(\tau) = Q_{X|Y}(\tau)$.
- ▷ **H.2** There exist positive definite matrix D_0 and D_1 such that:
  - $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_n y_n y_n' = D_0$,
  - $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_n f_n(x_n) y_n y_n' = D_1$

$$- \sum_n y_n y_n' = \prime(N).$$

Theorem 2. *Under **H.1** and **H.2**,*

$$\hat{\beta}(\tau) \xrightarrow{p} \beta(\tau).$$

A proof of the above theorem is available in **?**.

An other important result regarding the asymptotic estimate, is its asymptotic normality of the asymptotic error: $\hat{\beta}(\tau) - \beta(\tau)$, which is useful for hypothesis testing.

Theorem 3. *Under **H.1** and **H.2**,*

$$\sqrt{N}\left(\hat{\beta}(\tau) - \beta(\tau)\right) \xrightarrow{d} \mathcal{N}\left(0, \tau(1-\tau)D_0^{-1}D_1D_0\right).$$

These two theorems are used to build a statistical framework to test the validity of estimate parameters. This can be done through a *Wald test* or a *Rank test*.

As quantile regression is built on a L1 distance, the goodness of fit cannot be implemented through the standard R^2. Koenker and Machado (1999) propose notably an alternative measure based on the L1 distance **?**. However, we precise that an explanation of variability in the variance of the model based on the mean is not particularly appropriate for quantile regression.

Chapter 3

Intra-industry volatility spillovers around Earning Announcements

This contribution has been presented to the PhD Lunch Seminar of Paris 1, the 5th Workshop in Economics of Braga University, June 2017. I warmly thanks my PhD supervisor Philippe de Peretti for his advices on this work. I am also grateful to Susanna Martins for the review she made. Commendation is also extended to all others who in various ways positively contributed to this work.

This work has been presented at the following seminar or conferences:

- ▷ *Phd Lunch Seminar, University Paris 1, 04/2017*
- ▷ *5th PhD Students Workshop : Advanced in Quantile Regression, Minho University, 16/06/2017*
- ▷ *Thé des jeunes économétres, Paris, 11/10/2017*
- ▷ *Développements récents de l'économétrie appliquée à la finance, Université de Nanterre, 8/11/2017*

Abstract

In this paper we propose an analysis of intra-industry volatility spillovers around Earnings Dates. Since the seminal works of Ball and Brown (1968) and of Patell and Wolfson (1978), it has been shown that Earning Announcements generate shocks on asset returns and on asset implied volatility. In this paper, we find evidences of volatility spillovers around the Earning dates between the announcer and its co-sector firms. Following Ben Rejeb and Arfaoui (2016), we model volatility spillovers through Quantile Regressions. This approach has the advantage to distinguish the effect of volatility

shocks of the announcer, on lower and upper quantiles of the co-sector volatility distributions. We test the volatility spillover with the nature of the news, the earning surprise, the announcer capitalization, the sentiment of investors regarding the announcer, and on the publication order in a given earning season.

1 Introduction

According to the Efficient Market Hypothesis, the price of a financial asset represents all the available information. Yet, market inefficiencies have been observed around some particular scheduled events, leading to arbitrage opportunities **?**, **?**. Earning Announcements (EA) are known to impact both returns and volatility dynamics of the announcer. Ball and Brown (1968) are the first to identify a post-announcement move in the announcer returns: the Post-EA-Drift (PEAD) **?**. Several explanations arise. Around the announcements, stocks are under the scrutiny of investors, which leads to higher traded volumes and so stocks are under a strong buying pressure. Besides, the persistence of the PEAD have been first explained by transactions costs. The higher the transaction cost, the higher the friction in the market, and the longer it takes for the stock to recover is fair price **?**. However Korajczyk and Sadka (2003) conclude that transaction costs do not fully explain the PEAD-momentum **?**. Lastly, according to Vega (2006), the more the information (public or private) the investors have on a reporting firm, the smaller the abnormal drift the announcement day **?**. Other studies explain this phenomenon by the existence of a risk premium: firms reporting unexpectedly high earnings make investing in these firms more risky. Della Vigna and Pollet (2009) find that this effect is slower on Fridays, due to a lower investor attention; Sheng (2016) demonstrates that PEAD can be affected by macro announcements arising on the exact same date as the EA **? ?**.

Regarding the impact on the announcer's volatility, Patell and Wolfson (1978, 1981) are the first to identify a significant pre-announcement increase in the implied volatility (IV) **?**, **?**. This effect results from a higher uncertainty hitting the market before the EA and has been confirmed by other studies **?**, **?**, **?**. Notably, Perignon and Isakov (2001) discriminate the evolution of the IV by the type of news reported by the firm. They find that, when announcers reported bad news, the IV stays at a high level, translating investor uncertainty regarding the future of the firm **?**. Uctum et al. (2017) studied the impact of various events on the Historical Volatility (HV) of french stocks **?**. Whilst the IV represents the forward looking uncertainty on the market, the HV defines the current uncertainty on the market. Their results confirm that events affect durably the HV, and the delay of response illustrate the time needed by investors to analyse the new information. This can be of great interest in the framework of portfolio selection. Indeed, the volatility of the portfolio can be artificially decreased by picking stocks that do not publish their results in the upcoming months.

Firm's EA also affect peers. Ramnath (2002) references the existence of intra-industry information flows around EA. According to the author, financial investors use EA to revise their expectations on the whole sector or on upcoming announcements **?**. Thomas and Zhang (2008) confirm this intra-industry learning effect around EA **?**. Patton and Verardo (2012) identify an increase in the beta linking the announcer and the market around the EA. They differentiate the jump in beta by several variables such as the surprise effect or the market visibility of the announcer **?**. Pavor and Wilson (2014) explain this contagion effect through a risk premium point of view. The increase in the announcer stock price (higher market beta), is the consequence of an higher contribution to the systematic risk of the whole economy **?**. Indeed, they find that the return of the announcing firm has a higher predictive power of the upcoming earnings, than the market return, and so the EA conveys much more information, than the information brought by the market.

In this paper, we ask whether the announcer generates intra-industry volatility spillover. The existence of such effect could have several applications in option pricing, risk management, portfolio selection or in trading strategies. Volatility spillovers have already been studied around macro events **?**, **?**; but to our knowledge no studies have been devoted to volatility spillovers around EA. Following Patell and Wolfson (1979), Perignon and Isakov (2001), Patton and Verardo (2012) we raise a set of assumptions to identify and to explain such effect **?**, **?**, **?**. We focus on the nature of the news, the earning surprise, the announcer capitalization, the sentiment of investors regarding the announcer, and on the publication order in a given earning season. Instead of modelling volatility spillovers through GARCH-BEKK or DCC-GARCH, we use quantile regression **?**, **?**, **?**. Ben Rejeb and Arfaoui (2016) are the first to model volatility spillovers through quantile regression **?**. This technique permits to identify, with an higher accuracy, the effect of a move in the announcer volatility on the volatility distribution of co-sector firms. As a result, we can assess which quantile of co-sector firms volatility is more affected by EA.

Our paper is organized as follows. In Section 2, we present the set of assumptions we want to verify, in Section 3, we introduce the data set and the methodology, in Section 4 our results. In Section 5, we implement a simple volatility arbitrage strategy to illustrate market inefficiencies around EA.

2 Assumptions

The existence of intra-industry information transfers around EA is well documented and can be explained by at least two facts **?**, **?**. The information hitting the market after the EA may be interpreted as an indicator of the health of the whole industry: investors consider that the news reported by a single firm may be extended to its co-sector firms. Also, it has been proved that it exists a significant intra-industry learning effect. Financial analysts use the outcome of last announcements to re-adjust their expectations regarding up-coming ones **?**. As a result, announcing firms attract investors' attentions and those may modify their positions on co-sector firms. This may create spillover effects between the announcer and the rest of the sector. Following Patell and Wolfson (1979, 1981), Perignon and Isakov (2001), we want to validate assumptions on the intensity of the spillover effect around the EA **?**, **?**, **?**.

We set up the following assumptions.

▷ *H1-H2*: The spillover effect increases/decreases before/after the EA.

Around its EA, a firm attracts attention of investors, and the uncertainty existing on the outcome of the announcement affects its co-sectors. We expect to have pre-announcement and post announcement effects. An increase of the spillover effect some days before the EA, underlines the transmission of uncertainty on the outcome of the announcement **?**, **?**. This is illustrated by an higher volatility on the announcer log-returns, which affects the volatility of co-sector firms.

Whereas the pre-announcement is due to an increase on uncertainty, the post-announcement effect results from the announcement outcome. It is well-documented that this effect is slow and takes times before being dissipated **?**, **?**. For instance, the so-called *Post-Earnings-Announcement-Drift* (PEAD), defining the time delay of the announcer log-return to recover its long term level, can be explained by an higher risk premium, by the access to information or by the time it takes to investors to react to the new information **?**, **?**.

As a result, the pre-announcement spillover is *forward looking* while the post-announcement spillover is *backward looking*. The reaction delay of investors to the information hitting the market at the EA questions the Market Efficiency Hypothesis.

▷ *H3*: The spillover effect is stronger for bad news than for good news.

Black (1976) is the first to identify the so-called leverage effect: the volatility response to shocks on returns is known to be asymmetric: a drop in the stock value, makes the stock riskier (increase in volatility) and leads to a financial leverage. However, the asymmetric relation can also be explained by a time-varying risk premium, which is reflected by a reversal relation: an increase in the conditional volatility raises the expected returns, and declines the current stock value. The raise of the expected return is viewed as a premium for investors to hold risky assets. The relation between current volatility and past realized returns is modelled with a *news impact curve* **?**, **?**. We can expect the same kind of relation with volatility and outcomes of Earnings. Perignon and Isakov (2001) find that firms with bad announcements have an higher level of Implicit Volatility after the EA **?**. Bad announcements create more uncertainty for investors, which is illustrated by an increase in the IV just after the announcement. On the contrary, good news reassure investors and so the IV decreases after the EA. One can wonder whether bad news spread more to co-sector firms than good news. Even if volatility spillovers around particular macro or firm level news have already been studied, none of them distinguish volatility spillovers for good or bad news **?**, **?**. Indeed, when a firm reports a bad outcome, then the volatility of the stock immediately rises due to the *leverage effect*. The existence of linkages among firms belonging to the same sectors, questions whether the risk premium can be transferred to co-sector firms. This assumption is in line with the *learning effect* literature around EA **?**, **?**. We believe that firms with bad realized EPS carried more uncertainty on their future, and thus create more volatility spillover around the EA.

▷ *H4*: The spillover effect is stronger for large surprises than for small surprises.

Whatever the type of news (good/bad), the surprise, measured by the difference between the forecasted outcome of the announcement and the realized outcome, generates information transfers among the industry **?**, **?**. As we said above, investors will use large surprises to re-adjust their expectation either on the outcomes of up-coming announcers or more generally on the health of the industry. Omrane and Hafner (2011) studied volatility spillovers around macroeconomic events between major exchange rates and they find a significant increase in volatility spillovers from a currency to an other after an unscheduled outcome **?**. For instance, Patton and Verardo (2012) find that firms with large earning surprises manifest higher jumps in their beta **?**. To them, EA with a large surprise effect have a large information content. Yet, we assess that this information content will spread to co-sector firms, and force investors to re-adjust their expectations.

As a result, we believe that spillover effect is stronger for large surprises.

▷ *H5*: The spillover effect generated by large capitalized firms is stronger than the one generated by low capitalized firms.

Lo and McKinlay (1990) are the first to find statistical evidences of cross-correlation in stock returns, between large and small cap in the US stock market, and particularly returns of small cap are correlated with lagged returns of large cap **?**. Various explanation arises to explain such phenomenon: unsynchronised trading, differences in transaction costs (large cap have smaller bid-ask spreads), or lastly differences in term of signal quality (large cap benefit from a better coverage from analysts). Also, Conrad Gultekin and Kaul (1991) find evidence of volatility spillovers among large and small cap **?**. According to them, as stock price volatility is directly related to the rate of flow of information hitting the market, the asymmetry in volatility spillovers is consistent with the speed of response of stocks to new information: prices of large stocks respond immediately while prices of small stocks respond with a lag. Moreover, Harris and Pisedtasalasai (2005) use a multivariate GARCH to study volatility spillovers between large and small cap in the UK market, and they find evidences of volatility spillovers from small cap to large ones, even if these effects are weaker than the spillover effect generated by large caps **?**.

As a result, the weight of the company is a common discriminant for Earning analysis. Large capitalized companies are more visible and therefore benefit from better Earning coverages. Moreover, large cap influences the market due to their large weights, and generally belong to a key stock market index. As a result, large capitalized companies are expected to generate higher volatility spillover the day of their EA **?**.

▷ *H6*: Market Sentiment Effect

Investor sentiment refers to the degree of optimism or pessimism about stocks that is not directly explained by fundamental information. Recently, a lot of research have been devoted to the linkage between investor sentiment and market prices.

Antweiler and Murray (2004) are the first to investigate the impact of messages published on Yahoo!Finance and on Raging Bull on financial markets **?**. On 45 listed companies, they study the effect of the number of messages and of their contents on financial dynamics such as stock returns, traded volumes or stock volatility. They find that disagreements in posted messages increase the trading activity through an increase of the trading volume. Also, they find that posted messages predict volatility both at the daily and at the intra-day level. Masoud, Sameena and Wenhui (2013) built a predictive model for stock prices using a supervised learning algorithm on tweets **?**. They find that tweets have a significant predictive power regarding stock prices.

Livnat and Petrovits (2009) find that investor sentiment play a certain role around Earning Announcements through the study of the relation between investors sentiment and the PEAD **?**. Notably, the research on PEAD indicates that investors react immediately and in the same direction than the news content. However, taking into account investor sentiment lead to a different suggestion: in period of high sentiments preceding the EA, investors are overconfident, and so if the outcome of the EA is consistent with their believes, they under react to the EA, which leads to a small PEAD. On the contrary, if the outcome is different from their beliefs, investors will take time to analyse the new information, and the PEAD will be higher. Before the Earning Announcement, Livnat and Petrovits (2009) rank sentiments by their nature: low, medium and high, and find significant abnormal returns following low sentiment periods. They find that the upward move of stock prices following good announcements and low sentiment period, is greater than the upward move of stock prices following an high sentiment period.

As a result, we believe that the investor sentiment may affect volatility contagion. Similarly for bad news and bad surprises, we believe, that firms with a negative sentiment will generate more volatility contagion around their EA.

▷ *H7*: Learning effect.

Following Patton and Verardo (2012), we test whether its exists a learning effect in the EA process **?**. In their paper, they built a simple learning model in which investors use information from EA to extract information on the entire economy. Notably, they test if the evolution of investor's expectation may explain the jump in beta linking the firm and the market. Hence, for a given earning season, the outcome of firms that first report their EA, allows investors to revise their expectation on upcoming EA. As a result, within a given earning season, the more EA reported, the less uncertainty on upcoming EA. In this paper we test whether the learning effect explains the contagion in volatility. More precisely, we suppose that, for a given earning season, firms that first publish their announcements, create more contagion in volatility than the others. In this setting, the level of contagion between a firm reporting its result, and a pair, should decrease with the number of announcements done.

3 Methodology

3.1 Data

In this subsection, we introduce the data used in our study.

Earning Announcements

We consider 812 earning announcements between April, 1st 2017 and August, 22nd 2017 [1]. These events correspond to announcements of 406 companies from the S&P

[1] The size of our sample is larger than ones of previous studies on EA and volatility. Patell and Wolfson (1979): 83 events; Donders and Vorst (1996): 96; Donders, Kouwenber, Vorst (2000): 190

500, and they belong to 83 different sub industries. More precisely, within the time interval studied, each firm published its result for trimesters ending in March 2017 and June 2017, indeed, we have kept companies from the whole S&P 500, that publish their quarterly results on the following basis: January, March, June, September.

To distinguish good news from bad news, we use the *Earning Per Share* (EPS), that measures the effective gain or loss faced by investors at the Earning Date **?**, **?**. It can be defined as follows.

$$\text{Earnings Per Share} = \frac{\text{Net Income}}{\text{Number of Shares}}.$$

When the EPS is positive (resp. negative), stockholders face an increase (resp. decrease) in their stock value. For investors, anticipating the expected trimestrial result of companies permits for them to re-adjust their expectations on a particular company or on a whole industry **?**, **?**. As a result, forecasted EPS are available few months before the Earning Date and are continously re-adjusted until the EA. At a given date, the consensus of forecasted EPS is publicly available [2]. The consensus is defined as the median of all forecasted EPS provided by analysts. In our study, only the most recent consensus forecasted EPS is employed. One can notice that the evolution of all forecasted EPS, until the EA, can provide an interesting measure of the confidence of analysts on a given firm. Notably, Patton and Verardo (2012) measure a jump in the beta, linking a given stock and the market, through the standard deviation of all available forecasted EPS. They found that, the higher the uncertainty among forecasted EPS, the higher the jump in beta the day of the announcement **?**.

The publication time of each company is either before the market opening, or after the market closing[3]. When the publication time is after the market close, the announcement is considered to be done the next trading day. In this study, we don't discriminate events by calendar effects (Friday announcements), neither by the number of events published each day. EA data are collected on the Nasdaq website[4].

At the Earning Date, Earning surprise (the error between the forecast EPS and the realized EPS) is measured with the *EPS surprise*.

$$Surprise_{i,trim} = \frac{EPS_{i,trim} - E\hat{P}S_{i,trim}}{E\hat{P}S_{i,trim}} \tag{3.1}$$

events.

[2]Nasdaq provides, for each listed company, data related to EA.

[3]In our sample, 60 % of the EA are done before market opening.

[4] `http://www.nasdaq.com/quotes/earnings-surprise.aspx`

$EPS_{i,trim}$ is the EPS reported by firm i for trimester $trim$ while $\hat{EPS}_{i,trim}$ is the consensus EPS forecast.

The Earning Surprise measures whether the reaction of investors to the EA is proportional to the surprise of the announcer. Hence, early announcements gather attentions of investors and may generate overreactions, both in terms of returns and uncertainty. For each EA we have collected the realized and the last forecast *Earning Per Share* (EPS).

Table 3.1 presents descriptive statistics on the entire sample. On the whole sample, the average EPS reported for March 2017 is 1.09 while it is slightly greater for June 2017: 1.19. Similarly, the average surprise is smaller for March 2017: 9% while it is of 15% for June 2017. Moreover, on the whole sample only 3% of the events register negative EPS while 19% register a negative surprise.

$trimester$	$\bar{EPS}$	EPS_{min}	EPS_{max}	$\bar{Surprise}$	$Surprise_{min}$	$Surprise_{max}$
Mar2017	1.06	-1.07	9.88	0.09	-3.00	17.67
Jun2017	1.12	-1.46	15.14	0.15	-1.26	8.20

Table 3.1: EA statistics. EPS are realized Earning Per Share, and the surprise is calculated using Equation 3.1. EPS are expressed in dollar while surprises in percentages. $\bar{EPS}$ and $\bar{Surprise}$ refer respectively to the sector average EPS and to the sector average Surprise.

Company weight

For each company, we calculate the relative weight of a company in its sector. The weight is defined as the firm capitalization over the sub-industry capitalization, obtained from Nasdaq website.

Order of announcement

To test for learning effect, we sort companies of a same sub-sector by order of announcement in for given earning season. As a result, the firm that first reports her financial result regarding a trimester is labelled by 1, the second by 2...

Volatility

Between April 1st 2017 and August 22nd 2017 five minutes stock prices are collected from Bloomberg. Opening and closing effects are removed by cancelling the first and the last thirty minutes. More precisely, we consider five minutes stock prices with, for each time observation: the opening price, the closing price, the highest price and the lowest price.

One challenge is to find a robust estimate of intra-day volatility and various methods exists. We can enumerate: GARCH oriented approaches, such as the MCD-GARCH, that combines intra-day and daily patterns **?**; a combinaison of FIGARCH and Fourier Flexible Form **?**; however GARCH methods suffer from two drawbacks. They smooth intra-day volatility, and then they depend on a close/close approach. In this setting stock returns are computing on close prices which omits all the possible price variations within the interval $[t-1, t]$. Volatility estimators combining the opening price, the closing price, the highest price and the lowest price provide much more information on the behaviour of the price in $[t-1, t]$. The Garman-Klass estimator takes account of the opening, closing, highest and lowest price **?**. In this setting, the volatility estimator provided by Garman and Klass is:

$$\sigma_{d,t} = \sqrt{\frac{1}{2}\left(\log \frac{P^h_{d,t}}{P^l_{d,t}}\right)^2 - (2\log 2 - 1)\left(\log \frac{P^c_{d,t}}{P^o_{d,t}}\right)^2},$$

where $P^h_{d,t}$, $P^l_{d,t}$, $P^o_{d,t}$, $P^c_{d,t}$ refer respectively to the highest, lowest, opening and closing prices, during day d and time t.

Investor sentiment measure

Various proxies have been proposed to characterize investor sentiment: the number of press articles published, the Google Search Engine or tweets. **?**, **?**, **?**, **?**.

In this paper, we consider tweet mood indicators for sentiment analysis. Indeed Mao, Counts and Bollen (2011) and Zhang et al. (2011), find both that there exists a significant correlation between emotional tweets and market indicators: mood indicators are positively correlated with VIX **? ?**

As investor sentiment measure, we consider the daily Bloomberg variable *TWITTER_SENTIMENT_DAILY_AVG*, that is in $[-1, 1]$. A negative value refers to a negative mood while a positive value to a positive mood. This variable looks for tweets through a combination of company mentions, cash tags and major product and people mentions as well. The sentiment is computed using supervised machine learning algorithm which classifies every tweet for every company into positive/negative/neutral along with confidence measure. As a result, for each company and for each EA, we collect the twitter sentiment variable the day before the EA. In our sample, 33% of the events we study are preceded by a negative sentiment mood.

Table 3.2 provides a correlation analysis for the explanatory variable introduced above. A part from a strong significant correlation between the forecasted EPS and the realized EPS, there is weak correlation among the variables. For instance, we find a negative and significant correlation of -29% between the firm weight and the running order of announcements, which signifies that large caps report first their results.

	EPS	$\hat{EPS}$	Surprise	Twitter Sentiment	Weight	Order
EPS	1.00	0.99^{**}	-0.04^{**}	-0.10^{***}	0.11^{***}	-0.07^{***}
$\hat{EPS}$	0.99^{**}	1.00	-0.08	-0.10^{***}	0.11^{***}	-0.07^{***}
Surprise	-0.04^{**}	-0.08^{***}	1.00	0.01^{***}	0.05^{**}	-0.05^{***}
Twitter Sentiment	-0.10^{***}	-0.10^{***}	0.01	1.00	-0.03	0.04^{*}
Weight	0.11^{***}	0.11^{***}	0.05^{***}	-0.03	1.00	-0.29^{**}
Order	-0.07^{***}	-0.07^{***}	-0.05^{***}	0.04^{***}	-0.29^{**}	1.00

Table 3.2: Correlations between explanatory variables. $*$, $**$, $***$ signify the correlation is statistically different from zero at 90%, 95% and 99%. Weight is the relative weight of the company capitalization in the sub-industry. Order is the order of announcement in the sub-industry and for an earning season. EPS is the realized EPS and $\hat{EPS}$ is the consensus forecast EPS.

3.2 Measuring Volatility Spillovers

Most of the techniques employed in the econometric litterature to measure volatility spillovers are BEKK-GARCH or DCC-GARCH **?** **?**. BEKK-GARCH is one of the first model designed to include volatility spillovers in the equation of the conditional variance and has been used in intra-day volatility spillovers modelling **?** **?**. Its principal drawback relies on the large number of parameters to estimate. DCC-GARCH models are interesting alternatives, but are mainly focused on the modelling of the conditional correlation and not on the spillover effect.

Belgacem et al (2015) and Ben Omrane and Hafner (2015) test for volatility spillovers around macroeconomic news using DCC-GARCH **?**, **?**. To measure the increase in volatility spillover, they add a dummy variables in the conditional variance equation. It is not a suitable method for describing the dynamic of the spillover effect. We believe that a rolling window approach would be more appropriate[5].

Recently, Ben Rejeb and Arfaoui (2016) suggest to model volatility spillover through quantile regression **?**. In this context, the spillover effect can be estimated thanks to a simple quantile regression on volatility estimates. One advantage of quantile regression is to distinguish volatility spillovers in upper and lower quantiles, corresponding to high and low volatility regimes [6].

Standard Ordinary Least Squares model the variation in mean of a variable conditionally to the variation of an explanatory variable. Even if this technique is widely used, it does not explore the variation in all the conditional distribution, which is useful in many cases. In finance, it is well known that the correlation structure of assets is af-

[5] Also, the author simulated a two dimensional BEKK of 10 000 observations, and estimated it through a rolling window of 1000 observations. He find highly volatile results.

[6] For others applications of QR in Finance, we refer the reader to **?**, **?**, **?**, **?**, **?**, **?**, **?**

fected by changes in volatility regimes: correlation tends to increase when the market volatility increases **?**. It has also been proved that causality relations differ in quantiles **? ?**. Quantile regressions are used to model dependencies among random variables in given quantiles **?**. This technique has many advantages. First estimators are robust to the presence of extreme values. Second, it allows the strength of the relation to vary across quantiles. Last, contrary to conditional correlation, conditional quantiles are not necessarily symmetric.

The quantile regression is defined as follows:

$$Q_{Y|X}\left(\tau\right) = \inf\left\{y \in \Omega_Y | F_{Y|X}\left(y\right) \geq \tau\right\} = X'\beta(\tau). \tag{3.2}$$

Where $\tau \in [0,1]$, $F_{Y|X}\left(.\right)$ is the conditional distribution of Y given X, and $\beta(\tau)$ corresponds to regression coefficient of the quantile τ. Henceforth, when X moves by one unit, values of Y in quantile τ move by $\beta(\tau)$. In this particular example, the relation in quantiles is linear, but can be extended with ease to the non-linear case **?** as well as to the non-parametric case **?**.

The estimation of $\beta(\tau)$ in Equation 3.2 is done by solving the following weighted minimization problem.

$$\beta^{\star}(\tau) = \underset{\beta(\tau)\in\Lambda}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \left(\tau - 1_{y_i<\beta(\tau)x_i}\right) |y_i - \beta(\tau)x_i| \right\}, \tag{3.3}$$

for $Y = (y_1, ..., y_N)'$, $X = (x_1, ..., x_N)'$ and $N < \infty$. The presence of the $L1$ norm in Equation 3.3 leads to solving difficulties. A natural solution is to express the above equation into a linear programming problem **?**, **?**.

In our context, we want to insulate a possible volatility spillover among two companies. A volatility spillover is defined by a move in volatility at time $t+1$ of company j induced by a move in volatility at time t of company i. To insulate the volatility spillover, we consider the following model.

$$\begin{cases} Q_{\sigma_{j,d,t+1}|\sigma_{j,d,t},\sigma_{i,d,t},\sigma_{mkt,d,t}}\left(\tau\right) = X_{i,d}^{'}\beta_{i,d}(\tau) \\ \\ X_{i,d,t}^{'} = (1, \sigma_{j,d,t}, \sigma_{i,d,t}, \sigma_{mkt,d,t}) \\ \\ \beta_{i,d}(\tau) = (\alpha_{i,j,d}(\tau), \beta_{j,d}(\tau), \beta_{i,j,d}(\tau), \beta_{mkt,j,d}(\tau))^{'} \end{cases} \tag{3.4}$$

To insulate the effect of the announcer volatility ($\sigma_{i,d,t}$) on the volatility of the co-sector firm ($\sigma_{j,d,t+1}$), we incorporate the past realization of the co-sector firm volatility ($\sigma_{j,d,t}$)

as well as the past realization of the market volatility ($\sigma_{mkt,d,t}$). Model 3.4 is estimated through a rolling window between day $d-h$ and day d. If the length of the window is too small it can leads to estimation drawbacks, while if it is too large, the last observations will have a small weight and so the sequence of spillover coefficients will be too smooth. We select a rolling window of ten days, which corresponds approximatively to seven hundred fifty observations.

If the spillover coefficient, $\beta_{i,j,d}(\tau)$, is significant, then a move in volatility of firm i, will create a significant move at the next time step, in the $\tau - th$ quantile of company j. In other terms, if $\beta_{i,j,d}(\tau)$ is significant, there is an evidence of a spillover effect originated by company i on company j[7].

3.3 Testing for jumps in the spillover coefficient

As in Patton and Verardo (2012), to test for a jump in volatility transmission around the EA, we regress the vector of daily spillover coefficients estimated 7 days around the EA: $(\beta(\tau)_{i,j,d=m-3}, ..., \beta(\tau)_{i,j,d=m}, ..., \beta(\tau)_{i,j,d=m+3})'$ on a vector of daily dummies: $(\mathbf{1}_{d=m-3}, ..., \mathbf{1}_{d=m}, ..., \mathbf{1}_{d=m+3})'$ **?**. Which is similar to implement the following panel regression:

$$\beta(\tau)_{i,j,d} = \delta_{i,j,m-3}\mathbf{1}_{d=m-3} + ... + \delta_{i,j,m}\mathbf{1}_{d=m} + ... + \delta_{i,j,m+3}\mathbf{1}_{d=m+3} + \beta(\tau)_{i,j} + \varepsilon_{i,j,d} \quad (3.5)$$

In Equation 3.5, $\beta(\tau)_{i,j,d}$ corresponds to the spillover coefficient between the announcing firm (i) and a co-sector firm (j). The term $\beta(\tau)_{i,j}$ is the long term value of the spillover coefficient, computed on the whole sample. As a result, the panel regression captures the deviation of the spillover coefficients around the announcement from its long run level. The effect of EA on spillover coefficients can be detected by looking for changes in panel regression coefficients.
Henceforth, the spillover effect differs from its long run level at the EA if $\delta_{i,j,m=d}$ is significantly different from zero. If $\delta_{i,j,m=d}$ is negative (positive), it signifies that $\beta_{i,j,d}$ is below (above) its long term level. The use of quantile regressions to measure spillover effects in volatility suggests that the intensity of the spillover effect varies across quantiles. For instance, we assess that some part of the conditional distribution will be more affected than others by the announcement.
Moreover, for each publisher, we delete from the sample co-sector firms that publish their own results seven days around the EA (from three days before to three days after). When spillover coefficients are not significant at the 90% level, corresponding $\beta_{i,j,d}(\tau)$ are set to zero.

[7] Confidence intervals are computed using an inverted rank test, as proposed by Koenker (1994).

4 Results

▷ *H1-H2*: The spillover effect increases/decreases around the EA.

Figure 3.1 illustrates the behaviour of $\beta(0.25)$, $\beta(0.5)$ and $\beta(0.75)$ around the EA. For both quantiles, spillover coefficients increase around the EA. For instance, the day of the announcement, an increase in 1 point in the announcer volatility increases by 2.4% its co-sector's volatility. However, the level of contagion increases with the quantiles being respectively: 1.5% for $\tau = 0.25$, 1.8% for $\tau = 0.5$ and 2.4% for $\tau = 0.75$.

A strong heterogeneity of behaviours is observed among sectors and companies. For instance, Figure 3.2 represents the behaviour of $\beta(0.25)$, $\beta(0.50)$ and $\beta(0.75)$ for three particular sub sectors: Restaurants, Electrical Components & Equipment and Health Care Equipment. Restaurants is the sub sector having the sharpest decrease in betas. On average, the day of the announcement, an increase of 1 point in the volatility of the announcer, makes the volatility of co-sectors decreasing by 3.8% in the 75th quantile. On the contrary, the Electrical Components & Equipment sub sector has the highest level of contagion, for all quantiles. On average, the day of the announcement, an increase of 1 point in the volatility of the announcer, makes the volatility of co-sectors increasing by 10% in the 75th quantile. An other matter of interest is to analyse the contagion in the largest sub sector of the sample: Health Care Equipment. Around the announcement, betas face two sharp increases: a first five days before the EA, and a second one the day of the announcement. When a company from this sub sector publishes its results: an increase of 1 point of its median volatility, makes the median volatility of its co-sectors increasing by 6%.

The heterogeneity is also strong at the level of firms. Figure 3.3 represents the behaviour in betas for companies generating the most contagion (Quest Diagnostics Incorporated), the most negative contagion (Procter and Gamble), and a company representing the average contagion (Facebook). When Quest Diagnostics Incorporated publishes its results, a move of 1 point in its volatility makes the median volatility of its co sector, jumping by 15%. At the same quantile, when Facebook publishes its result, a move in 1 point in its volatility makes its co-sector's volatility jumping by 4%. Lastly, when Procter and Gamble reports its EA, a move in 1 point in its volatility makes its co-sector's volatility decreasing by 5%.

In Table 3.3, we observe that for both the median and quantile 75, the spillover coefficient is below its long run level before the EA. The spike in contagion is reach a day before the EA, where spillover coefficients at quantile 50 and 75 jump respectively by 0.31% and 0.1%. Two days after the EA, the spillover

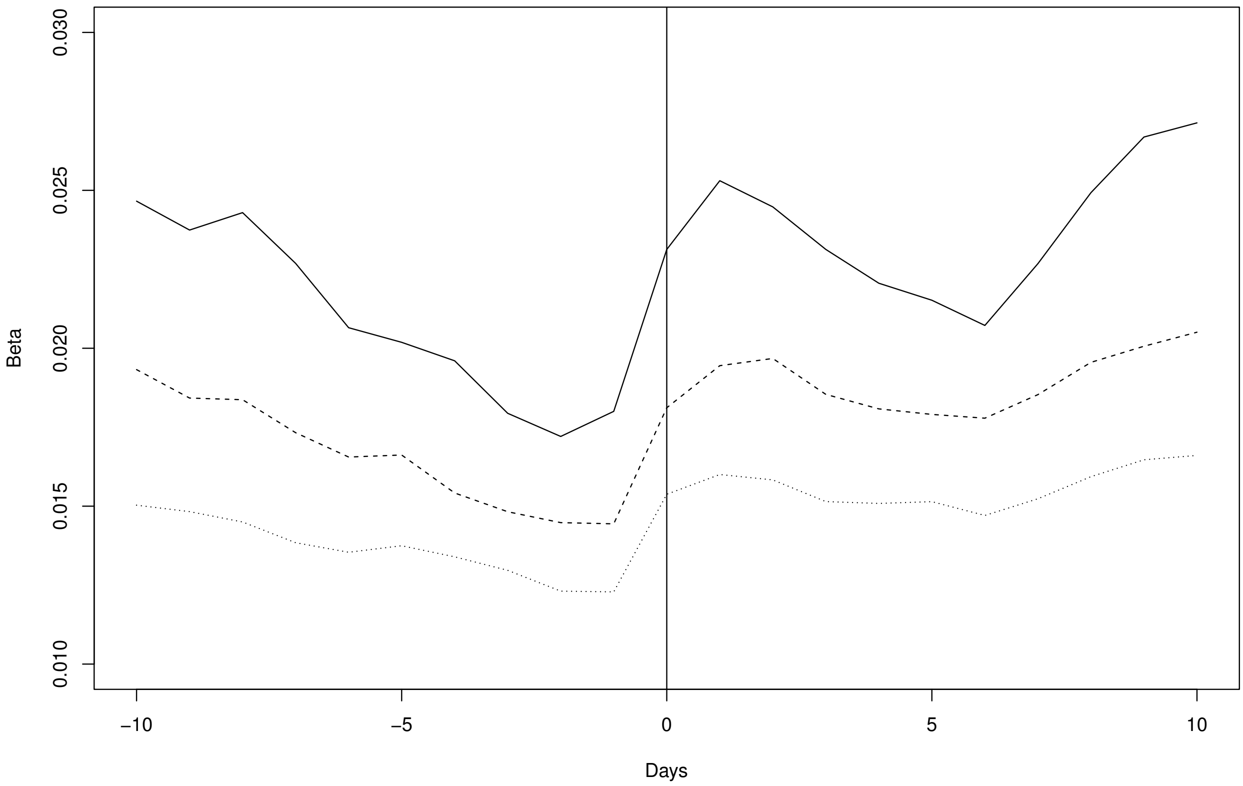


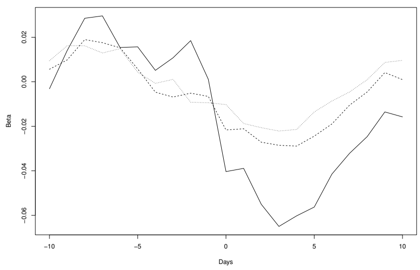
Figure 3.1: Average behaviour in $\beta(\tau)$ around the EA, for $\tau = 0.25, 0.5, 0.75$. Straight line stands for $\beta(0.75)$, Dashed line for $\beta(0.5)$ and Dotted line for $\beta(0.25)$. Day 0 is the EA. Betas are averaged over 33,681 observations

coefficient returns toward its long run level. Hence, there are evidences that spillover coefficients are affected by the EA.

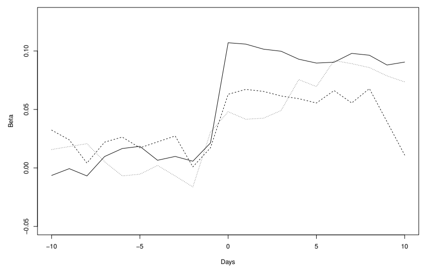
▷ *H3*: The spillover effect is stronger for bad news than for good news.

According to Perignon and Isakov (2002), when firms report bad news, their IV increases just after the EA **?**. We want to test whether bad news create more volatility spillover than good news.

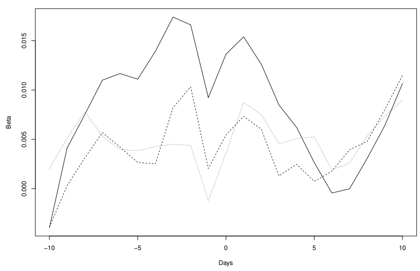
Figure 3.4 represents the behaviour of betas for firms reporting good news, average news and bad news. We observe that, for both quantile, the contagion generated by bad news is the strongest. For instance, when a firm reports a bad news, a move of 1 point in its volatility, increases the volatility in the 75th quantile of its co sectors by 4%, while good news only create a move of 2.5% in



(a) Restaurants

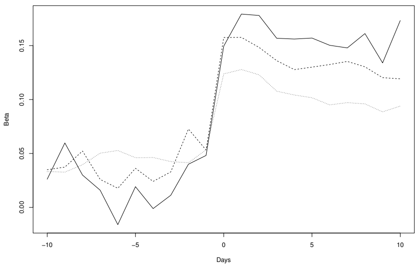


(b) Electrical Components & Equipment

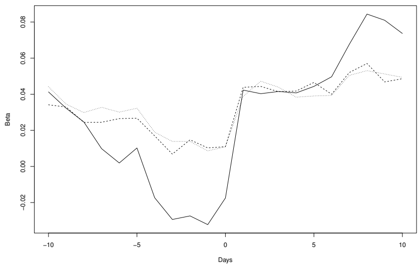


(c) Health Care Equipment

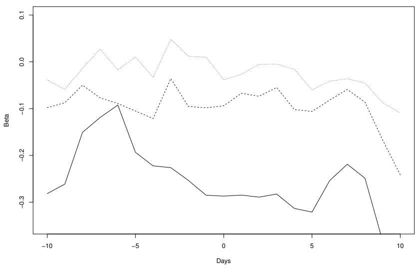
Figure 3.2: Average behaviour in $\beta(\tau)$ around the EA, for $\tau = 0.25, 0.5, 0.75$. Straight line stands for $\beta(0.75)$, Dashed line for $\beta(0.5)$ and Dotted line for $\beta(0.25)$. Day 0 is the EA. Restaurants, Electrical Components & Equipment and Health Care Equipment contain respectively: 4, 17 and 4 companies. For each sub-industry, betas are computed on respectively 20, 481 and 22 observations.



(a) Quest Diagnostics Incorporated



(b) Facebook



(c) Procter and Gamble

Figure 3.3: Average behaviour in $\beta(\tau)$ around the EA, for $\tau = 0.25, 0.5, 0.75$. Straight line stands for $\beta(0.75)$, Dashed line for $\beta(0.5)$ and Dotted line for $\beta(0.25)$. Day 0 is the EA. For each company, betas are averaged on the two earning announcements.

the same quantile. This confirms the assumption of the transmission of negative shocks through volatility.

To test for jumps in beta, we regroup EA by their realized EPS and we partition them into quartiles. Bad news are in q1 whereas good news are in q4.

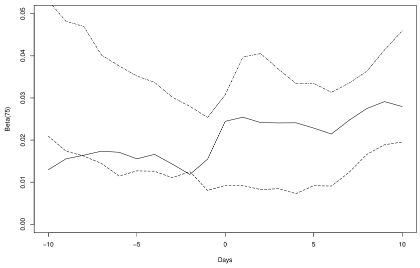
Results, in Table 3.4, show that spillover coefficients generated by bad news (q1, q2) are above their long run level while those generated by good news (q3, q4) are below their long run level. More particularly, spillover coefficients generated by bad news (q1) quickly exceed their long run value. For instance, at the median level and the day following the announcement (Day 1), the spillover coefficient for bad news (q1) is above its long run value by 0.43% (4) while for good news (q4), it is below its long run value by 0.22% (-1.72). At quantile 75, only bad news generate spillover coefficients above their long run value: 0.22% (2.13). However, we observe that for both quantiles and quartiles, spillover coefficients start to converge toward their long run values starting from a day before the earning announcement. More precisely, the spillover effect is significantly reinforced the day after the EA and is more diffuse on time. Our findings confirm the leverage effect, as the volatility spillover is stronger for bad outcomes than for good ones.

▷ *H4*: The spillover effect is stronger for large surprises than for small surprises.

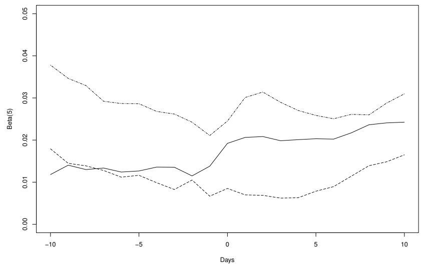
We test whether large surprises create more contagion among the sector.

Figure 3.5 represents the behaviour of betas for firms facing large positive (good) surprises, average surprises and large negative (bad) surprises. We observe that, for both quantile, the contagion generated by good surprises is the strongest. When the outcome of the announcement is well above the forecast, a move by 1 point in the announcer's volatility make the median volatility of its co-sectors jumping by less than 3%. Also for both quantiles bad surprises generate an high level of contagion.

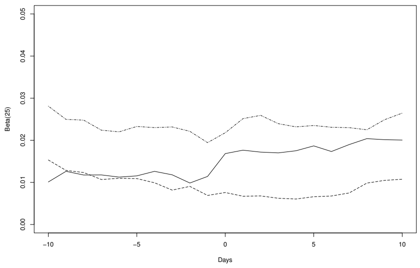
To test for jump in betas, we partition EA surprise effects into quartiles. Results are presented in Table 3.5. The quartile classifies EA surprise from large negative (q1) to large positive (q4). For $\tau = 0.50$ and $\tau = 0.75$, we find that big negative (q1) and big positive surprises (q4) generate spillover coefficients, the day after the EA, that is above the ones of other surprises. More precisely, for $\tau = 0.75$ and for bad surprises it jumps by 0.8% (6.24) the day after, for good surprises (q4), it jumps by 0.6% (5.58), and for small surprises (q2) and (q3), it jumps respectively by 0.03% (-0.24) and by 0.17% (1.39). Moreover, at the median level, it seems to exist a delay of reaction in the increase of the spillover coefficients. Notably, the spillover coefficient for good surprises (q4), is above its long run value a day after the EA while for bad surprises (q1) it becomes above the long run value, the day of the announcement. We notice also that for $\tau = 0.50$, the



(a) $\beta(0.75)$

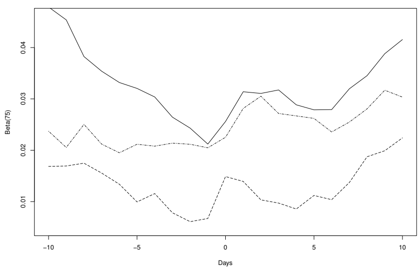


(b) $\beta(0.5)$

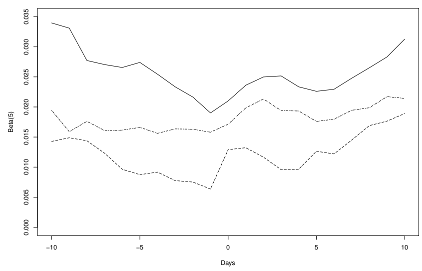


(c) $\beta(0.25)$

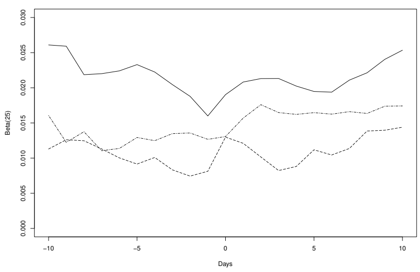
Figure 3.4: Behaviour in $\beta(\tau)$ around the EA, for firms reporting good news (thick lines), average news (dotted lines) and bad news (dot dashed lines). Days 0 is the EA day. Good news are news above quantile 2/3 of the EPS distribution, bad news are below quantile 1/3 of the EPS distribution, and average news are between quantile 1/3 and quantile 2/3 of the EPS distribution.



(a) $\beta(0.75)$



(b) $\beta(0.5)$



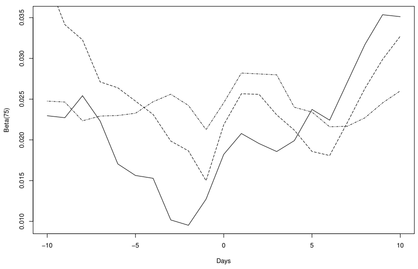
(c) $\beta(0.25)$

Figure 3.5: Behaviour in $\beta(\tau)$ around the EA, for firms reporting large surprises (thick lines), average surprises (dotted lines) and small surprises (dot dashed lines). Days 0 is the EA day. Good surprises are surprises above quantile 2/3 of the surprise distribution, bad surprises are below quantile 1/3 of the surprise distribution, and average surprises are between quantile 1/3 and quantile 2/3 of the surprise distribution.

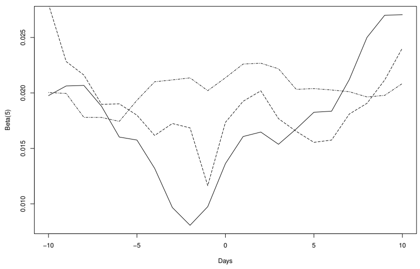
rise in coefficient spillovers for extreme good or bad surprises is above the rise generate by the nature of the outcome.

▷ *H5*: The spillover effect generated by large capitalized firms is stronger than the one generated by small capitalized firms.

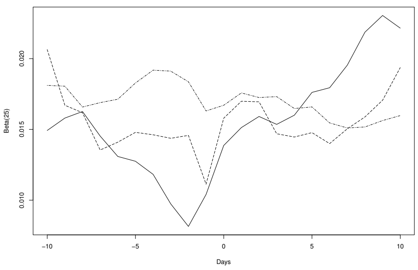
Figure 3.6 represents the behaviour of betas depending on the relative weight of the announcer in the sub-industry. More precisely, we consider large capitalized firms, with a relative capitalization above the two third of the whole relative capitalization distribution, average capitalized firms, with a relative capitalization between one third and two thirds of the whole relative capitalization distribution, and lastly, small capitalized firms, with a relative capitalization below the one third of the whole relative capitalization distribution. We observe that, for both quantile, the contagion generated by small capitalized firms is the strongest. This is in opposition with the results found in the literature suggesting that large



(a) $\beta(0.75)$



(b) $\beta(0.5)$



(c) $\beta(0.25)$

Figure 3.6: Behaviour in $\beta(\tau)$ around the EA, for large capitalized firms (thick lines), average capitalized (dotted lines) and small capitalized (dot dashed lines). Days 0 is the EA day. Large capitalized are capitalizations above quantile 2/3 of the capitalization distribution, small capitalizations are below quantile 1/3 of the capitalization distribution, and average capitalization are between quantile 1/3 and quantile 2/3 of the capitalization distribution.

caps directly influences small caps. For instance, when a small cap publishes its quarterly financial results, a move by 1 point in its volatility make the median volatility of its co-sectors jumping by more than 2.12%, while its median jumps by more than 1.5% when a large caps publishes its result.

To test for jump in betas, we partition firms into quartiles corresponding to their capitalization and to their popularity.

In Table 3.6, we observe that only small capitalized firms (q1 and q2) have spillover coefficients above their long run level. Indeed, the day of the announcement, small capitalizations (q1) have spillover coefficients at $\tau = 0.5$ that is 1.53% (13.14) above their long term value, while large capitalizations have spillover coefficients below their long run level by 0.96% (-6.96). This can be explained by the highest long run value of large capitalizations spillover coefficients. Indeed, large caps create more contagions in volatility all along the year while small caps seem to generate contagion in volatility only around their EA.

▷ *H6*: Market Sentiment Effect.

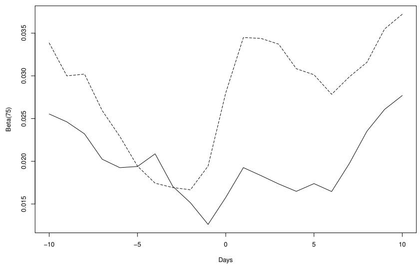
Figure 3.7 represents the behaviour of betas depending on the investor sentiment the day preceding the EA. We divide the sample in two sub-groups: the one containing companies with a negative investor sentiment, which corresponds to 33% of the sample, and companies with a positive investor sentiment. Our results find that firms associated with negative moods generate much more volatility spillovers. For instance when a firm, characterized by a bad investor mood, published its results, a move of 1 point in its volatility, the day of the announcement, makes the median volatility of its co-sectors rising by more than 3%, while when a firm under the good mood of investors, reports its result, a move of 1 point in its volatility, the day of the announcement, makes the median volatility of its co-sectors rising by 1.5%.

As a result, we observe in Table 3.7 that at $\tau = 0.50$, firms characterised by bad investor sentiments, have a spillover coefficient starting being above its long run level, a day before the earning announcement. For instance, the day before the EA, it is 0.14% (1.5) above its long run level, while firms with a positive investor sentiment the day before have a spillover coefficient that is below its long run level by 0.86% (-12.87). Notably, at all quantile levels, the spillover coefficient generate by firms with positive mood is always below its long run level. This underlines the role played by investors on the contagion in volatility.

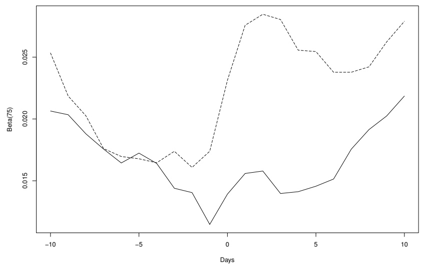
▷ *H7*: Learning effect.

Figure 3.8 represents the behaviour of betas depending on announcement order for a given earning season. We divide our results in three sub samples: the firms reporting first their EA, the firms reporting in second, and lastly later announcers. We observe evidences of learning effect, in the sense that early announcers generate more volatility spillovers than the rest of the sample. This may be explained by an higher focus of investors in early EA, as they use their outcomes to revise their expectation regarding upcoming EA. For instance, the day of the announcement, when the volatility of the announcer moves by 1 point, the median volatility of co-sectors move by 2.5% if it is the first announcer of the season, by 2% if it is the second announcer of the season, and by 1% if it is later announcers. Lastly, the pre-announcement effect at $\tau = 0.5$ is much longer for early announcers, than for others, as spillover coefficients start to increase 3 days before the EA, while for others, spillover coefficients start to increase a day before.

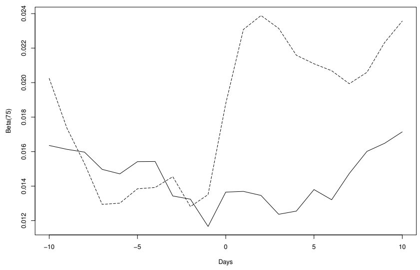
However, these results are mitigate when studying the behaviour of betas regarding their long run level. Indeed, we observe in Table 3.8 that there are no statistical evidence of an higher level of betas for early announcers. This can be explained by the nature of early announcers, that are in majority large caps. As



(a) $\beta(0.75)$

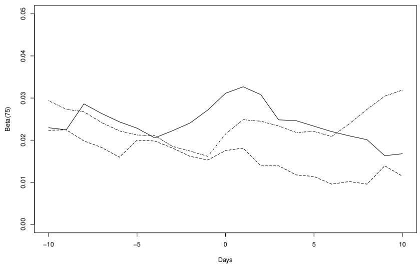


(b) $\beta(0.5)$

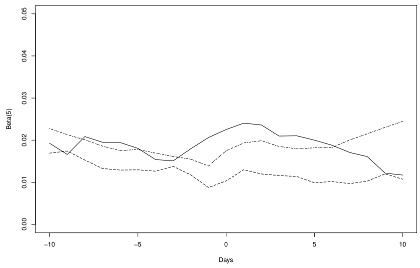


(c) $\beta(0.25)$

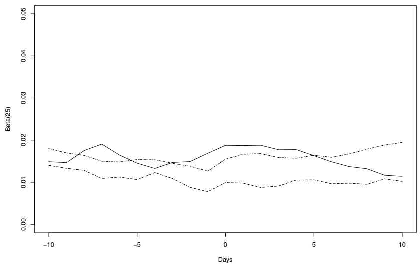
Figure 3.7: Behaviour in $\beta(\tau)$ around the EA for a positive investor sentiment the day before the EA (thick lines) and for a negative investor sentiment (dotted lines). Days 0 is the EA day. Investor sentiment is computed using the Bloomberg measure: average twitter daily sentiment.



(a) $\beta(0.75)$



(b) $\beta(0.5)$



(c) $\beta(0.25)$

Figure 3.8: Behaviour in $\beta(\tau)$ around the EA, for firms publishing first their results (thick lines), in second (dotted lines) and later announcers (dot dashed lines). Days 0 is the EA day.

large caps are under the scrutiny of investors, sub-sectors have an interest to let them publish their result first.

5 Application: Volatility Arbitrage

In the last section, we find statistical evidences of volatility spillover around EA and we want to see if we can built a profitable volatility arbitrage strategy using these results. A volatility arbitrage strategy consists in taking advantage of market inefficiencies by exploiting differences between the current implied volatility and the future historical volatility.

Before introducing our strategy, we provide a brief recall on option prices.

Let S_t be a daily stock price defined on the following probability space $(\Omega, \mathcal{F}_t, \mathbb{Q})$, where Ω is the set of realizations of S_t, that is $\mathbb{R}^+$ is our case, $\mathcal{F}_t$ is the associated filtration and $\mathbb{Q}$ is the risk neutral probability. In this setting, we assume that S_t follows a geometric Brownian motion:

$$dS_t = S_t \left(\mu dt + \sigma dW_t \right).$$

The term $\mu \in \mathbb{R}^+$ corresponds to the price trend, $\sigma \in \mathbb{R}^+$ is the stock volatility, and dW_t is a Brownian motion. In the following σ denotes the implied volatility, and σ_{HV} the historical volatility (both are assumed constant).

A call is an option that gives the right, and not the obligation to buy a stock in T at a pre-defined price K. The value of this option, (denoted hereafter $C(K, \sigma, r, t, T)$) is defined as the expected pay-off:

$$C(K, \sigma, r, t, T) = e^{-\int_t^T r(T-s)ds} \mathbb{E}\left[(S_t - K)^+ | \mathcal{F}_t \right].$$

Call options exchanges on the market are priced using implied volatility. An arbitrage occurs if call options are under priced, that is when estimates of future HV differ from the current IV. More precisely, as we are interested in rises of HV, we only consider cases where the future HV is greater than the current IV.

The strategy consists in buying a call option today, and hedging using the first derivative of the option price with respect to the stock price. This value is denoted Δ and is defined as follows.

$$\Delta = \mathcal{N}(d1),$$

where $\mathcal{N}(d1)$ is the probability, under the $\mathcal{N}(0,1)$ probability distribution to exercise the option at maturity and $d1 = \frac{log(S/K)+(r+\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$.

We observe that, when S goes up (goes down), the probability to exercise the option increases (decreases). According to Ahmad and Wilmott (2005) there are two manners of expressing a profit for such strategy. First, one can replace the IV in the Δ equation by the forecasted HV. However, this depends on the forecasted HV and the obtained profit is random as it depends on S_t fluctuations **?**. A second approach consists in hedging using the current IV which leads to a deterministic payoff over the life of the option:

$$\pi = \frac{1}{2}\left(\sigma_{HV}^2 - \sigma_{IV}^2\right)\int_t^T e^{-r(t-s)}S^2\Gamma ds, \tag{3.6}$$

Where Γ is the second derivative of the call value with respect to S.

See **?** for a complete proof. We notice that the profit of such strategy is particularly high when the strike is close to the underlying, which is the case with close to the money call options.

Regarding EA and volatility arbitrage, Donders and Vorst (1996) and Baik, Kang and Kim (2013). have implemented such strategies on the announcer **?**, **?**. More precisely, Donders and Vorst (1996) buy call options and short the underlying ten days before the EA. Each day, they delta hedge using the Black-Scholes framework. They close the position the day before the EA, when the IV is assumed to reach its maximum. They find an average return of 49.6% of the initial option value. However, their trading strategy faces a significant cost because of the daily hedging **?**. When they include a cost based on the bid-ask spread, they find a negative return of -2.34%. As Baik, Kang and Kim (2013) observe a decrease in the IV of Equity linked Warrants (ELW) on the Korean market, they build a short volatility arbitrage strategy. To short volatility is the opposite of the strategy exposed above. More precisely it consists in shorting a call option and buying Δ shares of the underlying stock. Ten days before the EA, they short one ELW and buy delta shares of the underlying stock. Also, they hedge using the Black Scholes framework and close the position ten days after. After having included liquity costs, they find a significant return per ELW of 11.4%. They also find that deep out the money options present the higher returns **?**.

In this part, we suppose that around the Earning Announcement of a particular firm, the volatility of its co-sector goes up. Hence, two days before the announcement, we go long to all co-sector volatilities, by buying a close to the money call option and by shorting delta stocks. Two days after the EA, we close the position. The profit in Equation 3.6 is approximated with delta hedging. For instance, at $t=-2$, we buy one call at C_{-2}, and we short Δ_{-2} stocks at a unit value S_{-2}. The borrowing amount at $t=-2$ is therefore $D_{-2} = C_{-2} - \Delta_{-2}S_{-2}$. At $t=-1$, to make our position delta-neutral, we sell (buy) $\delta_{-1}\delta_{-2}$ shares of stocks at price S_{-1} and we save $(\delta_{-1}\delta_{-2})S_{-1}$. We do it every day until $t = 2$. At $t = 2$, we close our positions by selling C and by paying back borrowed stocks. For each position, the profit is normalized by the stock value in $t = -2$.

Regarding the option maturity, as the sensitivity of an option price to the volatility (vega), decreases when the time to maturity decreases, we consider options with time to maturities greater than 15 trading days. We assume no transaction costs.

We focus on two sub-industries, during the Earning Season of March 2017: Internet & Software Services and Electric Utilities. We consider announcements done between April 19th and May 30th.

In the sample studied, we have daily option prices only for 8 companies in the Electric Utilities sub industry, and 6 announcements in the time interval considered. For the Internet & Software Services sub industry, we have daily option prices for 8 companies, and 4 announcements.

Figure 3.9 represents the dispersion of strategy returns per announcement day. We observe that for all of them, the average return is positive. Moreover Table 3.6 goes deeper in the performance analysis of our strategy, by computing a sequence of measures use in the financial industry.

- ▷ The *Average return* permits to have an estimate of the daily realised return, expressed in percentage. Statistic significance is provided through a *t test*: *, **, *** to significant at 90%, 95%, 99%. In Table 3.6, we observe that for two data sets, the average return is statistically different from zero at 95% level. Hence, for the Electric Utilities sub industry, the average return from being long of volatility when a co-sector publishes its earnings is 0.6% while it is 0.7% for the Internet & Software Services sub industry.

- ▷ The *Total return* gives the total return generated by the strategy It is measured in %. Without transaction costs, being long in volatility when a co-sector publishes its earning generates a total return of 28.17% for the Electric Utilities and 15.81% for the Internet & Software Services sub industry.

- ▷ The *Sharpe ratio* is a performance measure taking account of the risk carried by the strategy. The performance is defined as expected excess return while risk is defined by the standard deviation of strategy returns. Sharpe ratio is defined as follows.

$$Sharpe = \frac{\mathbb{E}\left[R - R_{rf}\right]}{\sigma_R}.$$

The risk free rate is the LIBOR overnight rate of April,12th 2017 and worth

0.92611%[8]. For both sub industries, Sharpe ratios are always below 1 and so, the risk of the strategy is always greater than the excess return.

▷ The *Sortino ratio* is a modified Sharpe ratio that penalizes downside risk. The downside risk is defined arbitrarily by returns that are below the *Minimum Accepted Return* (MAR). Hereafter, we select the risk free rate as the MAR.

$$Sortino = \frac{\mathbb{E}\left[R - R_{rf}\right]}{\sigma_{R|R<MAR}}.$$

By considering only downside risk, Sortino ratios are greater than Sharpe ratio. This suggests that negative strategy returns have a lower standard-deviation than all strategy returns. The strategy computed on the Internet & Software Services sub industry generates the higher Sortino ratio (3.41), and is less risky than the Electric Utilities which remains below 1 (0.65).

▷ The *Omega ratio* is an alternative to Sharpe ratio, that take considerations of the whole distribution of strategy returns. Omega ratio compare an area of gain to an area of loss. Again, gains are strategy returns above the MAR.

$$Omega = \frac{\int_{MAR}^{\infty} (1 - F(x))\, dx}{\int_{-\infty}^{MAR} F(x)dx}.$$

The area of gains generated on the Electric Utilities is larger than its area of losses, like its Omega ratio is 3.41. However, Internet & Software Services has an Omega ratio below one (0.81).

▷ The *Upside potential ratio* is a measure of performance relative to a minimum acceptable ratio. Again, in this measure, the risk is defined as standard deviation of strategy returns that are below the MAR. Contrary to the Sortino ratio, the numerator is only composed of upside deviations with respect to the MAR.

$$UPR = \frac{\mathbb{E}[(R - MAR)_+]}{\sigma_{R|R<MAR}},$$

where $()_+ : \mathbb{R} \to \mathbb{R}^+, x \to (x)_+ = \mathbf{1}_{x>0}x$.

For both sub industries, UPR is larger than 1. The Internet & Software Services has the highest UPR : 3.49.

[8] All the indicators are computed on a daily basis, and so we express the risk free rate in a daily basis, by dividing it by $\sqrt{252}$.

Hence, after an investigation of difference performance measures, we conclude for the profitability of a trading strategy on volatility around EA. Moreover, these results shed lights on market inefficiencies around some corporate market events such as EA.

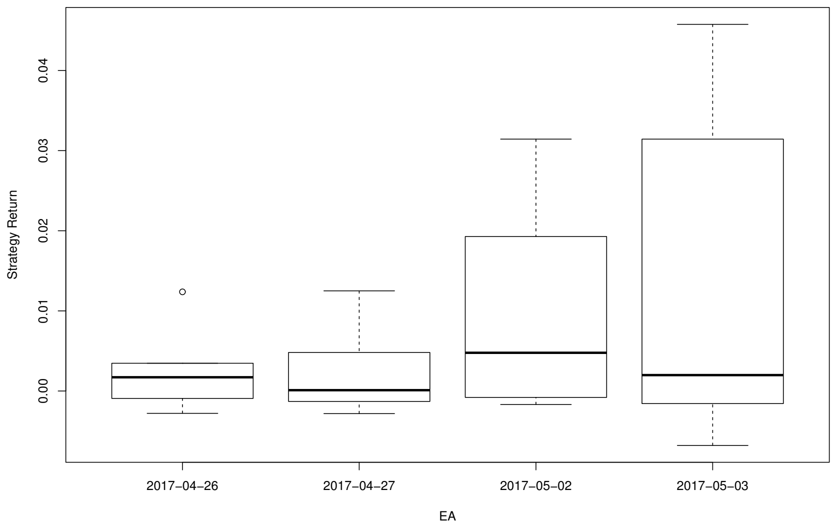
This simple illustration could be improved by several manner. First, we could implement the arbitrage strategy on all maturities and strikes. This will allow to measure the intra-industry spillover effect on the IV term structure as well as on the IV surface. Moreover, transaction and liquidity costs have to be added. Lastly, it would be interesting to implement such strategy on intra-day option data, to see whether the arbitrage opportunity is stronger at a lower time scale.

6 Conclusion

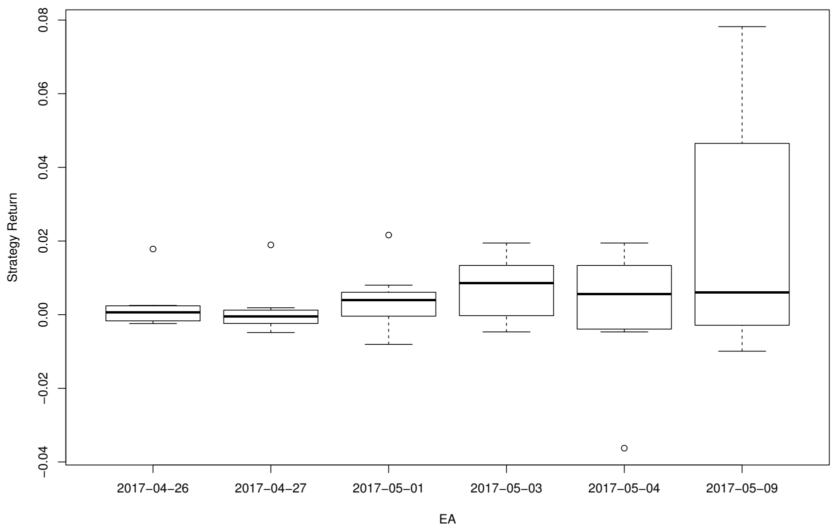
This article provides a first analysis of intra-industry volatility spillovers around EA and confirms existing results on intra-industry information spillover around EA. Using quantile regression, we find that both quantiles of the conditional volatility distribution are significantly affected by the EA. At the median level, the volatility spillover starts to increase two days before the EA, and go back to its long run level three days after it. In addition, the spillover effect varies across the nature of outcome, the intensity of the surprise the announcer capitalization, or by the market sentiment toward the announcer. By providing statistical evidences of volatility spillover around EA, this article strongly suggests the incorporation of this effect in portfolio management as well as in risk and pricing models.

Day	$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$	
-3	1e-04	(0.18)	-0.0028***	(-5.48)	-0.0049***	(-9.58)
-2	-6e-04**	(-1.96)	-0.0041***	(-8.02)	-0.0062***	(-12.14)
-1	-0.0015***	(-5.02)	-0.0052***	(-10.08)	-0.0073***	(-14.22)
0	0.0012***	(3.84)	-5e-04	(-0.9)	-0.0025***	(-4.97)
1	0.0022***	(7.31)	0.0031***	(5.97)	0.001**	(1.95)
2	0.0024***	(7.82)	0.0025***	(4.87)	4e-04	(0.85)
3	0.0014***	(4.51)	0.0015***	(2.86)	-6e-04	(-1.18)

Table 3.3: δ_m values. Hypothesis Testing. *H1-H2*: The spillover effect increases/decreases before/after the EA. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%. The regression is done on 32414 observations.



(a) Internet & Software Services



(b) Electric Utilities

Figure 3.9: Dispersion of strategy returns per announcement day. The Internet & Software Services has 4 announcements while Electric Utilities has 6 announcements. Boxes represent the dispersion of the returns.

Day	q1		q2		q3		q4	
			$\tau = 0.25$					
-3	0.0047***	(6.62)	-0.0027***	(-4.39)	0.0033***	(5.18)	-0.0013**	(-1.8)
-2	0.0033***	(4.68)	-0.0033***	(-5.3)	0.0047***	(7.46)	-0.0044***	(-5.88)
-1	0.0012**	(1.66)	-0.0056***	(-9.09)	0.0021***	(3.32)	-0.0022***	(-2.89)
0	0.0031***	(4.38)	-0.0043***	(-6.9)	0.0018***	(2.84)	0.0028***	(3.8)
1	0.0041***	(5.75)	4e-04	(0.66)	0	(0.04)	0.0023***	(3.06)
2	0.0049***	(6.91)	0.0016***	(2.52)	-0.001*	(-1.57)	0.001*	(1.4)
3	0.002***	(2.86)	0.001**	(1.67)	-0.002***	(-3.2)	0.0014**	(1.93)
			$\tau = 0.50$					
-3	-0.0022**	(-2.04)	-0.0077***	(-7.13)	0.0024**	(2.16)	-0.0066***	(-5.15)
-2	-0.0037***	(-3.44)	-0.01***	(-9.21)	0.0034***	(3.13)	-0.0111***	(-8.57)
-1	-0.004***	(-3.76)	-0.0143***	(-13.15)	-5e-04	(-0.47)	-0.0047***	(-3.64)
0	-3e-04	(-0.33)	-0.0096***	(-8.84)	-2e-04	(-0.21)	0.0012	(0.92)
1	0.0043***	(4)	0.0028***	(2.57)	5e-04	(0.47)	-0.0022**	(-1.72)
2	0.005***	(4.73)	0.0038***	(3.53)	-0.0012	(-1.08)	-0.0041***	(-3.19)
3	1e-04	(0.13)	0.0033***	(3.05)	-0.0027***	(-2.44)	-0.0027**	(-2.07)
			$\tau = 0.75$					
-3	-0.0042***	(-4)	-0.0106***	(-9.9)	5e-04	(0.48)	-0.0096***	(-7.57)
-2	-0.0057***	(-5.41)	-0.0128***	(-12.01)	0.0016*	(1.45)	-0.0141***	(-11.04)
-1	-0.006***	(-5.74)	-0.0171***	(-16)	-0.0024**	(-2.17)	-0.0077***	(-6.04)
0	-0.0024**	(-2.26)	-0.0124***	(-11.63)	-0.0021**	(-1.91)	-0.0018*	(-1.43)
1	0.0022**	(2.13)	-1e-04	(-0.07)	-0.0013	(-1.23)	-0.0052***	(-4.1)
2	0.003***	(2.88)	0.001	(0.91)	-0.003***	(-2.78)	-0.0071***	(-5.59)
3	-0.0019**	(-1.79)	5e-04	(0.42)	-0.0045***	(-4.15)	-0.0057***	(-4.44)

Table 3.4: δ_m values. Hypothesis Testing: *H3*: The spillover effect is stronger for bad news. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%.

Day	q1		q2		q3		q4	
			$\tau = 0.25$					
-3	0.0052***	(7.02)	-0.004***	(-5.02)	-0.0045***	(-6.42)	0.0042***	(6.39)
-2	0.005***	(6.81)	-0.0042***	(-5.29)	-0.0054***	(-7.7)	0.0033***	(4.99)
-1	0.0026***	(3.52)	-0.004***	(-4.95)	-0.0032***	(-4.56)	-3e-04	(-0.46)
0	0.0049***	(6.67)	-0.0031***	(-3.83)	0.0036***	(5.15)	6e-04	(0.87)
1	0.0085***	(11.51)	-9e-04	(-1.18)	0.0032***	(4.58)	0.0041***	(6.18)
2	0.0107***	(14.48)	-0.0025***	(-3.17)	0.0028***	(4.03)	0.0066***	(10.06)
3	0.0088***	(11.93)	-0.004***	(-4.94)	1e-04	(0.17)	0.0046***	(6.99)
			$\tau = 0.50$					
-3	9e-04	(0.73)	0.0016	(1.18)	-0.0138***	(-11.27)	0.0024**	(2.19)
-2	2e-04	(0.14)	-6e-04	(-0.44)	-0.0127***	(-10.32)	-6e-04	(-0.58)
-1	-5e-04	(-0.38)	-0.0035***	(-2.62)	-0.0094***	(-7.62)	-0.0028***	(-2.6)
0	0.0021*	(1.62)	-0.002*	(-1.46)	0.0036***	(2.92)	-0.0024**	(-2.24)
1	0.0103***	(7.93)	1e-04	(0.08)	0.0042***	(3.4)	0.0078***	(7.22)
2	0.0116***	(8.99)	-0.0014	(-1.07)	0.0015	(1.23)	0.0089***	(8.16)
3	0.0076***	(5.84)	-0.0036***	(-2.63)	-8e-04	(-0.66)	0.0074***	(6.83)
			$\tau = 0.75$					
-3	-0.0013	(-1.05)	0.0012	(0.86)	-0.0163***	(-13.35)	6e-04	(0.53)
-2	-0.0021*	(-1.64)	-0.001	(-0.77)	-0.0151***	(-12.39)	-0.0024**	(-2.26)
-1	-0.0028**	(-2.16)	-0.004***	(-2.95)	-0.0118***	(-9.69)	-0.0046***	(-4.3)
0	-2e-04	(-0.14)	-0.0024**	(-1.79)	0.0011	(0.91)	-0.0042***	(-3.93)
1	0.008***	(6.24)	-3e-04	(-0.24)	0.0017*	(1.39)	0.006***	(5.58)
2	0.0094***	(7.31)	-0.0019*	(-1.4)	-0.001	(-0.8)	0.007***	(6.52)
3	0.0053***	(4.13)	-0.004***	(-2.97)	-0.0033***	(-2.69)	0.0056***	(5.19)

Table 3.5: δ_m values. Hypothesis Testing: *H4*: The spillover effect is stronger for big surprises. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%.

Day	q1		q2		q3		q4	
			$\tau = 0.25$					
-3	0.0075***	(11.66)	0.0088***	(14.16)	-0.0088***	(-14.05)	-0.0077***	(-9.3)
-2	0.0076***	(11.7)	0.0084***	(13.53)	-0.0075***	(-12.04)	-0.0076***	(-9.14)
-1	0.0079***	(12.14)	0.0036***	(5.76)	-0.0115***	(-18.48)	-0.0045***	(-5.45)
0	0.0099***	(15.35)	0.0036***	(5.84)	-0.0059***	(-9.49)	-0.0036***	(-4.38)
1	0.0125***	(19.38)	0.0022***	(3.62)	-0.0036***	(-5.76)	-0.0032***	(-3.88)
2	0.0125***	(19.35)	0.0022***	(3.6)	-0.0036***	(-5.77)	-0.002***	(-2.42)
3	0.0119***	(18.39)	0.0025***	(3.97)	-0.0065***	(-10.35)	-0.0029***	(-3.52)
			$\tau = 0.5$					
-3	0.0072***	(6.18)	0.0146***	(14.35)	-0.0139***	(-13.63)	-0.0137***	(-9.92)
-2	0.0073***	(6.27)	0.0129***	(12.7)	-0.0149***	(-14.58)	-0.014***	(-10.11)
-1	0.0061***	(5.23)	0.0067***	(6.57)	-0.0189***	(-18.52)	-0.0111***	(-8.04)
0	0.0153***	(13.14)	0.0035***	(3.43)	-0.0113***	(-11.07)	-0.0096***	(-6.96)
1	0.0214***	(18.36)	0.0056***	(5.5)	-0.0039***	(-3.84)	-0.0059***	(-4.28)
2	0.0207***	(17.78)	0.0052***	(5.08)	-0.0021**	(-2.09)	-0.0074***	(-5.35)
3	0.019***	(16.31)	0.0062***	(6.09)	-0.0063***	(-6.21)	-0.0066***	(-4.78)
			$\tau = 0.75$					
-3	0.0053***	(4.59)	0.0129***	(12.73)	-0.0161***	(-15.98)	-0.0164***	(-11.93)
-2	0.0054***	(4.69)	0.0113***	(11.08)	-0.0171***	(-16.95)	-0.0166***	(-12.12)
-1	0.0042***	(3.64)	0.005***	(4.93)	-0.0211***	(-20.93)	-0.0138***	(-10.04)
0	0.0134***	(11.61)	0.0018**	(1.78)	-0.0135***	(-13.39)	-0.0123***	(-8.95)
1	0.0195***	(16.87)	0.0039***	(3.85)	-0.0061***	(-6.06)	-0.0086***	(-6.24)
2	0.0188***	(16.29)	0.0035***	(3.44)	-0.0043***	(-4.29)	-0.0101***	(-7.33)
3	0.0171***	(14.8)	0.0045***	(4.45)	-0.0085***	(-8.46)	-0.0093***	(-6.74)

Table 3.6: δ_m values. Hypothesis Testing: *H5*: The spillover effect is stronger for large caps. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%.

Day	Negative Sentiment		Positive Sentiment	
		$\tau = 0.25$		
-3	0.0029***	(5.31)	-0.001***	(-2.4)
-2	0.0012**	(2.17)	-0.0011***	(-2.83)
-1	0.0019***	(3.43)	-0.0026***	(-6.57)
0	0.0072***	(12.98)	-5e-04	(-1.17)
1	0.0115***	(20.77)	-3e-04	(-0.84)
2	0.0123***	(22.23)	-8e-04**	(-1.94)
3	0.0115***	(20.88)	-0.0022***	(-5.55)
		$\tau = 0.5$		
-3	-0.0011	(-1.15)	-0.0044***	(-6.48)
-2	-0.0014*	(-1.43)	-0.0062***	(-9.29)
-1	0.0014*	(1.5)	-0.0086***	(-12.87)
0	0.01***	(10.53)	-0.0053***	(-7.84)
1	0.0165***	(17.38)	-0.0014**	(-2.08)
2	0.0164***	(17.26)	-0.0025***	(-3.69)
3	0.0157***	(16.56)	-0.0038***	(-5.65)
		$\tau = 0.75$		
-3	-0.004***	(-4.27)	-0.0062***	(-9.3)
-2	-0.0043***	(-4.55)	-0.0081***	(-12.14)
-1	-0.0015*	(-1.59)	-0.0105***	(-15.75)
0	0.0071***	(7.52)	-0.0071***	(-10.68)
1	0.0136***	(14.43)	-0.0032***	(-4.87)
2	0.0134***	(14.31)	-0.0043***	(-6.49)
3	0.0128***	(13.6)	-0.0056***	(-8.47)

Table 3.7: δ_m values. Hypothesis Testing: *H6*: Investor sentiment. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%.

Day	Order=1		Order=2		Order>2	
			$\tau = 0.25$			
-3	-0.0026*	(-1.35)	0.0036**	(1.77)	1e-04	(0.19)
-2	-0.0029*	(-1.46)	0.0019	(0.93)	-6e-04**	(-1.88)
-1	-0.001	(-0.52)	8e-04	(0.4)	-0.0016***	(-5.04)
0	0.001	(0.53)	0.0025	(1.21)	0.0012***	(3.72)
1	7e-04	(0.37)	0.0011	(0.55)	0.0023***	(7.33)
2	5e-04	(0.26)	-6e-04	(-0.29)	0.0025***	(7.95)
3	-8e-04	(-0.41)	-6e-04	(-0.31)	0.0015***	(4.68)
			$\tau = 0.5$			
-3	-0.0076**	(-2.29)	0.0016	(0.46)	-0.0028***	(-5.31)
-2	-0.0072**	(-2.18)	7e-04	(0.2)	-0.0041***	(-7.88)
-1	-0.0054*	(-1.63)	6e-04	(0.17)	-0.0053***	(-10.03)
0	-0.0015	(-0.45)	0.0017	(0.5)	-5e-04	(-0.91)
1	0.0023	(0.7)	6e-04	(0.16)	0.0031***	(5.95)
2	0.0011	(0.33)	-0.0023	(-0.66)	0.0026***	(4.97)
3	-0.0048*	(-1.46)	-0.0032	(-0.94)	0.0017***	(3.18)
			$\tau = 0.75$			
-3	-0.0077***	(-2.34)	4e-04	(0.1)	-0.0049***	(-9.39)
-2	-0.0073**	(-2.23)	-5e-04	(-0.15)	-0.0063***	(-11.97)
-1	-0.0055**	(-1.68)	-6e-04	(-0.19)	-0.0074***	(-14.15)
0	-0.0016	(-0.49)	5e-04	(0.15)	-0.0026***	(-4.95)
1	0.0022	(0.67)	-7e-04	(-0.19)	0.001**	(1.95)
2	0.001	(0.3)	-0.0035	(-1.02)	5e-04	(0.97)
3	-0.0049*	(-1.51)	-0.0044*	(-1.3)	-4e-04	(-0.83)

Table 3.8: δ_m values. Hypothesis Testing: *H7*: Publication Order. Day 0 stands for the announcement day. T-values are in parenthesis and *, **, *** account for significance levels at 10%, 5%, and 1%.

Industry	Average Return	TR	Sharpe	Sortino	Omega	UPR
Electric Utilities	0.6 **	28.17	0.01	0.65	2.23	1.30
Internet Software & Services	0.7 ***	15.81	0.02	3.41	0.81	3.49

Table 3.9: Volatility Arbitrage Results. *, **, *** stands for significance levels at 90%, 95% and 99%. The Average return is the average return per position (in %), TR is the cumulated return on all the positions (in %), Sharpe, Sortino, Omega and UPR are calculated on all the positions.

Chapter 4

A denoising algorithm for an additive heteroscedastic noise and an application to the News Impact Curve modelisation.

This work has been co-written with Dr Matthieu Garcin, Natixis Asset Management. A different version of this work is available under the name *the News Impact Curve: A Variational Approach*. We thanks for their valuable comments: Christophe Boucher, Éric Gautier and Marco Mazzola. This acknowledgement does not make them accountable for any remaining mistake.

This work has been presented at the following seminar/conferences:

- ▷ *Developpements recents de l'econometrie appliquee a la Finance, Nanterre, 25/11/2015*
- ▷ *CFE, London, 12/2015*
- ▷ *Natixis Asset Management, Paris, 13/04/2016*
- ▷ *48e Journees de Statistique, Montpellier, 3/06/2016*
- ▷ *GDRE, Clermont-Ferrand, 8/07/2016*
- ▷ *World Finance Conference, New York, 28/07/2016*
- ▷ *Seminaire LEO, Universite d'Orleans, Orleans, 17/01/2017*

Abstract

In this paper we propose a new methodology for estimating a signal disrupted by a non-constant Gaussian noise. The challenge is therefore to get an accurate estimate of both the noise standard-deviation and of the unobserved signal. Our approach combines wavelet denoising techniques and a total variation based method. After having estimated the noise standard-deviation thanks to a total variation approach, the original signal is recovered through a standard wavelet shrinkage procedure. The algorithm we built can be easily extended to multidimensional signals and to other noise probability distributions. Last, the behaviour of the algorithm is illustrated through numerical simulations and an application in finance to the modelisation of volatility is proposed.

1 Introduction

During the last decade, noise removal techniques mainly focused on additive and multiplicative homogenous noises. Regarding additive denoising, two techniques have demonstrated their superiority: wavelets and total variation based. Wavelets is a recent theory well suited for noise removal, notably when the noise is homogenous and normally distributed. In this setting, the signal is distinguished from the noise among wavelet coefficients and most of the filters implemented depend on the noise standard-deviation. A first advantage of wavelets is due to the robust noise standard-deviation estimate developped by Donoho and Johstone (1994). A second advantage is the possibility of using the *Stein Unbiased Risk Estimate* in the filtering proces as it permits to have a robust estimate of the reconstruction error, between the signal estimate and the unobserved original signal **?**, **?**. Wavelet shrinkage is applied to diverse problems such as imagery, speech recognition, chaotic systems or in finance **?**, **?**, **?**, **?**. An alternative approach focuses on the Total Variation of the unobserved signal. Indeed, it assumes that the true signal is of bounded variations, and therefore, the problem consists in recovering the signal by minimizing an energy composed by a data fidelity term and by a data regularity constraint, that is the Total Variation semi norm. A great advantage of working with Total Variation is that it allows for discontinuities in the original signal and preserves rough edges in the denoising procedure. Rudin, Osher and Fatemi (1992) are the first to set up a denoising algorithm with this concept **?**. It has been generalized to other classes of estimators, such as *Maximum A Priori* estimators, and to several distributions **?**, **?**. Other additive noise denoising techniques exist: Kalman and Wiener filtering or Independent Component Analysis **?**, **?**. Multiplicative noises occur in domains such as in imagery with *active imaging systems* (Laser images, Satellite Aperture Radar images...) or in *finance* (heteroscedastic processes). A natural approach to recover the original signal consists in expressing the problem into an additive one, through a logarithmic transformation, and then to apply one of the methods exposed above. Recently, Aubert and Aujol (2008) propose a multiplicative denoising method for speckle noise, where the original signal is of bounded variations. This approach benefits from the advantage to not transform the original problem. After having demonstrated the existence and the unicity of the solution, they propose to recover the original signal by solving numerically a Euler-Lagrange equation **?**.

Most of the methods presented above, consider a noise with a constant variance. However, this assumption is too strong as much data will not show a constant variance. For instance, it is well known that the variance of speech signals change through times. In Finance, stock returns are characterized by their time dependent variance. Finally, in wavelet analysis, heteroscedasticity is introduced while transforming irregularly spaced observations into regularly spaced ones. One challenge with non-constant noise is to adapt the denoising procedure to the noise variance level: to eliminate signal as little as possible, when the noise variance is low, and to get ride as much noise as possible when the noise variance is large. To our knowledge, few studies are devoted to non-constant noises. In wavelet theory, threshold and filters have been adapted to such noises **?**, **?**, **?**, **?**. Notably, results regarding the *universal threshold* and the *Stein Unbiased Risk Estimate* threshold hold when a signal is disrupted by an non-constant noise. But, a remaining challenge consists in estimating the noise variance. Gao (1997) proposes a simple estimate based on a running window approach of the wavelet based standard-deviation estimator **?**. In total variation denoising, Gilboa, Sochen et al. (2006) develop a method where the regularization parameter is spatially varying **?**.
In this paper we propose an iterative algorithm to estimate successively the noise standard-deviation and the unobserved signal. Indeed, we consider that a signal disrupted by an additive and heteroscedastic noise can naturally be decomposed into both additive and multiplicative noise problems. The algorithm we propose, combines wavelet theory and a total variation approach. The signal is estimated with a standard wavelet approach while the noise standard-deviation is the solution of a Total Variation problem. The Total Variation part is an application of the algorithm proposed by Aubert and Aujol (2008), for a Gaussian noise **?**. The solution of such problem is computed numerically by solving a *Euler-Lagrange* equation. As soon as an estimate of the standard-deviation is obtained, the noise is removed in the wavelet basis using an appropriate noise dependent threshold. In this paper, we consider the case of a Gaussian noise with unit variance. After having presented the behaviour of the algorithm on simulated data, we propose a financial application to the modelisation of volatility. More precisely, we provide an application of this algorithm to the *News Impact* curve, describing the relation between past realized returns and current volatility.

The paper is organized as follows. In Section 2, we present the model and a short algorithm for its estimation. In Section 3, we run data simulations and in Section 4 we provide a financial application to volatility modelling.

2 Estimation Methodology

We consider the following model, for $x \in \Omega$ and $\Omega = [a, b] \subset \mathbb{R}$:

$$s(x) = y(x) + g(x)\varepsilon(x), \tag{4.1}$$

With $y \in L^2(\Omega)$ the original and unobserved signal, for all x, $g(x)\varepsilon(x) \sim \mathcal{N}(0, g(x)^2)$, and so: $\varepsilon(x) \sim \mathcal{N}(0, 1)$. The function g: $\Omega \to \mathbb{R}^+$, returns the noise standard-deviation and belongs to the space of Bounded Variation functions of Ω. As a result only $s \in L^2(\mathbb{R})$, is observed. The purpose of this paper is first to recover an estimate of the noise standard-deviation, that is to estimate the function g, and therefore to derive a robust estimate of the original signal y. In this particular setting, $g(x)\varepsilon(x)$ is the noisy part of the observed signal and in this paper we assess that $\varepsilon(x)$ are *i.i.d.* unit Gaussian random variables. Henceforth, the challenge consists in the estimation of both y and g. For this purpose, we will use wavelet and variational methods. On the one hand, wavelet denoising methods are well-designed to recover an original signal disrupted by an additive noise. Moreover, their theoretical framework permits to have an upper bound on the reconstruction error. On the other hand, recovering the function g is similar to recover a signal disrupted by a multiplicative noise: $s(x) - y(x) = g(x)\varepsilon(x)$. Such problem can be easily solved using a total variation approach.

2.1 Description

Challenges are double. First, we need to provide an accurate estimate of the noise standard-deviation, which can be expressed as a multiplicative denoising problem. Second, we employ the noise standard-deviation estimate to recover the original signal y. As we said above, we combine variational calculus and wavelets. The approach we propose is iterative, as estimation of y depends on g estimation. The variational part of our methodology is an extension of Aubert and Aujol (2008) algorithm to the case of multiplicative Gaussian noise, and the wavelet part is an application of wavelet shrinkage to a varying noise standard-deviation **?**.

Wavelet Part

The first preliminary step of our estimation algorithm is devoted to the decomposition of the signal y in a wavelet basis. This basis $(\psi_{j,k})$ of functions is obtained by dilatations and translations from a unique real mother wavelet, $\Psi \in L^2(\mathbb{R})$:

$$\psi_{j,k} : x \in \mathbb{R} \mapsto 2^{j/2}\Psi\left(2^j x - k\right),$$

where $j \in \mathbb{Z}$ is the resolution parameter and $k \in \mathbb{Z}$ is the translation parameter. As the observations are equispaced, we define the empirical wavelet coefficient $\langle y, \psi_{j,k} \rangle$ of

y, for the parameters j and k, by:

$$\langle y, \psi_{j,k} \rangle = \sum_x y(x)\psi_{j,k}(x). \tag{4.2}$$

Further details on wavelets and its use to denoise time series can be found in **??**.

We assume that we have already an estimate of g. Then, estimating y matches the quite classical problem of estimating a variable linearly disrupted by an inhomogeneous Gaussian noise. We can achieve it using wavelets filtering, like *SureShrink*, for example. More precisely, we have decomposed the signal y in a basis of wavelet functions. The coefficients of this decomposition are a noisy version of the pure coefficients $\langle y, \psi_{j,k} \rangle$. In order to get rid of that additive noise, we filter the coefficients and we build an estimate of y thanks to the inverse wavelet transform. Since the noise is Gaussian, we can use a *soft-threshold filter* or an *hard-threshold filter*. We recall that for the *soft-threshold filter* the filtered wavelet coefficients are $F_{j,k}(\langle y, \psi_{j,k} \rangle)$, where:

$$F_{j,k} : c \in \mathbb{R} \mapsto (c - \Lambda_{j,k})\mathbf{1}_{c \geq \Lambda_{j,k}} + (c + \Lambda_{j,k})\mathbf{1}_{c \leq -\Lambda_{j,k}},$$

and for the *hard-threshold filter*, the filtered wavelet coefficients are $F_{j,k}(\langle y, \psi_{j,k} \rangle)$, where:

$$F_{j,k} : c \in \mathbb{R} \mapsto c\mathbf{1}_{c \geq \Lambda_{j,k}},$$

for a level-dependent threshold $\Lambda_{j,k} = \lambda\sqrt{\langle g^2, \psi_{j,k}^2 \rangle}$ where λ is a parameter. Examples indeed show that a level-dependent threshold performs much more than a constant threshold **?**. The choice for λ may be arbitrary, but we prefer to optimize it, that is to choose the value of λ which minimizes an estimate of the reconstruction error. This is the aim of *SureShrink* **???**. The estimate of the reconstruction error is

$$\bar{\mathcal{S}} = \sum_j \sum_k \mathcal{S}_{j,k}(\langle y, \psi_{j,k} \rangle),$$

where

$$\mathcal{S}_{j,k} : c \in \mathbb{R} \mapsto \begin{cases} (\lambda^2 + 1)\langle g^2, \psi_{j,k}^2 \rangle & \text{if } |c| \geq \lambda\sqrt{\langle g^2, \psi_{j,k}^2 \rangle} \\ c^2 - \langle g^2, \psi_{j,k}^2 \rangle & \text{else,} \end{cases}$$

because $\langle g^2, \psi_{j,k}^2 \rangle$ is the variance of the empirical wavelet coefficient $\langle y, \psi_{j,k} \rangle$ **?**. Conditionally to y, $\bar{\mathcal{S}}$ is an unbiased estimate of the reconstruction error. Any basic optimization algorithm allows then to get the λ minimizing $\bar{\mathcal{S}}$. Thus, $\Lambda_{j,k}$ and $F_{j,k}$ for all j and k are now defined. We can hence write an estimate of the function x as:

$$\hat{y}(x) = \sum_j \sum_k F_{j,k}(\langle s, \psi_{j,k} \rangle)\psi_{j,k},$$

for $x \in \Omega$.

Variational Part

Suppose that a first estimate of y has been obtained, then Equation 4.1 can be expressed as a multiplicative denoising problem.

$$s(x) - y(x) = g(x)\varepsilon(x). \tag{4.3}$$

Knowing that ε are Gaussian random variables, the variational problem consists in finding g having the smallest variations over its support and returning ε, that are the *closest*, in the distribution sense, to $\mathcal{N}(0,1)$ random variables. A variational approach permits to reach this double objective.

Before introducing our methodology, we provide some recalls on functions with bounded variations (denoted BV functions).

Definition 13. *A function f is in $BV(\Omega)$, where Ω is an open subset of $\mathbb{R}$, if it is integrable, and there exists a Radon measure Df such that:*

$$< Df, \phi >= -\int_\Omega f(x)\phi'(x)dx, \ \textit{with } \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}).$$

When f is smooth enough the following relation between the distributional derivative and the derivative can be set:

$$Df(x) = \dot{f}(x).$$

The *total variation seminorm* of f is:

$$||f||_{TV(\Omega)} = \int_\Omega |Df| := \sup\left\{\int_\Omega f(x)\phi'(x)dx, \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}), \forall x \in \Omega, |\phi(x)| \leq 1\right\}$$

When f is smooth, the TV seminorm is equivalently the integral of its first derivative magnitude:

$$||f||_{TV(\Omega)} = \int_\Omega |\dot{f}(x)|dx.$$

As a result, the total variation approach consists in minimizing an energy composed by the TV seminorm and by a data fidelity term.

Aubert and Aujol (2008) set up a denoising model for a multiplicative noise. This model is built for speckle noise, that is present in images obtained through *Synthetic Aperture Radars*. In their framework, they consider pixels, that are positive. In the present paper, we apply their main results to the case of a Gaussian noise. The model consists in finding the function g in $BV(\Omega)$, that is the most probable under a given posterior distribution. This method is known as a *Maximum A Posteriori* approach (MAP). To derive the objective functional, we combine assumptions on g and on ε. For sake of simplicity, we set the following change of variables: $h(x) = s(x) - y(x)$, and Equation 4.3 becomes: $h(x) = g(x)\varepsilon(x)$. In the MAP approach, the aim is to maximize $P(G|H)$, which is obtained thanks to the *Bayes Rule* **?**[1].

Hence, we need to maximize $P(G|H) = \frac{P(H|G)P(G)}{P(H)}$, and so, a first challenge consists in calculating $P(H|G)$, which is obtained with the following proposition.

Proposition 1. *Assume G and $\mathcal{E}$ are independent random variables, with continuous probability functions p_G and $p_{\mathcal{E}}$. Then we have for $g > 0$:*

$$p_{\mathcal{E}}\left(\frac{h}{g}\right)\frac{1}{g} = p_{H|G}(h|g).$$

Proof. Let $\mathcal{A}$ an open subset in $\mathbb{R}$, we have:

$$\int_{\mathbb{R}} p_{H|G}(h|g)1_{h\in\mathcal{A}} = \frac{P\left(H \in \mathcal{A}|G\right)}{P(G)} = \frac{P\left(\mathcal{E} = \frac{H}{G} \in \frac{\mathcal{A}}{G}|G\right)}{P(G)}.$$

As G and $\mathcal{E}$ are independent: $\frac{P\left(\mathcal{E}=\frac{H}{G}\in\frac{\mathcal{A}}{G}|G\right)}{P(G)} = P\left(\mathcal{E} = \frac{H}{G} \in \mathcal{A}\right) = \int_{\mathbb{R}} p_{\mathcal{E}}(g)1_{\frac{h}{g}\in\mathcal{A}} = \int_{\mathbb{R}} p_{\mathcal{E}}(\frac{h}{g})1_{\frac{h}{g}\in\mathcal{A}}\frac{dh}{g}$ $\square$

As $\varepsilon \sim \mathcal{N}(0,1)$ then:

$$\varepsilon \to p_{\mathcal{E}}(\varepsilon) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\varepsilon^2}{2}},$$

and thanks to Proposition 1, $p_{H|G}(h|g) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(h/g)^2}{2}}\frac{1}{g}$.

A second challenge consists in selecting a relevant prior $P(G)$. An efficient one is given by the *Gibbs prior.*

$$p_G(g) = \frac{1}{Z}e^{-\gamma\phi(g)}.$$

[1] In the following lines, notations changes: $h(x)$, $\varepsilon(x)$, $g(x)$ become respectively h, ε, g.

Where Z is a normalizing constant, $\gamma > 0$ is a regularization parameter, and $\phi : \Omega \to \mathbb{R}$, a non-decreasing function.

Therefore, maximizing the quantity $P(G|H)$ is equivalent to minimize the log-likelihood:

$$-log(P(G|H)) = -P(H|G) - P(G) + P(H),$$

$$= C + \sum_n log(g(x_n)) + \frac{1}{2}\left(\frac{h(x_n)}{g(x_n)}\right)^2 + \gamma\phi(g).$$

For $C = \frac{1}{2}log(2\pi) + log(Z)$ and $\{x_1, ..., x_N\}$ a discretization of Ω.

By writing the continuous form of the above equation, and by setting $\phi(g) = ||g||_{TV} = \int_\Omega |\dot{g}(x)|dx$, we come up with the following functional to minimize:

$$J[g] = \int_\Omega log(g(x)) + \frac{1}{2}\left(\frac{h(x)}{g(x)}\right)^2 dx + \gamma \int_\Omega |\dot{g}(x)|dx. \tag{4.4}$$

In the next part, we set $\gamma = 1$, and introduce a new regularization parameter $\lambda_1 > 0$ such that the functional J to minimize becomes:

$$J[g] = \int_\Omega \lambda_1 \left(log(g(x)) + \frac{1}{2}\left(\frac{h(x)}{g(x)}\right)^2\right) dx + \int_\Omega |\dot{g}(x)|dx. \tag{4.5}$$

The parameter $\lambda_1 > 0$ is a smoothing parameter permitting to give more priority to the *closeness* part of the problem or to the variational constraint. When g is assumed to have high variations (resp. low), λ_1 will be set high (resp. low). Automatic procedures exist to select λ_1 according to some optimality criteria: *Cross-Validation*, *Stein Unbiased Risk Estimate* (SURE). However, as we are dealing with a multiplicative problem, therefore the standard SURE approach is not applicable any more.

Aubert and Aujol (2008) also provide a proof of both the existence and the uniqueness of a minimizer $g^\star$ to Problem 4.4.

The total variation constraint, $\int_\Omega |\dot{g}(x)|dx$ relies on the L1 norm, which is not differentiable every where. It can create *staircasing effects*. As suggested by Nikolova (2005), we replace the Total Variation term in Equation 4.6 by the following approximation **?**:

$$|\dot{g}(x)| \simeq \sqrt{\beta^2 + |\dot{g}(x)|^2}.$$

For $\beta > 0$.

A solution to Problem 4.5 can easily be derived thanks to the Euler-Lagrange equation. Before, we provide a new formulation of the problem expressed in Equation 4.5.

$$g^\star = \underset{g \in BV(\Omega)}{\text{argmin}} \int_\Omega H\left(x, g(x), \dot{g}(x)\right) dx. \tag{4.6}$$

Indeed, a minimizer of Problem 4.5 allows to reach an extremum of $\int_\Omega H\left(x, g(x), \dot{g}(x)\right) dx$ if and only if it verifies the Euler-Lagrange equation:

$$\forall x \in \Omega,\ 0 = \frac{\partial}{\partial g} H\left(x, g(x), \dot{g}(x)\right) - \frac{d}{dx}\frac{\partial}{\partial \dot{g}} H\left(x, g(x), \dot{g}(x)\right). \tag{4.7}$$

This proposition is a standard result of the variational theory. A proof can be found in **?**. This equation leads to a concise optimization problem, which finally enables to get an estimate of g.

In our context, we can simplify Equation 4.7 such that:

$$\forall x \in \Omega,\ 0 = \lambda_1 \left(\frac{1}{g(x)} - \frac{\left(s(x) - y(x)\right)^2}{g(x)^3} \right) - \frac{\frac{d}{dx}\dot{g}(x)}{\sqrt{\beta^2 + \dot{g}(x)^2}}.$$

An estimate of g can obtained by numerically solving the above partial differential equation.

To solve numerically Equation 4.7, we implement the following evolution equation:

$$\frac{d}{dj} g_j(x) = \frac{\frac{d}{dx}\dot{g}_j(x)}{\sqrt{\beta^2 + \dot{g}_j(x)^2}} + \lambda_1 \left(\frac{1}{g_j(x)} - \frac{\left(s(x) - y(x)\right)^2}{g_j(x)^3} \right),$$

which in a discrete setting leads to:

$$g_{j+1}(x) = g_j(x) + \delta \left(\frac{\frac{d}{dx}\dot{g}_j(x)}{\sqrt{\beta^2 + \dot{g}_j(x)^2}} + \lambda_1 \left(\frac{1}{g_j(x)} - \frac{\left(s(x) - y(x)\right)^2}{g_j(x)^3} \right) \right).$$

For $\delta > 0$ a parameter controlling for the speed of convergence. The quantity $\frac{d}{dx}\dot{g}(x)$ is approximated by a finite difference method: $\frac{d}{dx}\dot{g}(x) \simeq g_j(x_{n+1}) - 2g_j(x_n) + g_j(x_{n-1}))$. Lastly, the initial quantity g_0 is defined in the following subsection. To break the iterations, we propose to do it through the quantity $J[g]$: when it is stable enough, iterations are stopped.

In the next subsection, we present the algorithm proposed to estimate iteratively y and g.

2.2 Estimation Algorithm

The estimation of y and g is based on an iterative algorithm, since both the estimations require distinct techniques. However, some similar transformations of the data are used in each iterations. Therefore, they can be extracted from the iterative loop and they can be executed only once. They must be considered as preliminary steps of the algorithm. These steps relate both to the wavelet and to the variational approaches.

Let $\{x_1, ..., x_N\}$ be a discretization of Ω.

The algorithm can be summarized as in the following pseudo-code:[2]

```
1  WaveletCoefficients = GetWaveletCoefficients(y);
2  g_0 = Median(WaveletCoefficients)/0.6745 ;
3  for (i = 0; i <= NumberIteration1; i++){
4      FilteredWaveletCoefficients =
           GetFilteredCoefficients(WaveletCoefficients,NoiseAmplitude=g);
5      x = GetWaveletReconstruction(FilteredWaveletCoefficients);
6      for (j = 1; j <= NumberIteration2; n++){
7          for (n = min(Ω); n <= max(Ω) ; n ++){
8              g_{j+1}(x_n) = g_j(x_n) + delta * g_j(x_{n+1}) - 2 * g_j(x_n) + g_j(x_{n-1})
                   -lambda_1 * (g_j(x_n)^2 - (s(x_n) - y(x_n))^2)/g_j(x_n)^3); }}}
```

The second step of the algorithm consists in initializing the series of estimators. Depending on the *a priori* we have on the noise dynamic, we propose an ordinary estimate the noise standard-deviation, as usually done for wavelet denoising techniques with an homogeneous variance of the noise, $g_0 = M/0.6745$ **?**. M is the median of the absolute value of the wavelet coefficients of s at the finer scale. Indeed, $M/0.6745$ is a robust estimator for the noise standard-deviation. We can use also another initial estimate for the noise standard-deviation, such as a rolling window of the noise standard-deviation, suggested by Gao (1997) **?**. Similarly, for the first estimate of the trend, $\hat{y}$, other methods than wavelets can be implemented, such as the *A Maximum A Posteriori approach to remove asymmetric and leptokurtic noise* Goulet (2017).

[2] These lines of pseudo-code only present the main architecture of the algorithm. They refer to functions with explicit name, which also exist in many programming language but under another name. Specifically, $GetWaveletCoefficients$ creates a vector of wavelet coefficients, $Median$ calculates a median, $GetFilteredCoefficients$ applies a threshold filter to wavelet coefficients, $GetWaveletReconstruction$ computes an inverse wavelet transform and $GetOrderingIndexes$ provides the permutation allowing to sort in ascending order the coordinates of a vector. The loops iterate until *NumberIteration1* and *NumberIteration2*. A convergence criterion can be added so as to break the loop as soon as the likelihood of the model reaches a steady state.

From the partial differential equation derived above, the algorithm we proposed provide a numerical solution to our problem using standard finite-difference method. Moreover, thanks to the simple form of the Euler-Lagrange equation, this method could easily be extended to other probability distributions of the noise than the Gaussian law, such as Student-t or Poisson laws.

The algorithm we introduce is denoted hereafter the *wavelet-variational* approach.

3 Data Simulation

In the present subsection, we present and compare the ability of our denoising algorithm on simulated data. Following Donoho and Johnstone (1995), Gao (1997) and Delouille et al (2004) we select the HeavySine function **?**, **?**, **?**.

The HeavySine function is defined by:

$$y(x) = 4\sin(4\pi x) + \operatorname{sign}(x - 0.3) - \operatorname{sign}(0.7 - x),$$

for $x \in \Omega = [0, 1]$.

The particular challenge with the HeavySine signal is to preserve the two local singularities around $x = 0.3$ and $x = 0.7$. Indeed, the signal jumps at these points, which leads to large wavelet coefficients. In the denoising procedure, the challenge is therefore to preserve these two singularities.

Following Gao (1997) we propose two different standard-deviation dynamics:

1. $g(x) = \begin{cases} v + a(x - \omega_1)^2 \text{ if } 0 \leq x < \omega_1 \\ v \text{ else} \\ v + b(x - \omega_2)^2 \text{ if } \omega_2 \leq x < 1 \end{cases}$

2. $g(x) = v + (a - b(x - c)^2)^2$ if $0 < x < 1$.

These functions are depicts in Figure 4.1. We simulate the HeavySine signal on 512 observations and for a *Signal To Noise* ratio of 7.

For Model 1, this corresponds to :

▷ $\omega_1 = 0.3$,

▷ $\omega_2 = 0.7$,

- ▷ $a = 22$,
- ▷ $b = 22$,
- ▷ $v = 1e - 3$;

And for Model 2:

- ▷ $a = 0.96$
- ▷ $b = 6$,
- ▷ $c = 0.515$.

In a non-constant noise context, the Signal To Noise Ratio is defined by **?** :

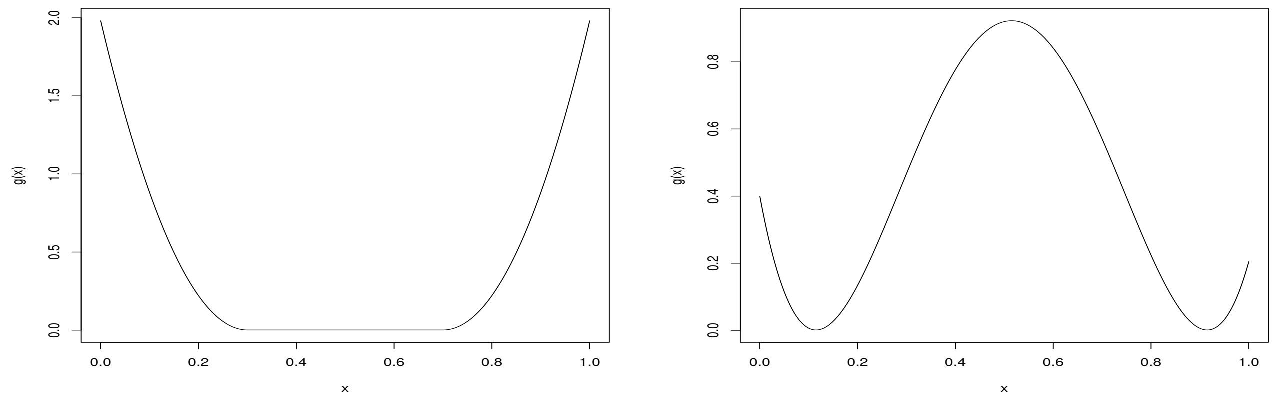
$$SNR = \frac{\sigma_y}{\displaystyle\int_\Omega g(x)dx}.$$

We select the *hard-threshold function*, as by decreasing the amplitude of largest coefficients, the *soft-threshold function* smooths the two singularities of the *HeavySine* signal. Moreover, we use the *Daubechies* wavelet family with four null moments. We simulate 1000 noisy signals, with 512 observations each. We compare our approach to cases where the standard-deviation is estimated trough:

- ▷ a rolling window on the wavelet coefficients at the finest scale. The width of the window is set to 50 **?**,
- ▷ a constant standard deviation.

For Model 1 and Model 2, the *Signal To Noise* ratio is set to seven. To compare performances of each estimation techniques, we compute Mean Square Error (MSE) and the Mean Absolute Deviation. Unlike MAD, MSE penalizes large deviations from the true unobserved signal. More precisely, as we are running 1000 simulations, we have consequently 1000 MSE and 1000 MAD. Average of these statistics as well as there standard deviation are provided in Table 4.1.

For Model 1, the MSE obtained with the *wavelet-variational* ($4.40e - 2$), is a bit more than two times smaller than the one obtained with the *rolling-window* ($1.05e - 1$). In terms of MAD, the ratio is slightly smaller, as the MAD obtained with the *wavelet-variational* ($1.01e-1$) is less than two times smaller than the MAD of the *rolling-window* ($1.95e - 1$). Moreover, as MAD are larger than MSE, we deduce that all estimates of



(a) Standard Deviation 1 (b) Standard Deviation 2

Figure 4.1: Two functions g. As in Gao (1997), they are simulated on 512 observations.

y still carry a weak amount of noise.

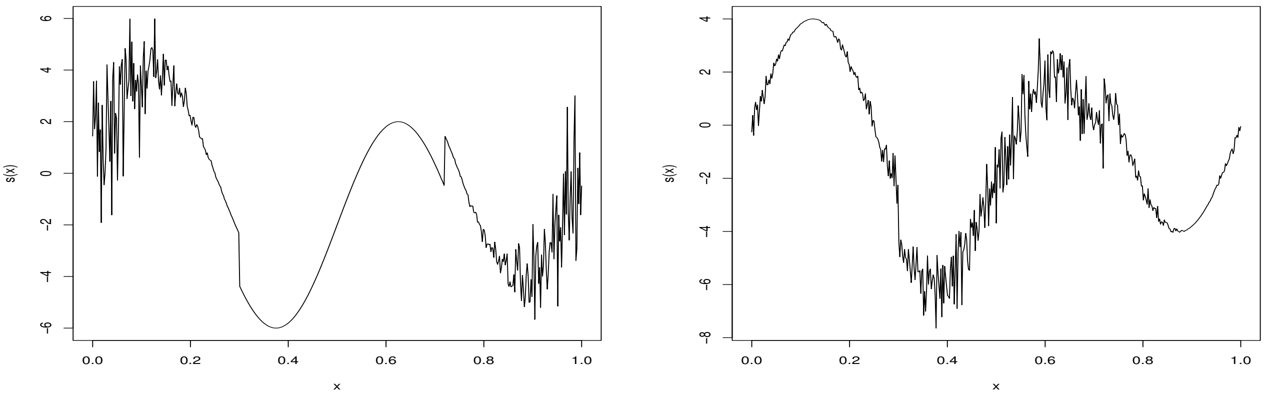
For Model 2, differences among all denoising techniques are smaller. In terms of MSE, the one obtained with the *wavelet-variational* ($9.50e-2$), is 25% weaker than the MSE obtained though the *rolling-window* technique ($1.19e-1$). In terms of MAD, differences among techniques are, again, smaller as the MAD for *wavelet-variational* ($1.70e-1$) estimates is 19% weaker than the MAD of *rolling-window* ($2.03e-1$) estimates.

As estimates of the true signal y, directly rely on estimates of the non-constant standard-deviation, we provide MAD and MSE computed between g and the estimates obtained using each technique. In Table 4.2, we observe that differences, in terms of MAD and MSE, across models is smaller than differences in MAD or MSE obtained for y estimates. This strongly supports the idea that the final error (the error between y and an estimate), increases with the number of steps in the denoising process. This also encourage improvements regarding the choice of the threshold in the trend estimation.

Graphical representations of some estimates are available in Figure 4.3 and in Figure 4.4.

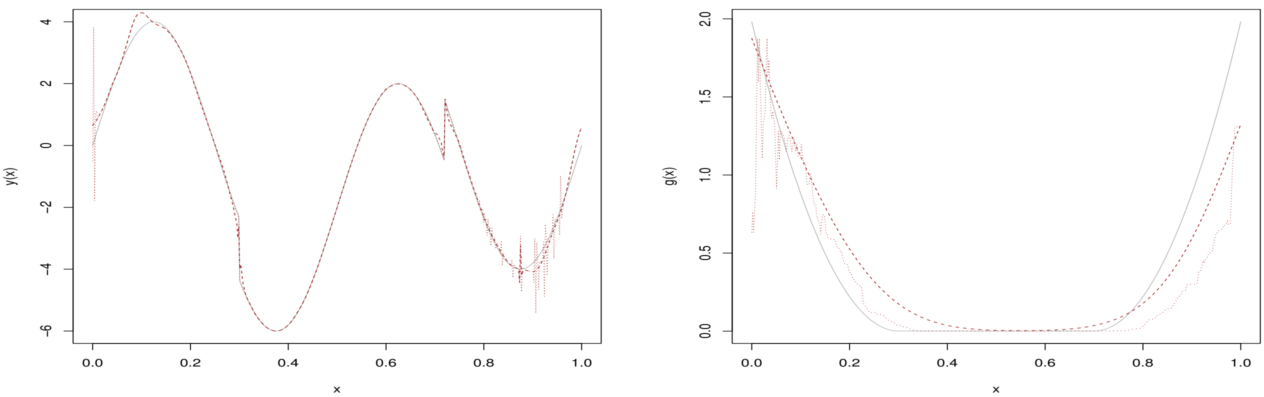
4 Financial Application : WV-ARCH and the News Impact Curve

For the last three decades, various models have been proposed for forecasting volatility. Volatility has many applications in finance and its forecast is indeed useful in risk



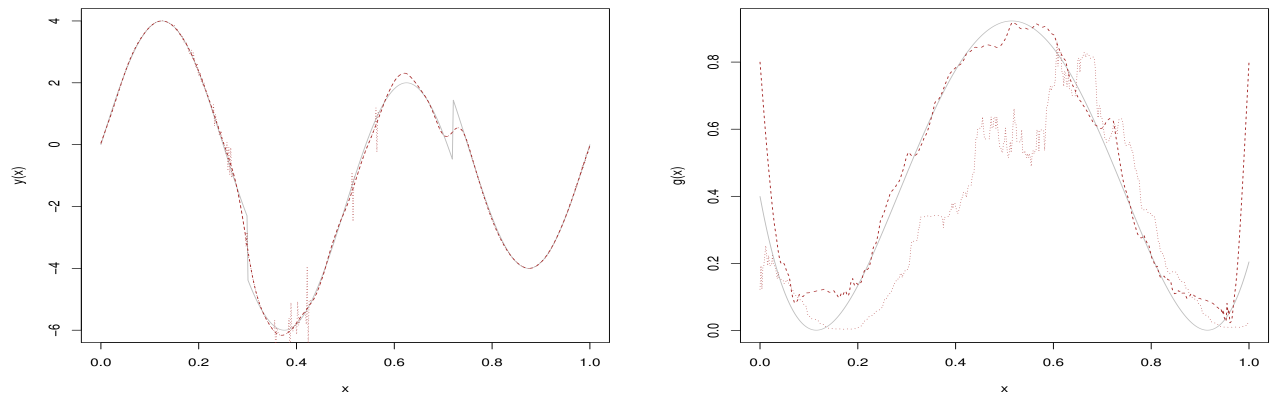
(a) Model 1 (b) Model 2

Figure 4.2: Two noisy signals corresponding to the two models for g. As in Gao (1997), they are simulated on 512 observations. The *Signal To Noise* ratio is set to 7.



(a) Model 1: y estimates. (b) Model 1: g estimates

Figure 4.3: Estimates of y and g for Model 1. In grey is original signals, in dashed brown, estimates using our approach and in dotted brown estimates using a running window. Estimates are obtained after two iterations of our algorithm.



(a) Model 2: y estimates.

(b) Model 2: g estimates

Figure 4.4: Estimates of y and g for Model 2. In grey is original signals, in dashed brown, estimates using our approach and in dotted brown estimates using a running window. Estimates are obtained after two iterations of our algorithm.

measurement, portfolio management, trading strategies or options pricing. One way of forecasting volatility is to take advantage of its interaction with returns. The *news impact curve* defines the relation between past returns and current volatility **?**. This relation has been well described and presents two notable characteristics. First, it is a function that decreases for negative past returns and then increases for positive past returns. It thus reaches its minimum around 0: the weaker the lagged returns in absolute value, the weaker the uncertainty concerning the next return. Second, it is an asymmetric function. Indeed, "bad news" (high negative returns) create more volatility than "good news" (high positive returns). Hence, volatility models have to take account of these two features. For example, ARCH models impose a symmetric and hyperbolic relation between past returns and current conditional volatility **?**.

Parameters can be added to the news impact curve to take into account returns asymmetry. GJR-GARCH **?** and E-GARCH **?** are such parametric asymmetric models. They are extensions of the GARCH model, which has been introduced by Bollerslev **?** and which depicts the self-dependence of the volatility across time. In other words, GARCH models consider volatility as a weighted sum of a news impact curve and of past volatility realizations. All these parametric hyperbolic news impact curves have some limits. First, adding parameters for better matching statistical properties of volatility increases the estimation complexity. As a consequence, their estimation requires a large number of observations to converge. Then, since self-dependence of the volatility results in its persistence, GARCH-oriented models may be slow to react to extreme market moves **?**.

Model	Method	MSE	MAD
1	WV	$4.40e-2$ $2.0e-2$	$8.90e-2$ $1.7e-2$
	Varying Wavelets	$1.05e-1$ $3.3e-2$	$1.29e-1$ $1.8e-2$
	Constant	$4.77e-1$ $6.6e-2$	$3.95e-1$ $<1e-4$
2	WV	$9.50e-2$ $1.31e-2$	$1.70e-1$ $1.29e-2$
	Varying Wavelets	$1.19e-1$ $1.4e-2$	$2.03e-1$ $1.2e-2$
	Constant	$2.24e-1$ $2.5e-2$	$3.22e-1$ $4.0e-3$

Table 4.1: Mean Squared Error associated for y. MSE and MAD provided are average over 1000 obtained measures. Standard deviation are below.

Model	Method	MSE	MAD
1	WV	$5.00e-2$ $1.8e-02$	$1.01e-1$ $1.6e-2$
	Varying Wavelets	$1.19e-1$ $2.0e-2$	$1.95e-1$ $1.5e-2$
	Constant	$4.54e-1$ $3.0e-3$	$3.96e-1$ $<1e-4$
2	WV	$1.00e-2$ $1.1e-02$	$7.50e-2$ $2.1e-2$
	Varying Wavelets	$4.20e-2$ $9.0e-3$	$1.61e-1$ $1.8e-2$
	Constant	$1.68e-1$ $7.0e-03$	$3.21e-1$ $4.0e-3$

Table 4.2: Mean Squared Error associated for g. MSE and MAD provided are average over 1000 obtained measures. Standard deviation are below.

Semi-parametric and non-parametric approaches overpass the constraint of a hyperbolic relation between volatility and past returns. Pagan and Schwert (1990) developed a smooth function, based on a Nadaraya-Watson estimator, to model conditional variance **?**. Gouriéroux and Montfort (1992) are the first to build a semi-parametric equation (QTARCH) to model conditional mean and conditional variance **?**. Their approach mixes Markovian theory and parametric step functions. Härdle and Tsybakov (1997) extended their work to a larger class of step functions **?**. Linton and Mammen (2005) built a semi-parametric estimate of the news impact curve using local linear estimators **?**. This last model differs from others by the inclusion of the volatility persistence. By iterating local linear regressions, one can get a fully nonparametric estimate too **??**. These approaches are based on the minimization of a quadratic error. However, the local linear models may lead to a negative volatility estimate. In this case, to improve the accuracy of the estimation, the tilted nonparametric approach introduces a weighted local linear method for the volatility in which the weights result from the minimization of an empirical likelihood of the observed returns **?**. Other nonparametric approaches for ARCH or GARCH models, consisting in minimizing a squared error, include polynomial splines **?** or neural networks **?**. Other approaches have been developed to take into account the long memory property of volatility and non-parametric news impact curves **???**.

Most of the mentioned methods are based on non-parametric regressions and estimators rely on the minimization of some quadratic error. The tilted nonparametric estimation takes into account the empirical distribution of the observed returns but does not relate to the distribution of residuals **?**.

In the present section, we propose an application of the algorithm introduced in section 2 to the modelisation of the News Impact Curve. An indirect contribution of this application is to show whether Signal Theory techniques are suitable in volatility modelling and can be use with success in other financial issues.

We consider the following discretized version of model 4.1, for $t \in \{0, ..., T\}$:

$$\left\{ \begin{array}{lcl} y_t & = & x_t + u_t \\ u_t & = & g(u_{t-l})\varepsilon_t \end{array} \right. \tag{4.8}$$

Model	**Equation**
ARCH(1)	$h_t = \omega + \alpha u_{t-1}^2$
NP-ARCH(1)	$\sqrt{h_t} = m(u_{t-1})$
GARCH(1,1)	$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1}$
GJR-GARCH (1,1)	$h_t = \omega + \alpha u_{t-1}^2 + \beta h_{t-1} + \theta 1_{u_{t-1}<0} u_{t-1}^2$

Table 4.3: Time series models.
Each model is defined by three equations: the first two are shared by all the models ($y_t = x_t + u_t$ and $u_t = \sqrt{h_t}\varepsilon_t$), the third one is specific and defines the volatility, as written in the table. All the innovation processes $(\varepsilon_t)_t$ are assumed to be standard Gaussian variables. In ARCH-NP, m is a kernel function.

In which y is the observed price return of an asset, x is an unknown deterministic function corresponding to the return of the fundamental asset value, $l \geq 1$ is an integer indicating the number of lags in the information and u_t is the noisy part of the observed price at time t. More precisely, for every t, the innovation ε_t is a unit Gaussian random variable which is independent of ε_s, for every $s \neq t$, and of x_u, for every u. Moreover, g is an unknown deterministic and positive function, corresponding to the standard-deviation of y_t conditionned to the past realizations of u_t. We can easily compute the first unconditional moment and the conditional variance of the model defined by equation (4.1):

$$\begin{cases} \mathbb{E}\left[y_t\right] = x_t \\ V\left[y_t|\mathcal{I}_t\right] = h_t, \end{cases}$$

where $\mathcal{I}_t = \{y_0, ..., y_{t-1}\}$.

As we said above, in this model, the conditional standard-deviation of returns is a function of past innovations. This model is a very general form of an ARCH-type model with time-varying trend. We could set a parametric form to g, to recover a standard ARCH model but we prefer to define a non-parametric framework both for the trend, x, and for the news impact curve, g. We denote our model the WV-ARCH. In this application, we only provide g for $l = 1$. The main challenge of the WV-ARCH model is to provide a relevant estimate of g. As the theoretical form g is unknown, we cannot compare graphically the estimate $\hat{g}$ to g. To overcome this issue, we compare each estimated news impact curve $\hat{g}$ to an estimate σ_t of the realized volatility. An estimate of the realized volatility for S&P 500, FTSE 100 and DAX is available on *Oxford-Man Institute realized library*[3]. The realized volatility is estimated using a Kernel method, which has the advange of being robust to intra-day noise **?**.

We consider three stock indexes: S&P 500, FTSE 100 and DAX. First, we present the results of the algorithm for the estimation of the news impact curve g. Then, we forecast the instantaneous volatility and we compare the obtained results with other conditional volatility models.

Figure 4.5 shows an example of the estimation of g for S&P 500 log-returns between the 3rd of January 2000 and the 24th of August 2007. For the three log-return series used, we obtain a relationship between the instantaneous innovations and the one-time-ahead volatility which captures both returns asymmetry and volatility clustering. The clustering effect occurs because high returns in absolute value are followed by high volatility. The asymmetry is striking in Figure 4.5, in which the distance between our non-parametric news impact curve and the asymptotic line of the ARCH hyperbolic news impact curve is much bigger on the left than on the right. For the particular case of stock returns, the returns asymmetry can be interpreted as the leverage effect.

To validate the model we designed, we compare it to the classical time series framework. We treat four well-known models, whose equations are given in Table 4.3: the ARCH model **?**, the non-parametric ARCH model based on Nadaraya-Watson kernel estimator **?**, GARCH(1,1) **?** and the GJR-GARCH(1,1) **?**. To improve the accuracy and the fairness of the comparisons, we replaced the constant drift term in all the time

[3] `http://realized.oxford-man.ox.ac.uk/documentation/econometric-methods`

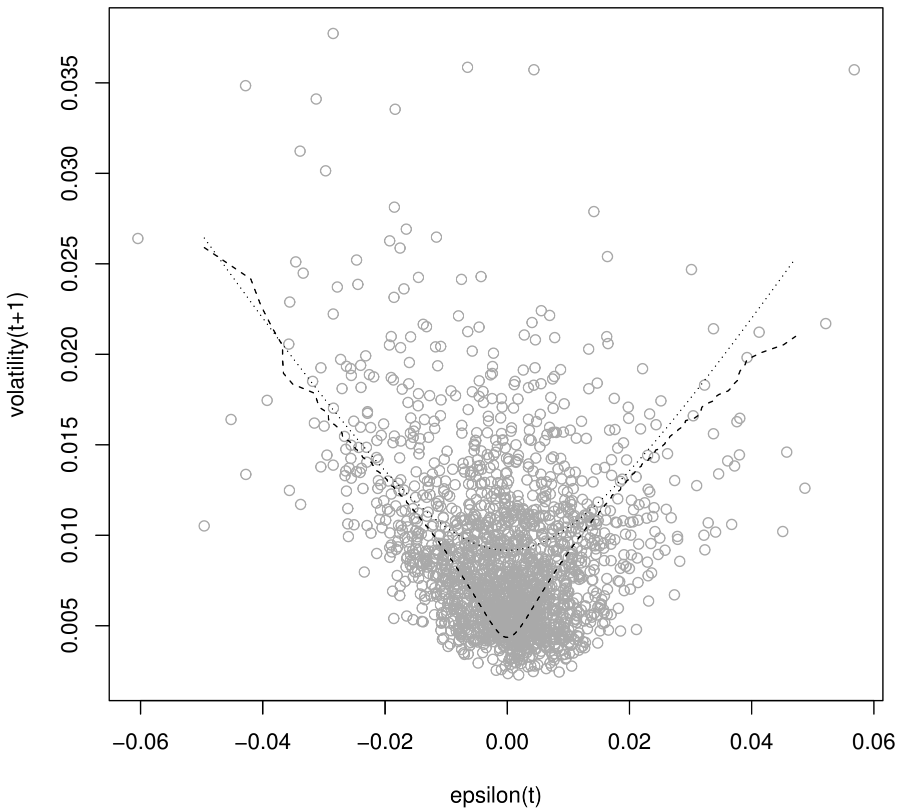


Figure 4.5: Estimates of g for S&P 500 log-returns. The bullet points are the realized volatility, the dashed grey line is the estimate of g after one iteration in the first loop indexed by i of the pseudocode presented in section 2. The dotted grey line is the news impact curve estimated for an ARCH(1) model. Estimation is done for the S&P 500 volatility, between the 3rd of January 2000 and the 24th of August 2007.

series model by a moving drift term, x_t, obtained by wavelet shrinkage.

First, we estimate the five models between the 3rd of January 2000 and the 24th of August 2007 (approximately 2000 observations) and we study the distribution of the innovations obtained. Second, we recursively forecast the volatility. Finally we study forecast errors.

4.0.1 In-sample estimation

To find the model that better reproduces the stylised facts of log-returns, we compare their residuals. The purpose of this study is to find whether our model better matches the statistical properties of financial assets log-returns. Various papers have studied these statistical properties **???**. We focus on five stylised facts: conditional heteroskedasticity, volatility clustering, leverage effects, leptokurtic and asymmetric distributions. A model reflecting those stylised facts should have residuals consistent with the distribution specified in the model. It is thus expected from the innovations of all these Gaussian models to have moments close to those of a unit Gaussian. In particular, we focus on the four first moments. We also expect that the innovations pass a Gaussian test. And finally, a significantly better model will have a higher likelihood. We gather those statistics for the five models on S&P 500, FTSE 100 and DAX in Table 4.4.

The estimates of the WV-ARCH model are obtained after only one iteration of our first loop, indexed by i, in the pseudocode provided in Section 2.[4] By doing so, all the tested models have the same estimated trend x_t, which does not depend on our non-parametric news impact curve g.

The innovations of the WV-ARCH model seem globally close to a standard Gaussian distribution. More precisely, among the three ARCH models, the innovations of the WV-ARCH model are the closest to a Gaussian variable, regarding the mean, the skewness, the kurtosis and the log-likelihood. Except for the FTSE 100, The p-value of the Kolmogorov-Smirnov test on the innovations is always higher for the WV-ARCH than for ARCH and NP-ARCH models. In particular, the normality is not rejected for the WV-ARCH, whereas it is always rejected for ARCH and NP-ARCH. The difference with innovations of GARCH-oriented model mitigates the superiority of the WV-ARCH model. This is due to the introduction of the persistence of the volatility, β. Across all the datasets, only the mean, the skewness and the log-likelihood unanimously state

[4] For a practical use, more iterations can provide a slightly higher accuracy in the estimation of the WV-ARCH model. However, by doing so, the drift incorporates iterated estimations of the news impact curve. Therefore, in order to make a fair comparison of all the models, we restrict to only one iteration so that the drift does not depend on the estimated non-parametric news impact curve. It can thus be used as a mutual drift for all the models.

Data	Model	Mean	Variance	Skewness	Excess Kurtosis	Log-lik.	K.-S. p-value
S&P 500	ARCH	$-5.28e-3$ 81.4%	1.005 93.8%	0.116 3.4%)	1.65 <0.1%	-2842	<0.1%
	NP-ARCH	$2.18e-2$ 32.0%	0.966 23.9%	0.170 0.2%	0.91 <0.1%	-2803	8.9%
	GARCH	$-7.45e-3$ 74.0%	1.006 91.4%	-0.144 0.8%	1.10 <0.1%	-2843	35.6%
	GJR-GARCH	$-2.44e-2$ 27.6%	1.006 93.0%	-0.263 <0.1%	1.20 <0.1%	-2843	28.6%
	WV-ARCH	$-1.32e-3$ 95.1%	0.929 1.7%	0.014 80.1%	-0.62 <0.1%	-2766	29.5%
FTSE 100	ARCH	$-5.56e-3$ 80.4%	1.005 95.0%	-0.078 15.6%	2.10 <0.1%	-2842	< 0.1%
	NP-ARCH	$8.61e-3$ 70.0%	0.996 82.9%	0.037 49.6%	1.08 <0.1%	-2833	3.3%
	GARCH	$-8.48e-3$ 70.5%	1.007 88.3%	-0.192 <0.1%	0.26 <0.1%	-2843	80.3%
	GJR-GARCH	$-1.60e-2$ 47.5%	1.006 91.4%	-0.230 <0.1%	0.26 <0.1%	-2844	84.9%
	WV-ARCH	$4.45e-4$ 98.4%	0.987 62.5%	-0.018 74.4%	-0.61 <0.1%	-2818	2.1%
DAX	ARCH	$-1.00e-2$ 65.5%	1.003 99.0%	-0.051 34.7%	1.84 <0.1%	-2840	< 0.1%
	NP-ARCH	$1.97e-2$ 37.3%	0.981 49.0%	0.206 <0.1%	1.55 <0.1%	-2818	< 0.1%
	GARCH	$-9.20e-3$ 68.2%	1.003 98.4%	-0.167 0.2%	0.36 0.4%	-2840	23.6%
	GJR-GARCH	$-2.22e-2$ 32.3%	1.004 96.3%	-0.237 <0.1%	0.36 0.3%	-2842	27.0%
	WV-ARCH	$-1.20e-3$ 95.8%	0.961 18.6%	0.015 77.8%	-0.64 <0.1%	-2798	14.8%

Table 4.4: Estimation results for S&P 500, FTSE 100 and DAX log-returns: four first moments of the innovations, as well as log-likelihood of the estimated model and p-value of a normality test for the innovations. Under the four first moments is indicated the p-value of the moment for the null hypothesis of a standard Gaussian distribution. Models are estimated on the first 2000 observations. For WV-ARCH, $N = 36$, and the parameters of the variational problem are $\lambda_1 = 4 \times 10^{-4}$, $\delta = 1 \times 10^{-4}$. For NP-ARCH, the bandwidth is set to 2×10^{-3}.

a higher closeness of the $\mathcal{N}(0,1)$ distribution with WV-ARCH innovations than with GARCH and GARCH-GJR innovations. This result highlights the ability of WV-ARCH to model log-returns asymmetry.

4.0.2 Out-of-sample forecasts

In this section, we compare the forecast ability of WV-ARCH model out-of-sample with the forecast ability of the four other models presented in the previous section.

Table 4.5 presents the results of DMW tests for the five models.

Data	Model	QLIKE	DMW *vs* WV-ARCH
S&P 500	ARCH	3.57	2.96***
	NP-ARCH	3.81	2.87***
	GARCH	0.27	2.13**
	GJR-GARCH	0.25	1.90**
	WV-ARCH	0.18	-
FTSE 100	ARCH	3.08	4.16***
	NP-ARCH	2.47	3.93***
	GARCH	0.31	2.64***
	GJR-GARCH	0.32	2.62***
	WV-ARCH	0.13	-
DAX	ARCH	2.97	3.19***
	NP-ARCH	1.50	4.19***
	GARCH	0.34	2.92***
	GJR-GARCH	0.35	3.04***
	WV-ARCH	0.12	-

Table 4.5: QLIKE losses and DMW statistics for the three series of log returns. *, ** and *** respectively signify rejecting the null hypothesis of equal losses for 90%, 95% and 99% confidence levels. Forecasts are done during 400 trading days, between the 27th of August 2007 and the 6th of March 2009. The benchmark is the WV-ARCH model.

Forecasts are done during 400 trading days, between the 27th of August 2007 and the 6th of March 2009, so that it includes periods of high and low volatility. Models are re-estimated at each time step using a rolling-window procedure. The size of the rolling window is set to 1000 trading days. In Table 4.5 we observe that, for each series of log-returns, the null hypothesis of equality of QLIKE loss function is always rejected and so, better predicting abilities of the WV-ARCH model are confirmed for all the datasets. Indeed, differences in terms of mean of QLIKE loss functions are significant. Among all other models, GARCH-oriented models have better results than ARCH and NP-ARCH. This can be explained by the different ways we include the volatility persistence in GARCH- (β, as in Table 4.3) and ARCH-oriented models (only implied by the rolling-window procedure). However, when comparing QLIKE losses, differences between GARCH and GJR-GARCH do not seem significant.

5 Conclusion

The algorithm we propose to recover a signal disrupted by a non-constant noise behaves well on simulated data as well as on financial data. This algorithm has the advantage to be easy to extend to other noise distribution and to other class of signals. Notably, we believe that this algorithm is well-suited for audio or video signals. However, the technique we propose can be improved by adding a selection procedure for the regularization parameter λ_1. For instance, Guan et al. (1998) propose a selection procedure based on unsupervised learning while Reeves et al.(1994) suggest a Cross Validation approach **?**, **?**.

The WV-ARCH enables to model the news impact curve and its asymmetry with a quite simple algorithm. This method has better estimation and forecast results than standard heteroscedastic models for simulated processes and financial data, without visible over-fitting. Moreover, the model is well specified since we get standard Gaussian innovations. Therefore, heavy tails or asymmetry are well described by the non-parametric news impact curve g rather than by the probability law of the innovations. Common alternative such as introducing more sophisticated probability laws to improve the basic ARCH model are thus avoided.

Chapter 5

A Maximum A Posteriori approach to remove asymmetric and leptokurtic noise

I am grateful for the fruitful comments of Jean-François Aujol, Khalid Jalalzai, Matthieu Garcin, Marco Mazzola and Olivier Guéant. This acknowledgement does not make them accountable for any remaining mistake.

This work has been presented at the following seminar/courses:

▷ *Phd Lunch Seminar, University Paris 1, 10/11/2016*

▷ *Hedge Fund course, Olivier Gueant, master IRFA, 03/2017*

Abstract

Total Variation methods are applied with success to denoising problems. These approaches recover an original signal by minimizing both its Total Variation and a closeness constraint with the true signal. Generally, the closeness constraint is a standard Mean Squared Error (MSE), which may be inappropriate for a noise with specific properties such as asymmetry or leptokurticity. In this paper, we apply the *Maximum A Posteriori* approach to asymmetric and leptokurtic noise probability distributions. The original signal is easily obtained thanks to the Euler-Lagrange equation and numerical simulations for Gaussian, Student, skewed Gaussian and skewed Student noises are provided. Lastly, we propose an application in Finance to the denoisement of an investment strategy based on pattern recognition.

1 Introduction

A common interest of many fields of sciences or engineering is to recover a unidimensional signal y from noisy observations s, where $\forall\ x \in \Omega,\ s(x) = y(x) + \varepsilon(x)$. In Signal Theory, various techniques are devoted to this objective. Estimators of y can be obtained through wavelet shrinkage **?**, **? ?**. The methodology consists in projecting the observations on a particular basis of functions and then distinguishing y from the noise with a threshold. Even if this technique is particularly innovative, it is still subject to the choice of several variables, such as the wavelet family, the threshold or the dimension resolution. Total variation (TV) techniques constitute an interesting alternatives to wavelets. These approaches consist in finding the function y that has the smallest TV over its domain and that is the "closest" to the true function y. In this setting, the function y is assumed to be of bounded variations (BV), and as enhanced by Chambolle (2004), BV spaces benefit from interesting properties such as the *lower semicontinuity* of the total variation or the possibility for BV functions to be approximated by *smooth functions*. Besides, derivatives of BV functions are defined in the sense of distributions, which allows for discontinuities in y. Thus, BV functions are only required to be integrable.

The Rudin-Osher-Fatemi model (denoted hereafter ROF model) is the first to consider this approach **?** and it has been applied with success to various additive denoising problems like in computer vision **?**, **?** or in finance to foreign exchange rate prediction **?**. Yet, the energy to minimize is composed of a data regularity term (the TV semi-norm) and of a data fidelity term (the MSE). Even if the MSE is well-suited for a Gaussian noise, like in some cases it approximately equals the noise variance, it is ill-suited when the noise is drawn from other probability distributions with features such as asymmetry or leptokurticity. Hence, an alternative is to replace the MSE by a criterion based on the noise probability distribution: as a result, the data fidelity term ensures to select y such as the noise is the most likely to follow a given probability distribution and a Bayesian interpretation through the *Maximum A Posteriori* (MAP) allows to reach this objective **?**. In the context of additive noise problems, the MAP has been extended to a Poisson noise **?** and to Laplace noise **?**, **?**. Recently, this method has been applied with success to Cauchy noise. **?**. To solve their functional, they use the *Primal Dual* algorithm. In addition, for multiplicative denoising, the MAP methodology has been adapted to a Beta noise, and recently to a Gaussian noise **?**, **?**.

Gaussianity is a traditional assumption for noise in Signal Theory and most of denoising procedures rely on it **?**, **?**. However, Gaussian assumption suffers from a lack of flexibility like it forces the noise to be symmetric and mesokurtik. It is well known that the distribution of the noise generally depends on the problem considered. Then a natural improvement consists in assessing that the noise distribution has more general characteristics. Notably, in this paper, we consider the symmetry and the leptokurticity properties of noise distributions. To our knowledge, noises with leptokurtic distributions have been implemented mainly in Kalman filtering literature **?**, **?**, **?**. In wavelet theory, a technique using a Bayesian threshold has been proposed **?**. Most of the mentioned papers consider Student t noise because it presents pleasant properties: well-defined moments and closed-form likelihood. In addition to the mesokurtik property, recent developments in asymmetric distributions, thanks to the founding works of **?**, **?**, **?**, **?** allow to consider asymmetric noises.

In this paper, we propose to apply the MAP approach to leptokurtic and asymmetric distributions, such as, the Student t distribution, the skewed Gaussian distribution and finally to the skewed Student t distribution. For each of these, Euler-Lagrange equations are derived with ease and numerical simulations are provided. Lastly, we propose a financial application of our model.

The paper is organized as follows. In Section 2, we present our methodology and the estimation technique. In Section 3, we provide some simulations. In Section 4, a financial application of our algorithm is done.

2 Model and Estimation

The problem consists in recovering a function y from noisy observations. Let $x \in \Omega$, an open subset of $\mathbb{R}$. However, y cannot be directly observed like it is perturbed by some additive noise, and: $s(x) = y(x) + \varepsilon(x)$ is observed instead. As a result, our problem is to recover y by setting assumptions on y as well as on the noise distribution. We precise that x are observed.

We consider the following model:

$$s(x) = y(x) + \varepsilon(x). \tag{5.1}$$

With, $s \in L^2(\mathbb{R})$, $\forall\ x$, $\{\varepsilon(x)\}$ are independent and identically distributed random variables satisfying $\mathbb{E}[\varepsilon] = 0$ and with finite first four moments. The function y belongs to space of Bounded Variations functions (BV).

Definition 14. *A function f is in $BV(\Omega)$, where Ω is an open subset of $\mathbb{R}$, if it is integrable, and there exists a Radon measure Df such that:*

$$< Df, \phi >= - \int_\Omega f(x)\phi'(x)dx, \ \textit{with}\ \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}).$$

When f is smooth enough the following relation between the distributional derivative and the derivative can be set:

$$Df(x) = \dot{f}(x).$$

The *total variation semi-norm* of f is:

$$||f||_{TV(\Omega)} = \int_\Omega |Df| := \sup \left\{ \int_\Omega f(x)\phi'(x)dx, \phi \in \mathcal{C}^\infty(\Omega, \mathbb{R}), \forall x \in \Omega, |\phi(x)| \leq 1 \right\}$$

When f is smooth, the TV semi-norm is equivalently the integral of its first derivative magnitude:

$$||f||_{TV(\Omega)} = \int_\Omega |\dot{f}(x)|dx.$$

Rudin, Osher and Fatemi propose to find an estimate to the denoising problem by solving the following equation:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \; ||y||_{TV(\Omega)} + \frac{\lambda}{2} \int_\Omega (s(x) - y(x))^2 \, dx.$$

The TV part of the objective functional is defined by the TV semi-norm and the data fidelity term by $\int_\Omega (s(x) - y(x))^2 dx$. The objective functional takes into account of any *a priori* on both ε and on y regularity. Notably, when λ is high, we look for a function y that primarly satisfies the closeness constraint. The presence of this penalty term is also traditional in other estimation fields. For instance, it is popular in spline regression **?**, local spline regression **?** and more generally in Tikhonov regularization **?**. The parameter λ can be fixed arbitrarly or it can be selected through an algorithm such as *Cross-Validation* **?**, (**?**). Other selection algorithm relies on neural networks **?**. Moreover, according to the *discrepancy principle*, when the noise is Gaussian, λ is chosen to match the noise variance **?**, **?**. However this principle is known to overestimate the mean squared error optimal choice of λ. Chambolle (2004) proposes an algorithm to find a value of λ so that $|ls - y||$ is approximatively the noise variance **?**. In this paper, we consider an automatic procedure selection for λ, based on *Stein Unbiased Risk Estimator*, presented below.

The mathematical study of the existence and uniqueness of a solution to Problem 5.2 is notably done in **?**.

Maximum A Posteriori

When the noise distribution has nice statistical properties such as symmetry or mesokurticity, a closeness measure based on L^2 distance appears relevant. However, when the noise distribution is more complex (explosive, asymmetric or leptokurtic), then this closeness measure is inappropriate[1]. A measure based on ε distribution may be more

[1] When the noise is locally explosive, a L1 norm can be used instead of the standard L2 norm **?**.

suitable, particularly if it takes account of the third and fourth moments of ε. Hence, a natural solution is to employ the MAP approach **?**, **?**.

The aim is therefore to select among a set of signals, the most probable signal under a posterior probability distribution, $P(Y|S)$. From Bayes's rule, we know that:

$$P(Y|S) = \frac{P(S|Y)P(Y)}{P(S)},$$

and as maximizing $P(Y|S)$ is equivalent to minimize $-P(Y|S)$, our problem consists in solving:

$$-\log\left(P(Y|S)\right) = -\log\left(P(S|Y)\right) - \log\left(P(Y)\right) + \log(P(S)).$$

As $\log(P(S))$ does not depend on Y, we focus on $-\log\left(P(S|Y)\right) - \log\left(P(Y)\right)$. The term $P(Y)$ contains any *a priori* we have on y behaviour, and as in a total variation context, $y \in BV(\Omega)$, the Gibbs's prior is selected:

$$P(Y) = \exp\left(-\mu \int_\Omega |\dot{y}(x)|dx\right),$$

where $\mu > 0$ stands for a regularization parameter.

A general form for MAP estimate is:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}}\ \mu \int_\Omega |\dot{y}(x)|dx - \log\left(P(S|Y)\right).$$

For instance, a MAP formulation of the ROF model is:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}}\ \mu|\dot{y}(x)|dx + \frac{1}{2\sigma^2}\int_\Omega \left(s(x) - y(x)\right)^2.$$

For additive noise problem, some adaptations of the MAP estimate to particular noises have already been proposed. Le, Chartrand and Asaki (2007) propose a MAP estimate for a signal disrupted by a Poisson noise **?**:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \int_\Omega |\dot{y}(x)|dx + \lambda \int_\Omega \left(y(x) - s(x)\right)\log y(x)dx.$$

Recently, Sciacchitano, Dong and Zeng (2015) adapt the MAP estimate for a signal disrupted by a Cauchy noise **?** and propose the following the functional to minimize:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \int_\Omega |\dot{y}(x)|dx + \frac{1}{2}\int_\Omega log\left(\gamma^2 + (y(x) - s(x))^2\right)dx.$$

Asymmetric and Leptokurtic distributions

In this paper, we propose to apply the MAP estimate to three probability distributions: the Student t distribution, the skewed Gaussian and the skewed Student t. While the Student t distribution has been widely studied, different approaches have been introduce to characterize skewed Gaussian and skewed Student t distributions.

If ε follows a Student t distribution of degree of freedom ϑ then its probability distribution $p_{\mathcal{E}}$ is:

$$\varepsilon \to p_{\mathcal{E}}(\varepsilon) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{\varepsilon^2}{k}\right)^{-\frac{k+1}{2}}$$

If ϑ is strictly greater than 1, then $\mathbb{E}[\varepsilon] = 0$. Moreover, when $\vartheta > 2$, the variance has the following formulation: $\mathbb{V}[\varepsilon] = \frac{\vartheta}{\vartheta-2}$. $\Gamma()$ is the Gamma function.

O'Hagan and Leonard (1976) are the first to propose a parametrization of the skewed Gaussian distribution which has been simplified by Mudholkar and Hutson (2000) **?**, **?**. The simplified distribution form of the distribution is the product of a standard $\mathcal{N}(0,1)$ density and a distorted $\mathcal{N}(0,1)$ cdf. If ε follows a skewed Gaussian distribution, with location parameter $\xi \in \mathbb{R}$, with scale parameter $\omega > 0$ and of distortion parameter $\alpha \in \mathbb{R}$, then its probability distribution $p_{\mathcal{E}}$ is:

$$\varepsilon \to p_{\mathcal{E}}(\varepsilon) = \frac{1}{\omega\pi} e^{-\frac{(\varepsilon-\xi)^2}{2\omega^2}} \int_{-\infty}^{\alpha\left(\frac{\varepsilon-\xi}{\omega}\right)} e^{-\frac{t^2}{2}} dt$$

Theodossiou (1998) developped the skewed generalized Student t distribution **?**. From this general expression can be derived various other skewed and leptokurtic distributions such as the skewed Gaussian, skewed Student t or skewed Laplace. Even though it is flexible, the great number of parameters make its estimation difficult and its extension to multivariate appears complicate. Azzalini and Capitanio (1999, 2003) built a skewed Student distribution based on an hidden truncation that is to combine an original Student distribution with a perturbated Student cdf **?**, **?**. This form has the advantage to be easily extented to the multivariate case. However, this forms suffers also from estimation drawbacks, like the fact that its likelihood function has a non-zero probability to be infinite. Gomez et al. (2007) simplifies the skewed Student distribution proposed by Azzalini and Capitanio (2003), by replacing the perturbated Student cdf by a perturbated $\mathcal{N}(0,1)$ cdf **?**. This simplification leads to a basic parameter estimation method through an Expected-Maximum algorithm. If ε follows a skewed Student t distribution with location parameter $\xi \in \mathbb{R}$, with scale parameter $\omega > 0$, degree of freedom $\vartheta > 2$ and of distortion parameter $\alpha \in \mathbb{R}$, then its probability distribution $p_{\mathcal{E}}$ is:

$$\varepsilon \to p_{\mathcal{E}}(\varepsilon) = 2\frac{\Gamma(\vartheta+1)}{\Gamma(\vartheta/2)\sqrt{2\pi}} \left((1 + \frac{(\varepsilon-\xi)/\omega^2}{\vartheta})\right)^{\frac{\vartheta+1}{2}} \int_{-\infty}^{\alpha\left(\frac{\varepsilon-\xi}{\omega}\right)} e^{-\frac{t^2}{2}} dt$$

In this work, we consider Gomez et al. formulation of the skewed Student distribution and the Mudholkar and Hutson formulation of the skewed Gaussian.

Probability Density Functions (PDF) of the Gaussian, skewed Gaussian, Student t and skewed Student t are presented in Figure 5.3. For skewed distribution, we set the location parameter to 0 and so, the condition $\mathbb{E}[\varepsilon] = 0$, for all ε is violated. Indeed, the distortion in the pdf, resulting from the asymmetry, introduces a bias. For example in the skewed Gaussian case, when the location parameter is null[2]:

$$\mathbb{E}[\varepsilon] = \omega \frac{\alpha}{\sqrt{1+\alpha^2}} \sqrt{\frac{2}{\pi}}.$$

We observe that in the particular case of the symmetry ($\alpha = 0$), the nullity of the expectation is recovered.

As a result, we come up with the three following applications of the MAP estimator.

The Student t:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \ \mu \int_\Omega |\dot{y}(x)| dx - \log\left(\frac{\Gamma(\vartheta+1)}{\Gamma(\vartheta/2)\sqrt{2\pi}}\right) + \int_\Omega \frac{\vartheta+1}{2} \log\left((1 + \frac{(s(x)-y(x))^2}{\vartheta})\right) dx.$$

The skewed Gaussian:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \ \mu \int_\Omega |\dot{y}(x)| dx + \log(\omega\pi) + \int_\Omega \frac{(s(x)-y(x))^2}{2\omega^2} - \log(\Phi(\alpha\left(\frac{s(x)-y(x)}{\omega}\right)))dx.$$

The skewed Student t distribution:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \ \mu \int_\Omega |\dot{y}(x)| dx - \log\left(2\frac{\Gamma(\vartheta+1)}{\Gamma(\vartheta/2)\sqrt{2\pi}}\right) + \int_\Omega \frac{\vartheta+1}{2} \log\left((1 + \frac{(s(x)-y(x))/\omega^2}{\vartheta})\right)$$
$$+\log(\Phi\left(\frac{\alpha}{\omega}(s(x)-y(x))\right) dx.$$

In the following, we set $\mu = 1$ and reintroduce a regularization parameter λ before the data regularity term.

For numerical resolutions, we regularize the L1 norm of the TV term in the above equations by $\sqrt{\beta^2 + |\dot{y}(x)|^2}$, for $\beta > 0$. This allows to avoid the *staircasing effect* due to the non-differentiability of the L1 norm in zero **?**, **?**.

[2] However, we can consider also asymmetric noises without setting $\xi = 0$. To have a noise distribution verifying $\mathbb{E}[\varepsilon] = 0$, it is tantamount to selecting the location parameter, in skewed distributions, such as $\mathbb{E}[\varepsilon] = 0$.

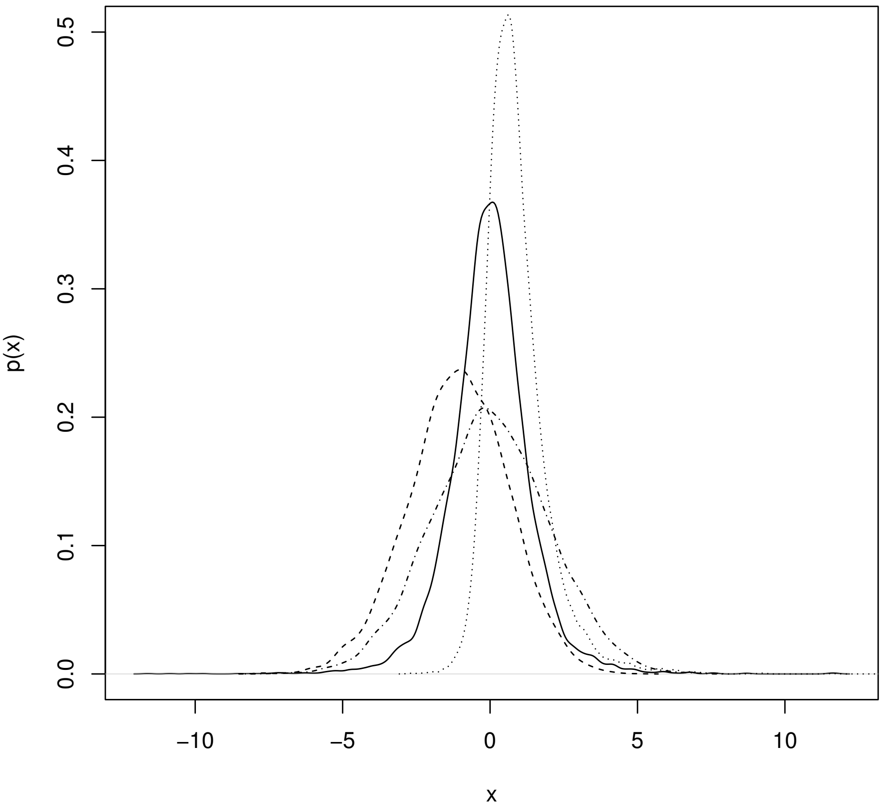


Figure 5.1: Comparison of the PDFs for Gaussian (dotted-dashed), skewed Gaussian (dashed), Student t (line), skewed Student t (dotted). For Gaussian type distributions: $\omega = 2$ and $\alpha = -0.9$. For the Student t type: $\vartheta = 4$, $\omega = 1$ and $\alpha = 2$.

The *Stein Unbiased Risk Estimator* (SURE) is known to provide an unbiased estimate of the theoretical risk, between the unknown signal y and one estimate $\hat{y}$ **?**. Aside from the fact that this method is useful to compare a variety of models, it is popular for automatic parameter selection. For instance, it has been apply to parameter selection in image denoising by Ramani, Blu and Unser (2008) and by Duval, Aujol and Gousseau (2010) **? ?**. Also, various extensions of SURE have been made such as Stein Unbiased Gradient Risk Estimate (SUGAR) or Generalized Stein Unbiased Gradient Risk Estimate (GSURE). In this part, we briefly recall the SURE estimate for the Gaussian noise and we propose an extension to the skewed Gaussian distribution.

Let $\hat{y}$ be an estimate of y, and we are interested in computing the following empirical error:

$$\mathbb{E}\left[(s(x)-\hat{y}(x))^2\right] = \mathbb{E}\left[((y(x)+\varepsilon(x))-\hat{y}(x))^2\right]$$

$$= \sigma^2 - 2\mathbb{E}\left[\varepsilon(x)\left(y(x)-\hat{y}(x)\right)\right] + \mathbb{E}\left[(y(x)-\hat{y}(x))^2\right].$$

Switching the empirical error ($\mathbb{E}\left[(s(x)-\hat{y}(x))^2\right]$) with the theoretical error ($\mathbb{E}\left[(y(x)-\hat{y}(x))^2\right]$), we come up with the following estimate:

$$\mathbb{E}\left[(y(x)-\hat{y}(x))^2\right] = -\sigma^2 + \mathbb{E}\left[(s(x)-\hat{y}(x))^2\right] + 2\mathbb{E}\left[\varepsilon(x)\left(y(x)-\hat{y}(x)\right)\right].$$

As the two first terms of the above equations are known, the challenge is therefore to estimate the quantity $2\mathbb{E}\left[\varepsilon(x)\left(y(x)-\hat{y}(x)\right)\right]$, which in the Gaussian case, comes easily thanks to the *Stein's Lemma*.

Lemma 1 (Stein's Lemma)**.** *If $X \sim \mathcal{N}(0,\sigma^2)$ and if g a weakly differentiable function, therefore:*

$$\mathbb{E}\left[Xg(X)\right] = \sigma^2\mathbb{E}\left[g'(X)\right].$$

The proof comes easily with an integration by part of the left-hand term of the above equation Several methods exists to implement the right-hand term: Monte-Carlo, differentiation or finite-difference. For a complete review, we refer the reader to **?** and to **?**.

An other challenge is to apply the Stein Unbiased Risk Estimate (SURE) methodology to the distributions we study here. In a recent article, Pal, Ching-Hui and Li (2008) adapts the Stein's lemma to the skew-normal case **?**.

An illustration of the use of SURE in the automatic parameter selection procedure is done in Figure 5.2.

Various techniques exist in the literature to find a solution to the above equations. We can enumerate *Dual Formation*, *Split Bregman Methods*, *Euler-Lagrange equations*. For a complete review we refer the reader to **?**, **?**. In this paper, we derive it through

Euler-Lagrange equations like they are applied to various TV problems and benefit from stable convergence properties and from an easy implementation **?**, **?**, **?**, **?**, **?**. Expressing the TV denoising into a variational calculus is tantamount to solving:

$$\underset{y \in BV(\Omega)}{\operatorname{argmin}} \; J[y] = \int_{\Omega} H_\theta \left(x, y\left(x\right), \dot{y}\left(x\right)\right) dx. \tag{5.2}$$

Where θ is the vector of parameters of a given probability distribution.

In other terms, denoising is similar to find the function y, solution of a dynamic problem. In our context: $H_\theta\left(x, y(x), \dot{y}(x)\right) = -\lambda \log P\left(s(x) - y(x)|\theta\right) + |\dot{y}(x)|$.

A minimizer of the functional J must satisfies the following *Euler-Lagrange* equation:

$$\frac{\partial}{\partial y} H_\theta\left(x, y(x), \dot{y}\left(x\right)\right) - \frac{d}{dx}\frac{\partial}{\partial \dot{y}} H_\theta\left(x, y(x), \dot{y}\left(x\right)\right) = 0, \tag{5.3}$$

with Neumann boundary conditions. Discussions and applications of Euler-Lagrange equation can be found in **?**, **?**, **?**. We precise that the term $\frac{d}{dx}\frac{\partial}{\partial \dot{y}} H_\theta\left(x, y(x), \dot{y}\left(x\right)\right)$ is identical for all noise distributions and it equals:

$$\frac{d}{dx}\frac{\partial}{\partial \dot{y}} H_\theta\left(x, y(x), \dot{y}\left(x\right)\right) = \frac{d}{dx}\left(\sqrt{\beta^2 + \dot{y}\left(x\right)^2}\right) = \frac{\beta^2 \frac{d^2}{dx^2} y(x)}{\left(\beta^2 + \dot{y}(x)\right)^{\frac{3}{2}}}.$$

Forms of $\frac{\partial}{\partial y} H_\theta$ for the four studied noise distributions are available in Table 5.2.

However, Total Variation methods have two notable drawbacks: the *staircasing effet* and the *selection of the tuning parameter* λ.

General Algorithm

The solution to Equation 5.3 is approached using a simple descent gradient algorithm, which is well-suited to our problem and the following discretized version of Equation 5.3 is implemented **?**, **?**, (**?**), **?**.

$$\begin{aligned} \tfrac{d}{dj} y &= \tfrac{\beta^2 \frac{d^2}{dx^2} y_j(x)}{\left(\beta^2 + \dot{y}_j(x)\right)^{\frac{3}{2}}} - l(s(x) - y_j(x)) \\ y_{j+1} &= y_j + \delta\left(\tfrac{\beta^2 \frac{d^2}{dx^2} y_j(x)}{\left(\beta^2 + \dot{y}_j(x)\right)^{\frac{3}{2}}} - l(s(x) - y_j(x))\right) \end{aligned} \tag{5.4}$$

Where $l(s(x) - y_j(x)) = \frac{\partial}{\partial y_j} H_\theta\left(x, y_j(x), \dot{y}_j\left(x\right)\right)$. The parameter $\delta > 0$ controls for the speed of convergence.

The solution of Equation 5.1 is approached numerically by the sequence of estimates $\{y_j\}_{j<J}$. The descent gradient algorithm is summarized in Table 5.1. An important

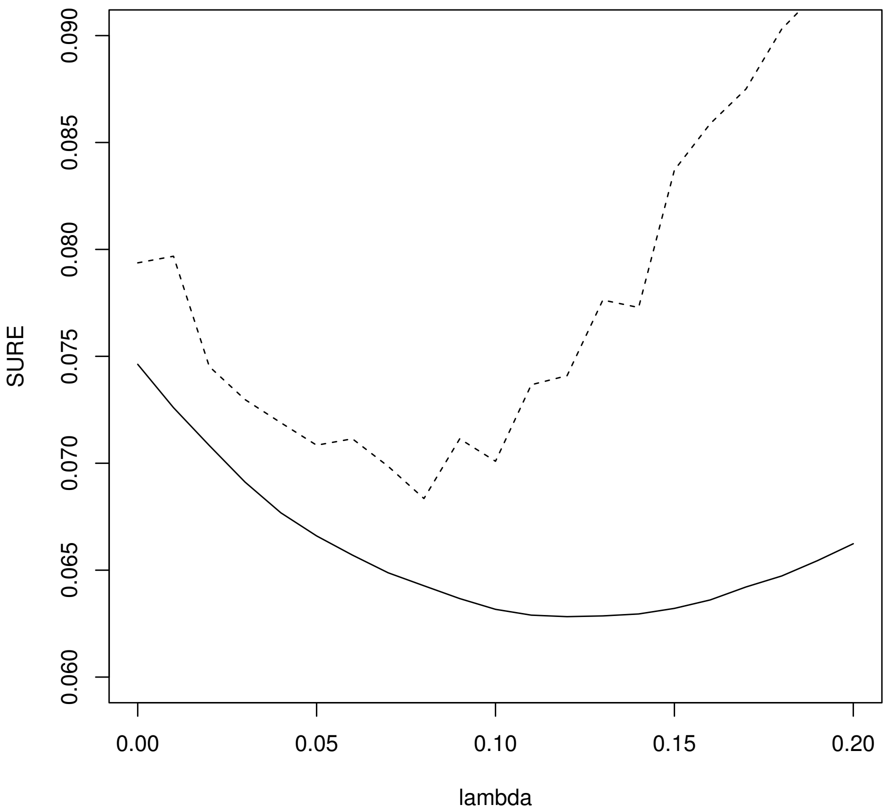


Figure 5.2: Estimation of SURE using a finite-difference approach. The dotted line corresponds to the SURE while the straight line is the true MSE.

topic is about the selection of an appropriate stopping criterion for the algorithm. In this work, we propose to use the quantity $J[y]$ and when this quantity is stable enough, the numerical resolution is stopped.

Initialize
for $n = 1, ..., N$, **do:**
$y_0(x_n) = s(x_n)$
end for n

Solve
for $j = 1, ..., J$, **do:**
for $n = 1, ..., N$, **do:**
$\frac{\beta^2 \frac{d^2}{dx_n^2} y_j(x_n)}{(\beta^2 + \dot{y}_j(x_n))^{\frac{3}{2}}} = \frac{\beta^2 (y_j(x_{n+1}) - y_j(x_n))}{(\beta^2 + (y_j(x_{n+1}) - y_j(x_n))^2)^{\frac{3}{2}}} + \frac{\beta^2 (y_j(x_n) - y_j(x_{n-1}))}{(\beta^2 + (y_j(x_n) - y_j(x_{n-1}))^2)^{\frac{3}{2}}}$
$y_{j+1}(x_n) = y_j(x_n) + \delta \left(l\left(s(x_n) - y_j(x_n)\right) + \frac{\beta^2 \frac{d^2}{dx_n^2} y_j(x_n)}{(\beta^2 + \dot{y}_j(x_n))^{\frac{3}{2}}} \right)$
end for n
end for j

Table 5.1: Algorithm to solve Equation 5.3. $l(s(x_n) - y_j(x_n)) = \frac{\partial}{\partial y_j} H_\theta\left(x, y_j\left(x_n\right), \dot{y}_j\left(x_n\right)\right)$ and $\lambda > 0$, $\delta > 0$, $\beta > 0$. J is the number of iterations and N the length of the signal.

3 Simulations

In this subsection, we test the accuracy of our methodology against ROF and Wavelet approaches. Wavelet theory consists in projecting a signal y on an orthonormal basis **?**. Henceforth Equation 5.1 is decomposed as follows:

$$\sum_j \sum_k \langle s(x), \psi_{j,k} \rangle = \sum_j \sum_k \langle y(x), \psi_{j,k} \rangle + \sum_j \sum_k \langle \varepsilon, \psi_{j,k} \rangle,$$

where $\psi_{j,k}$ stands for the mother wavelet at resolution $j \in \mathbb{Z}$ and at location $k \in \mathbb{Z}$. Wavelet regression consists in recovering y by removing the noise influence in the wavelet coefficient $\langle s(x), \psi_{j,k} \rangle$. This can be done with a filtering function F and an estimate $\hat{y}$ of y can be obtained:

$$\sum_j \sum_k \langle \hat{y}(x), \psi_{j,k} \rangle = \sum_j \sum_k F\left(\langle s(x), \psi_{j,k} \rangle\right)$$

Distribution	$\frac{\partial}{\partial y} H_\theta\left(x, y(x), \dot{y}\left(x\right)\right)$
Student	$\left(\frac{(\vartheta+1)}{\vartheta}\frac{(y(x)-s(x))}{1+\frac{1}{\vartheta}(s(x)-y(x))^2}\right)$
Skewed Gaussian	$\frac{(y(x)-s(x))}{\omega^2}+\frac{\alpha\phi\left(\alpha\left(\frac{s(x)-y(x)}{\omega}\right)\right)}{\omega\Phi\left(\alpha\left(\frac{s(x)-y(x)}{\omega}\right)\right)}$
Skewed Student	$\left(\frac{(\vartheta+1)}{\vartheta}\frac{(y(x)-s(x))}{1+\frac{1}{\vartheta}(s(x)-y(x))^2}\right)\ \frac{(y(x)-s(x))}{\omega^2}+\frac{\alpha\phi\left(\alpha\left(\frac{s(x)-y(x)}{\omega}\right)\right)}{\omega\Phi\left(\alpha\left(\frac{s(x)-y(x)}{\omega}\right)\right)}$

Table 5.2: Values of $\frac{\partial}{\partial y}H_\theta$. Where $\Gamma()$ is the Gamma function. $C_3 = log(\omega\pi)$, ϕ and Φ the density function and the cdf of a $\mathcal{N}(0,1)$ random variable.

Various filter functions exist : *WaveShrink*, *VisuShrink*[3].

We study the accuracy of our method on the HeavySine function that is commonly used to compare denoising methods **?**, **?**, **?**, **?**. This function is a sinusoidal with two jumps and is defined as follows.

$$y(x) = 4\sin(4\pi x) + \operatorname{sign}(x-0.3) - \operatorname{sign}(0.72-x). \tag{5.5}$$

However, we do not observe y directly, but a noisy version:

$$s(x) = y(x) + \varepsilon(x). \tag{5.6}$$

As a result, we distinguish the three following cases:

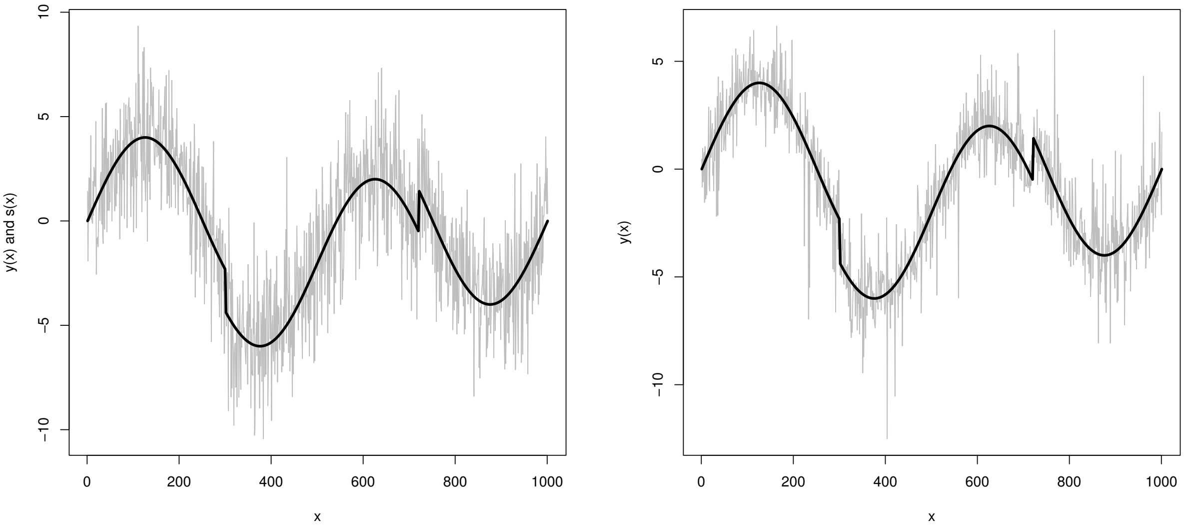
- ▷ ε follows a Student t distribution with $\vartheta = 5$
- ▷ ε follows a skewed Gaussian distribution with $\omega = 2$ and $\alpha = -2$
- ▷ ε follows a skewed Student t distribution with $\vartheta = 5$, $\omega = 2$ and $\alpha = 2$

Figure 5.3 represents the true function, $y(x)$ disrupted by a noise from the three different families. As expected, Student and skewed Student noises have spikes representing distributions leptokurticity, and skewed Gaussian and skewed Student are translated from the true signal.

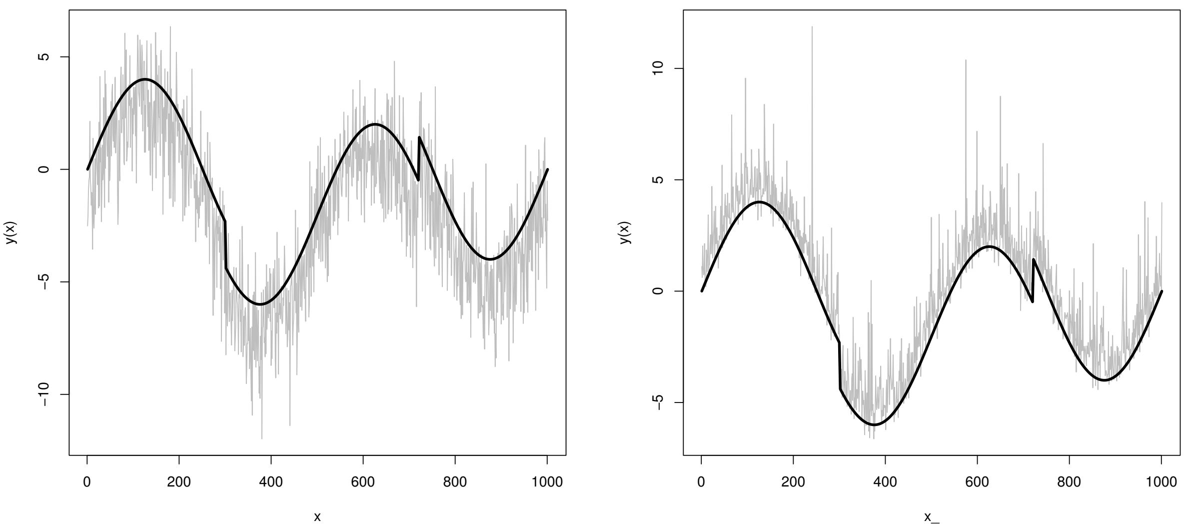
θ estimation

An ambitious task of parametric distribution is the estimation of parameters and in

[3]For a complete review, we refer the reader to **?**, **?**, **?**, **?**.



(a) Gaussian noise and Student noise



(b) Skewed Gaussian noise and skewed Student noise

Figure 5.3: Representations of $s(x)$ and $y(x)$ for different noise distributions. In black $y(x)$ and grey $s(x)$.

this paper, we propose a two steps estimation procedure described in Table 5.3. The first step consists in obtaining an estimate of the noise, thanks to a free-noise smoother, such as ROF. As soon as a first estimate of the noise has been obtained: ε_0, we estimate θ on it. We precise that this procedure can be repeated until $\hat{\theta}$ is stable enough.

To estimate Skewed Gaussian and Skewed Student parameters, we implement the method developed by Azzalini (2013) **?**. It consists in combining a Maximum Likelihood Approach with Dynamic programming [4].

1: **First Smoothing**
2: Get y_0 and compute $\varepsilon_0 = s - y_0$
3: Estimate θ on ε_0

Table 5.3: Algorithm to estimate θ, the vector of parameters of the noise probability density function. The procedure can be repeated until the estimate $\hat{\theta}$ is stable enough.

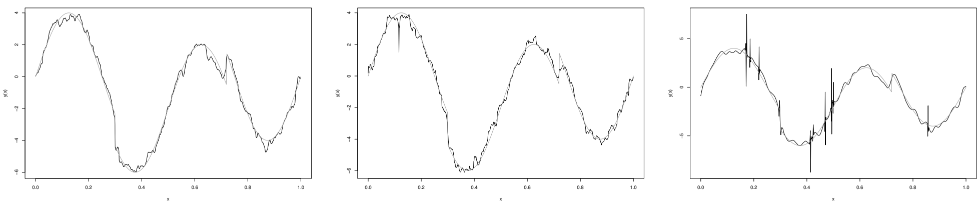
Results

For each noise distribution, we compare the associated MAP methods with a wavelet smoother and the ROF model[5]. For sake of fairness, all variational models (MAP and ROF) have the same parameters : $J = 600, \lambda = 0.2, \beta = 0.01$. To compare the smoothing accuracy we use two common measures: Mean Squared Error (MSE) and Mean Absolute Deviation (MAD). The first one has the particularity to weight more large errors. y estimators are depicted in Figure 5.4. In Table 5.4, we observe that, whatever the noise distribution, MAP methods provide always an estimator closest to the original signal. Moreover, the error difference between methods increases with the noise distribution complexity.

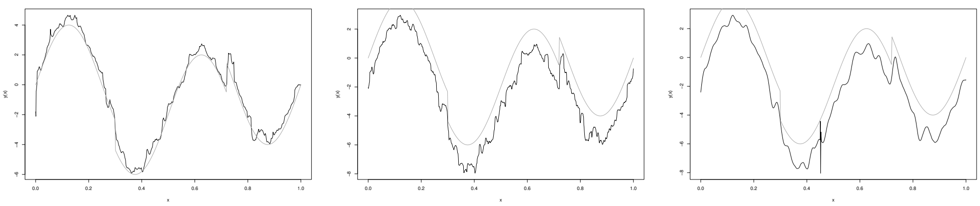
Lastly, we compare distributions of Signal to Noise ratios, estimated on the 1000 simulations. They are available in Figure 5.5. Whatever the underlying noise process, the SNR pdf of MAP approaches is always unbiased, whereas for wavelets and ROF, SNR pdf are biased.

[4] The method is available with *selm* function in the R package *sn*

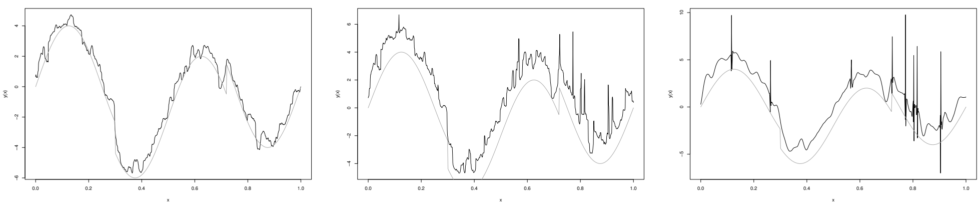
[5] We use the Daubechies wavelet family with four null moments. The resolution dimension is four. Denoising signal is obtained thanks to a universal threshold and an hard filter.



(a) Student noise



(b) Skewed Gaussian noise

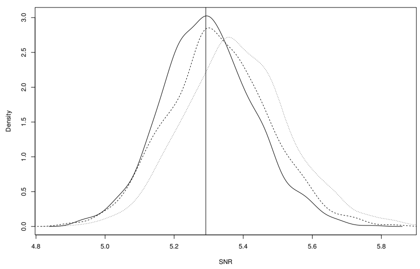


(c) Skewed Gaussian noise

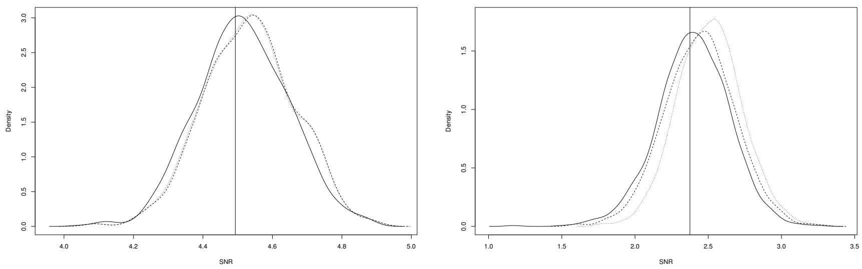
Figure 5.4: Representations of $\hat{y}$ estimates against y for different noise distributions. For each noise, the left hand image is the MAP estimator, image at the center is the ROF estimator, and the right hand image is the wavelet estimator. The black line is the estimator while the grey line is the original signal. For TV models, parameters are identical.

Model	Wavelet Regression	ROF Model	MAP
Student Noise			
MAD	0.26 0.02	0.21 0.02	0.18 0.01
MSE	0.18 0.08	0.08 0.02	0.05 0.01
Skewed Gaussian Noise			
MAD	1.43 0.04	1.42 0.04	0.32 0.13
MSE	2.16 0.13	2.12 0.13	0.40 0.10
Skewed Student Noise			
MAD	1.70 0.06	1.69 0.06	0.57 0.05
MSE	3.45 0.5	3.45 0.36	0.47 0.08

Table 5.4: Mean Absolute Errors and Mean Squared Errors and associated standard deviations computed on one thousand simulations.



(a) SNR for Student noise



(b) SNR for skewed Gaussian noise and skewed Student noise

Figure 5.5: True SNR (black vertical line) versus SNR empirical probability distribution functions, computed on the 1000 simulations. Solid line is the SNR empirical pdf for MAP model, dashed line is the SNR empirical pdf for the ROF model, dotted line is the SNR empirical pdf for the wavelet model.

4 Application to Technical Analysis

Technical Analysis (TA) is an investment strategy for which the decision is based on the analysis of technical variables, such as price dynamic, trading volume or volatility. Unlike quantitative investment that mainly depends on algebra, TA is based on geometry. It is consequently often considered as a "visual" approach. According to Pring (2002), "the art of Technical Analysis, for it is an art, is to identify a trend reversal at a relatively early stage, and ride on that trend until the weight of evidences shows or proves that the trend has reversed" **?**. TA is a key method in Finance. Mehnkhoff (2010) documented the use of TA in the hedge fund industry **?**. He shows that TA is a valuable source of information and has therefore become more popular than Fundamental Analysis in the decision making process of investments at weeks horizon[6].

The range of TA strategies is large. Originally, TA is implemented on financial variables such as stock prices, trading volumes or volatility. In this application, we only consider TA of stock prices, a subpart of pattern recognition methods[7].

A developing literature is devoted to pattern recognition of stock prices dynamic. Lo et al. (2000) test a TA investment strategy based on 10 well known patterns and find promising results **?**. Dawson and Steeley (2003) test the same strategy in the UK market and show that it is profitable **?**. Zapranis and Tsinaslanidis (2012) who implement the exact strategy in the US market propose a pattern recognition procedure independent of subjective criteria, such as the width of the time window **?**. El-Ansary and Mohssen (2017) look for the presence of traditional patterns in the Egyptian stock market. They show that this market is highly predictable **?**. Finally, Bulkowski (2005, 2011) lists 47 standard stock patterns **?**. Zhu and Zhou (2009) documented the founding principle of TA **?**. They suggest that TA brings considerable value in the investment process when stock returns are predictable. Besides, when uncertainty is high, they advocate the use of a mixed approach combining TA and traditional asset allocation methods.

Most of the mentioned articles implement pattern recognition on filtered prices. Observed prices are composed of finer movements (often referred as noise) that make direct implementation of pattern recognition algorithm ineffective. To avoid loss of crucial

[6]The predictive capacity of TA is especially powerful when financial prices are under the influence of non-fundamental behaviours.

[7]The term *patterns* refers to any repeated dynamics in the stock price behaviours. As soon as a pattern is identified, the underlying stock price dynamic can easily be forecasted

information, the following problem is considered:

$$s(t) = y(t) + \varepsilon(t), \tag{5.7}$$

where s is the observed stock price, y is the underlying and unobserved price trend, and ε is a noise process. To estimate y, Lo et al. (2000), Dawson and Steeley (2003), Arevalo et al. (2017) either use a non-parametric regression with a Gaussian kernel, or an exponential moving average **?**, **?**, **?**. In this application, we propose an estimate of y using the MAP techniques introduced above, which has two advantages. First, by considering that the underlying price is L^1, we allow it to have singularities such as local sharp variations. Unlike a kernel filtering that will smooth singularities, and so will delete relevant information, MAP approaches will preserve them. Second, we assume that the noise probability distribution can be asymmetric and leptokurtic. This permits to the noise to be distributed from more complex distributions such as skewed Gaussian or skewed Student. As a result we believe than estimates of y obtained through MAP approaches will keep relevant information on the non-linear trend of the underlying price. In the next paragraphs, we first define the considered patterns and the data. We provide the identification algorithm. Finally, we display the results of our strategy.

Patterns

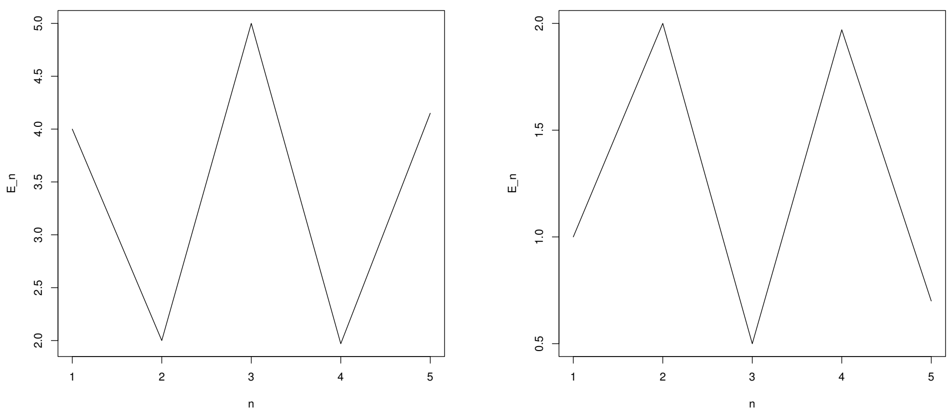
Following Lo et al. (2000) and Dawson and Steeley (2003), we are interested in finding the following patterns in the denoised dynamics of stock prices (y). These patterns directly depend on local extrema of the denoised prices. All of them deal with a combination of successive local minima and local maxima. Each of the three patterns introduced has its inverse. For example, the *Head and Shoulders* suggests to short the stock at the end of the pattern, while the *Inverse Head and Shoulders* suggests to long the stock at the end of the pattern.

Let $\{E_1, ...E_n\}$ be a collection of local extrema collected on $\{y(t), y(t+h)\}$, for $h > 0$.

Definition 15 (Head and Shoulders)**.** *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is an head and shoulders if and only if:*

$$\begin{cases} E_1 \text{ is a maximum} \\ E_1 < E_3 \text{ and } E_5 < E_3 \\ \text{E1 and E5 are within 1.5 percent of their average} \\ \text{E2 and E4 are within 1.5 percent of their average.} \end{cases}$$

Definition 16 (Inverse Head and Shoulders)**.** *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is an inverse head and shoulders if and only if:*



(a) Head and Shoulders (b) Inverse Head and Shoulders

Figure 5.6: Head and Shoulders patterns

$$\begin{cases} E_1 \text{ is a minimum} \\ E_1 > E_3 \text{ and } E_5 > E_3 \\ \text{E1 and E5 are within 1.5 percent of their average} \\ \text{E2 and E4 are within 1.5 percent of their average.} \end{cases}$$

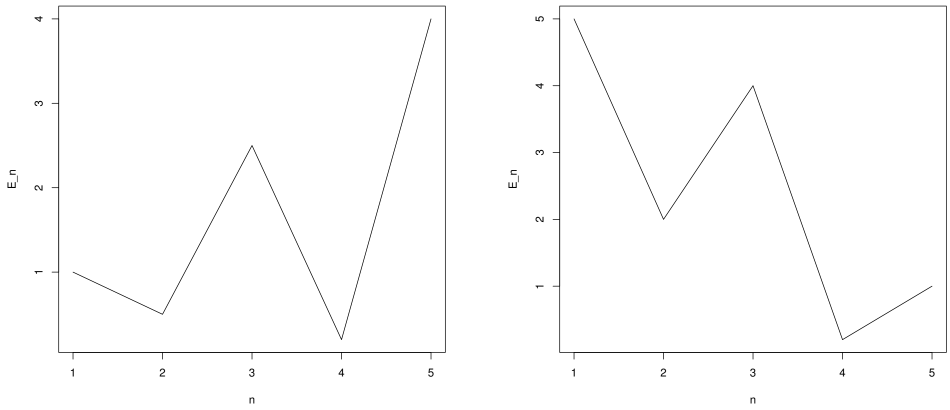
Definition 17 (Broadening). *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is a broadening if and only if:*

$$\begin{cases} E_1 \text{ is a maximum} \\ E_1 < E_3 < E_5 \\ E_2 > E_4. \end{cases}$$

Definition 18 (Inverse Broadening). *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is an inverse broadening if and only if:*

$$\begin{cases} E_1 \text{ is a minimum} \\ E_1 > E_3 > E_5 \\ E_2 > E_4. \end{cases}$$

Definition 19 (Triangle). *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is a triangle if and only if:*



(a) Broadening (b) Inverse Broadening

Figure 5.7: Broadening patterns

$$\begin{cases} E_1 \textit{ is a maximum} \\ E_1 > E_3 > E_5 \\ E_2 < E_4. \end{cases}$$

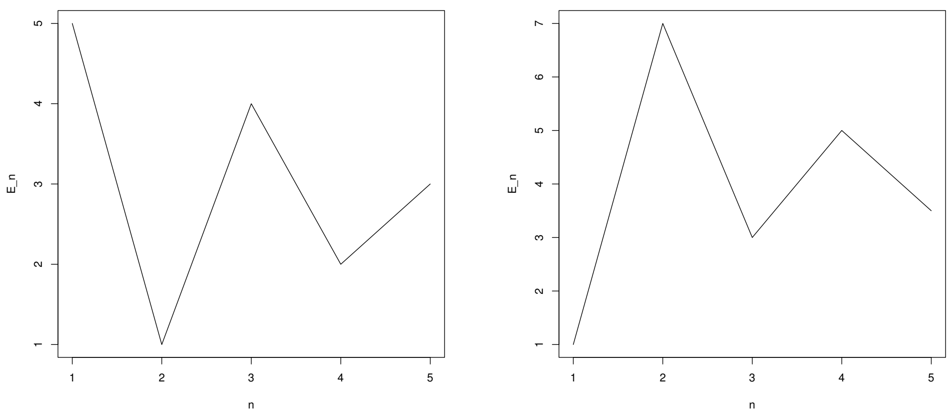
Definition 20 (Inverse Triangle)**.** *The collection of five consecutive extrema:* $\{E_1, ...E_5\}$ *is an inverse triangle if and only if:*

$$\begin{cases} E_1 \textit{ is a minimum} \\ E_1 < E_3 < E_5 \\ E_2 > E_4. \end{cases}$$

Data

We implement the pattern recognition on American Airlines (AA) and on the Kroger company (KG), between February, 29th 2016 and September, 9th 2016, that is on 135 trading days.

We consider 1 minute stock prices for two reasons. First, high frequency prices are disrupted by the so-called *micro-structure* noise, that moves the price away from its fundamental value. The origin of micro-structure noise is well-documented and it is generated by three mains phenomena: the bid-ask bounce, the discreteness of price change and the trading latency **?**, **?**. All these reasons legitimate the need to de-noise the observed prices. Second, pattern recognition strategies directly depends on the memory of the studied price process. The simple estimation of an Hurst exponent on



(a) Triangle (b) Inverse Triangle

Figure 5.8: Triangle patterns

daily stock prices and on 1 minute stock prices confirms the idea that intra-day data are much uncertain than daily data [8]. We consider 1 minute closing prices extracted from Bloomberg[9].

Identification Algorithm

The purpose of the application is to identify patterns, belonging to the library of patterns we introduced above, within a pattern detection window. As we are considering intra-day prices, we consider only intra-day patterns. Suppose we have the following sliding window during day d:

$$\{s_d(t), ...s_d(t+h)\},$$

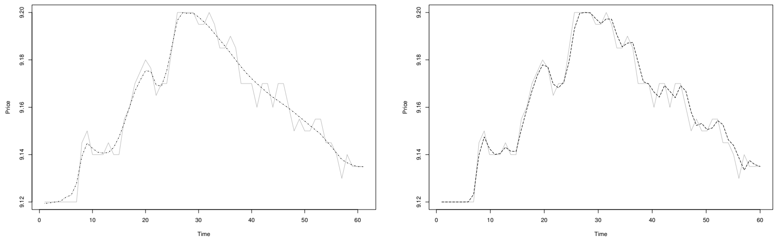
where h is the window length [10]. We recall that s are closing prices observed at a minute frequency. The first step of the algorithm consists in estimating the unobserved non-linear trend y on this sliding time window [11]. We get the following sequence of

[8] For AA, estimations of the Hurst exponent through the R/S method gives: 0.74 for daily data and 0.82 for 1 minute data while for KG, it returns 0.77 for daily data and 0.84 for 1 minute data.

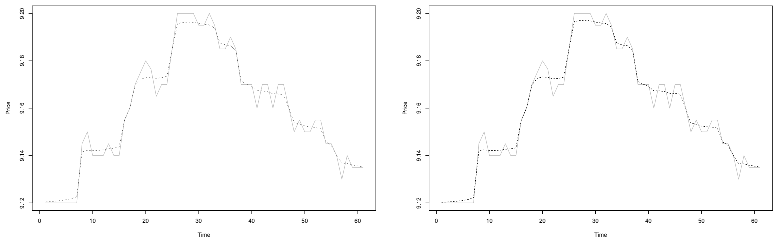
[9] Closing price are the last observed price in a 1min interval.

[10] Lo et al. (2000) and Andrew and Steeley (2003) both use h=38.

[11] When noise removing techniques are parametric, parameters are estimated on the previous trading day. To prevent the estimate y from border effects, we extend the signal to the left and to the right, by respectively the five first and five last observations.



(a) Wavelets and Kernel



(b) MAP skewed Gaussian and MAP skewed Student noise

Figure 5.9: American Airlines, 15th, March 2016 between 12h56 and 13h56. Comparison of the original stock price value (grey) with the four denoising methods.

estimates:

$$\{y_d(t), ...y_d(t+h)\}\,.$$

The next step consists in computing numerically local extrema in this sample. We look for times $\tau \in \{t, ..., t+h\}$, where the following relation holds:

$$\operatorname{sign}(\hat{y}'(\tau)) = -\operatorname{sign}(\hat{y}'(\tau+1)).$$

If $\operatorname{sign}(\hat{y}'(\tau)) = 1$ and $\operatorname{sign}(\hat{y}'(\tau+1)) = -1$, then $y(\tau)$ is a local maximum while if $\operatorname{sign}(\hat{y}'(\tau)) = -1$ and $\operatorname{sign}(\hat{y}'(\tau)+1) = 1$, then $y(\tau)$ is a local minimum.

We get the following sequence of local extrema: $\{E_1, ...E_n\}$, where $n < h$. We suppose that the last observed price $y(t+h)$ is a local extrema, and we are thus interested in recovering a known pattern on the five last local extrema:

$$\{E_{n-4}, ...E_n, y(t+h)\}\,.$$

If a known pattern is identified, we buy (short) a stock price in $t+h$, and we hold the stock for δ minutes, where $\delta = (t+h) - n$[12]. For instance, if there is five minutes between the realization of the extrema E_n and the last observed price $y(t+h)$, then the buy (short) position is hold for five minutes. We are aware that other applications of TA pattern recognition, suggest to hold a position until a target price has been reached[13]. However, the target price does carry any information of the frequency of oscillation within the pattern, which we think contents a great deal of informations.

When the holding time is up, the following profit is made:

- ▷ For a long position:

$$r_{t+h} = \log \frac{s(t+h+\delta)}{s(t+h)}.$$

- ▷ For a short position:

$$r_{t+h} = \log \frac{s(t+h)}{s(t+h+\delta)}.$$

If a pattern has been detected at $t+h$, then nothing is done between $t+h$ and $t+h+\delta$, and so the rolling window jumps to:

$$\{y_d(t+\delta), ...y_d(t+h+\delta)\}.$$

If no patterns has been detected at $t+h$, then t increments by one.

Results

For American Airlines and Kroger Cie, we build estimates of y with the following four methods:

- ▷ Gaussian Kernel smoothing [14],
- ▷ Gaussian Wavelet smoothing,
- ▷ Skew Gaussian MAP estimate,

[12]All the six patterns considered are described in Figure 5.6, 5.7, 5.8

[13]Generally this target price is defined as the last observed extremum plus/minus the highest distance between the extrema in the pattern.

[14]For Kernel smoothing, the regularization parameter is selected through Cross Validation.

- ▷ Skew Student MAP estimate.

Figure 5.9 provides an illustration of the behaviour of the four different estimates for a slicing window computed on American Airlines. We observe that MAP oriented methods best match the objective that is to smooth the noisy prices while preserving local singularities. Indeed, the wavelet estimate (top left) fails in estimating the two plateaus in the right part of the graph. On the contrary, Kernel smoothing, keep to much information. Only MAP approaches succeed in deleting local oscillation while preserving their information content.

We recall that the strategy directly depends on the number of patterns detected within a given day, and so, if within a given day, ten patterns are identified, then we have only ten returns for the strategy. Consequently, the intra-day PNL will be too discontinuous and so we consider daily returns, R_d:

$$d \in \{1, ..., D\}\,, \quad R_d = \sum_{n=1}^{N(d)} r_n.$$

$N(d)$ is the number of patterns detected during day d, and D in the number of days on which the strategy is implemented. We recall that in our application, $D = 135$.

When implementing investment strategies, a challenge consists in selecting performance measures taking into account both the excess returns, with respect to a given benchmark, as well as the downside risk of the strategy. We select the following measures.

- ▷ The *Average return* permits to have an estimate of the daily realised return, expressed in percentage. Statistic significance is provided through a *t test*: *, **, *** to significant at 90%, 95%, 99%. In Table 5.5, we observe that for two data sets, only MAP oriented models provide daily returns significantly different from 0, at the 99% confidence level. In addition, strategy daily returns for MAP approaches are always greater than others. For American Airlines, the average daily return for skew Gaussian MAP is 0.247%, while it is 0.231% for skew Student. For other methods, these returns are lower: 0.17% for Wavelets and 0.03% for Kernel smoothing. For the Kroger companies, differences among daily strategy returns is less important.
- ▷ The *Total return* gives the total return generated by the strategy on the 135 days. It is measured in %. In Table 5.5, we observe that on the two data sets, MAP oriented methods always have the highest total return. Moreover, Skew Gaussian seems to slightly outperform Skew Student in term of total return.
- ▷ The *Sharpe ratio* is a performance measure taking account of the risk carried by the strategy. The performance is defined as expected excess return while risk is

defined by the standard deviation of strategy returns. Sharpe ratio is defined as follows.

$$Sharpe = \frac{\mathbb{E}\left[R - R_{rf}\right]}{\sigma_R}.$$

The risk free rate is the LIBOR overnight rate of February,29th 2016 and worth 0.3710%[15]. When computing the Sharpe ratio for all models, we observe that Sharpe ratio computed on strategy returns generated by MAP approaches in approximatively two times bigger than the Sharpe ratio computed on strategy returns generated by a Wavelet approach. However, all Sharpe are below 1 and so, the risk of the strategy is always greater than the excess return.

▷ The *Sortino ratio* is a modified Sharpe ratio that penalizes downside risk. The downside risk is defined arbitrarily by returns that are below the *Minimum Accepted Return* (MAR). Hereafter, we select the risk free rate as the MAR.

$$Sortino = \frac{\mathbb{E}\left[R - R_{rf}\right]}{\sigma_{R|R<MAR}}.$$

By considering only downside risk, Sortino ratios are greater than Sharpe ratio. This suggests that negative strategy returns have a lower standard-deviation than all strategy returns. For American Airlines and for Wavelets and Skew Gaussian MAP, the Sortino ratio is twice larger than the Sharpe ratio. In terms of Sortino, a pattern recognition strategy set up on a stock price denoised through a Skew Gaussian MAP is the less risky.

▷ The *Omega ratio* is an alternative to Sharpe ratio, that take considerations of the whole distribution of strategy returns. Omega ratio compare an area of gain to an area of loss. Again, gains are strategy returns above the MAR.

$$Omega = \frac{\int_{MAR}^{\infty} (1 - F(x))\, dx}{\int_{-\infty}^{MAR} F(x) dx}.$$

Both for American Airlines and for the Kroger Cie, all models generate area of gains larger than area of losses as Omega ratios are always above 1. For American Airlines, the best models, according to this criterion are Gaussian Skew MAP and Student Skew MAP, with Omega ratios of respectively 2.05 and 1.82. However, for the Kroger Cie, the Kernel approach generates the higher Omega ratio (1.81).

[15] All the indicators are computed on a daily basis, and so we express the risk free rate in a daily basis, by dividing it by $\sqrt{252}$.

▷ The *Upside potential ratio* is a measure of performance relative to a minimum acceptable ratio. Again, in this measure, the risk is defined as standard deviation of strategy returns that are below the MAR. Contrary to the Sortino ratio, the numerator is only composed of upside deviations with respect to the MAR.

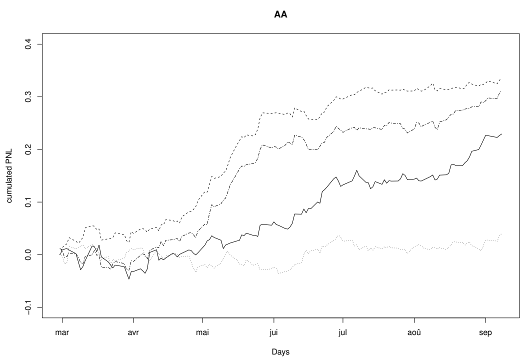
$$UPR = \frac{\mathbb{E}[(R - MAR)_+]}{\sigma_{R|R<MAR}},$$

where $()_+ : \mathbb{R} \to \mathbb{R}^+, x \to (x)_+ = \mathbf{1}_{x>0}x$.

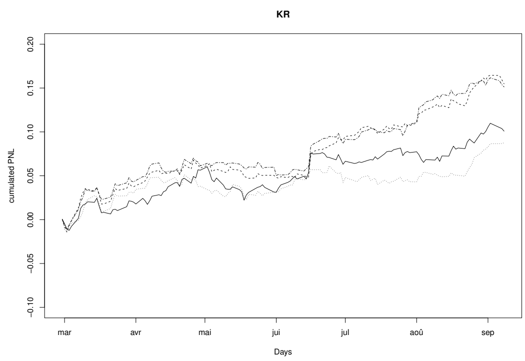
On all data sets and on all models, only the Skew Gaussian approach, for American Airlines, generate a UPR greater than 1 (1.12). On the same data set, Skew Student MAP has an UPR near 1 (0.98) and which is above Kernel and Wavelet UPR (0.59) and (0.78). Again on the Kroger Cie data set, UPR for MAP models are the highest and near one. Skew Gaussian MAP is also the less risky in terms of UPR (0.90). It is followed by the Skew Student MAP (0.88), by Wavelets (0.73) and lastly by the Kernel (0.63).

However, these results are directly linked to the behaviour of the pattern recognition algorithm, implemented on the denoised stock prices. Table 5.6 provides two highly informative measure: the daily average number of patterns detected by each technique ($\bar{N}(d)$), as well as the average time duration between the two last extrema ($\bar{\delta}$). Notably, they provide information on the ability of a denoising method to over fit the stock price. For instance, if the number of patterns detected is larger and if the duration between the two last extrema is weak, then it is a clue of over fitting. Again, in this application, a relevant denoising technique should eliminating smallest details, that hide crucial information, but it should preserve local tendency to be able to identify patterns. On the contrary, we have to identify techniques that will smooth to much the stock price and so, will delete crucial information. These techniques will be characterized by a low number of detected patterns, with an high duration.

Among all the techniques, the Kernel approach suffers from over fitting like for the American Airlines stock, it detects on average 25 patterns each day, with an average duration of 4 minutes. On the Kroger Cie data set, it detect on average 24 patterns each day, with an average duration of 3 minutes. On the contrary, the Wavelet approach seems to smooth to much the stock prices like it identifies on average between 15 and 17 patterns each day on American Airlines and on the Kroger Cie. Moreover, among all models, its duration is always the highest with respectively 5.80 minutes and 5.5 minutes on American Airlines and on the Kroger Cie. MAP approaches always carry variables between these two extrema, which encourage the idea that the estimates of non linear trends obtained through a Total Variation approach, succeed the best the dilemma between removing the noise and keeping the crucial information.



(a) American Airlines



(b) The Kroger Company

Figure 5.10: Cumulated strategy daily returns for the four implemented denoising models. Thick line, dotted line, dashed line and dotted-dashed line are respectively cumulated PNL for Wavelets, Kernel smoothing, MAP skew Gaussian, MAP skew Student.

Lastly, we believe that this strategy could be improved by adding transaction costs carried by the bid-ask spread. For instance, a more conservative approach would be to buy at the ask price, and to sell at the bid price[16].

Stock	Model	$\bar{R}$	TR	Sharpe	Sortino	Omega	UPR
AA	Kernel	0.03	4.07	0.01	0.02	1.06	0.59
	Wavelets	0.17 **	22.94	0.17	0.29	1.61	0.78
	MAP Skew Gaussian	0.247 ***	33.36	0.32	0.59	2.05	1.12
	MAP Skew Student	0.231 ***	31.15	0.28	0.48	1.82	0.98
KR	Kernel	0.065 **	8.76	0.10	0.15	1.82	0.63
	Wavelets	0.075 **	10.08	0.11	0.21	1.27	0.73
	MAP Skew Gaussian	0.115 ***	15.47	0.20	0.40	1.19	0.90
	MAP Skew Student	0.112 ***	15.12	0.20	0.38	1.61	0.88

Table 5.5: Performance Analytics. $\bar{R}$ is the average daily return of the strategy (in %), TR, is the total return of the strategy (in %), UPR is the Upside Potential Ratio.

[16]Conservative in the sense that the bid-ask spread is lost.

Stock	Model	$\bar{N}(d)$	$\bar{\delta}$
AA	Kernel	24.70	4.10
	Wavelets	15.90	5.80
	MAP Skew Gaussian	19.50	4.20
	MAP Skew Student	19.10	4.30
KR	Kernel	24.30	3.70
	Wavelets	17.30	5.50
	MAP Skew Gaussian	19.10	4.30
	MAP Skew Student	17.80	4.60

Table 5.6: Pattern Analytics. $\bar{N}(d)$ is the average daily number of patterns detected. $\bar{\delta}$ is the average trade duration.

5 Conclusion

In this article we apply the MAP framework to Student, skewed Gaussian and skewed Student noise. It is therefore possible to take account of the asymmetry and the leptokurticity of the noise in the denoising procedure. Even if we considered only unidimensional signals, we believe that this methodology can be extended to higher dimension signals, such as images. However when higher dimensional signals are generated by irregularly spaced observations, the main challenge would be to define a robust integration path. As shown, the estimation algorithm is easy to implement and has promising results. Finally, we implemented it in an industry oriented problem, where the goal is to denoise stock prices for pattern recognition purposes. Empirical results indicate both the well-behaviour of our methodology as well as its concrete superiority in terms of profit made. An other further research would be to implement the modification of Stein's lemma, by Pal (2008) for a skewed Gaussian noise, to derive an optimal regularization parameter selection for skewed Gaussian noise.

Chapter 6

Conclusion

The first part of this dissertation contributes to a growing literature in finance: volatility spillovers. Using quantile regressions to measure volatility spillovers, we bring empirical evidences of contagion in volatility around the Earning Announcement. For instance, the day of the announcement, a move by one point in the announcer's historical volatility, make the median volatility of its co-sectors rising by 2.5%. We also notice that the spillover effect varies across the nature of outcome: when the announcing firm reports a bad new, a move of one point in its volatility makes the median volatility of its co-sectors rising by 3.5%, and only 2.2% for good news. The intensity of the volatility spillover varies also across other variables such as the capitalization of the announcer, the sentiment of the market across the announcer or the order of publication within a given earning season. As a result, we gather strong evidences of volatility spillovers around EA. We believe that these findings are of great importance in the asset management industry: one could decrease the volatility of an equity portfolio by removing firms generating volatility spillovers. Also, a simple volatility strategy sheds light on the existence of volatility arbitrage opportunities around EA.

The second part focuses on noise properties and contributes to the signal theory literature in two ways. First, we relax the assumption of an *identically* distributed noise, by considering problems where the noise variance is non-constant. To remove such noise, we develop an algorithm, where the noise variance and the original signal are estimated iteratively. We illustrate the well-behaviour of the algorithm on simulated data and we show that it achieves the two challenges face by the denoising procedure with success. We can obtain both a robust estimator of the noise standard-deviation and a robust estimator of the underlying signal. This algorithm is well suited for volatility estimation in finance. We therefore applied it to volatility estimation. When comparing it with standard parametric and non-parametric techniques of volatility estimation, we find that our approach is well-suited for this task. It over performs standard non-parametric approaches. This conclusion is true for both in-sample estimation than for out-sample forecasts.

Second, following existing extensions of Maximum A Posteriori models to cases where the noise is non-Gaussian, we propose MAP formulations to cases where the noise is asymmetric or leptokurtic. In our setting, the MAP formulations of the total variation denoising problem is solved using *Euler-Lagrange* equations. Simulations highlight the well-behaviour of the MAP formulation, like when the noise is leptokurtic or asymmetric, a signal estimator obtained thanks to the appropriate MAP approach is always the closest to the true signal. We propose an application to an industry oriented problem. We denoise an intra-day stock price, to recognize known patterns in order to invest. We observe that, when the price is denoised through a skewed Gaussian MAP or a skewed Student MAP, the profit generated by the pattern recognition algorithm is always the highest. This suggests that MAP approaches are the best to remove the noise while preserving local extrema.

Following this research, my next objective is to built an *R package* on Total Variations techniques for unidimensional signals. This package will propose a variety of denoising techniques based on Total Variation as well as reproducible examples. This package could facilitate the use of these techniques in the industry as well as in the academic world.

Further research is divided into three main directions.

First, an on going project studies the change in the memory of a time continuous process depending on time scale selected. The intuition behind this assertion, is that, high frequency data, are more predictable than low frequency data. This approach will use the estimate of time dependent Hurst exponent, developed by Garcin (2017). This estimate is obtained by solving a ROF functional.

Second, Pal (2008) applies Stein's lemma to the case of a skewed Gaussian noise. I am interested in using this result to derive the Stein Unbiased Risk Estimate (SURE) when the noise is skewed Gaussian. This could have several applications such as the choice of the regularization parameter through a SURE minimization.

Lastly, a more data oriented project will focus on the impact of denoising in a classification problem. For instance, we are interested in the stability of some machine learning algorithm to the presence of noise in the original data.

List of Figures

List of Tables