Mateus Gonzalez de Freitas Pinto

Long memory in high frequency time series using wavelets and conditional volatility models

São Paulo, Brazil
February 2021

Mateus Gonzalez de Freitas Pinto

Long memory in high frequency time series using wavelets and conditional volatility models

Universidade de São Paulo - USP
Instituto de Matemática e Estatística
Programa de Pós-Graduação em Probabilidade e Estatística

Supervisor: Prof. Dr. Chang Chiann

São Paulo, Brazil
February 2021

Ad maiorem Dei gloriam.

Agradecimentos

Agradeço em primeiro lugar a Deus, que me concedeu os dons necessários para desenvolver este trabalho.

À minha mãe Erika, pelo incondicional apoio que não se cansou de me dar ao longo desta caminhada. Ao meu avô João, em memória de minha avó Elenice e à minha tia Glaucia, pelo suporte contínuo que me deram nestes anos de pós-graduação.

Sou também muito grato à minha orientadora, professora Chang Chiann, pelas incansáveis explicações e contínuas reuniões, e pelas excelentes sugestões que me deu no desenvolvimento desta dissertação, tornando este processo de aprendizado muito melhor. Agradeço-lhe também pelo incentivo e confiança para que prosseguisse meus estudos, conciliando-os com o trabalho.

Agradeço aos professores da banca, professora Silvia Lopes e professor Michel Montoril pelos comentários e sugestões feitos no dia de minha defesa de mestrado, que foram incorporados nesta versão final da dissertação.

Agradeço ainda ao professor Guilherme de Oliveira Lima, pelas boas conversas e observações feitas durante este trabalho, além de todo apoio e incentivo. Sou também grato ao professor Alexei Veneziani, que durante a graduação me auxiliou em meus estudos e me deu grande suporte para que ingressasse no programa de mestrado.

A meus colegas de trabalho da B3, agradeço pelo companheirismo, pela compreensão e pelo suporte, em me permitirem assistir às aulas do mestrado, mesmo durante o horário comercial. Em especial aos amigos: Fábio Ciribelli, meu professor dentro da B3, por todo o companheirismo e pelos excelentes conselhos nestes anos trabalhando juntos; Julio Thomaz, pelas recomendações e pelo apoio durante o período dos meus estudos; Luiz França, pelas agradáveis conversas e pelas ideias que delas surgiam durante os cafés. Agradeço ainda à minha superintendente Paula Martins, pelo apoio durante o período de estudos, aos demais colegas da Superintendência Risco de Crédito e, em geral, a todas as equipes da Diretoria de Administração de Riscos, pelo apoio que sempre encontrei em vocês.

Aos meus colegas e amigos do IME, junto dos quais pude passar excelentes momentos de aprendizado e convivência, e que tornaram estes anos mais agradáveis.

Por fim, agradeço aos professores do IME, por me auxiliarem a desenvolver maior criticidade e rigor metodológico para elaborar esta dissertação e pelo grande comprometimento com a excelência do ensino e pesquisa.

"Nolite ergo solliciti esse in crastinum.
Crastinus enim dies sollicitus erit
sibi ipse: sufficit diei malitia sua."
(Matthaeus 6:34)

Resumo

O objetivo desta dissertação é descrever uma metodologia para modelagem da volatilidade de dados financeiros de alta frequência, considerando suas particularidades e fatos estilizados. Os modelos ARFIMA e FI(E)GARCH são utilizados para modelar a longa persistência das séries na média e na variância condicional, respectivamente, quando isto for observado. A fim de contemplar não-normalidade, assimetria e curtose são utilizadas as distribuições t de Student Assimétrica e Distribuição Generalizada de Erros (GED) para o termo de inovações dos modelos supracitados. A limiarização de ondaletas é utilizada para identificação e separação dos "*jumps*" intradiários de forma não-paramétrica. A aplicação deste procedimento é apresentada utilizando séries financeiras reais de retornos de ações em alta frequência para ativos negociados no mercado à vista na bolsa de valores brasileira, além de séries de taxas de câmbio de criptomoedas, comparando o modelo semiparamétrico proposto a uma abordagem tradicional sem remover os "*jumps*".

Palavras-chaves: memória longa, FIGARCH, volatilidade, ondaletas, dados intradiários, dados de alta frequência, retornos.

Abstract

The goal of this dissertation is to describe a methodology for modelling the volatility of high frequency financial data, considering its features and stylized facts. In order to account for the long-range dependence in conditional mean and conditional variance, ARFIMA and FI(E)GARCH models are used respectively, when observed. To account for the non-normality, skeweness and kurtosis, features observed in the the distribution of the log-returns in high frequency, the Skewed Student t and the Generalized Error Distribution (GED) are adopted for the innovation term of the aforementioned models. Wavelet shrinkage is used in a non-parametric identification and separation of the intraday jumps from the time series data. The application of this procedure is presented using real high frequency asset returns from the Brazilian Exchange and OTC, as well as exchange rates from cryptocurrencies traded in Crypto Exchanges.

Key-words: long memory FIGARCH, volatility, wavelets, intraday data, high frequency data, asset returns.

List of Figures

List of abbreviations and acronyms

$\overline{A}$	closure of a given set A
ACF	abbreviation for "autocorrelation function"
PACF	abbreviation for "partial autocorrelation function"
ARIMA	abbreviation for "autoregressive integrated moving average process"
ARFIMA	abbreviation for "autoregressive fractionally integrated moving average"
GARCH	abbreviation for "Generalized autoregressive conditional heteroskedasticity"
FI(E)GARCH	abbreviation for "Fractionally integrated (exponential) generalized autoregressive conditional heteroskedasticity"
APARCH	abbreviation for "Asymmetric Power ARCH model".
TARCH	abbreviation for "Threshold ARCH model".
GJR-GARCH	abbreviation for "Glosten–Jagannathan–Runkle generalized autorregressive conditional heteroskedasticity".
X-GARCH	abbreviation for "GARCH model with cross-sectional volatility", sometimes abbreviated as GARCH-X as well.
M-GARCH	abbreviation for "GARCH in the mean" model, sometimes abbreviated as GARCH-M as well.
GPH procedure	abbreviation for the Geweke-Porter-Houdak procedure to estimate the long memory parameter d.
ISML	abbreviation for "International Mathematics and Statistics Library" for Fortran.
NSPE	Abbreviation for "Nonseasonal ARMA Parameter Estimation", a subroutine written for Fortran in the IMSL.
BFGS	abbreviation of "Broyden–Fletcher–Goldfarb–Shanno algorithm" used for numerical optimization.
SOLNP	algorithm for nonlinear optimization using augmented Lagrange method.

GOSOLNP	quasi-global SOLNP, with random initialization and multiple restarts.
MLE	abbreviation for "maximum likelihood estimation".
HFD	abbreviation for "high frequency data".
MSE	abbreviation of mean squared error.
RMSE	abbreviation of root mean squared error.
CTS	abbreviation for "calendar time sampling".
D2	orthogonal Daubechies Wavelet with 2 vanishing moments, also referred to as Haar wavelets.
D8	orthogonal Daubechies Wavelet with 8 vanishing moments.
$\mathcal{L}_2(\mathbb{R})$	the set of all square integrable functions in the real line
r.v.	abbreviation for "random variable".
i.i.d.	abbreviation for "independent and identically distributed".
GED	abbreviation for "Generalized Error Distribution".
$X \sim \mathcal{N}(\mu, \sigma^2)$	The r.v. X is normally distributed, with mean μ and variance σ^2.
$X \sim \mathcal{WN}$	The r.v. X is distributed as a white noise.
$\Gamma(z)$	The gamma function, given by $\Gamma(z) = \int_0^\infty e^{-x} x^{z-1} dx$.

Contents

Introduction

Volatility is a measure of return variability over a certain period of time and it is a feature of great interest when it comes to modelling financial data, be it among theoretical researchers or practitioners. Risk analysts and other market players are usually interested in modelling and understanding the volatility, understood as the covariance structure of asset returns. This is even more evident particularly for intraday data (also called high frequency data), because they tend to display a distinct pattern of behaviour from those of lower frequencies, which was noticeable ever since the beginning of study of these classes of data.

At first, one might expect no difference between modelling high frequency financial data and other types of time series. However, these time series usually display stylized facts which need to be taken into account in order to describe its behaviour. Such peculiarities are partially due to the price formation and market microstructure effects, which are usually shadowed in lower frequencies and other type of time series, but also depend on the liquidity of the financial markets in which the assets are traded (ZIVOT; WANG, 2005).

In general, as stated by Zivot and Wang (2005), Tsay (2005) and Dacorogna et. al. (2001), high frequency spot prices tend to be non-stationary, with stochastic trend, and the high frequency financial returns tend to be highly autorocorrelated, be it in the conditional mean or conditional variance. In addition, these data tend to display the so-called "intraday jumps", which are abnormal returns in specific periods of time, to which there is no consensus about how to be dealt with. Nonetheless, high frequency data are not broadly available in free platforms, which makes it difficult for them to be analyzed and studied. These traits make the analysis of high frequency financial data a topic of great interest, because of the distinctiveness with which the data is analyzed.

The main goal of this dissertation is to present a method for modelling volatility of high frequency financial data, particularly considering the spot market asset returns and discuss the underlying theory of models and methods used to account for the stylized facts. Some of these particularities are: long-range dependence in the conditional mean and conditional variance, intraday jumps (anomalous returns), volatility clusters, fat-tailedness, non-normality. Apart from that, the data tend to be skewed and the data is available in the "tick-by-tick" form, meaning that the quantities of interest are not usually displayed in homogeneously spaced time, but are irregularly spaced in time.

In order to account for the non-normality, skeweness and kurtosis features observed in the the distribution of the log-returns in high frequency, different distributions such

as the Skewed Student t and the Generalized Error Distribution (GED) will be used for the innovation term of the aforementioned models. A procedure using wavelet shrinkage is described and applied in order to identify and separate the intraday jumps from the time series data. The jumps themselves are treated as additive outliers in the time series data, which is corroborated by the empirical analysis. Apart from that, high frequency data require a preprocessing which includes not only cleaning the data from spurious information (such as cancelled orders), but also generating the variables of interest from the "tick-by-tick" series, which is irregularly spaced. It is necessary then to apply a procedure to turn the irregular data into an homogeneous series, called regularizing procedure, which will be discussed in details.

This dissertation is divided in four chapters. The first chapter presents some time series foundations in order to model these asset returns, and it is subdivided in models for the conditional mean and conditional variance, apart from other technicalities which are useful to be discussed.

The second chapter discusses the wavelet theory for signal processing, with the required concepts for the practical application, together with computational examples of wavelet families and wavelet shrinkage.

The third chapter presents a theoretical discussion on modelling the financial intraday data, its stylized facts and preprocessing, as well as the methodology which will be used in practical applications.

The fourth chapter consists in a chapter of results, which displays the procedure for modelling asset returns for three different assets negotiated in the Brazilian Exchange and OTC (B3 S.A.- Brasil, Bolsa, Balcão) sampled every minute, which uses the techniques discussed in the previous sections, as well as exchange rates of cryptocurrencies traded in Crypto Exchanges.

As for the results chapter, the computational applications are performed using different softwares. In general, applications are developed in R language (R Core Team, 2013) together with Ox Language (DOORNIK, 2007) and OxMetrics-G@RCH^{TM} distributed by Timberlake Consultants and described in details with software implementation by Laurent (2018) and the S modules S+FinMetrics and S+Wavelets, which were distributed by Insightful Co. and currently by Tibco Spotfire S-PLUS®.

Finally, there is the chapter of conclusion and discussion of further work, where the previous applications and theory are discussed.

1 Foundations

> All models are wrong, but some are useful
>
> George E. P. Box

1.1 Some basic concepts of time series

This section provides the basic concepts for time series analysis, which are useful for the further sections and chapters in this dissertation. Several concepts here introduced are based on Morettin and Toloi (2006), Morettin (2014) and Shumway and Stoffer (2016), Hamilton (1994), Zivot and Wang (2005) and Vidakovic (1999). Other important references are cited in the chapter.

Definition 1.1.1. *A function $f : \mathbb{R} \longrightarrow \mathbb{R}$ is called square integrable on the real line if:*

$$\int_{-\infty}^{\infty} (f(x))^2 dx < \infty.$$

The set of all square integrable functions is called $\mathcal{L}_2(\mathbb{R})$.

For the set of all square integrable functions, it holds that (VIDAKOVIC, 1999):

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)dx, \tag{1.1}$$

and thus the norm is given by:

$$||f(x)|| = \sqrt{\int_{-\infty}^{\infty} (f(x))^2 dx}.$$

In fact, the set $\mathcal{L}_2(\mathbb{R})$ with the inner product defined in 1.1 forms a Hilbert space (VIDAKOVIC, 1999).

Definition 1.1.2. *Let T be an arbitrary set. A stochastic process $\{x(\omega, t), \omega \in \Omega, t \in T\}$ is a set of random variables defined in the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where $\mathcal{F}$ is a σ-algebra (also σ-field) of subsets of Ω and $\mathbb{P}$ is a measure.*

If $T = \mathbb{Z}$, then the process is called discrete. An assumption henceforth made is that the random variables X_t is real. Hence, for every $t \in T$, $X(t, \omega)$ is a random variable defined over Ω, with a given $f_X(x)$ probability density function, which is assumed to be existent.

Definition 1.1.3. *A time series is any set of real numbers indexed or ordered in time. In fact, for a given stochastic process* $\{x(\omega,t), \omega \in \Omega, t \in T\}$*, for every* $\omega \in \Omega$ *fixed, a time series may be understood as a function* $x(.,\omega)$ *of* $t : T \longrightarrow \mathbb{R}$*, called trajectory of the stochastic process.*

Definition 1.1.4. *Let* $\{X_t\}_{t\in T}$ *be a real stochastic process. Let also* $F(\cdot)$ *be the cumulative distribution function of* $\{X_t\}_{t\in\mathbb{Z}}$*. The mean function of* X_t *is given by:*

$$\mu(t) = E(X_t) = \int_{-\infty}^{\infty} x_t dF(X_t).$$

Definition 1.1.5. *Let* $\{X_t\}_{t\in T}$ *be a real stochastic process. The autocovariance function is given by:*

$$\gamma(t+\tau, t) = E(X_{t+\tau}X_t) - E(X_{t+\tau})E(X_t),$$

where $E(\cdot)$ *is the expectation operator, for* $\tau = 1, 2, \ldots$.

The autocorrelation function (ACF) is given by:

$$\rho(\tau) = \frac{cov(X_t, X_{t+\tau})}{\sqrt{Var(X_t)Var(X_{t+\tau})}}.$$

Particularly, if $\tau = 0$ in definition 1.1.5, then:

$$\gamma(t,t) = Var(X_t) = E(X_t^2) - E^2(X_t), \tag{1.2}$$

which is the variance of the stochastic process at the time t.

Definition 1.1.6. *A real stochastic process* $\{X_t\}_{t\in T}$ *is called strictly stationary when the whole family of its finite dimensional distributions is invariant under a common translation of the time arguments, that is:*

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) = F(x_1, \ldots, x_n; t_1 + \tau, \ldots, t_n + \tau),$$

with $\tau \in \mathbb{Z}$*, for every* $t_1, \ldots, t_n \in T$.

Also, $\{X_t\}_{t\in T}$ *is called second-order stationary or wide-sense stationary if:*

1. $E(X_t) = \mu(t) = \mu, \forall t \in T$;
2. $E(X_t^2) < \infty, \forall t \in T$;
3. $\gamma(t, t+\tau) = Cov(X_{t+\tau}, X_t) = \gamma(\tau)$.

For more details, see Brillinger (2001).

If a process is strictly stationary, then all of the multivariate distribution functions for subsets of variables must agree with their counterparts in the shifted set for all values of the shift parameter τ (SHUMWAY; STOFFER, 2016).

Second-order stationary processes are "stable" in the sense that their mean is constant and their variance, even though not necessarily constant, is finite. It means that the traits and statistical properties of the process in $X_{t+\tau}$ are the same as in X_t, regardless of which $t \in T$ is chosen. Also, some properties hold to stationary processes regarding the autocovariance function and the autocorrelation function introduced in definition 1.1.5.

Another important definition is the one of ergodicity. In straightforward terms, a stationary sequence is ergodic if it satisfies the strong law of large numbers (FRANCQ; ZAKOIAN, 2019). Formally, consider the definition that is provided by Francq and Zakoian (2019):

Definition 1.1.7. *Let $\{Z_t, t \in \mathbb{Z}\}$ be a real valued and strictly stationary process. It is called ergodic if, and only if, for any Borel set B and any integer k:*

$$\frac{\sum_{t=1}^{n} \mathbb{1}\left(Z_t, Z_{t-1}, Z_{t-2} \ldots, Z_{t-k}\right)}{n} \rightarrow \mathbb{P}\left(\left(Z_t, Z_{t-1}, Z_{t-2} \ldots, Z_{t-k}\right) \in B\right), \tag{1.3}$$

with probability 1, where $\mathbb{1}(.)$ denotes the indicator function.

On what regards the next propositions and definitions, it is assumed that, for the process $\{X_t\}_{t\in\mathbb{Z}}$, $E(X_t) = 0$ without loss of generality. Whenever a process is called stationary hereinafter, one may assume it is referred to second-order stationarity, unless stated otherwise.

Proposition 1. *Let $\{X_t\}_{t\in\mathbb{Z}}$ be a real discrete zero mean stationary process with autocovariance function of $\gamma(\tau) = E(X_t X_{t+\tau})$. Then the following properties hold:*

1. $\gamma(0) \geq 0$;
2. $\gamma(-\tau) = \gamma(\tau)$;
3. $|\gamma(\tau)| \leq \gamma(0)$;
4. *for $a_1, a_2 \ldots a_n \in \mathbb{R}$ and $\tau_1, \tau_2 \ldots \tau_n \in \mathbb{Z}$, it holds that:*

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_j a_k \gamma(\tau_j - \tau_k) \geq 0.$$

Similar properties hold for the autocorrelation function of a given process. One can just divide the autocovariance function by the quantity $\gamma(0)$ in order to find which properties hold for the $\rho(\tau)$ function.

Definition 1.1.8. *A second-order real process $\{X_t\}_{t\in T}$ is called continuous in square mean in a given point t_0 if:*

$$\lim_{t\to t_0} E(|X_t - X_{t_0}|^2) = 0.$$

It is denoted $X_t \to X_{t_0}$.

The convergence in the square mean is related to the properties of convergence of the autocovariance function. It means, essentially, that on average, X_t is close to X_{t_0}, but for some values, X_t is allowed to be far away from X_{t_0}. It is a less strong type of convergence than the uniform convergence and it is also called convergence in $\mathcal{L}_2$, the set of all square integrable functions. (BOGGESS; NARCOWICH, 2009).

Proposition 2. *Continuity of $\gamma(\tau)$ for $\tau = 0$ implies in continuity for $\gamma(\tau), \forall\tau \in \mathbb{R}$.*

Proposition 3. *If $\gamma(\tau)$ is continuous, then the process X_t is continuous in square mean.*

The proof for the last two propositions is straightforward can be found within Morettin and Toloi (2006) or Shumway and Stoffer (2016).

Definition 1.1.9. *Let $\{y_t\}_{t\in T}$ be a sequence of random variables and let $I_t = \{y_t, y_{t-1}, \ldots\}$ a set of all the previous information based on the past of $\{y_t\}_{t\in T}$. The sequence $\{y_t, I_t\}$ is a martingale if:*

1. $I_{t-1} \subset I_t$;
2. $E(|y_t|) < \infty$;
3. $E(y_t|I_{t-1}) = y_{t-1}$.

An example of a martingale sequence is the random walk model, which is given by:

$$y_t = y_{t-1} + \epsilon_t, \epsilon \sim \mathcal{WN}(0, \sigma^2), \tag{1.4}$$

and it is usually denoted $y_t \sim \mathcal{RW}$.

Definition 1.1.10. *Let $\{e_t\}_{t\in T}$ be a sequence of random variables and let $I_t = \{y_t, y_{t-1}, \ldots\}$ a set of all the previous information based on the past of $\{e_t\}_{t\in T}$. The sequence $\{e_t, I_t\}$ is a martingale difference sequence if:*

1. $I_{t-1} \subset I_t$;
2. $E(e_t|I_{t-1}) = 0$.

If a process $\{y_t\}_{t\in T}$ is a martingale, as in definition 1.1.9, one can get a martingale difference sequence by:

$$e_t = y_t - (y_t|I_{t-1}),$$

which is an uncorrelated process by construction, that follows also the law of iterated expectations. These processes are useful for modelling the conditional variance of financial time series (ZIVOT; WANG, 2005).

Definition 1.1.11. *A real stationary process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is called a white noise, denoted by $\epsilon_t \sim \mathcal{WN}$ if the r.v. ϵ_t are uncorrelated, that is: $cov(\epsilon_t, \epsilon_s) = 0, t \neq s$.*

The next definition below is necessary for the upcoming theorem, which is widely used in time series analysis and provides a good tool for analyzing stationary processes (PALMA, 2007):

Definition 1.1.12. *Let $\mathbf{\Omega_t} = \{y_s : s < t\}$ be the past of the process $\{y_t\}_{t\in T}$ at the time t and $\mathbf{\Omega_{-\infty}} = \bigcap_{t=-\infty}^{\infty} \mathbf{\Omega_t}$. A process is called deterministic or perfectly predictable or singular if $t \in \mathbf{\Omega_{-\infty}}$ for all $t \in \mathbb{Z}$. On the other hand, a process is called purely non-deterministic or also regular if $\mathbf{\Omega_{-\infty}} = \{0\}$.*

An interpretation is that a stationary process is purely non-deterministic if the (sufficiently far) linear past is of no use in terms of predicting future values of the process (FRANCQ; ZAKOIAN, 2019). From that, Palma (2007) presents the following theorem:

Theorem 1. *[Wold's Decomposition] Any stationary process is the sum of a regular process and a singular process, as in Definition 1.1.12. These two processes are orthogonal and the decomposition is unique.*

For the proof see Brockwell and Davis (2009). The Wold decomposition theorem stated that any stationary process $\{y_t\}_{t\in T}$ can be decomposed into a linear process or infinite order moving average (FULLER, 1996), sometimes referred to as Wold expansion or Wold form, with representation:

$$y = \mu + \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}, \tag{1.5}$$

with:

$$\psi_0 = 1, \qquad \sum_k \psi_k^2 < \infty,$$

$$\epsilon_t \sim WN(0, \sigma^2).$$

In the Wold form of equation 1.5 and recalling definition 1.1.11, it is possible to show that:

$$\begin{aligned}
E(y_t) &= \mu, \\
\gamma(0) &= Var(y_t) = \sigma^2 \sum_{k=0}^{\infty} \psi_k^2, \\
\gamma(j) &= cov(y_t, y_{t-j}) = \sigma^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+j}, \\
\rho(j) &= \frac{\sum_{k=0}^{\infty} \psi_k \psi_{k+j}}{\sum_{k=0}^{\infty} \psi_k^2}.
\end{aligned}$$

It means that any ergodic and stationary process follows a pattern in the autocorrelation function, determined by the Wold representation of a time series and, in order to ensure convergence of the linear process representation with well behaved properties, it is necessary to further restrict the behavior of the moving-average weights, which are the following:

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty, \tag{1.6}$$

$$\frac{\partial y_{t+s}}{\partial \epsilon_t} = \psi_s. \tag{1.7}$$

Equation 1.6 is called summability property, whereas equation 1.7 is called impulse-response property, and it holds that $\sum_{s=0}^{\infty} \psi_s < \infty$ for any ergodic and stationary process y_t (ZIVOT; WANG, 2005). The process in 1.5 is called causal, in the sense that y_t does not depend on any future value of ϵ_t. The last two conditions are those necessary for the convergence of the infinite series (BERAN et al., 2013).

Definition 1.1.13. *The backshift operator, denoted by B, is such that, for a given time series X_t:*

$$B^p X_t = X_{t-p}. \tag{1.8}$$

Using the backshift operator, it is possible to write the Wold representation of a time series as:

$$y_t = \mu + \Psi(B)\epsilon_t, \tag{1.9}$$

where:

$$\Psi(B) = \sum_{k=0}^{\infty} \psi_k B^k, \psi_0 = 1. \tag{1.10}$$

Given a process y_t which is second-order stationary, with Wold representation as in 1.9 considering $\mu = 0$ without loss of generality, one might ask which would be the best linear prediction $\hat{y}_{t+k}$ of y_{t+k} given vector of the past information $\mathbf{\Omega}\boldsymbol{s}$, with $s \leq t$ and $k > 0$. The prediction with minimum mean squared error (MSE):

$$E\left(y_{t+k} - \hat{y}_{t+k}\right)^2 = E\left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t+k-j} - \sum_{j=0}^{\infty} \psi^*_{j+k} \epsilon_{t-j}\right)^2, \tag{1.11}$$

will be given by:

$$\hat{y}_{t+k} = E(y_{t+k}|\mathbf{\Omega}\boldsymbol{s}). \tag{1.12}$$

Wold representation is theoretically interesting due to its properties. In fact, computing the prediction of k-steps ahead of a time series, under assumptions of second-order stationarity and a representation of a process in the Wold form, will require only general results on orthogonal projections in Hilbert spaces (BERAN, 2017). Yet this is rather complicated to achieve in terms of estimation and computational modelling. Box and Jenkins have proposed a methodology in 1976 which is extensively described in Box, Jenkins and Reinsel (2008). This class of stationary and ergodic processes is the autoregressive-moving average processes, shortly denoted ARMA.

Definition 1.1.14. *Let $\{X_t\}_{t\in T}$ be a real stationary process with mean μ and let $Z_t = X_t - \mu$. The autoregressive integrated moving average process of order (p, d, q), hereinafter denoted as $ARIMA(p, d, q)$ is defined as:*

$$\Phi(B)\Delta^d Z_t = \Theta(B)\epsilon_t, \tag{1.13}$$

where $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$, B is the backshift from definition 1.1.13, $\Delta = 1 - B$, $d \in \mathbb{Z}_+$ and:

$$\Phi(B) = (1 - \sum_{i=1}^{p} \phi_i B^i), \tag{1.14}$$

$$\Theta(B) = (1 - \sum_{i=j}^{q} \theta_i B^j). \tag{1.15}$$

If $d = 0$, $\{y_t\}_{t \in T}$ is said to follow an $ARMA(p, q)$ process.

The equation 1.14 is called the autoregressive polynomial, whereas 1.15 is called moving average polynomial. For the process to be stationary and invertible, the roots of $\Phi(z) = 0$ and $\Theta(z) = 0$ must lie outside the unit circle, and also share no common or cancelling factors (SHUMWAY; STOFFER, 2016). The operator Δ is called difference operator, and Δ^d the d-th difference of the time series.

The ARIMA models have been used extensively to model the dynamics of the conditional mean of a given time series, and it was proposed initially by Box and Jenkins (1970).

The method for univariate series consists of four steps: identification, specification, estimation and diagnostic checking (also referred to as residual analysis) (BOX; JENKINS; REINSEL, 2008). All of the steps of this procedure are described in time series manuals, such as Morettin and Toloi (2006), Enders (2010) or Shummway and Stoffer (2016), with details for computational implementation with R subroutines.

The concept of invertibility means that the time series $\{y_t\}_{t \in T}$ can be represented by finite-order convergent autoregressive process. The invertiblity of the process is important because the use of the ACF and PACF for model identification assume that the time series can be represented by an autoregressive model (ENDERS, 2010). The estimation is done via conditional maximum likelihood and it is described in details by Hamilton (1994).

Proposition 4. *Let $\{y_t\}_{t \in T}$ be an ARMA process, as in definition 1.1.14. The spectrum of the ARMA is given by:*

$$f_Y(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1 - \sum_{j=1}^{q} \theta_j e^{-i\lambda j}|^2}{|1 - \sum_{k=1}^{p} \phi_k e^{-i\lambda k}|^2}, -\pi \leq \lambda \leq \pi. \tag{1.16}$$

1.2 Long memory processes for the conditional mean

1.2.1 The pure long memory process

There is a more general class of processes which assumes the $d \in \mathbb{R}$, and not necessarily integers, which is called ARFIMA or FARIMA. These are referred to also as long memory processes, or also as processes with strong dependence or long-range dependence.

A brief introduction to the history of long memory processes is provided by Beran (1994). Even though these models are extensively used in financial econometrics, they have been developed in the 1950's in studies for climatology and hydrology and were lately extended to other fields of knowledge, such as economics, astronomy, chemistry, geosciences and statistics. The phenomenon was noted by Hurst (1951), Mandelbrot and Wallis (1968) with the first heuristics to estimate the self-similarity parameter. Granger identified applications of self-similar processes to economics in his pioneer work in 1966, where he identified that at a sufficiently good approximation, the spectral density of an economic time series is typically a function with a pole at the origin. Later Granger and Joyeux (1980) and also Hosking (1981) have introduced the fractional ARIMA, also referred to as ARFIMA or FARIMA, which will be presented in this section.

1.2.1.1 Definition and identification

Overall, long memory processes tend to look stationary. If one looks at a short range or period of time, the process seems to have cycles and/or local trends, which does not apply when one analyzes the whole series. Additionally, the variance of the sample mean seems to decay to zero at a rate slower than $1/n$ and the sample autocorrelations tend to decay slowly to zero. All of these features are not compatible with Markovian processes or ARIMA processes (BERAN, 1994).

The $ARMA(p, q)$ is considered a short memory process, in the sense that their ACF $\rho(j)$ of the process decays rapidly (or exponentially) towards zero. In fact:

$$|\rho(j)| \leq Cr^j, j = 1, 2, \ldots$$

Since the ACF contains the same information as the spectral density, because:

$$f(\lambda) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda}, \tag{1.17}$$

where $\lambda \in [-\pi; \pi]$ is the Fourier frequency. So it is possible to show that:

$$f(\lambda) \to C_f \lambda^{\alpha-1}, \lambda \to 0. \tag{1.18}$$

And from the previous equation, it is possible to prove that the long memory process is such that the ACF decays hyperbolically towards zero, i.e. :

$$\rho(j) \sim Cj^{-\alpha}, j \to \infty, \tag{1.19}$$

where $C > 0$ and $\alpha \in (0, 1)$. In fact, it decays so slowly that their autocorrelations are not summable:

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty, \tag{1.20}$$

which leads to the definition below (MORETTIN, 2016).

Definition 1.2.1. *Let $\{y_t, t \in \mathbb{Z}\}$ be a process with ACF denoted by $\rho(j)$. The process is called a long memory process if the quantity:*

$$\lim_{n\to\infty} \sum_{j=-n}^{n} |\rho(j)|, \tag{1.21}$$

is not finite.

Another definition is given by Beran (1994):

Definition 1.2.2. *Let y_t be a stationary process for which there exists a real number $\alpha \in (0, 1)$ and a constant $c_\rho > 0$ such that:*

$$\lim_{k\to\infty} \frac{\rho(j)}{[c_\rho j^{-\alpha}]} = 1. \tag{1.22}$$

Then the process for which it holds is called a "stationary process with long memory" or "long range dependence" or even "strong dependence". Alternatively, there can be said that the process is a stationary process with long memory if for a given $\beta \in (0, 1)$:

$$\lim_{\lambda\to 0} \frac{f(\lambda)}{[c_f |\lambda|^{-\beta}]} = 1. \tag{1.23}$$

It is worth to note that all the three statements in the definitions above are equivalent. It is also worth to note that the definitions of long-range dependence are asymptotic definitions, and ultimately describe the behaviour of the autocorrelations as $j \to \infty$, and determines only the rate of convergence, not the absolute size of it (BERAN, 1994).

Beran (1994) also showed that, for a long-range dependence process, it holds:

$$\lim_{n\to\infty} Var\left(\sum_{i=1}^{n} y_i\right) / [c_\gamma n^{2H}] = \frac{1}{H(2H-1)}, \tag{1.24}$$

where:

$$c_\gamma = 2c_f \Gamma(2-2H) \sin(\pi H - \pi/2). \tag{1.25}$$

Hosking (1981) showed that the long memory process can be modelled parametrically by extending an integrated process to a fractionally integrated process, using the fractional difference operator.

Definition 1.2.3. *Let $\Delta = (1 - B)$ be the difference operator. The $\Delta^d = (1 - B)^d$ is called the fractional difference filter, with fractional integration parameter d. For any real number $d > -1$:*

$$(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-1)^k B^k, \tag{1.26}$$

where $\binom{d}{k}$ is the binomial coefficient.

When $|d| > 1/2$, the process is called non-stationary. When $0 < d < 1/2$ the process is called stationary, have long memory and it is referred to as *persistent*. When $-1/2 < d < 0$ the process is stationary, has shorter memory and is referred to as *anti-persistent* (ZIVOT; WANG, 2005). Some times a process where $0 < d < 1/2$ is referred to as a "long memory process", whereas $-1/2 < d < 0$ is referred to as "intermediate memory process", because $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$ (BROCKWELL; DAVIS, 2009).

1.2.1.2 Statistical tests for long memory and estimation

When dealing with a pure long memory process, there are several statistical tests for the long memory or long range dependence. One of the best-known is the rescaled range, or range over the standard deviation (R/S), defined by:

$$Q_T = \frac{1}{s_T} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) \right], \tag{1.27}$$

where $\overline{y} = 1/T \sum_{i=1}^{T} y_i$ and $s_T = \sqrt{1/T \sum_{i=1}^{T} (y_i - \overline{y})^2}$ (ZIVOT; WANG, 2005) and here T denotes the sample size. A more robust estimator to long range dependence is the modified R/S, defined as:

$$\tilde{Q}_T = \frac{1}{\tilde{\sigma}_T(q)} \left[\max_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) \right], \tag{1.28}$$

where $\tilde{\sigma}_T(q)$ is the square root estimator of the long run variance of Newey-West, with bandwidth q, that is given by:

$$\tilde{\sigma}_T^2(q) = S_T^2 \left(1 + \frac{2}{T} \sum_{j=1}^{q} w_{qj} r_j \right), \tag{1.29}$$

where $w_{qj} = 1 - j/(q + 1)$, $q < T$ and r_j are the sample autocorrelation of the series. It is suggested by Newey and West to use $q = [4(T/100)^{2/9}]$ (MORETTIN, 2016).

There are also the KPSS statistic and the rescaled variance procedure, which have slightly simple software implementations and are alternatives to the rescaled range.

These other procedures of estimation of the fractional difference parameter are described in details by Beran (2013).

An estimating procedure for the d parameter is proposed by Geweke and Porter-Hudak in 1983, and is based on the spectral density. Let $(1-B)^d X_t = \epsilon_t$ be the pure long memory process. Considering the fractional difference filter as an input filter, the spectral density of the process is given by:

$$f_X(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_\epsilon(\lambda), \tag{1.30}$$

where $f_\epsilon(\lambda)$ is the spectral density of ϵ_t. From that:

$$\ln f_X(\lambda) = \ln f_\epsilon(0) - d \ln|1 - e^{-i\lambda}|^2 + \ln\left(\frac{f_\epsilon(\lambda)}{f_\epsilon(0)}\right). \tag{1.31}$$

In practice, to estimate the d parameter from that, it is necessary to replace λ for the Fourier Frequency: $\lambda_j = 2\pi j/T$ and use the periodogram as an estimation of the spectral density. The GPH procedure is considered a semiparametric alternative to other estimates of the long memory term.

There are some useful results and properties for the ARFIMA(0,d,0) which can be extended for the ARFIMA(p,d,q). For more details, see Hosking (1981), Sowell (1992) and Beran et al. (2013).

1.2.2 The ARFIMA(p,d,q) model

1.2.2.1 Definition and identification

The idea behind using ARFIMA is to model both the long memory and short run dynamics of a stationary time series. Empirical studies suggests that there is evidence of long memory in financial volatility series and in the mean of the process (ZIVOT; WANG, 2005). One can consider then the definition below:

Definition 1.2.4 (Fractional ARIMA)**.** *Let $\{y_t\}_{t=1,2,\dots T}$ be a stationary process and let $-0.5 \leq d \leq 0.5$. The process y_t is said to follow a fractionally integrated ARMA, or shortly denoted $ARFIMA(p,d,q)$ if it is stationary and satisfies the difference equation:*

$$\Phi(B)\Delta^d y_t = \Theta(B)\epsilon_t, \tag{1.32}$$

where $\epsilon_t \sim \mathcal{WN}(0,\sigma^2)$, B is the backshift from definition 1.1.13, $\Delta = 1 - B$, and $\Phi(B)$ and $\Theta(B)$ are polynomials as in equation 1.14 and 1.15.

The ARFIMA(p,d,q) is stationary and invertible if the roots of the autoregressive polynomial and the moving-average polynomial as in 1.14 and 1.15 lie out of the unit

circle, share no common or cancelling factors and $-0.5 < d < 0.5$. In fact, Hosking (1981) showed that $-0.5 < d$ is an invertibility condition and $d < 0.5$ is a stationarity condition.

There can be considered two interpretations for the ARFIMA models according to Beran (1994). One is to understand the model as an ARMA process after passing the process through the fractional difference filter, meaning:

$$(1 - B)^d y_t = \Phi^{-1}(B)\Theta(B)\epsilon_t. \tag{1.33}$$

And the other is to understand the model as obtained by passing a fractional white noise through the ARMA filter meaning:

$$y_t = \Phi^{-1}(B)\Theta(B)(1 - B)^{-d}\epsilon_t. \tag{1.34}$$

Using the representation in equation 1.34, Hosking (1981) showed that the ARFIMA model can also be expressed as:

$$y_t = \Psi(B)\epsilon_t, \tag{1.35}$$

with:

$$\Psi(B) = \Phi^{-1}(B)\Theta(B)(1 - B)^{-d}, \tag{1.36}$$

where $\Psi(B)$ is the transfer function of the filter, which represents the past effects on the values of the time series, and is stationary if $\Phi(B) = 1 + \sum_j \psi_j B^j$ converges, where each ψ_j are the impulse response coefficients. Another representation is:

$$\pi(B)y_t = \epsilon_t, \tag{1.37}$$

with:

$$\pi(B) = \Phi(B)\Theta^{-1}(B)(1 - B)^d. \tag{1.38}$$

The spectral density of the ARFIMA(p,d,q) process is given by (BROCKWELL; DAVIS, 2009):

$$f_Y(\lambda) = \frac{\sigma^2}{2\pi} \frac{|1 - \sum_{j=1}^{q} \theta_j e^{-i\lambda j}|^2 |1 - e^{-i\lambda}|^{-2d}}{|1 - \sum_{k=1}^{p} \phi_k e^{-i\lambda k}|^2}, -\pi \leq \lambda \leq \pi, \tag{1.39}$$

due to the fact that $\epsilon_t \sim \mathcal{WN}(0, \sigma^2)$. The evaluation of the spectral density of the ARFIMA(p,d,q) is performed by considering input and output filtering, similarly to the process for the ARIMA case. In fact,

$$f_Y(\lambda) = |1 - e^{-i\lambda}|^{-2d} f_{ARMA}(\lambda), \tag{1.40}$$

where $f_{ARMA}(\lambda)$ is the spectral density of the correspondent $ARMA(p, q)$ process, as in Proposition 4.

From equation 1.39, it is possible to show that, since $\lim_{\lambda \to 0}|1 - e^{-i\lambda}| = 1$, then (BERAN et al., 2013):

$$\lim_{\lambda \to 0} f_Y(\lambda) = |\lambda|^{-2d} \frac{\sigma^2}{2\pi} \frac{|1 - \sum_{j=1}^{q} \theta_j e^{-i\lambda j}|^2}{|1 - \sum_{k=1}^{p} \phi_k e^{-i\lambda k}|^2}. \tag{1.41}$$

Sowell (1992) evaluated the autocovariance function of the process $\gamma_Y(\tau)$ using the spectral density, and it is given by:

$$\gamma_Y(\tau) = \sigma^2 \sum_{l=-q}^{q} \sum_{j=1}^{p} \psi(l) \xi_j C(d, p + l - s, p_j), \tag{1.42}$$

where:

$$\psi(l) = \sum_{s=\max[0,l]}^{\min[q,q-l]} \theta_s \theta_{s-l}, \tag{1.43}$$

$$\xi_j = \left(p_j \prod_{i=1}^{p} (1 - p_i p_j) \prod_{m \neq j} (p_j - p_m) \right)^{-1}, \tag{1.44}$$

and

$$C(d, h, p) = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{p^{2p}}{1 - pe^{-i\lambda}} - \frac{1}{1 - p^{-1}e^{-i\lambda}} \right) (1 - e^{-i\lambda})^{-d} (1 - e^{i\lambda})^{-d} e^{-i\lambda h} d\lambda. \tag{1.45}$$

1.2.2.2 Estimation

The methods described before for estimation of the long memory parameter of a pure fractionally integrated process have the advantage of being simple to implement computationally. However, their main disadvantage is that they are not the most reliable choice when it comes to statistical inference. This is due to the fact that some of the procedures depend on "visual inspection" or lack robustness. These estimates are therefore to be used generally as a fast approach to check on whether there is presence or not of long memory in the time series. The utmost dilemma with the heuristics of the methods presented before is that they are focused primarily on the long-memory parameter d and they

ignore the other aspects of the data, and other underlying structures of dependence. More elaborate techniques are the parametric and broadband estimation, where the complete dependence structure is modelled (BERAN et al., 2013).

One of them is the maximum likelihood estimator (MLE). Considering $\boldsymbol{X} = (X_1, X_2, \dots X_T)$ observations following an ARFIMA(p,d,q), the estimation for the parameters unknown parameters $d, \boldsymbol{\phi} = (\phi_1, \dots, \phi_p), \boldsymbol{\theta} = (\theta_1, \dots, \theta_q)$ can be done by maximum likelihood estimation. For $\boldsymbol{\beta} = (d, \boldsymbol{\phi}, \boldsymbol{\theta})$, assuming that the errors are distributed as a Gaussian white noise, then the likelihood function is given as:

$$L(\boldsymbol{\beta}, \sigma^2) = (2\pi\sigma^2)^{-T/2}\sqrt{(r_0, \dots r_{T-1})}\exp\left[-\frac{1}{2\sigma^2}\sum_{j=1}^{T}(X_j - \hat{X}_j)^2/r_{j-1}\right], \tag{1.46}$$

where $\hat{X}_j, j = 1, 2, \dots T$ are the one-step ahead predictors and $r_{j-1} = \sigma^{-2}E(X_j - \hat{X}_j)^2, j = 1, 2, \dots T$. The maximum likelihood estimators can be found by maximizing $L(\boldsymbol{\beta}, \sigma^2)$ with respect to $\boldsymbol{\beta}$ and σ^2. The estimators will be (BROCKWELL; DAVIS, 2009):

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{j=1}^{T}\frac{(X_j - \hat{X}_j)^2}{r_{j-1}} = \frac{S(\hat{\boldsymbol{\beta}})}{T}, \tag{1.47}$$

and $\hat{\boldsymbol{\beta}}$ is the value of $\boldsymbol{\beta}$ which minimizes:

$$l(\boldsymbol{\beta}) = \ln\left(\frac{S(\boldsymbol{\beta})}{T}\right) + \frac{1}{T}\sum_{j=1}^{T}\ln r_{j-1}, \tag{1.48}$$

or also denoted as:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}}\left[\ln\left(\frac{S(\boldsymbol{\beta})}{T}\right) + \frac{1}{T}\sum_{j=1}^{T}\ln r_{j-1}\right]. \tag{1.49}$$

If X_t is not a Gaussian process, then minimizing the Gaussian-likelihood will still lead to a consistent estimator of the unknown parameters $\boldsymbol{\beta}$, under fairly general conditions, or quasi-likelihood estimators can be used. The main problem with the MLE presented in equation 1.49 is the fact that the estimation requires an inversion of an $T \times T$ covariance matrix and, for large values of T, it is a complicated task numerically and for d values closer to 0.5 may lead the covariance matrix to singularity (BERAN et al., 2013).

Because of the aforementioned particularities, there is an alternative semiparametric procedure which consists on approximating the log-likelihood using the spectral density of the ARFIMA and the periodogram of the data and also wavelet-based estimators. For a comprehensive review, see Beran (2013). Apart from these methods, there is

a wavelet version of the Whittle estimator, which is not disucssed in this chapter, but is developed with details by Moulines et al. (2008).

When using numerical optimization, another problem is the starting values which must be given for the program. Since the log-likelihood function is not globally concave, the result of the numerical optimization procedure will depend strongly on the starting values. One procedure is to estimate d using the rescaled range or GPH procedure and set it as a starting value, and use the NSPE subroutine from the IMSL to estimate the ARMA parameters and the variance. Another procedure is to choose a grid of d values and, for each d, calculate the series $(1-B)^d y_t$ and estimate the ARMA parameters and the innovation variance by NSPE: the starting parameters will be those associated with the smallest estimate of the innovation variance (SOWELL, 1992).

There are alternatives to the procedure that Sowell (1992) described to estimate the parameters computationally, namely determining the d parameter using the Hyndman-Khandakar algorithm (2008) and determining the orders of p and q according to the algorithm described by Haslett and Raftery (1989), which is an "automatic" way of estimating the parameters in the model.

Beran et. al. (2013) provide an overview of the MLE method for estimation of the long memory parameter d. The MLE appears to have better finite sample properties than the Whittle estimator based on wavelets, even though both of them have the same limiting variance. Conversely, it performs poorly when the data are contaminated, for which there are robust versions of the MLE. Both the Whittle and the ML estimators are asymptotically normal. From a numerical point of view, the Whittle estimator and the AR-based estimators are more tractable, due to the increasing computational complexity that involves the computation of ML estimators.

There have been developed several other methods for the maximum likelihood estimation, which are extensively discussed by Palma (2007) in details, yet are not presented in this chapter. For more details regarding the estimation procedures using Cholesky's decomposition, Dubrin-Levinson's algorithm, Haslett-Raftery procedure and the state state approach and its features, see Palma (2007).

In general, after fitting the conditional mean and conditional variance models, it is necessary to evaluate the serial autocorrelation in the residuals, to see if there are still remaining autocorrelations to incorporate in the model. An approach was proposed by Box and Pierce (1970) and another by Ljung and Box (1978). Let the null hypothesis be $H_0 : y_t \sim \mathcal{WN}(0, \sigma^2)$ with the alternative hypothesis being that H_0 is false. To test that, the Box-Pierce test uses the Q-statistic statistics given by:

$$Q(k) = T \sum_{j=1}^{k} \hat{\rho}_j^2, \tag{1.50}$$

where $\hat{\rho}_j$ is the sample autocorrelation at lag j. Under null hypothesis, $Q(k) \sim \chi^2_k$, a chi-squared distribution with k degrees of freedom. However, for finite samples, the statistics in equation 1.50 may not approximate well the χ^2_k. To account for that, Ljung and Box (1978) proposed a modified Q-statistic, which is given by:

$$MQ(k) = T(T+2)\sum_{j=1}^{k}\frac{\hat{\rho}_j^2}{T-j}, \tag{1.51}$$

which is referred to thereupon as the Ljung-Box test for serial autocorrelation. Laurent (2018) referred to these tests as tests for goodness-of-fit. For large samples, both tests are almost equivalent.

1.3 Error Distributions

Apart from the ARFIMA and GARCH models with normal distribution, one can use other distributions, for better fitting the data. Apart from the Student-T distribution, there are two predominantly used distributions. One of them is the Generalized Error Distribution, also GED, proposed by Nelson (1991). The other one is the asymmetric Student-t, also skewed Student-t distribution, proposed initially by Fernandez and Steel in 1998, with different parameterizations by Hansen in 1994 and also by Lambert and Laurent in 2001, with a more suitable and convenient parameterization for time series analysis (MORETTIN, 2016). These will be introduced in the consequent subsections.

1.3.1 Generalized Errors Distribution

There are mainly two types of generalized error distributions, developed as symmetric distributions with variation in kurtosis. The main interest, as above stated, is in the tail behaviour and they are used, therefore, when the errors show no normal behaviour.

Definition 1.3.1 (GED-I)**.** *The random variable X is said to follow a Generalized Error Distribution of first type (short GED-I), denoted $GED(0,1)$ if its probability density function is given by:*

$$f(x) = \frac{\nu \exp\{-0.5|x/\lambda|^\nu\}}{\lambda 2^{(\nu+1)/\nu}\Gamma(1/\nu)}, \tag{1.52}$$

where:

$$\lambda = \left[\frac{2^{-2/\nu}\Gamma(1/\nu)}{\Gamma(3/\nu)}\right]^{1/2}, \tag{1.53}$$

and $\Gamma(.)$ denotes the gamma function.

The parameter ν rules the behaviour of the tails of the distribution. For $\nu = 2$, it is the standard normal distribution. For $\nu = 1$ it is the double exponential or Laplace distribution. For $\nu < 2$, the distribution has heavier tails than the normal and for $\nu > 2$ the distribution has lighter tails (MORETTIN, 2016).

Nadarajah (2005) and Vasudeva and Kumari (2013) summarize and show interesting properties for the GED-I distribution. It is showed that $E(X) = 0$, $E(X^2) = 1$, $E(X^k) < \infty, k \in \mathbb{Z}_+$ and also that the r.v. X is distributed as a GED-I if $Y = \left(\frac{\Gamma(1/3)X^2}{\Gamma(1/\nu)}\right)^{\nu/2}$ has a gamma distribution with parameter $1/\nu$, meaning that the p.d.f. of Y will be given by:

$$h(y) = \frac{1}{\Gamma(1/\nu)} e^{-y} y^{1/\nu - 1} \mathbb{I}_{(0,\infty)}. \tag{1.54}$$

Vasudeva and Kumari (2013) also show that for $X_1 \dots X_n$ i.i.d GED-I r.v. with parameter ν, then $\sum_{i=1}^{N} |X_i|^\nu$ is a Gamma$(1/(2\lambda^\nu), n/\nu)$. Another parameterization is the Generalized Error Distribution is also called GED-II, which will not be addressed here but is described by Vasudeva and Kumari (2013).

1.3.2 Skewed Student *t* Distribution

To account for the excess skeweness and kurtosis, Fernandez and Steel (1998) proposed the Student-t distribution adding a parameter of skeweness. This procedure allows the introduction of skeweness in any continuous unimodal and symmetric distribution about zero, by changing the scale at each side of the mode. Hansen (1994) also has a parameterization for the Skewed-t distribution. Lambert and Laurent (2001) re-express the skewed-Student density in terms of the mean and the variance, reparametrizing them in such a way that the innovation process has mean zero and unit variance (LAURENT, 2018).

Recalling the Student-t distribution normalized to have unit variance:

$$f(x|\eta) = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\Gamma\left(\frac{\eta}{2}\right)} \left(1 + \frac{x^2}{\eta - 2}\right)^{-(\eta+1)/2}, \tag{1.55}$$

where $2 < \eta < \infty$. Hansen (1994) stated that the equation 1.55 is a restrictive parametric family, allowing only variation in the location, scale and tail thickness. In order to generalize it, Hansen (1994) proposed the following density function:

Definition 1.3.2. *[Hansen Skewed Student-t] A random variable X is distributed according to Hansen's Skewed Student-t distribution if:*

$$f(x|\eta,\lambda) = \begin{cases} bc\left(1+\frac{1}{\eta-2}\left(\frac{bx+a}{1-\lambda}\right)^2\right)^{-(\eta+1)/2}, & \text{if } x < -a/b \\ bc\left(1+\frac{1}{\eta-2}\left(\frac{bx+a}{1+\lambda}\right)^2\right)^{-(\eta+1)/2}, & \text{if } x \geq -a/b \end{cases}, \tag{1.56}$$

where $2 < \eta < \infty$ and $-1 < \lambda < 1$. The constants a and b and c are given by:

$$a = 4\lambda c\frac{\eta-2}{\eta-1},$$

$$b = 1 + 3\lambda^2 - a^2,$$

$$c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)}\Gamma\left(\frac{n}{2}\right)}.$$

It can be showed that this distribution in definition 1.3.2 has zero mean and unit variance. The above distribution specializes to the Student-t of equation 1.55 when $\lambda = 0$. Also, the Hansen distribution has a single mode $-a/b$, which is of opposite sign of the parameter λ. This parameterization allows for skeweness and is easily fitted to GARCH-type of models, because the innovations must have zero mean and unit variance, as it will be showed in the further chapters (HANSEN, 1994).

The parameterization of Lambert and Laurent (2000) is an extension of Fernandez and Steel (1998) parametrization of the Skewed Student t distribution the following:

Definition 1.3.3. *[Lambert & Laurent Skewed Student-t] A random variable X is distributed according to a skewed Student-t distribution if its p.d.f. is given by:*

$$f(x|l,\nu) = \frac{2}{l+1/l}\Big[sg(l(sx+m)|\nu)\mathbb{I}_{(-\infty,-m/s)}(x+m/s) + sg((sx+m)/l|\nu)\mathbb{I}_{[-m/s,+\infty)}(x+m/s)\Big],$$

where $g(.|\nu)$ denotes the Student-t distribution with ν degrees of freedom and

$$m = \frac{\Gamma((\nu+1)/2)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\nu/2)}(l-1/l),$$

$$s = \sqrt{(l^2+1)/(l^2-1) - m^2},$$

and l is the asymmetry parameter and $\Gamma(.)$ denotes the Gamma function.

The reparameterization of the Skewed Student-t density in definition 1.3.3 is necessary for the GARCH models, which assume the innovation process to have zero mean and unit variance. There is a straightforward relation between the Student-t quantiles and the Skewed Student-t ones. Also, it is possible to notice that the density $f(x|1/l, \nu)$ is the mirror of $f(x|l, \nu)$ with respect to the zero mean, meaning $f(x|1/l, \nu) = f(x|l, \nu)$. Therefore the sign of $\log(l)$ indicates the direction of the skeweness: in fact, if $\log(l > 0)$ the third moment is positive and the density is skewed to the right; else, if $\log(l) < 0$ the third moment is negative and the density is skewed to the left (LAURENT, 2018).

1.4 Long memory in the conditional variance

1.4.1 Conditional heteroskedasticity: the GARCH model

1.4.1.1 The ARCH model

The conditional mean models are usually extended for financial time series to the GARCH family models, due to the fact that the ARFIMA models alone are insufficient to describe the dynamics of some time series. The conditional mean models discussed in the previous sections are the ARFIMA family proposed by Box and Jenknins, as in definition 1.1.14 and its long-memory counterpart ARFIMA or FARIMA, as in definition 1.2.4: in both cases, it was assumed that the innovation process was a Gaussian white noise with constant and finite variance (BERAN et al., 2013). This, however, might not always be necessarily true.

Especially when dealing with financial data, a stylized fact is that these data tend to show a pattern of volatility on clusters: it means that there are periods where the volatility is higher and periods where the volatility might not be so high (ZIVOT; WANG, 2005). To account for that, it is necessary to consider heteroskedastic models. As an alternative, Engle (1982) introduced the autoregressive conditional heteroskedasticity model, shortly henceforth referred to as ARCH. This is a which assumes that the series can be decomposed in two different parts: an innovation term with mean zero and unit variance and an autoregressive term.

Definition 1.4.1. *Let $\{r_t, t \in \mathbb{Z}\}$ be a real-valued time series. Let also $\epsilon_t \sim i.i.d\ D(0, 1)$, where $D(.)$ is a distribution with zero mean and unit variance. An ARCH(m) model is given by:*

$$\begin{aligned} y_t &= \mu + r_t \\ r_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2. \end{aligned} \tag{1.57}$$

A convenient approach would be to assume that $\epsilon_t \sim i.i.d\ \mathcal{N}(0,1)$, because it is simple to bring about convergence results and to write the log-likelihood function. The normal distribution however is not the only possibility: to account for stylized facts of financial time series, it can be used a Student- t distribution, for instance, or other distributions which better suit the data (ZIVOT; WANG, 2005).

The ARCH model was originally proposed to assess the Friedman Conjecture that the unpredictability of inflation was a primary cause of business cycles and after the original publication of Engle (1982) the ARCH model proved interesting for applications in financial econometrics (ANDERSEN et al., 2009).

The ARCH model allows the error terms to be uncorrelated, however dependent. From equation 1.57, one notes that, if the past squared shocks $\{\alpha_i\}_{i=1}^m$ are large, it implies that also the conditional variance σ_t^2 will be large, which is coherent to the idea that in financial markets, large shocks tend to be followed by large shocks, which is even more evident when dealing with intraday data (TSAY, 2005).

For more details regarding the ARCH(m) models, see Zivot and Wang (2005) and Tsay (2005) for identification and estimation procedures. Hamilton (1994) describes with details the procedure for quasi maximum-likelihood methods to estimate the ARCH(m) parameters.

1.4.1.2 The GARCH model

Tsay (2005) discussed a few weaknesses of the ARCH models: first, it is assumed that the volatility reacts the same way to random shocks, meaning they have the same effects, be positive or negative. Second, the model is rather restrictive in terms of conditions positivity of the conditional variance and finite fourth moments. Finally, ARCH models tend to overpredict the conditional variance, because of the slow response to isolated shocks in the time series. Another problem according to Zivot and Wang (2005) is that these models tend to be not much parsimonious, because sometimes high order m is necessary in the ARCH(m) models in order to fully treat all the serial autocorrelations of the series.

To account for the aforesaid weaknesses, other models were developed. An alternative approach is the GARCH model, which was proposed by Bollerslev (1986). It is an extension which allows a more parsimonious model to describe the serial correlations in the conditional variance. Nevertheless, according to Andersen et al. (2009), the unconditional ACF for the r_t^2 decays too rapidly compared to the typically observed in the real financial time series, to which the GARCH model is more adequate.

Definition 1.4.2. *Let $\{r_t, t \in \mathbb{Z}\}$ be a real-valued time series adjusted to have zero mean without loss of generality. Let also $\epsilon_t \sim i.i.d\ D(0,1)$, where $D(.)$ is a distribution with*

mean zero and unit variance. The process $\{r_t, t \in \mathbb{Z}\}$ *is a GARCH(m,n) process if:*

$$\begin{aligned} r_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2 + \sum_{j=1}^{n} \beta_j \sigma_{t-j}^2. \end{aligned} \tag{1.58}$$

where $\epsilon_t \sim D(0,1)$, $\sum_{i=0}^{\max(m,n)} \alpha_i + \beta_i < 1$, $\beta_j \geq 0$, $\alpha_i \geq 0$ *for all* $i = 1, \ldots m-1$ *and* $j = 1, \ldots n-1$, $\beta_n > 0$ *and* α_m.

From the equation 1.58 it is understood that $\alpha_i = 0, i > m$, $\beta_j = 0, j > n$. Sufficient condition for the positivity of conditional variance is that all of coefficients are positive (TSAY, 2005). Necessary and sufficient conditions for positivity the conditional variance and stationarity are provided by Nelson and Cao (1992).

Some points are worth to note on equation 1.58 and the corresponding. The sequence $\{\epsilon_t, t \in \mathbb{Z}\}$ is sometimes referred to as the driving noise sequence. Notice that the GARCH(m,0) is the ARCH(m) process, but there is no possibility for the GARCH(0,n) to exist alone, since the volatility σ_t would be disconnected from the observed process r_t and the model would be senseless (ANDERSEN et al., 2009).

To discuss further properties of GARCH, Tsay (2005) proposed the following representation of the model: let $\eta_t = r_t^2 - \sigma_t^2$ so that $\sigma_t^2 = r_t^2 - \eta_t$. From that, it is possible to rewrite the model in equation 1.58 as:

$$r_t^2 = \alpha_0 + \sum_{i=1}^{\max(m,n)} (\alpha_i + \beta_i) r_{t-i}^2 + \eta_t - \sum_{j=1}^{n} \beta_j \eta_{t-j}, \tag{1.59}$$

which is similar to an ARMA representation. The difference is that $\{\eta_t, t = 1, 2, \ldots T\}$ is a martingale difference as in definition 1.1.10. A result is that:

$$E(a_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{\max(m,s)} (\alpha_i + \beta_i)}, \tag{1.60}$$

whenever holds that $\sum_{i=0}^{\max(m,n)} \alpha_i + \beta_i < 1$. From the equation 1.58 it is possible to realize that large values of σ_{t-i}^2 and $r_{t-i}^2 t$ for some i, j will imply the volatility cluster pattern, as in ARCH models (TSAY, 2005).

Equation 1.58 can also be rewritten in a more compact form as:

$$\sigma_t^2 = \alpha + \alpha(B) r_t^2 + \beta(B) \sigma_t^2, \tag{1.61}$$

where B is the backshift operator from definition 1.1.13 and:

$$\alpha(B) = \sum_{i=1}^{m} \alpha_i B^i, \tag{1.62}$$

$$\beta(B) = \sum_{j=1}^{n} \beta_j B^j. \tag{1.63}$$

Recall the model as in equation 1.61. If all of the roots of the polynomial $|1-\beta(B)|=0$ lie outside the unit circle, then:

$$\sigma_t^2 = \alpha_0 \left(1 - \beta(B)\right)^{-1} + \alpha(B) \left(1 - \beta(B)\right)^{-1} r_t^2, \tag{1.64}$$

which might be seen as an ARCH(∞) representation of the GARCH model (LAURENT, 2018).

In order to evaluate the strict stationarity of an GARCH process, Bougerol and Picard (1992) propose the Markov GARCH matrix representation, which defines a first order vector autoregressive model, with a positive and i.i.d matrix coefficients. The result depends on this random matrix representation and on the spectral radius of an iteration of the sequence of matrices. For more details and proof for these theorems, see Francq and Zakoian (2019), who also provide several other remarks on stationarity and ergodicity of GARCH models.

For practical purposes one might consider the second-order stationarity conditions for a GARCH(m,n) model, as in Francq and Zakoian (2019).

Theorem 2. *If there exists a GARCH(m,n) model as in 1.58 which is second-order stationary then:*

$$\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{n} \beta_j < 1, \tag{1.65}$$

Moreover, if this holds, then the unique strictly stationary solution to 1.58 is a weak white noise.

For the proof, see Nelson and Cao (1992).

Another desirable feature is the causality of the GARCH(m,n) process, meaning that the volatility σ_t should depend only on the past observations of ϵ_t, not on future values. Andersen et. al. (2009) stated however that the strictly stationary solutions are automatically causal for GARCH processes. Bougerol and Picard (1992) proved also that a strictly stationary GARCH(m,n) process is also ergodic in the sense of definition 1.1.7.

The estimation in the case of the GARCH(m,n) depends on the ϵ_t term, which is very flexible in terms of which distributions can be used. To account for that, Laurent (2018) describes the log-likelihood function of the GARCH model considering several

different cases. Let $\boldsymbol{\theta}$ denote the vector of unknown parameters. For $\epsilon_t \sim \mathcal{N}(0,1)$ the log-likelihood is given by:

$$l(\boldsymbol{\theta}|r_1 \dots r_T) = -1/2 \sum_{t=1}^{T} \left(\log(2\pi) + \log(\sigma_t^2) + \epsilon_t^2 \right). \tag{1.66}$$

For $\epsilon_t \sim t_\nu$ is a Student t with ν degrees of freedom as in equation 1.55, then:

$$\begin{aligned} l(\boldsymbol{\theta}|r_1 \dots r_T) = T \Bigg(& \log \Gamma \left(\frac{\nu+1}{2} \right) - \Gamma \left(\frac{\nu}{2} \right) - \frac{\log(\pi(\nu-2))}{2} + \\ & -\frac{1}{2} \sum_{t=1}^{T} \left[\log(\sigma_t^2) + (1+\nu) \log \left(1 + \frac{\epsilon_t^2}{\nu-2} \right) \right] \Bigg). \end{aligned} \tag{1.67}$$

Assuming ϵ_t to follow a generalized errors distribution as in definition 1.3.1 with $0 < \nu < \infty$, then:

$$l(\boldsymbol{\theta}) = \sum_{t=1}^{T} \left[\log \left(\frac{\nu}{\lambda_\nu} \right) - 0.5 \left| \frac{\epsilon_t}{\lambda_\nu} \right| - (1+\nu^{-1}) \log 2 - \log \Gamma \left(\frac{1}{\nu} \right) - 0.5 \log \sigma_t^2 \right], \tag{1.68}$$

where:

$$\lambda_\nu = \sqrt{\frac{\Gamma(1/\nu) 2^{-2\nu}}{\Gamma(\nu/3)}}.$$

And finally one can consider ϵ_t to follow a skewed Student t as in the parameterization of Lambert and Laurent (2000) in definition 1.3.3. The log-likelihood then becomes:

$$\begin{aligned} l(\boldsymbol{\theta}) = T \Bigg[& \log \Gamma \left(\frac{\nu+1}{2} \right) - \log \Gamma \left(\frac{\nu}{2} \right) - 0.5 \log(\pi(\nu-2)) + \log \left(\frac{2}{\xi + 1/\xi} \right) + \log(s) \\ & - 0.5 \sum_{t=1}^{T} \left\{ \log \sigma_t^2 + (1+\nu) \log \left[1 + \frac{(s\epsilon_t + m)^2}{\nu-2} \xi^{-2I_t} \right] \right\} \Bigg], \end{aligned} \tag{1.69}$$

where:

$$I_t = \begin{cases} 1 \text{ if } z \geq -m/s \\ -1 \text{ if } z < -m/s \end{cases}, \tag{1.70}$$

and ξ is the asymmetry parameter, ν is the degrees of freedom of the distribution and:

$$m = \frac{\Gamma(\frac{\nu-1}{2}) \sqrt{\nu-2}}{\sqrt{\pi} \Gamma(\nu/2)} \left(\xi - \frac{1}{\xi} \right), \tag{1.71}$$

and:

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}. \tag{1.72}$$

Although the GARCH(m,n) is rather general and the properties hold for $m, n \in \mathbb{Z}_+$, the most used model in practical applications for financial time series is the GARCH(1,1) model, which has been widely discussed (ANDERSEN et al., 2009). The previous theorems hold for $m = n = 1$, but the GARCH(1,1) has more specific properties, which are shown by Tsay (2005) and Bauwens, Hafner and Laurent (2012).

Whenever $\sum_{i=1}^{m} \alpha_i + \sum_{j=1}^{n} \beta_j = 1$ it results on a particularly interesting case, which is referred to as IGARCH model, or integrated GARCH, as proposed by Engle and Bollerslev (1986). This is a particular case in which the AR representation of a GARCH model has a unit root and, similarly to the ARIMA model, the impact of the squared shocks is persistent in the conditional variance (TSAY, 2005). The IGARCH model has a unique strictly stationary solution (with infinite variance) if the distribution of η_t has an unbounded support and no mass at 0 (FRANCQ; ZAKOIAN, 2019). For more details, see Tsay (2005).

1.4.1.3 Short memory extensions of the GARCH model

There are several extensions to the GARCH model, which account for several features of financial time series, or stylized facts. One of them was proposed by Nelson (1991) and is called the Exponential GARCH, also referred to shortly as EGARCH. This model is more flexible in the sense that it has less strict positivity constraints and allows what is called the "asymmetrical leverage effects in returns" (ZIVOT; WANG, 2005). Consider then the next definition:

Definition 1.4.3. *Let $\{\epsilon_t, t = 1, 2, \ldots T\}$ such that $\epsilon_t \sim D(0, 1)$ for a given distribution $D(.)$. Then r_t is called an EGARCH(m,n) model if $r_t = \sigma_t \epsilon_t$:*

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^{n} \alpha_i g(\epsilon_{t-i}) + \sum_{j=1}^{m} \beta_j \ln \sigma_{t-j}^2, \tag{1.73}$$

where $\sigma_t > 0$ and:

$$g(\epsilon_{t-i}) = \theta \epsilon_{t-i} + \gamma \left(|\epsilon_{t-i}| - E\left(|\epsilon_{t-i}| \right) \right), \tag{1.74}$$

and $\alpha_i, \beta_i, \gamma, \theta \in \mathbb{R}$ are parameters to be estimated.

Some other remarks are useful regarding the EGARCH model. Since the model considers log-volatility, the positivity of all coefficients is a constraint which can be relaxed in order to achieve positivity of the conditional variance. It is also worth to note that:

$$\sigma_t^2 = e^{\alpha_0} \prod_{i=1}^{n} \exp\left\{\alpha_i g(\epsilon_{t-i})\right\} \prod_{j=1}^{m} \left\{\sigma_{t-j}^2\right\}^{\beta_j}, \tag{1.75}$$

which produces a volatility model with the EGARCH which has a multiplicative dynamics. Again in equation 1.75 the positivity of α_i and β_j are not necessary for positivity of the conditional variance, since the exponents can be either positive or negative in the equation (FRANCQ; ZAKOIAN, 2019).

Black (1976) examined what is called in the financial literature by the "asymmetrical leverage effect", which stated that it is expected for the negative shocks to have a larger effect in the return series than the positive shocks. An interpretation to this, according to Zivot and Wang (2005), is that "bad news" tend to have a higher impact in financial markets than "good news". The EGARCH model incorporates this feature, and the asymmetrical leverage effect on log-returns can be seen clearly when (TSAY, 2005):

$$g(\epsilon_t) = \begin{cases} (\theta + \gamma)\epsilon_t - \gamma(E|\epsilon_t|) \text{ if } \epsilon_t \geq 0 \\ (\theta - \gamma)\epsilon_t - \gamma(E|\epsilon_t|) \text{ if } \epsilon_t < 0 \end{cases}. \tag{1.76}$$

For mature markets, it is expected for the γ parameter in equation 1.76 to be negative. However, this might not necessarily be true, especially in not so mature financial markets or during "anomalous" periods. For a theoretical discussion of this issue, see Black (1976) and Figlewski and Wang (1976).

Proposition 5. *Let $\{\epsilon_t, t \in \mathbb{Z}\}$ be a sequence of i.i.d r.v. with $E(|\epsilon_1|) < \infty$. Let $\{g(\epsilon_t), t \in \mathbb{Z}\}$ be defined as in equation 1.76 where $\gamma \neq 0$ and $\theta \neq 0$. Then $\{g(\epsilon_t), t \in \mathbb{Z}\}$ is a strictly stationary and ergodic process. If $E(\epsilon_1^2) < \infty$, then the process is a white noise with zero mean and constant variance which depend on γ and θ.*

For the proof and details in proposition 5, see Lopes and Prass (2013).

A difference from the EGARCH to the standard GARCH is that the conditional variance is written as a function of the past standardized innovations (i.e. divided by the standardized residuals) instead of the past innovations. This happens because $g(\epsilon_t)$ is a strong white noise with (FRANCQ; ZAKOIAN, 2019):

$$Var(g(\epsilon_t)) = \theta^2 + \gamma^2 Var(|\epsilon_t|) + 2\theta Cov(\epsilon_t, |\epsilon_t|). \tag{1.77}$$

The stationarity condition for the EGARCH process is provided by the following theorem:

Theorem 3. *Assume $g(\epsilon_t)$ is not almost surely equal to 0 and and that the polynomials $\alpha(z) = \sum_{i=1}^{n} \alpha_i z^i$ and $\beta(z) = 1 - \sum_{j=1}^{m} \beta_j z^j$ have no common or cancelling factors and $\alpha(z)$ not null. Under these conditions, the EGARCH process r_t has a strictly stationary solution and non-anticipative solution if the roots in $\beta(z)$ lie outside the unit circle.*

For the proof, see Franq and Zakoian (2019).

A more complicated feature of an EGARCH model is the invertibility. The proof for the invertibility conditions for an EGARCH(1,1) is due to Straumann and Mikosch (2006), and it usually gets more sophisticated to evaluate invertibility as the EGARCH orders increase. In practice, Franq and Zakoian (2019) state that the invertibility of an EGARCH model is important when fitting the model to real time series, because if it holds then one can recover the volatility from the past observations.

There are several other extensions for the GARCH models which are not discussed in this chapter, because they will not be used in the practical applications in the further chapters. For example, there are extensions such as APARCH, TARCH, GJR-GARCH, X-GARCH, M-GARCH and so on. For a comprehensive discussion of the underlying aspects of the extensions, see Franq and Zakoian (2005), Tsay (2005) and Andersen et. al. (2009).

1.4.2 Long memory extensions of the GARCH model

Another feature for financial time series is that they often show a pattern of long memory in the conditional variance (ZIVOT; WANG, 2005). As it happens with the conditional mean models as described previously, the so-called self-similarity can also be identified in the conditional variance, by evaluating the ACF and the PACF of the squared log-returns. The long memory GARCH processes are extensions to these long memory processes in the conditional mean, which allows one to describe the self-similarity pattern of behaviour sometimes noticed in the conditional variance of financial time series (BAUWENS; HAFNER; LAURENT, 2012).

All the previous ARCH, GARCH and EGARCH models are short memory processes, as stated by Lopes and Prass (2013). Due to the aforementioned condition, Baille, Bollerslev and Mikkelsen (1996) proposed the Fractionally Integrated GARCH, or also FIGARCH, which can be extended for a FIEGARCH when modelling $\log \sigma_t^2$ proposed by Bollerslev and Mikkelsen (1996).

Consider once more the GARCH model as:

$$\begin{aligned} r_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{m} \alpha_i r_{t-i}^2 + \sum_{j=1}^{n} \beta_j \sigma_{t-j}^2, \end{aligned} \tag{1.78}$$

which can be written in the ARMA form using the backshift operator of definition 1.1.13:

$$\Phi(B)r_t^2 = a_0 + \beta(B)u_t, \tag{1.79}$$

where: $u_t = r_t^2 - \sigma_t^2$ and it holds:

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_m B^m, \tag{1.80}$$

$$\beta(B) = 1 - \beta_1 B - \beta_2 B^2 - \ldots - \beta_q B^q, \tag{1.81}$$

with $q = \max(m, n)$ and $\phi_i = \alpha_i + \beta_i$.

In fact, the equation 1.79 is an ARMA(m,q) process in terms of r_t^2 and u_t. However in this case, the difference from this ARMA model defined in equation 1.79 with that defined by Box and Jenkins (1970), as in definition 1.1.14, is that the disturbance term u_t is a martingale difference sequence, in the sense of definition 1.1.10, whereas in def. 1.1.14 the disturbance term is a white noise (ZIVOT; WANG, 2005). That stated, the process u_t can be understood as an innovation term.

Intuitively, the next step for considering the Fractionally Integrated GARCH is to apply the fractional difference operator of order d to the ARMA representation of the GARCH process as in equation 1.79:

$$\Phi(B)(1 - B)^d r_t^2 = a + \beta(B)u_t. \tag{1.82}$$

Equation 1.82 is an ARFIMA(m,d,q) process. Apart from having that the roots of polynomials $\Phi(z) = 0$ and $\beta(z) = 0$ lie out of the unit circle, it is also necessary for $0 \leq d \leq 1$ the process $(1 - B)^d r_t^2$ follows a stationary ARMA(m,q) process. In fact, more specifically, the process is strictly stationary and ergodic if $0 < d < 1$ and it is second order stationary if $|d| < 0.5$ (XEKALAKI; DEGIANNAKIS, 2010).

Moreover, if $d = 1$, the model is simply the IGARCH presented in the previous section, and if $d = 0$, the model is the conventional GARCH(m,n) discussed during the previous section extensively.

In details from the previous steps in the representation of the FIGARCH(m,d,n), the model for σ_t^2 will lead to the following definition (BAILLE; BOLLERSLEV; MIKKELSEN, 1996):

Definition 1.4.4. *Let* $\{r_t, t \in \mathbb{Z}\}$ *be a stochastic process defined as* $r_t = \epsilon_t \sigma_t$*, in which* $\epsilon_t \sim D(0,1)$ *and:*

$$\left(1 - \sum_{j=1}^{n} \beta_j B^j\right) \sigma_t^2 = \alpha + \left(1 - \sum_{j=1}^{n} \beta_j B^j - \left(1 - \sum_{i=1}^{m} \alpha_i B^i\right)(1-B)^d\right) r_t^2, \tag{1.83}$$

where $0 < d < 1$ *and* B *is the backshift operator defined in 1.1.13 and* $D(0,1)$ *is a distribution with mean zero and unit variance. Then the above process* $\{r_t, t \in \mathbb{Z}\}$ *is called a Fractionally Integrated GARCH, shortly denoted FIGARCH(m,d,n).*

The FIGARCH equation in definition 1.4.4 can be rewritten, still in terms of the conditional variance, in a more compact form:

$$\sigma_t^2 = \alpha(1 - \beta(B))^{-1} + \left(1 - (1 - \beta(B))^{-1} \Phi(B)(1-B)^d\right) r_t^2, \tag{1.84}$$

$$\sigma_t^2 = \alpha(1 - \beta(B))^{-1} + \left(1 - (1 - \beta(B))^{-1} [1 - \alpha(B) - \beta(B)](1-B)^d\right) r_t^2, \tag{1.85}$$

or also:

$$\sigma_t^2 = \alpha(1 - \beta(B))^{-1} + \lambda(B) r_t^2, \tag{1.86}$$

using the backshift operator of definition 1.1.13, where $\alpha(B) = 1 - \sum_{i=1}^{m} \alpha_i B^i$ and $\beta(B) = 1 - \sum_{j=1}^{n} \beta_j B^j$ and $\lambda(B) = 1 - \sum_{k=1}^{\infty} B^k \lambda_k = 1 - (1 - \beta(B))^{-1} \Phi(B)(1-B)^d$. As in the GARCH case, $D(.)$ is usually the normal or Student t distribution, but this can be extended to a Skewed-t or to the GED (LAURENT, 2018).

Although complicated to show, Baille, Bollserlev and Mikkelsen (1996) claim that, even if the process FIGARCH in equation 1.85 is not second-order stationary, it is ergodic and strictly stationary.

For the representation in equation 1.86, Lopes and Mendes (2006) show the following proposition:

Proposition 6. *Let* $\{r_t, t \in \mathbb{Z}\}$ *be a real-valued FIGARCH(m,d,n) process for* $d \in [0,1]$ *as in definition 1.4.4 and 1.86. Then the coefficients* λ_k *for* $k \in \mathbb{Z}_+$ *are given by* $\lambda_0 = 0$ *and:*

$$\lambda_n = \sum_{i=1}^{m} \beta_i \lambda_{n-i} + \alpha_n + \delta_{d,n} - \sum_{j=1}^{\max(m,n)} \gamma_j \delta_{d,n-j} \ , \ \textit{if } 1 \leq n \leq r, \tag{1.87}$$

$$\lambda_n = \sum_{i=1}^{m} \beta_i \lambda_{n-i} + \delta_{d,n} - \sum_{j=1}^{\max(m,n)} \gamma_j \delta_{d,n-j} \ , \ \textit{if } 1 \leq n \leq r, \tag{1.88}$$

where:

$$\gamma_j = \begin{cases} \alpha_j, & \text{if } r > s \\ \alpha_j + \beta_j, & \text{if } r = s \\ \beta_j, & \text{if } r < s \end{cases}, \tag{1.89}$$

and the coefficient $\delta_{d,n}$ *is such that* $\delta_{d,n} = d(\Gamma(k-d))/(\Gamma(k+1)\Gamma(1-d))$, $\delta_{d,0} = 1$ *and for all* $k \geq 1$ *obtained iteratively as:*

$$\delta_{d,n} = \delta_{d,n-1}\left(\frac{k-1-d}{k}\right), \tag{1.90}$$

Proposition 6 is of great importance, because from it, one is able to characterize any FIGARCH(m,d,n) process and to obtain the coefficients λ_k in an iterative procedure (LOPES; MENDES, 2006).

The structure of definition 1.4.4 results in a memory structure which is particular for the long memory variance processes, evaluating the impulse response function. Indeed, the impact of a random shock will lie somewhere in between the typical exponential decay from a GARCH process and the infinite persistence of the IGARCH process (CAPORIN, 2002).

As stated by Xekalaki and Degiannakis (2010), in contrast to the IGARCH or to the GARCH model, where respectively either the shocks are indefinitely persistent or exponentially dissipated, the FIGARCH model provides an intermediate behaviour where the response of the conditional variance to the past shocks decay at an hyperbolic rate, with reasonable persistence. Meaning: the hyperbolic decay of the ACF of the series as described by Beran (2013) would be expected in the ACF or PACF of the squared series whenever a FIGARCH model is the data generating process of the series.

In that case, holding that the parameter $d \in (0,1)$ in equation 1.85, the polynomials $\Phi(B)$ and $\beta(B)$ model the short-run dynamics of the volatility, whereas the d parameter will describe the long-run behaviour of the volatility (ZIVOT; WANG, 2005).

Francq and Zakoian (2019) proved that a condition for the existence of the FIGARCH when $d \in (0,1)$, even as $d \to 1$, is that the distribution of ϵ^2 is non-degenerate.

Another interesting feature is that the memory parameter operates in a different fashion for the FIGARCH process than for the ARFIMA: as $d \to 0$, the memory increases, since the long memory parameter acts directly on the squared errors, not on σ_t^2. This changes the interpretation of the effect of a random shock in a FIGARCH model: it is, in

fact, the innovation in the ARFIMA representation of the FIGARCH process:

$$r_t^2 = \alpha + \beta(B)\left((1-B)^{-d}\Phi(B)\right)^{-1} u_t. \tag{1.91}$$

Because of the formulation in 1.91, this can be interpreted as an unexpected volatility of the variance. This happens since the r_t^2 can be a proxy for the conditional variance and the status at the time t depending on $t-1$ may be viewed as a one-step ahead for the conditional variance (CAPORIN, 2002).

The stationary solution of the FIGARCH model can be rather complicated, and necessary and sufficient conditions for the positivity of the conditional variance are not straightforward. Some questions on strict stationarity still remain open. For other results of FIGARCH models, see Caporin (2002) and Zaffaroni (2000).

As it happened with the EGARCH model, the FIGARCH does not incorporate the asymmetric leverage effects which are sometimes observed in financial time series. To account for that, the FIEGARCH(m,d,n) model was proposed. The same logical procedure can be applied, writing the EGARCH as an ARMA process and applying the fractional difference operator. It yields to the definition below (BOLLERSLEV; MIKKELSEN, 1996):

Definition 1.4.5. *Let $\{r_t, t \in \mathbb{Z}\}$ be a stochastic process defined as $r_t = \epsilon_t \sigma_t$, in which:*

$$\log \sigma_t^2 = \alpha_0 + \frac{1-\sum_{i=1}^{m}\alpha_i B^i}{1-\sum_{j=1}^{n}\beta_j B^j}(1-B)^{-d}\left(\theta\epsilon_{t-1} + \gamma(|\epsilon_{t-1}| - E(|\epsilon_{t-1}|))\right), \tag{1.92}$$

where $\epsilon_t = r_t/\sigma_t$ is the standardized errors and B is the backshift operator and $0 < d < 1$. Then the process $\{r_t, t \in \mathbb{Z}\}$ is a fractionally integrated EGARCH, shortly denoted FIEGARCH(m,d,n).

One can also write the equation of definition 1.4.5 in a more compact form using the backshift operator of definition 1.1.13 as:

$$\log(\sigma_t^2) = \alpha_0 + \frac{\alpha(B)}{\beta(B)}(1-B)^{-d} g\left(\epsilon_{t-1}\right). \tag{1.93}$$

As in the EGARCH model, the FIEGARCH model always has positive conditional variance. It is assumed, as for the FIGARCH, that the polynomials $\alpha(B)$ and $\beta(B)$ have no common roots nor cancelling factors. Also, for the FIEGARCH model in definition 1.4.5 to be well defined, it is necessary that $\beta(z) \neq 0$ whenever $|z| \leq 1$ (LOPES; PRASS, 2013).

There are other parameterizations for the FIEGARCH model, which are in a certain sense equivalent. For instance, the one of Laurent (2018) for the OxMetrics-G@RCH^{TM}, described in his OxMetrics-G@RCH^{TM} manual, and the parameterization of Zivot and Wang (2005) which is used in S+FinMetrics.

An example for the FIGARCH model is the FIGARCH(1,d,1), which is given by:

$$\sigma_t^2 = \alpha(1 - B\beta_1)^{-1} + (1 - (1 - B\beta_1)^{-1}(1 - B\phi_1)(1 - B)^d)r_t^2. \tag{1.94}$$

For this particular model, apart from $\alpha > 0$, it is shown that the coefficients in equation 1.94 must hold some properties to ensure the positivity of the conditional variance (LOPES; MENDES, 2006):

$$\beta_1 \leq \phi_1 \leq \frac{2-d}{3}, \tag{1.95}$$

$$d\left(\phi_1 - \frac{1-d}{2}\right) \leq \beta_1(d + \alpha_1), \tag{1.96}$$

where ϕ_1 comes from the polynomial representation $\Phi(B) = 1 - B\phi_1$. For the FIGARCH(0,d,0), the only restriction for positivity will be $\alpha > 0$ a $0 \leq d \leq 1$. For the FIGARCH(1,d,0), Baille, Bollerslev and Mikkelsen (1996) show that it is necessary and sufficient that $0 \leq \beta_1 < d \leq 1$ for the conditional variance to be positive.

Several other different theoretical results not only on stationarity and ergodicity of the FIEGARCH model, but in general for these models are proved by Lopes and Prass (2013). Conrad and Haag (2006) also have discussed inequality constraints for the FIGARCH(p,d,q) models for $p \leq 2$. An interesting result is that even if all parameters are non-negative, the conditional variance can be negative, whereas also negative parameters (except for d) can generate a model with positive conditional variance almost surely (CONRAD; HAAG, 2006).

It is possible to combine ARIMA and ARFIMA models with the GARCH family models. The following example is due to Palma (2007). The ARFIMA(p,d,q)-GARCH(1,1) model is given by:

$$\begin{aligned} y_t &= \sum_{j=0}^{\infty} \psi_j r_{t-j}, \\ r_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2, \end{aligned} \tag{1.97}$$

where $\sum_j \psi_j B^j = \Psi(B) = \Phi^{-1}(B)\Theta(B)(1-B)^{-d}$ and $\epsilon_t \sim \mathcal{N}(0,1)$. This would be similar if one wanted to consider a FIGARCH/FIEGARCH model, considering their equation.

1.5 Numeric Methods for Optimization

As the model incorporates features such as long memory and conditional heteroskedasticity, the log-likelihood tends to become rather complicated, and the surface

for optimization can get rough (LAURENT, 2018). Due to that, it is necessary to use numerical methods to optimize the log-likelihood functions and find the estimated parameters and the Hessian matrix. Some of the numerical methods widely used are BFGS method and Simulated Annealing.

BFGS is a quasi-Newton algorithm, which is gradient-based and it tends to be computationally fast. There are alternatives such as BFGS with bounds, which considers bounds for the parameter space. The algorithm is robust and converges quadratically with a smaller cost per iteration than the Newton's method (NOCEDAL; WRIGHT, 2006). The BFGS method is used in OxMetrics-G@RCH^{TM} and is implemented in Ox Language as a default to optimize the log-likelihood function of ARFIMA-FI(E)GARCH models, and the software implementation for that is described by Laurent (2018).

Simulated-annealing is a stochastic global optimization method which uses only the function values in order to optimize, but it is relatively slower than BFGS. It is generally an alternative when functions are not differentiable and the optimization surface is rough. The simulated annealing method can potentially distinguish different local optima. There are several implementations for this method, but the one which will be used follows exactly the description made by Goffe, Ferrier and Gary (1994), implemented in Ox language by Charles Bos and with software implementation instruction provided by Laurent (2018). Another alternative Simulated Annealing method is described by Belisle (1992).

For more details regarding these methods, software implementation and technical issues see Nocedal and Wright (2006).

There are other methods computationally implemented for numerical optimization in other programming languages. For instance, in R one can consider the "rugarch" package, developed and maintained by Ghalanos (2020), which includes the SOLNP and GOSOLNP, which are numerical optimization algorithms for general non-linear optimization using the augmented Lagrange Multiplier Method.

The original SOLNP algorithm was developed by Ye (1987), whereas the implementation for this algorithm and other extensions in R are performed by Ghalanos and Theussl (2015). The thesis published by Ye (1987) provides a method to solve a general nonlinear problem and was originally implemented for MatLab. An extension for the SOLNP, referred to as GOSOLNP, with multiple restarts and random initialization is implemented by Ghalanos and Theussl (2015) in R, based on Ye (1987) and Hu, Shonkwiler and Spruill (1994). The main advantage of GOSOLNP over SOLNP is that the algorithm is reinitialized several times, in order to avoid getting stuck in local mininma or maxima, which is useful especially when the objective function is non-smooth or has many local minima.

Another method implemented in R is the Subplex routine (abbreviated in R as

SBPLX) based on Rowan (1990), which is a variant of Nelder-Mead and seems to be much more efficient than the Robust Nelder-Mead algorithm and can also be used in rough surfaces for optimization with constraints. The rugarch package for R by Ghalanos (2020) allows one to use the Subplex subroutine to estimate the coefficients for GARCH models. For more details, see Rowan (1990).

2 Wavelet Methods for time Series

> Many things of interest to mathematicians or engineers are very complex and have several complicated structures layered on top of one another. As we do in other things in life, we like to take things apart into more elementary building blocks so we can unravel this rich structure.
>
> Ingrid Daubechies

Wavelts are alternative to other base systems that can represent functions of class $\mathcal{L}_2(\mathbb{R})$ (as in definition 1.1.1), such as orthogonal polynomials, sines and cosines, Walsh-Fourier and others. The main idea is the same as in Fourier analysis, to approximate functions as a linear combination of wavelets, which may not necessarily form an orthogonal set. The development of wavelets can be linked to Haar's work (1911), but it was with Morlet (1981), Meyer (1990) and Daubechies (1992) and in the decade of 1980's that the wavelet theory developed significantly (MEYER, 1993). Vidakovic (1999), Nason (2008) and Mallat (2009) developed wavelets applications in time series analysis and signal processing. In Portuguese, Morettin (2014) is also a reference for applications in time series.

This chapter consists four sections. In the first, the fundamental aspects of the wavelet theory are introduced, focusing on the applications for time series analysis and signal processing. A second section introduces some examples of wavelet orthogonal families, which will be used in the applications of the further chapters. A third section describes the method known as wavelet thresholding or also wavelet smoothing, which is one of the main applications of wavelets in signal processing and in non-parametric estimation. The fourth section introduces an example, based on Bruce and Gao (1996) and Nason (2008), of wavelet decomposition, multiresolution analysis and wavelet thresholding.

2.1 Wavelet Theory

Wavelets are functions that satisfy properties such that they can form a bases to represent all function in $\mathcal{L}_2(\mathbb{R})$. Wavelets need not necessarily to form an orthogonal

bases. One advantage in the usage of wavelets with respect to other bases is that wavelets are localized in time and scale, meaning that a discontinuity in a function $f(t)$ will only influence the wavelets that are near it and only those wavelet coefficients associated to the wavelets influenced that overlaps the discontinuity. This is different from the Fourier bases, which is localized in frequency, but not in time, meaning that a discontinuity in $f(t)$ might influence the whole Fourier decomposition. Wavelet transform is also efficient, because the computational complexity related to the Discrete Wavelet Transform is $O(n)$, whereas the one related to the Fast Fourier Transform is of order $O(n \log n)$ (NASON, 2008).

The wavelet bases are generated from binary re-scales and translations of one function, called thus mother wavelet, or wavelet function. Formally, let $\psi_{a,b}(t)$, $a \in \mathbb{R} \setminus \{0\}, b \in \mathbb{R}$ be a family of functions defined as translations and re-scales of a single mother function $\psi(t) \in \mathcal{L}_2(\mathbb{R})$, which is given by:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left(\frac{t-b}{a} \right). \tag{2.1}$$

The function ψ is called mother wavelet or wavelet function if it fulfills the admissibility condition:

$$C_\Psi = \int_{\mathbb{R}} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty, \tag{2.2}$$

where $\Psi(\omega)$ is the Fourier transformation of $\psi(t)$ (VIDAKOVIC, 1999).

The following parameterization with $a = 2^{-j}, b = k2^{-j}, k, k \in \mathbb{Z}$ is the one which will be adopted for discretization of continuous wavelet transform, choosing discrete values of a, b to still have an invertible transformation. Yet, other choices of a, b can be made, leading to different transformations (such as the non-stationary discrete wavelet transform) (VIDAKOVIC, 1999). The equation 2.1 becomes then:

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), j, k \in \mathbb{Z}. \tag{2.3}$$

Another possible way to write the parameterization is $a = 2^j$ and $b = 2k^j$, $j, k \in \mathbb{Z}$, which brings about almost the same results in the wavelet decomposition of a signal. The difference is in which levels correspond to the highest/lowest detail orders.

Wavelets may form a non-orthogonal bases of $\mathcal{L}_2(\mathbb{R})$, however the orthogonal wavelets are preferred for the particular application in time series analysis which will be developed in the further chapters, due to the fact that it is possible for one to reconstruct the original signal from the wavelet transform coefficients. An orthogonal wavelet transform such as Fourier transform is concise and each coefficient is calculated as the

inner product of the signal and the corresponding function in the wavelet bases (MORETTIN, 2014). Using the parameterization in equation 2.3, the set produced by all re-scaling and translations of the mother function form an orthogonal bases (VIDAKOVIC, 1999).

On what concerns the following properties, an orthonormal wavelet bases will be considered, generated by a mother function ψ:

$$\{\psi_{j,k}(t)\}_{j,k\in\mathbb{Z}}, \tag{2.4}$$

meaning that

$$\langle\psi_{j,k}, \psi_{l,m}\rangle = \delta_{j,l}\delta_{k,m}, \forall j,k,l,m \in \mathbb{Z},$$

such that for any $f \in \mathcal{L}_2(\mathbb{R})$ it holds:

$$f(t) = \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} c_{j,k}\psi_{j,k}(t). \tag{2.5}$$

Convergence in equation 2.5 is to be understood as convergence in the square mean, as in definition 1.1.8. Equation 2.5 is called a series of wavelets in $f(t)$ and wavelet coefficients are given by:

$$c_{j,k} = \langle f, \psi_{j,k}\rangle = \int_{-\infty}^{\infty} f(t)\psi_{j,k}dt. \tag{2.6}$$

Proposition 7. *Let ψ and $\psi_{j,k}$ be as in equation 2.3 and form an orthonormal bases of $\mathcal{L}_2(\mathbb{R})$. Then following properties hold:*

1. $\int_{-\infty}^{\infty}\psi(t)dt = 0$;
2. $\int_{-\infty}^{\infty}|\psi(t)|dt < \infty$;
3. $C_\psi < \infty$ *as in 2.2;*
4. $\int_{-\infty}^{\infty}|\psi(t)|^2dt = 1$ *or analogously* $\int_{-\infty}^{\infty}|\Psi(\omega)|^2d\omega = 2\pi$, *where* $\Psi(\omega)$ *is the Fourier transform of* $\psi(t)$, *as in equation 2.2;*
5. *For a given* r, *the first* $r-1$ *moments must be null, i.e.:*

$$\int_{-\infty}^{\infty} t^j\psi(t)dt = 0, j = 0, 1, \ldots r-1,$$

and also:

$$\int_{-\infty}^{\infty}|t^r\psi(t)|dt < \infty.$$

Definition 2.1.1. *Let $V_n, n \in \mathbb{Z}$ be a sequence of closed subspaces in $\mathcal{L}_2(\mathbb{R})$. They form a multiresolution analysis (MRA) if holds that:*

$$\ldots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \ldots, \tag{2.7}$$

$$\mathcal{L}_2(\mathbb{R}) = \overline{\bigcup_{j=-\infty}^{\infty} V_j}, \tag{2.8}$$

$$\bigcap_{j=-\infty}^{\infty} V_j = \{0\}. \tag{2.9}$$

It is: sequence of subspaces forms a MRA if they lie in an containment hierarchy, have a trivial intersection and an union that is dense in the space of all square integrable functions.

The containment hierarchy in equation 2.7 is constructed in such way that the V-spaces are self-similar, meaning:

$$f(2^j t) \in V_j \Leftrightarrow f(t) \in V_0, \tag{2.10}$$

and that there exists a function $\phi(t) \in V_0$ such that integer-translates spans the space V_0, meaning:

$$V_0 = \{f \in \mathcal{L}_2(\mathbb{R}) | f(t) = \sum_k c_k \phi(t-k)\}, \tag{2.11}$$

and for which the set $\{\phi(t-k), k \in \mathbb{Z}\}$ is an orthonormal bases.

Assuming $\int_{\mathbb{R}} \phi(t)dt \neq 0$, since $V_0 \subset V_1$, the scaling function $\phi(t) \in V_0$ can be represented as:

$$\phi(t) = \sum_{k \in \mathbb{Z}} h_k \sqrt{2} \phi(2t-k), \tag{2.12}$$

for $h_k, k \in \mathbb{Z}$ coefficients. The previous equation 2.12 is called dilation equation, and it is used to generate the mother wavelet function $\psi(t)$. The vector $h = \{h_n, n \in \mathbb{Z}\}$ is called wavelet filter. Its coefficients h_n which show in equation 2.12 of the scaling function represent in fact a low-pass filter. The filters are normalized, so that $\sum_k h_k = \sqrt{2}$ and they are orthogonal, in the sense that $\sum_k h_k h_{k-2l} = \delta_l$ (VIDAKOVIC, 1999).

Also, due to self-similarity, the set $\{\phi_{j,k} = 2^{j/2}\phi(2^j t - k)\}$ is a bases of V_j. Whenever the sequence of subspaces satisfies the MRA properties, there exists an orthonormal bases for $\mathcal{L}_2(\mathbb{R})$:

$$\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), j, k \in \mathbb{Z}\}, \tag{2.13}$$

such that the set in equation 2.13 is an orthonormal bases for a set W_j, satisfying for $j \in \mathbb{Z}$:

$$V_{j+1} = V_j \bigoplus W_j, \tag{2.14}$$

that is: W_j is the orthogonal complement of V_j in the space V_{j+1}.

Because $W_0 \subset V_1$, it can be represented, for some coefficients $g_k, k \in \mathbb{Z}$ as:

$$\psi_{j,k}(t) = \sum_{k \in \mathbb{Z}} g_k \sqrt{2} \phi(2t - k), \tag{2.15}$$

where the coefficient set $g = \{g_k,\ k \in \mathbb{Z}\}$ is a high-pass filter related to h_k by $g_k = (-1)^k h_{1-k}$. Both h and g form together what is know as the "quadrature mirror filters", and the relation between them is known as the "quadrature mirror relation" in the signal processing literature (VIDAKOVIC, 1999). The following theorems summarize the aforementioned discussion:

Theorem 4. *Let $\{V_j\}_{j \in \mathbb{Z}}$ be a multiresolution analysis with scaling function $\phi(t)$. Then for any $j \in \mathbb{Z}$, the set:*

$$\{\phi_{j,k}(t) = 2^{j/2} \phi(2^t - k), k \in \mathbb{Z}\}, \tag{2.16}$$

forms an orthonormal bases for V_j.

Theorem 5. *Let $\{V_j, j \in \mathbb{Z}\}$ be a multiresolution analysis with scaling function $\phi(t)$. Then the relations below hold:*

$$\phi(t) = \sum_k h_k \phi(2t - k),$$

where:

$$h_k = 2 \int_{-\infty}^{\infty} \phi(t) \overline{\phi(2t - k)} dt.$$

In addition:

$$\phi(2^{j-1} t - l) = \sum_k h_{k-2l} \phi(2^t - k).$$

Theorem 6. *Let $\{V_j, j \in \mathbb{Z}\}$ be a multiresolution analysis with scaling function:*

$$\phi(t) = \sum_k h_k \phi(2t - k), \tag{2.17}$$

with coefficients h_k as given in the theorem 5. Let also W_j be the span of $\{\psi(2^j t - k), k \in \mathbb{Z}\}$, where:

$$\psi(t) = \sum_k (-1)^k \overline{h_{1-k}} \phi(2t - k). \tag{2.18}$$

Then $W_j \subset V_{j+1}$ is the orthogonal complement of V_j in V_{j+1}. In addition, $\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k\}_{k\in\mathbb{Z}}$ is an orthonormal bases for W_j.

A proof for the previous theorems can be found within Vidakovic (1999), or also a straightforward proof can be found within Bogges and Narcowich (2009).

It follows from equation 2.14 that:

$$V_j = \bigoplus_{k=-\infty}^{j-1} W_k, \tag{2.19}$$

so that the wavelets set $\{\psi_{j,k}, j, k \in \mathbb{Z}\}$ form an orthogonal bases of $\mathcal{L}_2(\mathbb{R})$. Hence, for a given function $f \in \mathcal{L}_2(\mathbb{R})$, $\exists\ J$ so that $f_J \in V_J$ approximates f and for a given $g_{J-1} \in W_{J-1}$, $f_{J-1} \in V_{J-1}$:

$$f_J = f_{J-1} + g_{J-1}, \tag{2.20}$$

and by finite induction:

$$f \simeq f_J = f_0 + \sum_{n=0}^{J-1} g_n. \tag{2.21}$$

The equation 2.21 is called the wavelet decomposition of the f function. In fact, since:

$$\mathcal{L}_2(\mathbb{R}) = \bigoplus_{j\in\mathbb{Z}} W_j = V_0 \oplus \bigoplus_{j\geq 0} W_j, \tag{2.22}$$

meaning that any square integrable function can be approximated arbitrarily close in the $\mathcal{L}_2(\mathbb{R})$ norm by sums of elements in the bases W_j, for each $j \in \mathbb{Z}$. It is possible, consequently, to write a wavelet decomposition, as:

$$f(t) = \sum_{j,k} d_{j,k}\psi_{j,k}(t) = c_{00}\phi(t) + \sum_{j\geq 0}\sum_k d_{j,k}\psi_{j,k}(t), \tag{2.23}$$

with coefficients given by:

$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt, \tag{2.24}$$

$$c_{j,k} = \int_{-\infty}^{\infty} f(t)\phi_{j,k}(t)dt, \tag{2.25}$$

which is the inner product between the wavelets functions or scale function and the f function expanded (MORETTIN, 2014).

The following definition summarizes the multiresolution analysis using wavelet decomposition.

Definition 2.1.2. *Let $f \in \mathcal{L}_2(\mathbb{Z})$. The equation 2.23 is called the wavelet decomposition of the function f, with wavelet coefficients given by equation 2.24 and scale coefficients given by 2.25.*

Definition 2.1.3. *Let $\{X_t, t = 1, 2 \ldots T\}$ be a time series or a signal, with $T = 2^J$, for a given $J \in \mathbb{N} - \{0\}$ (said dyadic sample). The discrete wavelet transform (hereinafter DWT) is given by:*

$$X_t = c_{0,0}\phi_{0,0}(t) + \sum_{j=0}^{J}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t). \tag{2.26}$$

Consider the functions below:

$$S_0(t) = c_{0,0}\phi_{0,0}(t), \tag{2.27}$$

$$D_j(t) = \sum_k d_{j,k}\psi_{j,k}(t). \tag{2.28}$$

The equation 2.27 is called smooth signal, whereas the equation 2.28 is called detail signal of level j. Then, according to equation 2.22, it follows that:

$$f(t) \approx S_0(t) + D_1(t) + \ldots + D_J(t) \tag{2.29}$$

and because the terms at different scales represent components of the signal $f(t)$ at different resolutions, the approximation above is called in wavelet literature as "Multiresolution Decomposition" of the signal (BRUCE; GAO, 1996). Multiresolution analysis is also discussed in details by Daubechies (1992).

2.2 Examples of Orthogonal Wavelet Families

Wavelets may not necessarily form an orthogonal bases of $\mathcal{L}_2(\mathbb{R})$, yet there is a reason for choosing orthogonal wavelets in some of the applications in time series analysis: the orthogonal wavelet decomposition of a signal can be reconstructed without by means

of an antitransform, just as it happens with the expansion of a function in the Fourier bases. This is due to the result for wavelets which is similar to the Parseval's relation, meaning that there is a "conservation of energy" which allows the procedure of wavelet reconstruction (MORETTIN, 2014).

2.2.1 Haar Wavelet

Most of the orthogonal wavelets have no closed mathematical form. However, one of the simplest and oldest examples of wavelets is the Haar wavelet, illustrated in figure 1, has a mathematical closed form, given by:

$$\psi(t) = \begin{cases} +1, & \text{if } 0 \leq t < 1/2 \\ -1, & \text{if } 1/2 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \tag{2.30}$$

Whereas the father wavelet or scale function is given by:

$$\phi(t) = \mathbb{I}_{[0,1]}(t), \tag{2.31}$$

where $\mathbb{I}_A(t)$ denotes an indicator function, meaning a function of t that is unitary if $t \in A$ and null in the contrary case.

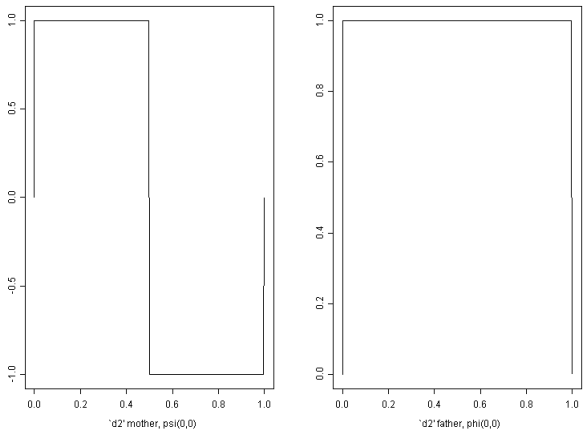

Figure 1: Haar Mother and Father Wavelets

Using equation 2.3, it leads to:

$$\psi(t) = \begin{cases} +2^{j/2}, & \text{if } 2^{-j}k \leq t < 2^{-j}(k+1/2) \\ -2^{j/2}, & \text{if } 2^{-j}(k+1/2) \leq t < 2^{-j}(k+1)\,, \\ 0, & \text{otherwise} \end{cases} \tag{2.32}$$

and the scaling equation 2.12 in this case becomes:

$$\phi(t) = \phi(2t) - \phi(2t-1) = \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t) + \frac{1}{\sqrt{2}}\sqrt{2}\phi(2t-1), \tag{2.33}$$

and thus $l_0 = l_1 = 1/\sqrt{2}$, implying that $h_0 = -h_1 = 1/\sqrt{2}$.

2.2.2 Other orthogonal wavelets

Other wavelets which are extensively used in signal processing literature are Daubechies Wavelets, which can be divided in the "extremal phase" wavelets (short Daublets or D) and the "least asymmetric" (short Symmlets or S) and the so-called "coiflets". All of them were introduced by Daubechies (1992) and are not symmetric. In fact, the Haar wavelet is part of the Daubechies Extremal-Phase wavelet family, and is also referred to as $D2$ (or $d1$, depending on the indexation).

The Daubechies scaling functions are found applying an iterative procedure, and the wavelet is constructed by the quadrature-mirror filtering relation in equation 2.15. The numbers associated to the Daubechies wavelets refer to the number of vanishing moments: each Daubechies wavelets has a number of vanishing moments equal to half the number of coefficients. This is the reason why sometimes Haar wavelet is referred to as $D2$ or $d1$: it has one vanishing moment, and two coefficients. Daublet 4, or $D4$ has four coefficients and, thus, two vanishing moments, and sometimes can be denoted as $d2$. To avoid misinterpretation, hereinafter the notation for Daubechies wavelets DN will have N referring to twice the number of vanishing moments. The figure 2, 3, 4 and 5 illustrate respectively D4, D8, D10 and S8 wavelets.

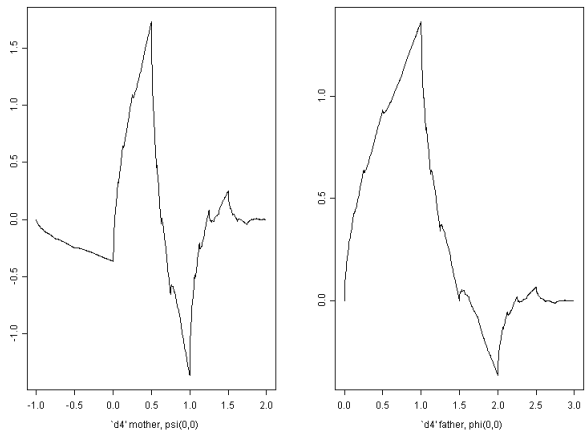

Figure 2: Daubechies Extremal Phase D4 Mother and Father Wavelets

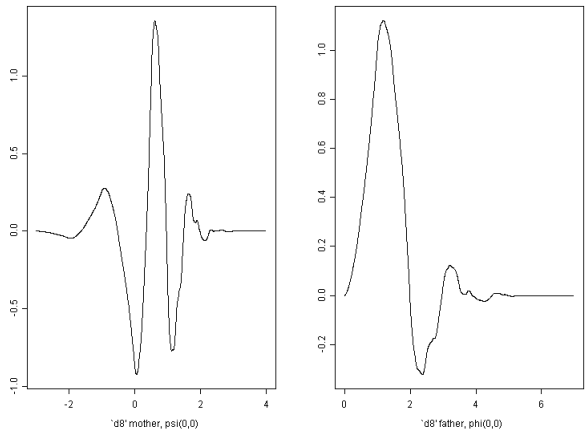

Figure 3: Daubechies Extremal Phase D8 Mother and Father Wavelets

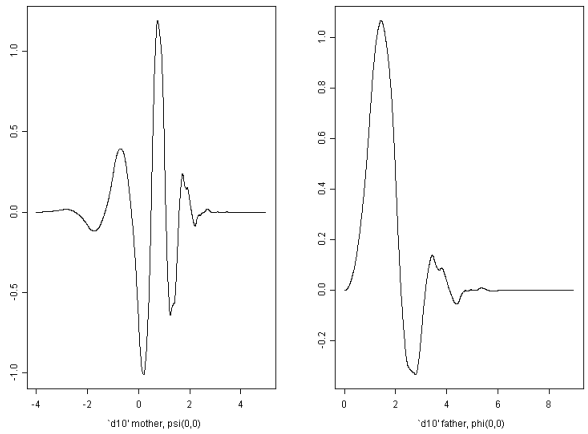

Figure 4: Daubechies Extremal Phase D10 Mother and Father Wavelets

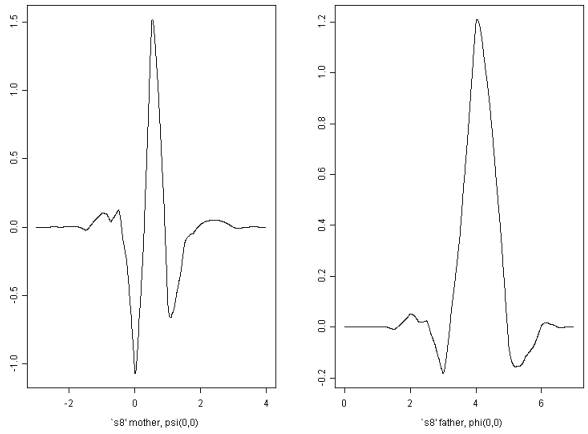

Figure 5: Daubechies Least Asymmetric S8 (Symlet) Mother and Father Wavelets

There are several computer softwares to generate wavelets computationally. Nason (2008) and Bruce and Gao (1996) give straightforward instructions on how to generate wavelet plots computationally with already existing softwares for that purpose. Other wavelet examples can be found within Daubechies's work (1992), Mallat's (2009) and Morlet's (1981). The logic behind the derivation of the Daubechies Wavelet families coefficients can be found in Frazier (1999) with step-by-step instructions to obtain the coefficients computationally using linear algebra.

2.3 Wavelet Shrinkage

One of the main advantages of the wavelet decomposition is the procedure known as wavelet shrinkage (also wavelet thresholding). It consists, in passing a filter through one or several specific levels $j = 0, 1, 2, \dots J-1$ of the decomposition and, when the wavelet family is orthogonal, reconstruct the signal with the corresponding antitransform. This is due to the fact that for a wavelet transform of a function f with coefficients $c_{j,k}$, it holds:

$$\int_{\mathbb{R}} f^2(t)dt = \sum_j \sum_k c_{j,k}^2, \tag{2.34}$$

which is the Parseval's identity, analogous to the. It means that the wavelet transform preserves the "energy", which is concentrated in a few coefficients due to its sparsity (MORETTIN, 2014).

The procedure for time series smoothing is the following (BRUCE; GAO, 1996):

1. Apply the discrete wavelet transform, described in 2.1.3.
2. Shrinking the desired wavelet coefficients towards zero using a desired threshold and policy.
3. Apply the inverse discrete wavelet transform, also called wavelet reconstruction.

There are several shrinkage functions which can be used. Two different functions of thresholding will be addressed for now, choosing different parameters, which is related to the trade off between variance and bias. One of them is called "Hard Threshold", which is given by:

$$\delta_\lambda^H(t) = \begin{cases} 0 & \text{if } |t| \leq \lambda \\ t & \text{if } |t| > \lambda \end{cases}, \tag{2.35}$$

The other threshold is called "Soft Threshold", given by:

$$\delta_{\lambda}^{S}(t) = \begin{cases} 0 & \text{if } |t| \leq \lambda \\ sgn(t)(|t| - \lambda) & \text{if } |t| > \lambda \end{cases}, \tag{2.36}$$

The hard threshold (equation 2.35) has a smaller bias yet tends to have a bigger variance, whereas the soft threshold (equation 2.36) tends to have a smaller variance, but a bigger bias (MORETTIN, 2014). The Figure 6 illustrates both of the functions above described. There are several other shrinkage functions, which are shown in details by Donoho and Johnstone (1995), Nason (2008) and Bruce and Gao (1996).

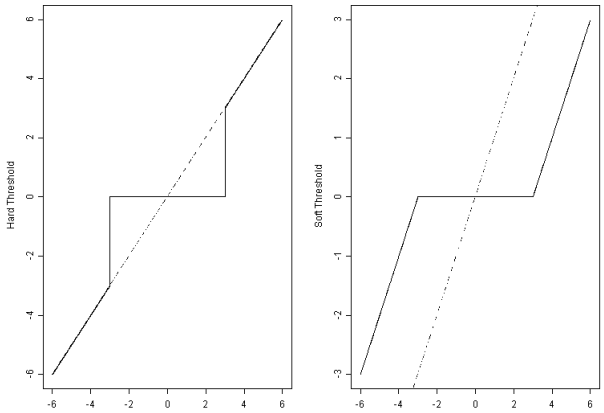

Figure 6: Hard and Soft Threshold Functions

The choice of λ must also be addressed in the thresholding. There are two main choices for λ: the so called "universal policy", the other called "Adaptive SureShrink". The "universal policy" is given by:

$$\lambda_n = \lambda_{j,n} = \sigma\sqrt{2\log n}, \tag{2.37}$$

and is called like this due to the fact that it does not depend on the scale. The value of σ must also be estimated from the data. The Adaptive SureShrink on the other hand is given by:

$$\lambda_j \underset{0 \leq t \leq \sqrt{2\log(n_j)}}{\arg\min} SURE(y_j, t), \tag{2.38}$$

with:

$$SURE(y_j, t) = n_j - 2\sum_{k=1}^{n_j} I(|y_{j,k}| \leq t\sigma_j) + \sum_{k=1}^{n_j} \left(\left(\frac{y_{j,k}}{\sigma} \right)^2 \wedge t^2 \right). \tag{2.39}$$

3 Methods for High Frequency Data

> Forecasts vary in horizon, from a few seconds up to a few days in financial markets, compared to from one to several months for macro variables. We have to provide uncertainty intervals around the central forecasts to indicate the extent to which we are unclear about the future.
>
> Clive Granger

In this chapter, some methods for dealing with intraday time series are presented. In the first section a general overview is discussed together with methods for preprocessing and cleaning high frequency financial databases. The second section presents stylized facts of intraday financial time series and time series methods for these type of data, whereas also the methodology for this dissertation is presented.

3.1 Preprocessing of intraday Data

3.1.1 A brief overview of market microstructure

The usage of methods for high frequency data (HFD) has increased in the last 20 years, mainly because of financial markets. Generally, as stated by Dacorogna et. al. (2001), stock markets are the main source of high frequency data. The original HFD are the tick-by-tick prices, which means that prices are shown not by time, but by each transaction, which might generate hundreds of thousands of prices per day of trade. Formally, a "tick" can be understood as any "time stamp", in which the quoted quantities are displayed in a data base.

The availability of these data has increased greatly due to the technological advances, as electronic trading becomes more and more the rule in financial markets worldwide, gradually substituting the floor-based trading, which makes possible to record market activity on high frequency. One of the main advantages of the electronic trading is the possibility of almost live transactions, where instruments are negotiated almost instantaneously, with little delay. This has lead to big and comprehensive databases for both practitioners and theoretical researchers. From these information it is possible to extract

even the whole book of a trade day, for instance, which is handful for the analysis of the market microstructure (HAUTSCH, 2012).

An example, which is used in this dissertation, is high frequency data of intraday prices of stocks in the Brazilian Stock Exchange and OTC ("B3 - Brasil, Bolsa, Balcão" in Portuguese). Morettin (2016) provided several examples of financial time series using HFD of the Brazilian Exchange and OTC. Other examples are provided by Dacorogna et. al. (2001), Hautsch (2012) and by Yan and Zivot (2003) using American stock exchanges data. Another interesting famous database which is freely available is the New York Stock Exchange Trades and Quotes, shortly referred to as NYSE TAQ. For more details, see Calvori et al. (2018).

As an example consider the intraday price of WEGE3 series, sampled with regular spacing of every minute in figure 7. There are two series plotted: the first one starts in 2020-01-02 at 10:03 and ends on 2020-04-30 17:30, comprehending a period of 4 months of trade, while the second one comprehends the period of a single day, 2020-01-02 from 10:03 to 17:54, which is the trading hours in Brazilian Exchange and OTC disconsidering opening and close auction time.

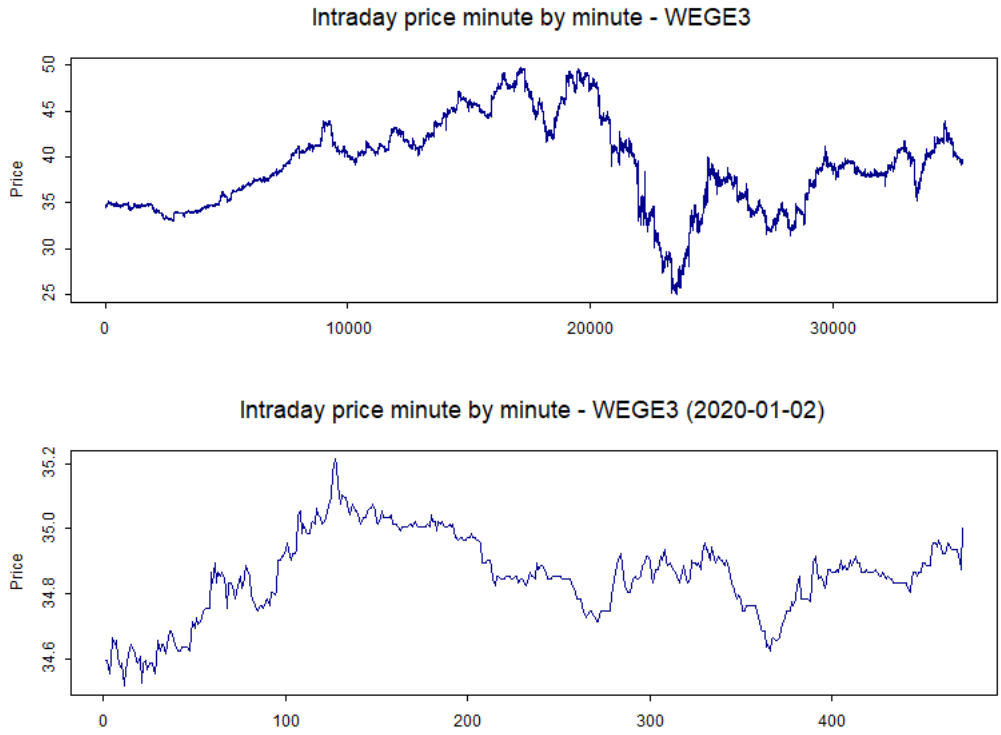

Figure 7: WEGE3 Intraday price series during four months and one day

It is worth to note that, while financial time series tend to display high volatility in longer periods (such as this four month series), the intraday series also displays a

noticeable volatile behaviour. However, even though prices oscillate, they do so within a certain range. Price jumps as seen in the series for the full range of four months usually are overnight price jumps.

While practitioners are keen to use these financial HFD for their economic decision-making process, it is also interesting from the theoretical point of view, because of properties of market microstructure which are unobservable or impalpable in daily pattern of stocks, but also for the different set of techniques which have to be developed for analyzing these financial data. As one analyzes the higher frequencies financial data, one gets closer and closer to the price formation and adjustment processes described by the market microstructure theory (RUSSELL; ENGLE, 2010).

Financial markets provide several different types of data, which might come from different sources. Dacorogna et. al. (2001) name some of them: spot market, future exchange rates, OTC interest rate, commodities, future rates, commodity futures and so on. The primary focus of this dissertation is on the spot equity prices, using the data of the Brazilian Stock Exchange and OTC sold by an authorized vendor, Nelogica [1].

The buy side of a market is typically driven by investors, borrowers or speculators, whereas the sell side consists of dealers, brokers or broker-dealers representing well-known firms. The formation of prices in electronic trade platforms of exchanges worldwide is based mostly in a rule-based order-matching system. All systems use price priority as primary order precedence rule, which stated basically that traders who offer the most competitive prices have priority in negotiation (HAUTSCH, 2012).

There are several other details which refer to the market microstructure and price formation in stock exchanges worldwide. For more details on the microstructure foundations of financial markets, see Hautsch (2012), and for more details on market microstrucutre theory see O'Hara (1998). The quantities of interest are usually prices during the intraday negotiation. However this might vary: for instance, practitioners might be interested in the trading volume or duration when dealing with High Frequency Trading models (HAUTSCH, 2012). Engle and Russel (1998) proposed a model for the durations (understood here as the time between two different trades) similarly to the ARCH process. For a more detailed review of this topic, see Degiannakis and Floros (2015) and Hautsch (2012).

3.1.2 Data cleaning and interpolation

The main difference between intraday financial data and the daily financial time series is that the observations are irregularly spaced: usually the daily financial time series displays the closure spot price of stock, whereas the financial intraday data display

[1] For more information, see Nelogica's web-page: https://www.nelogica.com.br/

the tick-by-tick prices of negotiation during a trade day (DACOROGNA et al., 2001). This generates a theoretical problem, because financial time series model usually assume regularly spaced data in time. Consider then the following definitions:

Definition 3.1.1. *A regularly spaced time series are called homogeneous, whereas irregularly spaced time series are called inhomogeneous.*

There is a main approach to deal with inhomogeneous time series, which is to apply an operator (called "regularizing operator") which will turn it into an homogeneous time series. In most of cases, the procedure then is to sample the inhomogeneous time series in a regular time interval, such that the sampled time series is homogeneous. The basic intraday financial variables, such as price, returns, realized volatility and spread, are usually inhomogeneous originally, and it is necessary to apply a regularizing operator to them.

In this case, a homogeneous time series is an artifact constructed artificially from the raw data. Consider $z_j = z(t_j)$ a time series which assumes value z_j at the time t_j, where the t_j points are not equally spaced. The way to turn these inhomogeneous time series into an homogeneous is an "interpolation" method, which will construct a series with values at $t_0 + i\Delta t$ times regularly spaced, with starting time t_0, i is an index and Δt denotes the regularly space between the observations by each time unit (DACOROGNA et al., 2001). Degiannakis and Floros (2015) referred to this method of construction of intraday time series as "calendar time sampling", abbreviated to CTS.

The time $t_0 + i\Delta t$ will be then bracketed by two times of the raw inhomogeneous series:

$$j' = \max(j | t_j \leq t_0 + i\Delta t) \text{ and } t_{j'} \leq t_0 + i\Delta t < t_{j'+1}, \tag{3.1}$$

meaning that the interpolation is between inhomogeneous times $t_{j'}$ and $t_{j'+1}$. Generally speaking, there are two main methods used for interpolation. One is the linear interpolation, namely:

$$z(t_0 + i\Delta t) = z_{j'} + \frac{t_0 + i\Delta t - t_{j'}}{t_{j'+1} - t_{j'}} \left(z_{j'+1} - z_{j'}\right), \tag{3.2}$$

and another method is the previous-tick interpolation:

$$z(t_0 + i\Delta t) = z_{j'}, \tag{3.3}$$

which is proposed initially by Wasserfallen and Zimmermann (1985). In simple terms, the previous-tick interpolation means that one considers the last available price as the price in the interpolation.

Zumbach and Muller (2000) propose other interpolation techniques, which use convolution filters or exponential moving average in order to obtain an homogeneous time series. Andersen and Bollerslev (1997) also propose a method for linear interpolation using the logarithm of bid and ask prices of different time stamps.

Some of the main advantages and disadvantages of the financial time series interpolation are discussed by Dacorogna et al. (2001) in more details. In general, for spot stock prices (prices in general) it is recommended to use the previous-tick interpolation. Yan and Zivot (2003) also use the previous-tick interpolation for stock prices. An advantage is that the interpolation by repetition (or previous-tick interpolation) respects the principle of causality, because only past values are used in order to determine the price in the instant $t_0 + i\Delta t$. A main disadvantage of this type of interpolation is that there is a risk of a spurious jump at the end of a big gap. Assuming that the stock evaluated and the market are sufficiently liquid, the presence of these referred to spurious jumps tend to be negligible. The linear interpolation, on the other hand, might be interesting to be used for rates. Since this dissertation deals with the intraday stock prices in the spot market, the regularizing operator used is the one in equation 3.3. In practice, the difference between the linear interpolation and the interpolation by repetition can only be noticed if the sampling rate is small, whereas when the sampling rate increases, both methods will be quite similar.

Consider as for an example some of them trades of ITUB4 in the day of 2019-10-30, at 10:04 in Table 1. There were more trades during the minute, however these observations are omitted. Using the previous-tick interpolation, the price for the whole minute would be the last one, namely 36.460. If using the average price interpolation, the price at timestamp 10:04 would be 36.460 (in bold). The average price within the minute is 36.45817, and thus the difference is would be not as much significant.

Table 1: ITUB4 trades - observations in 2019-10-30 at 10:04

Date	Instrument	Price	Time
2019-10-30	ITUB4	36.450	10:04:00.001
2019-10-30	ITUB4	36.470	10:04:00.001
2019-10-30	ITUB4	36.420	10:04:00.001
...	...	...	...
2019-10-30	ITUB4	36.450	10:04:49.872
2019-10-30	ITUB4	36.450	10:04:51.888
...	...	...	...
2019-10-30	ITUB4	36.460	10:04:56.309
2019-10-30	ITUB4	36.460	10:04:56.309
2019-10-30	**ITUB4**	**36.460**	**10:04:56.310**
2019-10-30	ITUB4	36.460	10:05:00.000

Hautsch (2012) discusses other underlying features of preprocessing of financial data. In fact, an important aspect is the data cleaning: sometimes there are spurious registers, which are obvious data errors and should be filtered out. Typically this happens because of either wrong recording or a delayed recording of trade or quote information. As the electronic trading becomes more and more common, these types of errors tend to be much less common.

Hautsch (2012) presents several different features which indicate an error in the database and should be cleaned. Some recording problems most easily identified if transaction prices show severe jumps between consecutive prices which are reverted instantly. Another problem are entries with a quote or transaction price equal to zero or being negative, or outside the regular trading hours. Dacorogna et al. (2001) mention that orders which have been cancelled should be discarded from the dataset. For more details on data cleaning, see Dacorogna et al. (2001) and Hautsch (2012).

There are several software implementations for these methods of intraday preprocessing of financial time series. For more details, see Yan and Zivot (2003) in S+ or the R package "highfrequency" developed and with detailed manual by Boudt et. al. (2020).

3.1.3 Methods used for intraday time series modelling

3.1.3.1 The concept of risk

One of the main goals of practitioners or theoretical researchers when studying financial intraday time series is either forecasting or to statistically describe the risk of a certain quantity. Here the term "risk" refers to the degree variability of a quantity of interest, such as price return for stocks in spot market or the returns of an index/market future interest rate. This is important in practice because from such features one can develop risk models.

Formally, a risk model can be understood as a quantitative approach to estimate numerically the "risk" and describe its dynamics based on observed data. Such models are not only useful to practitioners, but also to theoretical studies which aim to provide means to understand the volatility of stock markets and its behaviour (VIENS; MARIANI; FLORESCU, 2011). This dissertation emphasizes the risk modeling approach, using the methods described in the previous chapters and does not emphasize the forecasts of high frequency financial time series.

While volatility was assumed in the 1970's to be a constant parameter for risk models, it is currently almost consensual that it is a time-varying variable, which is not directly observed, but rather estimated from the data. There are several procedures for estimating volatility, namely: implied volatility models from option pricing, stochastic volatility, realized volatility and the GARCH framework (DEGIANNAKIS; FLOROS,

2015). For more details regarding these models, see Tsay (2005) and Zivot and Wang (2005).

The analysis of volatility in financial time series is relevant not only for theoretical researchers, but also among portfolio managers, investors and others market players. Volatility can be understood as a measure of the variability of financial returns over a given period of time and in statistics it is usually measured as the variance or the standard deviation (TSAY, 2005). This is a cardinal variable in risk management, asset pricing and portfolio selection, and understanding its behaviour is almost inescapable when dealing with financial data, not only for forecasting, but also in terms of statistical inference about underlying aspects of the financial data which one aims to comprehend and describe. Financial analysts and researchers are concerned with modelling volatility, in order to understand the covariance structure of asset returns (DEGIANNAKIS; FLOROS, 2015).

One of the main problems when it comes to risk modelling is that the models and its forecasts depend on an implicit choice of a family of distributions from which the inference or forecasts are made. Hence the choice of such distributions is crucial for the financial risk modelling (VIENS; MARIANI; FLORESCU, 2011).

3.1.3.2 Prices and returns

When dealing with the variable "Price", one might refer to different things in different contexts and markets (if future or spot market). For instance, there are bid and ask prices, transaction prices and middle prices (DACOROGNA et al., 2001). In the context of this dissertation, price refers to the transaction price for a given asset in the spot market, namely: the value for which a "bidder" was willing to buy and an "asker" was willing to sell, also called bid-ask pair. This is coherent with daily financial time series, because usually when dealing with the spot "prices" of stocks, one is usually referring to the close price, or to the adjusted closure price. In this case, in a same minute for instance, there happen to several transaction prices. The "price" then used for computing the quantities of interest in this dissertation is the last available transaction price in the spot market, because this again is coherent to our previous assumptions.

Once the time series is homogeneous, there are several ways to compute financial returns. Tsay (2005) defines two types of returns of assets, which are more commonly used: the simple return and the log-return or compound returns. Let P_t denote a price at the time t, and assume that the series $\{P_t\}$ for $t = 1, 2, \ldots T$ is homogeneous after a procedure of interpolation. Then the simple return and the log-return which are respectively given

by:

$$R_t = \frac{P_t}{P_{t-1}} - 1, \tag{3.4}$$

for $t = 2, \ldots T$ and:

$$r_t = \ln(1 + R_t) = \ln(P_t) - \ln(P_{t-1}). \tag{3.5}$$

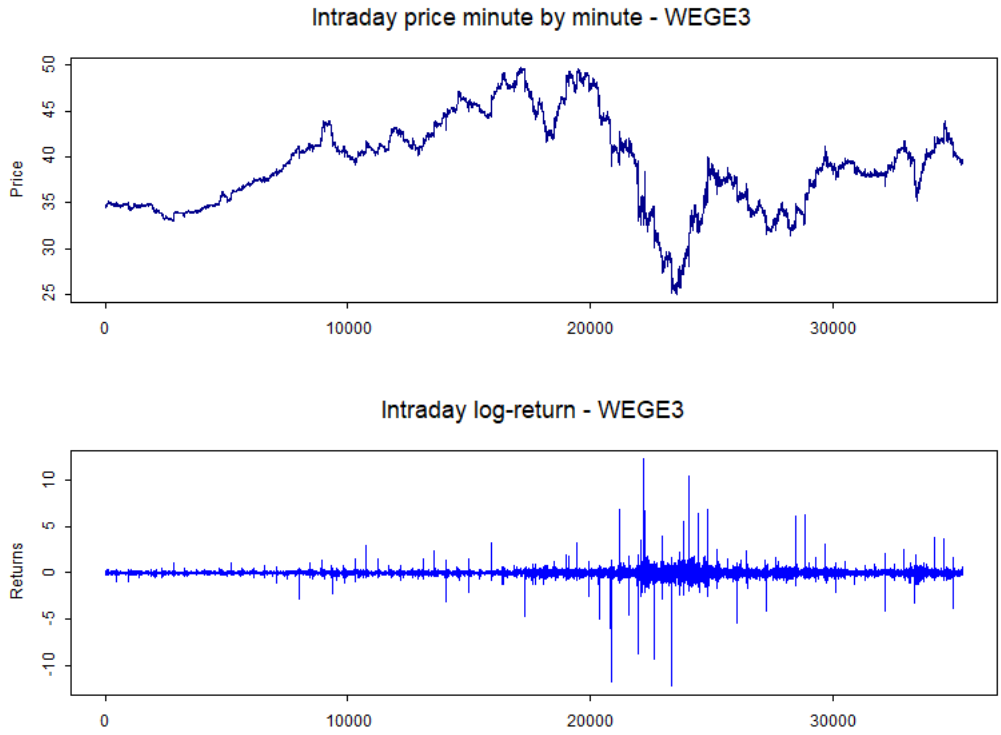

Figure 8: WEGE3 Intraday price and log-returns

As for an example, consider again the intraday prices of WEGE3 during the period of 2020-01-02 10:03 to 2020-04-30 17:30 sampled every other minute. Using the formula of the log-returns in equation 3.5, one can compute the log-returns for that series. Figure 8 displays the intraday prices of WEGE3 computed log-returns multiplied by 100, so the second graphics from figure 8 displays the series $100 \times r_t = 100 \times (\ln(P_t) - \ln(P_{t-1}))$. This transformation is usually useful because a change in the $100 \times r_t$ corresponds to approximately the percent change in prices. So for instance the price jumps which are close to 10 in figure 8 correspond approximately to 10% changes in prices.

In this dissertation the usage of log-returns is preferred over simple returns because, as Tsay (2005) stated, in order to compute the cumulative return R in a period from

timestamp $t = t_0$ to timestamp $t = T$, it is possible to compute it as:

$$R = \sum_{t=t_0}^{T} r_t, \tag{3.6}$$

which makes calculations simpler in terms of risk evaluation in financial markets.

It is worth to note that the financial returns are, by construction, an homogeneous time series. This is because for one to apply the difference operator Δ to the series $\{\ln(P_t)\}$ for $t = 1, 2, \ldots T$, the series must be homogeneous.

In this dissertation, when referring to financial returns, it will follow the equation 3.5. Log-returns are also referred to as continuously compound returns, and they present some advantages over the simple returns in equation 3.4 when it comes to spot market prices of assets. For more details, see Tsay (2005) and Jorion (2011).

In order to measure volatility for risk models in financial markets, some transformations might be applied to the financial data. For instance, a general type of transformation is the absolute log-return or the squared log-return, meaning:

$$x_t = |r_t| = |\ln(P_t) - \ln(P_{t-1})|, \tag{3.7}$$

or:

$$x_t = r_t^2 = (\ln(P_t) - \ln(P_{t-1}))^2 . \tag{3.8}$$

Andersen et. al. (2009) stated that the absolute log-returns represents the volatility of the returns. The absolute value of the log-return series is widely used as a transformation to the homogeneous data in order to measure volatility in applied finance. For instance, Kaizoji (2006) studied the properties of the absolute log-returns to analyze if it follows a power-law as theoretically assumed in market microstructure studies and Sibbertsen (2001) used this transformation in order to measure the long-memory effect in German financial data. In practice, the squared log-returns are used more for rates and foreign exchange and absolute log-returns are more used for spot market prices.

Morettin (2016) stated that, in general, to evaluate the realized volatility, one might consider a transformation applied to the log-returns as:

$$v_t = \left(\frac{1}{k} \sum_{j=0}^{k-1} |r_{t-j}|^p \right)^{1/p}, \tag{3.9}$$

where $p > 0$ and usually $p = 1$ or $p = 2$. In this case, volatility can be understood as a mean of k past returns with the proper transformations depending on p. Dacorogna et al. (2001) stated that the choice of p, in this case, must be careful, because a larger

value gives more weight to the tails of the distribution and provided a general discussion regarding this matter.

There are several other transformations that can be applied to the data. For instance, Parkinson (1980) proposed the price range to a time series partitioned in equidistant time points, using the highest and lowest log-prices, and the Garman and Klass range (1977). See Degiannakis and Floros (2015) and Andersen et. al. (2009) for a detailed review.

3.2 Stylized facts of high frequency financial data

There are several stylized facts of intraday financial time series, which are discussed in details by Dacorogna et. al. (2001) and Hautsch (2012). The aim of this section is to summarize some of these features.

Formally, according to Cont (2002), stylized facts can be understood as statistical features that are commonly observed in real market data across different markets and different periods of time, the "common denominators" in financial markets.

The reasons for these stylized facts to be more apparent in lower frequencies are still not much clear, but there is a consensus in literature that it has to do with the effects in price formation and market microstructure. Dacorogna et al. (2001) for instance stated that, at higher frequencies, the middle price is subject to the microstrcture effects, which overshadows the properties of lower frequencies financial series.

3.2.1 Empirical findings

3.2.1.1 Fat-tailedness and non-normality

Dacorogna et. al. (2001) stated that the distribution of the log-returns are increasingly fat tailed as the frequency of the data increases. These patterns are more noticeable as the frequency gets higher, and they are often shadowed in lower frequencies. The monthly returns display much less volatility clustering than the daily returns and this is more evident with intraday returns. Due to this, the second moment most probably exists, whereas the fourth moment tends to diverge and the distributions tend to be less stable.

As for the distributions of the log-returns, it is noticeable that financial returns are commonly not well described by normal distribution, even in lower frequencies, because of the prevalence of extreme values that tend to occur in financial data. An alternative for low frequencies usually is in the Student t distribution, as shown for instance by Tsay (2005), the models seem to fit well. However for higher frequencies, Viens, Mariani and Florescu (2011) recommended for practical risk management and theoretical purposes

the usage of the so-called hyperbolic distributions, such as the Skewed Student t, since for higher frequencies data tend to be skewed. Cont (2002) stated that, as Δt increases towards lower frequencies, then result is that the data tend to be more closer to the normal distribution - what he refers as "aggregational gaussianity".

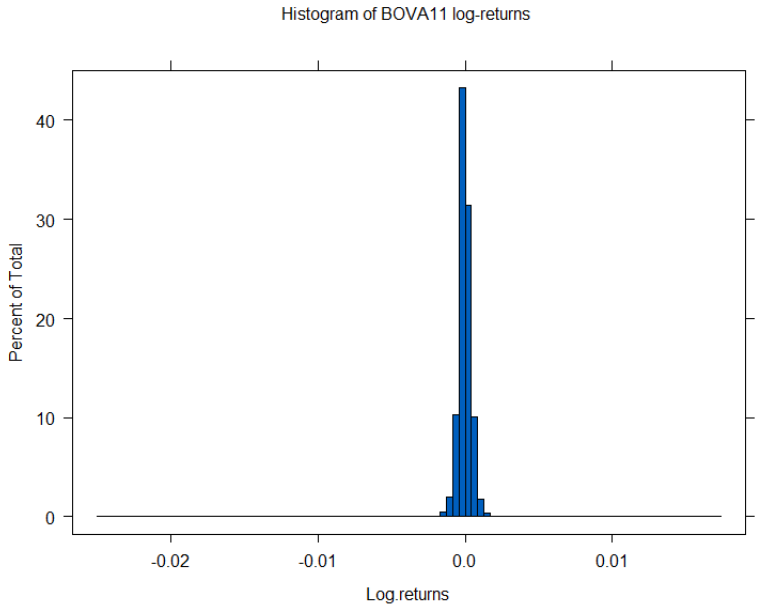

Figure 9: Histogram of BOVA11 log-returns

As for an example of non-gaussianity of log-returns and absolute-log-returns, consider the histogram of the log-returns of BOVA11 from 2020-01-02 to 2020-02-19, corresponding to approximately one month of trade, displayed in figure 9. It is clear from that figure the the data is non-gaussian, having fatter tails and a behaviour closer to a generalized error distribution. This is more easily noticeable in figure 10, comparing the log-returns with the quantiles of the normal distribution. This indicates that the log-returns have much fatter tails and displays characteristics of non-normality.

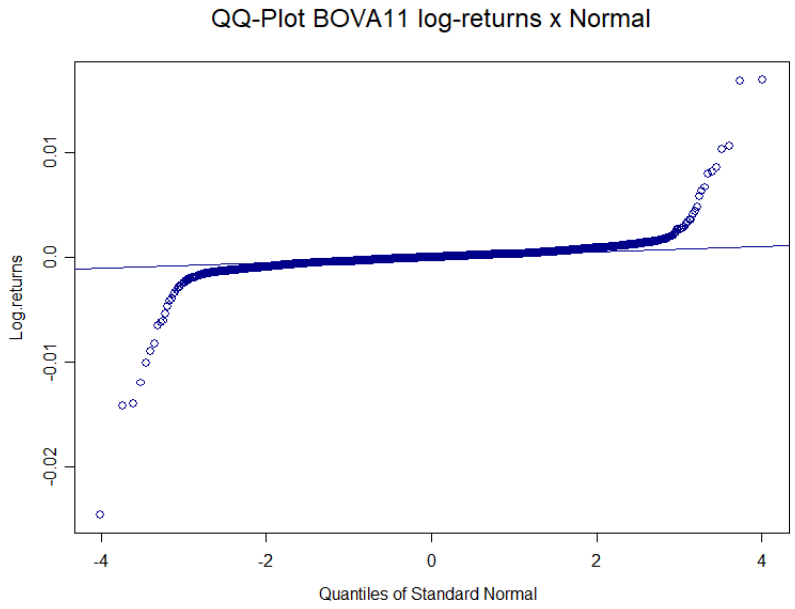

Figure 10: QQ-plot of BOVA11 intraday log-returns

As for the fat tailedness, Dacorogna et. al (2001) asserted that this is partially explained by what they refer as the "weekly clusters of volatility". Lambert and Laurent (2000) proposed the Skewed Student t distribution or generalized error distribution to account for these fatter tails and non-normality. Some other examples of the usage of Skewed Student t distribution for high frequency financial data are Huang et. al. (2012), Shao, Lian and Yin (2009) and Worthington and Higgs (2005). Still regarding these features of the distribution of these financial time series, Kercheval and Liu (2011) the (log) return series tend to display leptokurtic and skewed distribution, which is coherent to the usage Skewed T distribution.

A theoretical reason might by what Cont (2002) calls the "gain/loss asymmetry", meaning that drawdowns in prices are larger than upward movements. However, according to Malmsten and Teräsvirta (2004), this stylized fact still has no closed theoretic and economic explanation.

Another example of non-normality can be seen in figure 11, from the histogram of the absolute log-returns of BOVA11 from 2020-01-02 to 2020-02-19. Not only this displays the pattern of non-normality and fat-tailedness, but it is also displays the behaviour of skeweness. This shape of the histogram usually appears when evaluating the absolute log-returns, because of the transformation made to the data.

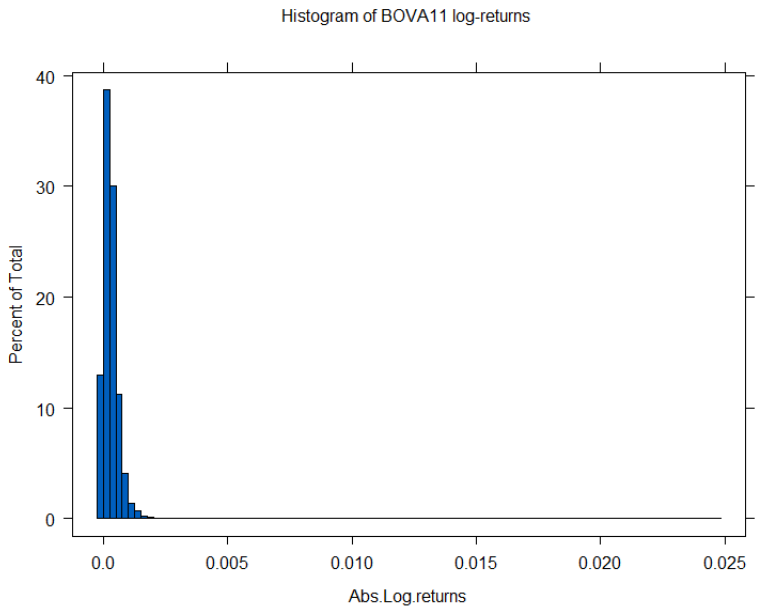

Figure 11: Histogram of BOVA11 absolute log-returns

3.2.1.2 Intraday jumps

Another stylized fact mentioned by Laurent (2018) is the presence of intraday jumps. In order to remains consistent with the no-arbitrage semimartingale setting, these intraday jumps must be unknown in terms of timing and magnitude, subject only to weak regularity conditions (ANDERSEN et al., 2009). These jumps are disconsidered in the estimation of the volatility model, according to Laurent (2018), because of the impact they have on the returns. Cont (2002) said that these bursts/jumps tend to happen due to the high degree of variability observed in the series. As for now, the literature does not offer a consensus on how to identify price jumps properly and how to deal with them (HANOUSEKA; KOČENDA; NOVOTNÝ, 2011).

As for an example, consider the log-returns of the series of absolute log-returns of WEGE3 from 2020-01-02 to 2020-02-19 in figure 12. There are clearly values which seem atypical or disparate as compared to the rest of the series: for instance one might say from visual inspection that values bigger than 1.5 (or price changes bigger than 1.5% within the same minute) represent atypical values. In this dissertation, wavelets will be used in order to identify these pattern of jumps and to provide a way to deal with them without having to disconsider such observations.

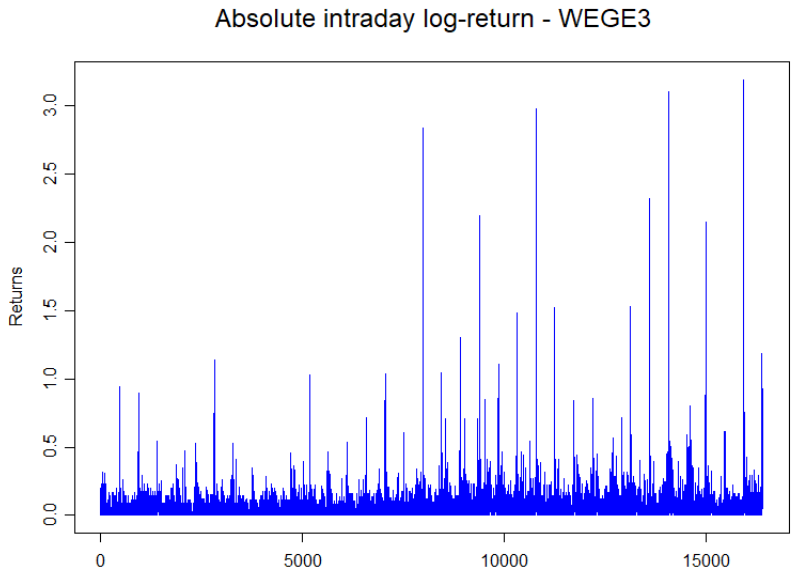

Figure 12: Absolute log-returns of WEGE3

3.2.1.3 Volatility clustering

Another stylized fact of intraday financial returns is that they tend to present heteroskedastic behaviour, as stated by Andersen et. al. (2009). Volatility clustering and non-Gaussian behavior in financial returns is typical in intraday data as well: larger price changes tend to be followed by larger price changes and the other way around. This results, as Cont (2002) affirmed, that higher volatility events cluster in time. A solution to this is to use an ARMA process with GARCH errors as stated by Kercheval and Liu (2011), which can be easily extended to the ARMA model with FI(E)GARCH errors to incorporate features of long-range dependence.

3.2.1.4 Long-range dependence

The presence of long memory is a stylized fact of financial time series. When observed in economics, it means in fact that a shock to the underlying variables have a long lasting effect. The persistency can occur in the first or in high order moments of the time series (ZIVOT; WANG, 2005).

Usually the log-returns display little or no serial correlation, but the squared and absolute log-returns are positively autocorrelated, due to the market microstructure properties. The absence of statistically significant linear correlations in returns is more noticeable in liquid markets, assumed to be due to the "efficient market hypothesis" (CONT, 2002).

Kaizoji (2006) declared that while the standard methods indicates a strong evi-

dence of long-range dependence, this might be an effect artificially produced by trends or structural breaks. Sibersten (2001) affirmed that slowly decaying trends and structural breaks can easily be confused with long memory dependence and therefore this requires a more subtle and careful analysis.

Russel and Engle (2010) claimed that this dependence in the high frequency financial data is due to the price discreteness and to the the fact that there is often a spread between the price paid by buyer and seller initiated trades. While these features of long-memory in the conditional mean and conditional variance are present for high frequency financial data, under temporal aggregation the range of dependence in the returns tends to decrease. This long-range dependence is also a noticeable pattern for the trade volume and duration, see Russel and Engle (2010).

To deal with the long-range dependence, long-memory models such as ARFIMA and FIGARCH presented in the previous chapters are good alternatives. The FIGARCH models and the subsequent models such as FIEGARCH, since their proposition in 1996, have been extensively used for practical applications, due to their versatility and to the parameters' interpretation, which lead to theoretically interesting analysis of the financial data. For example, comparing the estimation using ARCH models and FIGARCH models when trying to determine the volatility of exchange rates, it was noticed that the IGARCH and GARCH models tend to underestimate the effects of central bank interventions, for instance, in the volatility (XEKALAKI; DEGIANNAKIS, 2010).

3.2.1.5 Specific stylized facts of cryptocurrencies

Cryptocurrencies are novel investments. An interesting fact is that cryptocurrencies are effectively unregulated and there are several stock exchanges in which they are negotiated. The trading is unregulated, meaning that prices reflect the market's uncertainty and this leads to volatilities which tend to be higher in the minute-by-minute and hour-by-hour series (HU; PARLOUR; RAJAN, 2018).

Bitcoins were the first well-known cryptocurrencies, which use blockchain technology to record and decentralize the ledger of ownership and transactions solve the double-spend problem. According to Hu, Parlour and Rajan (2018), there are two categories of tradable cryptocurrencies alternative to the Bitcoins, referred to as "altcoins": the coins and tokens. For instance, Litecoin (LTC) is a cryptocurrency which is, technologically, almost identical to bitcoin, according to Borri (2018): however, it uses a blockchain system which is based on its own protocol (HAFERKORN; DIAZ, 2015). On the other hand, Ether (ETH) is a cryptocurrency that is transacted through Ethereum blockchain platform (ZHANG et al., 2018).

Hu, Parlour and Rajan (2018) claimed that the returns displays the usual time series properties that are similar to stock returns. For instance, it is possible to notice

the persistence behaviour in the daily cryptocurrency market, as pointed by Caporeale, Gil-Alana and Plastun (2018), whereas Cocco, Concas and Marchesi (2014) discussed the volatility clustering observed in cryptocurrency returns. Non-normality is also observed in the cryptocurrencies log-return series (PHILLIP; CHAN; PEIRIS, 2017).

An important stylized fact for cryptocurrencies which makes them different from other high frequency financial time series is that cryptocurrencies are traded 24/7, every day of the week, including holidays, as stated by Borri (2018). This means that there can be no overnight jumps, since the market never opens and never closes, being functional and operational 365 days of the year. This leads to a different pattern of ultra high frequency algorithms, as appointed by Petukhina, Reule and Härdle (2020).

Borri (2018) divides cryptocurrencies exchanges into two different types: the ones in which one can trade only cryptocurrency pairs (for instance BTC for LTC) and the ones in which one can trade both cryptocurrencies for actual currencies and other pairs, and vice-versa. There are several cryptocurrencies exchanges worldwide, the most important ones being in Asia, North America and in the UK. Some of the exchanges make their data freely available for students and practitioners who wish to test their high-frequency trading algorithm, via API and other online platforms. There are also companies which make this data available freely and easily accessible [2] for those who want to model these series. Spithoven (2019) provided more detailed information of cryptocurrency government and also Giudici, Milne and Vinogradov (2019).

3.2.2 Examples

Consider, for the next example, the trades of VALE3 in 2019-07-01, from 10:13:00 to 10:13:59, after cleaning the cancelled orders from the database. During this minute, we can observe 280 transaction prices (which means on average 4.7 transactions per second). It is worth to note that this database only reflects the transaction price: there were certainly more bids and asks during this minute, but this dataset only reflects the transactions which actually occurred. The Table 2 shows the 10 first prices for this dataset.

It is noticeable in first place that, even within the same second, there can be more than one trade, meaning that, for liquid assets, a single trade day could have thousands of transaction prices in the higher frequencies (for example, every second). In total, for this single asset VALE3, during the trade day of 2019-07-01 there were 62.616 transactions, disconsidering those which were cancelled. This provides an example of how sophisticated these datasets can become as one aims to use several trade days of a single asset. This number of transactions in trade day might vary, depending on the liquidity of a given asset in the market.

[2] For instance https://www.cryptodatadownload.com/

A second phenomenon to notice is that these prices are irregularly spaced, because the time difference varies between different transactions. Another interesting feature to notice is that the range of the transaction prices is not wide: for this particular range of time, 10:13:00 to 10:13:59, the minimum price was 53.24, whereas the maximum price was 53.29.

Table 2: First 10 prices of VALE3 at 10:13 in 2019-07-01

Trade day	Symbol	Price	Time
2019-07-01	VALE3	53.240	10:13:00.287
2019-07-01	VALE3	53.250	10:13:02.254
2019-07-01	VALE3	53.250	10:13:02.256
2019-07-01	VALE3	53.270	10:13:02.261
2019-07-01	VALE3	53.260	10:13:02.261
2019-07-01	VALE3	53.270	10:13:02.267
2019-07-01	VALE3	53.270	10:13:02.272
2019-07-01	VALE3	53.270	10:13:02.274
2019-07-01	VALE3	53.260	10:13:03.923
2019-07-01	VALE3	53.270	10:13:04.644

Changes are not significantly higher, and this indicates that, for certain cases, the sampling frequency chosen might affect the return in the time interval, because higher frequencies tend to display smaller values of log-returns. In fact the minimum price for the whole trade day of 2020-07-01 is 53.10 and the maximum is 53.79. Therefore the intraday jumps do not necessarily mean that the prices actually had "big" change, but that marginally the percent change was significantly higher than average enough to cause distortions.

Intraday datasets are massive. As an example, the trade day of 2019-07-01 has 1.431.728 transactions, but some trade days, however, can have as much as 1.600.000 transactions. This means that the procedure of filtering and applying a regularizing operator to these asset prices might take hours, depending on the characteristics of the computer used.

For example, consider 10 rows of the dataset filtered by only the transactions of the asset ITUB4 in Table 3. In practice, the quantities of interest are Price, date, time and an indicator of whether the order was cancelled or not. Originally from the database, the cancelled orders were denoted by 0, and hence our main interest lies in orders labeled with 1. Other information, such as the traded volume (expressed as Qty.), the transaction number and the brokers from buy and sell parts are irrelevant for this research and therefore are not mentioned or discussed.

Table 3: Example of the dataset - 10 transactions of ITUB4

Date	Symbol	Trans.nr.	Price	Qty.	Time	Ind.null
2019-07-01	ITUB4	840	36.770	1500	10:09:15.610	1
2019-07-01	ITUB4	250	36.770	500	10:09:15.610	1
2019-07-01	ITUB4	190	36.770	300	10:09:15.610	1
2019-07-01	ITUB4	270	36.770	500	10:09:15.610	1
2019-07-01	ITUB4	310	36.770	100	10:09:15.610	1
2019-07-01	ITUB4	30	36.770	5700	10:09:15.610	1
2019-07-01	ITUB4	300	36.770	100	10:09:15.610	1
2019-07-01	ITUB4	210	36.770	100	10:09:15.610	1
2019-07-01	ITUB4	230	36.770	1100	10:09:15.610	1
2019-07-01	ITUB4	160	36.770	2200	10:09:15.610	1

4 Applications

> We have the duty of formulating, of summarizing, and of communicating our conclusions, in intelligible form, in recognition of the right of other free minds to utilize them in making their own decisions.
>
> R. A. Fisher

4.1 The data

The goal of this chapter is to present empirical applications to real high frequency asset returns, using the methodology and the models described in the preceding chapters.

The data used in this practical application consists in the "tick-by-tick" inhomogeneous time series after applying a regularizing operator in order to sample the data every minute. The chosen assets were: FLRY3, WEGE3, VALE3 and BOVA11. The first three are ordinary shares and the last one is an exchange traded fund (ETF). A model was also adjusted for the SMLL Index, which works as a proxy for the behaviour of small caps in the Brazilian Exchange and OTC and for the high frequency cryptocurrency exchange rate, ETH/USD traded in Gemini Exchange. These assets were chosen because they are liquid, meaning that, as in the case of the Table 2, there are several trades within the same minute, and because the stock are from companies in different sectors and industries, which means that they might reflect some of their sector's particularities. The ETH/USD exchange rate used in the case of cryptocurrencies is the high frequency minute-by-minute time series, and refers to the ETH/USD traded in the Gemini exchange, from New York city.

The dataset has all trades in the spot market from 2020-01-01 to 2020-04-30, which is four months of trade of the Brazilian exchange and OTC. In practice it is necessary to filter the assets of interest in each trade day and apply the regularizing operator, in order to get an homogeneous time series. The dataset is massive and it is computationally complicated to be dealt with and, therefore, it is necessary to choose a time interval, which will be specified in each result section for each chosen asset.

The datasets of minute-by-minute high-frequency time series of the stocks and ETF traded in Brazilian Exchange and OTC, as well as the SMLL Index, were provided

by Nelogica [1] , an authorized vendor of the Brazilian Exchange and OTC.

The information of the ETH/USD minute-by-minute high frequency time series for the year of 2020 is provided by the Gemini Exchange, and then processed and made freely available by CryptoDataDownload [2], which also produces daily risk reports and metrics for cryptocurrencies, such as the VaR report for cryptocurrencies, liquidity and volume reports, and many others.

4.2 Methodology

Degiannarkis and Floros (2015) and Viens, Mariani and Florescu (2011) stated that the main challenges for volatility estimation models are dealing with the price level and with the microstructure noise. As said before, there is no consensus on how to deal with identification and modelling of price jumps. While some practitioners end up discarding these jumps, some authors proposed on the other hand to deal with these particular issues using Lévy processes.

In this dissertation, a method is proposed in order to deal with these intraday jumps using wavelet decomposition. Wavelets transform has already been used, for instance, in spike detection, as Nenadic and Burdick (2005) and for EGC jump detection in Lannoy, Decker and Verleysen (2008). Wavelets have also been used to generate a test for jump detection, done by Xue, Gençay and Fagan (2014) and Chen, Lai and Sun (2019). Nonetheless, the usage of wavelets for non-parametric regression and smoothing is described extensively with examples by Bruce and Gao (1996).

In the case of these intraday jumps, for the four stocks analyzed (WEGE3, FLRY3, VALE3 and BOVA11) in the spot market, it is noticeable that the higher values of the absolute log-returns happen either due to open or close auctions, which are part of the trading regulations in the Brazilian Exchange and OTC (hereinafter B3). The methodology for auctions and trading tunnels is detailed in the Brazilian Exchange and OTC's Trading Procedure (2019a) and the Trading Tunnel calculation methodology (2019b).

The methodology for modelling the financial high frequency data is as it follows:

1. Filter the dataset for every trade day in order to obtain the quantity of interest, in this case the tick-by-tick prices of a given asset.

2. Preprocess the data, cleaning the cancelled orders and the "NA" values from the database, which correspond to register errors.

[1] For more information, see Nelogica's web-page: https://www.nelogica.com.br/
[2] For more information, see CryptoDataDownload's webpage https://www.cryptodatadownload.com/

3. Apply a regularizing operator to the tick-by-tick inhomogeneous time series, in order to make it homogeneous, equally sampled with $\Delta t = 60$ seconds.

4. Evaluate the presence of spurious price jumps and, if necessary, apply a correction to the data.

5. Generate the log-return series $r_t = 100 \times (\ln(P_t) - \ln(P_{t-1}))$ for $t = 2, 3, \ldots T$ from the homogeneous time series and obtain the absolute returns $|r_t|$.

6. Apply the discrete wavelet transform to $|r_t|$ and use the wavelet shrinkage in order to obtain a "smooth" trend, which will indicate the intraday jumps, using a specified metric for the shrinkage.

7. Separate the $|r_t|$ series in two parts: the jump series resulting from the wavelet shrinkage procedure (hereinafter s_t) and the return series discounting the jumps identified with wavelets, hereinafter $x_t = |r_t| - s_t$.

8. The resulting series x_t is modelled according to a ARFIMA-FI(E)GARCH model, in order to account for the stylized facts usually observed in these high frequency financial time series.

9. Compare the results of this methodology with an approach without removing the jumps in the returns.

The data used is already preprocessed, either by Nelogica or by CryptoDataDownload. In the case of Nelogica, since the data is not publicly available, but it is paid and customized to the necessities of the customer, it was asked for the company to perform the steps from 1 to 4 in the data, meaning that the data used is already homogeneously sampled in regular time intervals of one minute, and that the interpolation method used was the previous-tick interpolation. A consequence is that only the steps from 5 to 9 will be described in the methodology section. CryptoDataDownload also makes the last tick information for the minute-by-minute price available freely in the their web-page for the years of 2017 to 2020, therefore, in that case, only steps from 5 to 9 will be described in the third section of this chapter.

About the methodology, a few comments are still worth done. For the calendar time sampling, the time interval chosen is $\Delta t = 60$ seconds. The reason for that meets practical issues: while in the higher frequencies the stylized facts of financial time series get more and more noticeable, intraminute data is much more complicated to model computationally, because of the length of the variables. Even with 1 minute sampled data, the estimation of the parameters of the FI(E)GARCH models using simulated annealing method may take up to 6 hours until convergence. Due to that, the time series of prices every minute is chosen, since the data still displays the stylized facts of high frequency data while taking

much less time to compute the model's parameters than if it was considered a higher frequency.

The regularizing operator chosen was the previous-tick interpolation because, as mentioned in the third chapter, these methods are coherent to the price formation in the market. The cancelled orders are already removed from the data.

The idea behind using the wavelet shrinkage procedure is in fact to separate the "noisy" part of the log-return series from a smooth trend of the data. The shrinkage procedure is applied in every level of the wavelet decomposition, using the universal threshold and a hard policy. From this, one concludes that whatever wavelet family is used, there will be little difference between the results of the decomposed series, mainly because the details are passing through the threshold.

In every model of these time series, the wavelet used is the Haar wavelet, because smaller wavelets tend to perform better empirically in order to identify discontinuities, and are generally used in series with breaks or jumps. A brief discussion of this matter is, for instance, proposed by Downie (2011).

There are several alternatives of shrinkage function used to generate the σ quantity in the universal policy of equation 2.37. One can use either the usual L_p norm (denoted as LP), usually with $p = 1$ or $p = 2$, or the median of absolute deviation from the median (denoted as MAD), or even the usual standard deviation (denoted as SD), as in Bruce and Gao (1996). The choice is heuristic, and will rely on the dataset used and which one fits each data better in each case.

Due to the stylized facts of high frequency financial time series, it is expected that the series after removing the jumps present the long memory patterns in the conditional mean and conditional variance. it is also expected to find the pattern of non-normality and heavier tails, and, in some cases, skewness. To account for that, ARFIMA models will be used for the conditional mean, whereas FI(E)GARCH models will be used for the conditional variance. It is likely that low orders of the long-memory GARCH models, for example FIGARCH(1,d,1) or FIGARCH(1,d,2), will account for the serial correlation structure of these time series, whenever there is indication of long memory in the conditional variance.

One aims to show that the methodology above described is coherent and versatile, in the sense that it is serves not only for a specific purpose/class of assets, but for a different range of high frequency financial time series.

In order to compare these models to the traditional methodology, both models will be fitted and then compared in terms of the root mean squared error (RMSE). When referring to the traditional methodology, it is meant the fit of ARFIMA-GARCH models without the need of removing the jumps using the wavelet decomposition and thresholding

procedure. In order to evaluate the RMSE of the traditional ARFIMA-GARCH, one can consider:

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left((r_t-\hat{r}_t)^2-\hat{\sigma}_t^2\right)^2}, \tag{4.1}$$

where $\hat{r}_t$ denotes fitted values of the ARFIMA model, r_t denotes the log-returns and $\hat{\sigma}_t^2$ denotes the estimated conditional variance. For the proposed methodology using the wavelet decomposition and shrinkage in order to identify and remove the jumps, we consider:

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left((x_t-\hat{x}_t)^2-\hat{\sigma}_t^2\right)^2}, \tag{4.2}$$

where $x_t = r_t - s_t$ denotes the series after removing the identified jumps using the wavelet procedure, $\hat{x}_t$ denotes fitted value of the ARFIMA model for the return series and $\hat{\sigma}_t^2$ denotes the estimated conditional variance.

The choice of the RMSE instead of other metrics relies on the fact that the mean is more sensitive to outliers as compared, for instance, with the median. So choosing metrics such as a median absolute deviation or median errors could perhaps reduce the effect of the presence of the jumps in the series.

4.3 Why remove the jumps?

The methodology herein proposed for modelling financial high frequency time series proposed a separation of the jump series from the log-returns. A question that might arise is why not to use a traditional approach instead removing the jump series, which is what this subsection aims to answer.

A jump in a time series behaves, essentially, as an outlier in the time series in terms of magnitude. Laurent (2018) stated that a jump should be abnormally big, however what constitutes an "abnormally big" return will depend much on the volatility of the data generating process. This is more complicated when dealing with heteroskedastic processes.

It is a well-known fact that outliers can influence the sample autocorrelations $\hat{\rho}(k)$ from a given time series and even affect it drastically, which might cause identification problems, as discussed thoroughly by Chang (1982) and Guttman and Tiao (1978). In fact, Chang, Tiao and Chen (1988) mentioned that the presence of an outlier might bring up several problems in terms of identification, such as bias in the estimation of the ACF/PACF and even for the ARMA parameters of a given series. Tsay (1988) also provided a discussion on outliers effects in time series

Chan (1995) describes outliers as nonrepetitive intervention in the time series, meaning that it is possible to separate a contaminated series Z_t into two parts:

$$Z_t = Y_t + \eta_t(T, \omega), \tag{4.3}$$

where Y_t denotes the outlier-free series and $\eta_t(T, \omega)$ an exogenous intervention effect located in $t = T$ with magnitude ω. For instance, one can consider an additive outlier of the form:

$$\eta_t(T, \omega) = \omega I_t^{(T)}, \tag{4.4}$$

where $I_t^{(T)}$ denotes the indicator function:

$$I_t^{(T)} = \begin{cases} 0 & \text{if } t \neq T \\ 1 & \text{if } t = T \end{cases}, \tag{4.5}$$

meaning that in equation 4.4, the parameter T indicates where the outlier happens in time and ω indicates the magnitude of the effect of this outlier. Consider then the next theorem:

Theorem 7. *Let $Z_1 \dots Z_N$ be an observed time series generated from the Additive Outlier model, as in equations 4.3 and 4.4. Then:*

$$\lim_{N \to \infty} \{\operatorname{plim}_{\omega \to \infty} \hat{\rho}(k)\} = 0.$$

The proof for theorem 7 can be found in Chan (1995).

The main consequence of theorem 7 is that the sample ACF will have a limiting behaviour of 0 whenever the magnitude of an additive outlier is much bigger than the values in the observed time series. In fact, Chan (1995) stated that large additive outliers could completely wash away the information in the sample ACF, which would be dominated by the behaviour of the outliers. As a consequence, this would mislead the researcher or practitioner into identifying a model that is not coherent to the actual data generating process when both N and ω are large.

This is precisely the case of these financial high frequency time series used in this practical applications section. Therefore, an approach without removing jumps would mislead the researcher and might even induce one into identifying a model that is not coherent to the DGP of the series, not bringing up the stylized facts of financial time series which are expected.

Other problems that come with the presence of anomalous values or abnormally large values in the series will show up in the maximum likelihood estimation. For instance,

Mendes (2000) stated that the presence of outliers in the GARCH models will generate maximum likelihood estimates which are biased for the GARCH parameters. For the case of GARCH models with additive outliers, Doornik and Ooms (2005) propose a model which is analogous to the one presented by Chen (1995), but also other GARCH models which might have the additive outlier term in the conditional volatility, with the solution to include binary or "dummy" variables in order to remove such effects.

For the case of GARCH models, Carnerno, Peña and Ruiz (2004) also show analogously that outliers can lead either into the detection of spurious conditional heteroskedasticity or hide ARCH effects, in an similar fashion to the result for the ARMA models presented by Chen (1995). In fact, it is shown that outliers can hide legitimate heteroskedasticity when they are isolated or with larger sizes. This was also identified and discussed by Mendes (2000).

Laurent (2018) stated also that, in times of higher volatility, an abnormal return is bigger than an abnormal return in times of low volatility. This is a reason to justify the usage of non-parametric estimates, here wavelets, in order to identify these jumps in the model. With this, the identification of outliers or jumps is less arbitrary and relies on a non-parametric estimate. The goal is to then assess whether this model was interesting to be used, by comparing the RMSE in equations 4.1 and 4.2, and whether the results mentioned in this section will show, meaning: to verify whether the jumps may "hide" or wash away the behaviour of the ACF/PACF for the time series and whether removing these jumps from the series would lead to more accurate estimates of the volatility.

4.4 Practical applications

In this section, the practical applications are described. Each of the 6 aforementioned time series will be separated and discussed in different subsections. The structure of each subsection is displayed as it follows:

1. An introduction, describing the time series used;
2. The description of the wavelet shrinkage procedure;
3. Conditional mean and conditional variance models;
4. A final section comparing the RMSE as in equations 4.1 and 4.2.

The fit of the proposed methodology is described thoroughly in each subsection, whereas the fit of the traditional GARCH models proposed is just mentioned in the end of each subsection, for means of comparison between the fit of the models.

The main goal of this subsection is to provide examples that the intraday and overnight jumps might hide the true heteroskedasticity and dependence structure by generating biased estimators of $\hat{\rho}(k)$, which might lead to misidentification of the model and maybe to a poorer fit in terms of the RMSE.

4.4.1 Models for WEGE3

The first time series used is the intraday price of WEGE3, homogeneously sampled minute-by-minute. The series starts in 2020-01-02 10:03:00 and ends in 2020-02-20 10:08:00, which corresponds to approximately one and a half month of trade days (36 days) and $T = 16,384$ points.

The symbol WEGE3 corresponds to the ordinary shares of WEG S.A. traded in the Brazilian Exchange and OTC (B3) since 2007. WEG S.A. is a corporation from south Brazil, engaged mainly in the capital goods sector with solutions in electrical machines, automation and paints for different sectors. The company has operations not only in Brazil, but in the Americas, USA, Europe, Asia and Africa, with commercial activities carried out in more than 135 countries.

It is possible to notice in Figure 13 that the absolute log-return series displays the traits expected due to stylized facts of financial time series: it is possible to notice volatility clusters and presence of jumps, in the sense of anomalous returns from one time stamp to the subsequent.

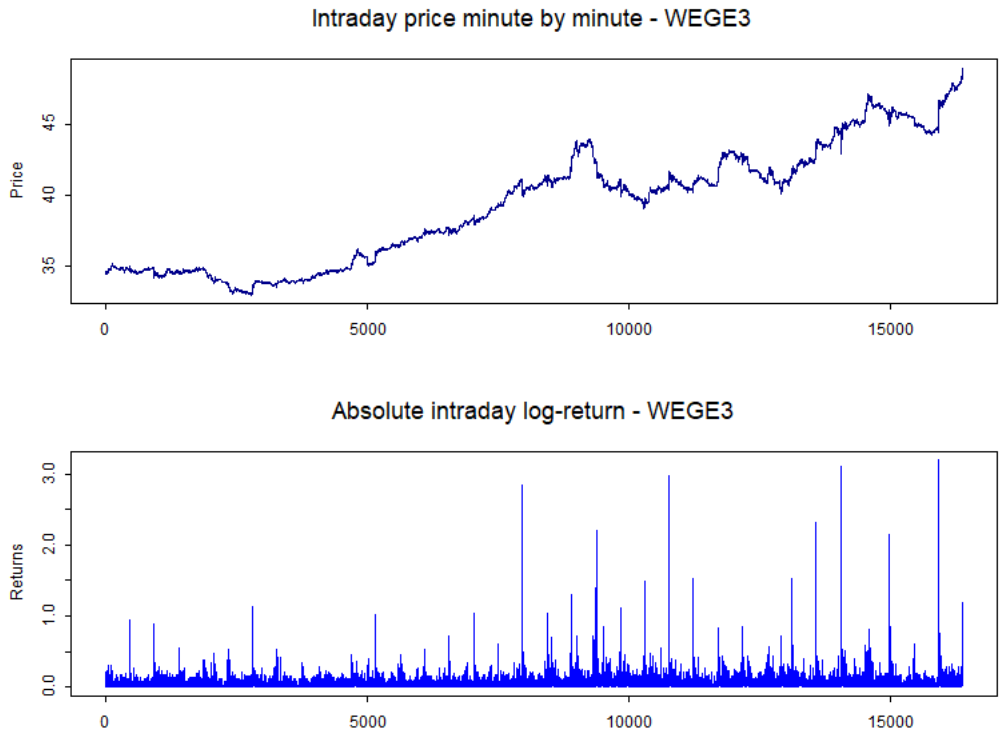

Figure 13: WEGE3 Intraday price series and absolute log-returns

It is worth noting that the series of log-returns in Figure 13 is given by:

$$r_t = 100 \times (\ln(P_t) - \ln(P_{t-1}))\,,$$

which means that the largest returns are around 3% from one time stamp to the other. Usually they correspond to overnight price changes. Table 4 displays the larges absolute 100× log-returns for WEGE3. It is possible to see that 19 out of 20 correspond to overnight changes in prices, only the 15^{th} line corresponding to a price change which was not overnight, but intraday, and this was most likely due to a close auction.

Figure 14 displays the absolute log-returns with the identified pattern of jumps and the series after the procedure to remove the jumps removed from the original series.

These jumps were obtained using the wavelet shrinkage procedure described in the methodology section, using the universal rule. The shrinkage function used was the median of absolute deviation from the median (hereinafter MAD), in the parameterization proposed by Bruce and Gao (1996).

Figure 15 displays, on the left, the discrete wavelet transform of the WEGE3 log-returns using the Haar wavelet and, on the right, the wavelet coefficients after the threshold procedure is applied, using the MAD function. It is worth to note that, as

Table 4: WEGE3 absolute log-returns - top 20

Ticker	Date	Time	Price	Abs.logret
WEGE3	20200219	100800	45.8408	3.1923
WEGE3	20200213	100800	42.9106	3.0996
WEGE3	20200204	101000	41.6193	2.9795
WEGE3	20200127	100800	40.0002	2.8398
WEGE3	20200212	100600	43.7648	2.3190
WEGE3	20200130	101300	41.1922	2.1942
WEGE3	20200217	101300	44.8873	2.1454
WEGE3	20200211	101000	41.8875	1.5293
WEGE3	20200205	100800	40.8644	1.5185
WEGE3	20200203	101400	39.7320	1.4860
WEGE3	20200217	101400	45.5528	1.4717
WEGE3	20200213	102000	43.5066	1.3793
WEGE3	20200217	100200	45.8706	1.3297
WEGE3	20200129	100800	42.2749	1.3007
WEGE3	20200219	174800	48.3738	1.1839
WEGE3	20200110	100300	33.4742	1.1340
WEGE3	20200131	101500	40.0698	1.1093
WEGE3	20200128	100700	41.1227	1.0441
WEGE3	20200123	101000	38.2421	1.0336
WEGE3	20200117	100600	35.9972	1.0262

expected, the shrinkage procedure cleans the "noisy" part of the series, which is clearly seen in the wavelet coefficients after the thresholding of the DWT.

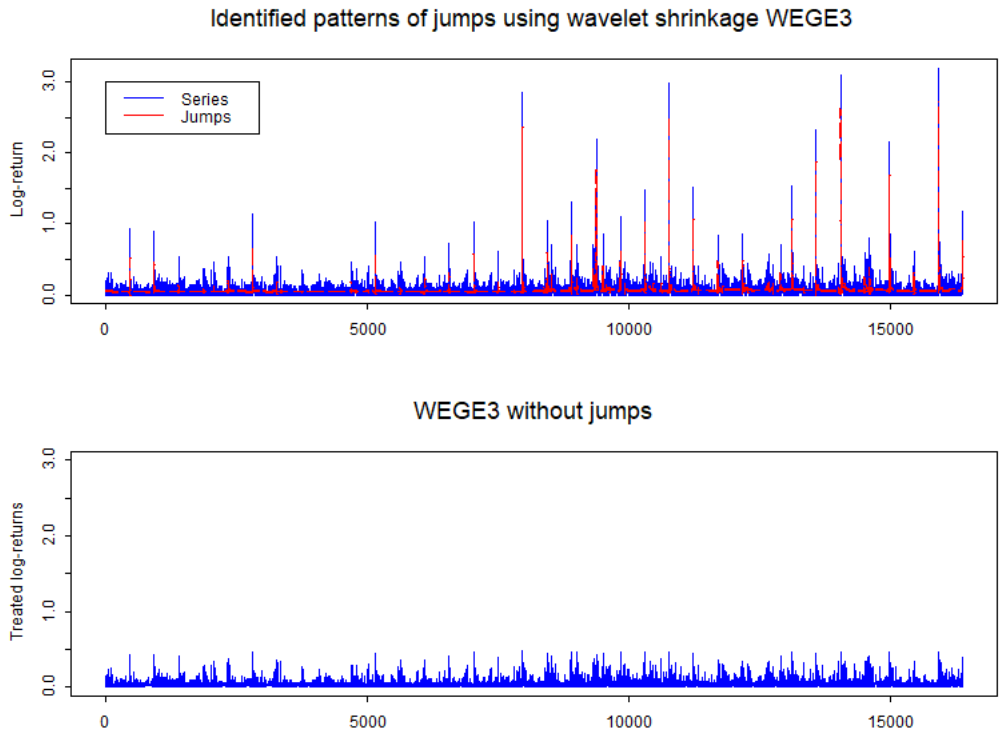

Figure 14: WEGE3 Identified jump series using MAD shrinkage function

From that, in order to obtain the jump series, it is applied the inverse discrete wavelet transform to the data. In Figure 14, the identified pattern of jumps correspond to the wavelet reconstruction of the shrinked DWT procedure of the log-return series as in Figure 15, as a non-parametric estimate.

It is noteworthy that not only the values close to 2.5% or 3% (which by visual inspection seemed anomalous) were identified, but also other values close to 1% price changes were identified as jumps as well. This provides a non-parametric and less arbitrary identification procedure for these jump series.

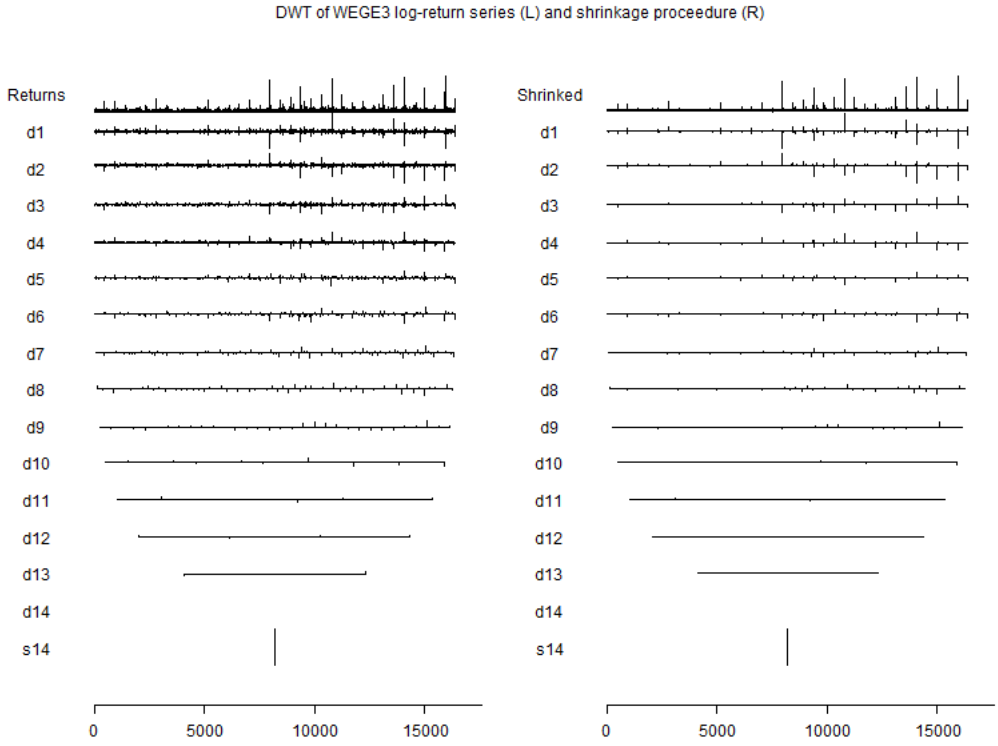

Figure 15: DWT and Shrinakge procedure - WEGE3 log-returns

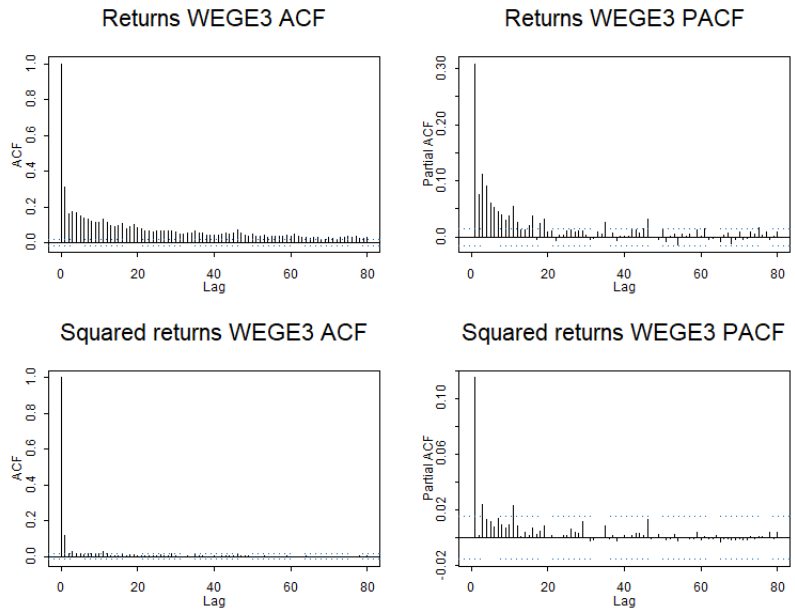

Figure 16: ACF and PACF of WEGE3 of log-returns without removing jumps

Figure 16 and 17 exhibit the ACF and PACF of the log-return series without removing the jumps or anomalous price changes and after removing the price jumps with

the wavelet procedure, respectively. From that, it is possible to see that, as Mendes (2000) and Carnerno, Peña and Ruiz (2004) describes, the presence of such outliers overshadows the true serial correlation behaviour in the log-returns time series, making it appear that there is no long memory in the conditional variance, because most of the structure of the serial correlation lie within the 95% confidence interval, making it statistically insignificant at the 5% level. However, after removing the price jumps, the outshadowed serial correlation structure reappears, as it can be seen in Figure 17.

In order to compare the proposed methodology with the traditional approach, one can consider two models: a model only removing the jumps using DWT and another model without removing the jumps, hereinafter referred to simply as "traditional approach". In the end of this section, the RMSE of these models will be compared in order to evaluate the results.

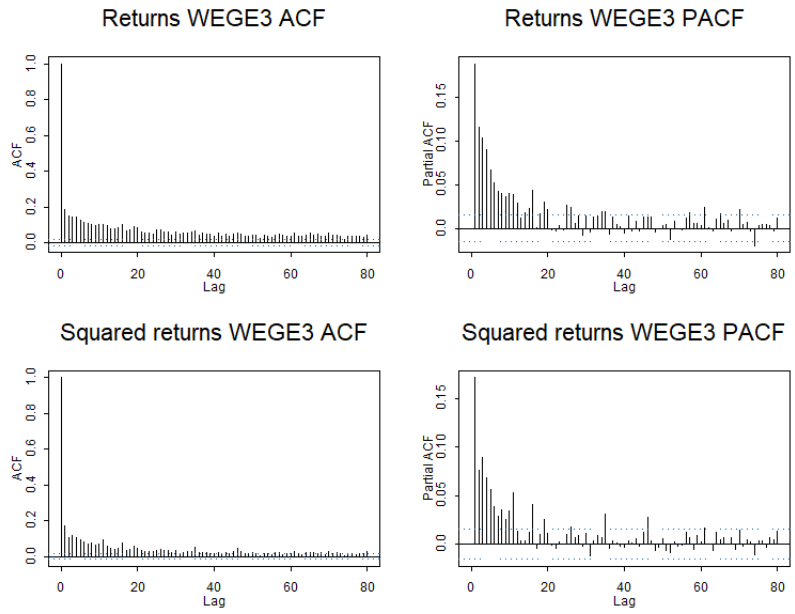

Figure 17: ACF and PACF of WEGE3 treated log-return series

Figure 17 shows the ACF and PACF for the log-return series after removing the identified jumps with the wavelet procedure, hereinafter only referred to as the returns, and for the squared returns. From that, it is possible to see that the ACF and PACF present the traits of a long-range dependence both in the return series and squared return series: the ACF has an hyperbolic decay, and there are non-null partial autocorrelations at the significance level of 5% in lags close to 30. This means essentially that long memory models for the conditional mean and conditional variance should be considered. A conditional mean model which can be considered is the ARFIMA(1,d,1). This model is chosen as the most parsimonious while ensuring that there are no significant autocorrelations

at the significance level of 5%. The parameters for the ARFIMA model and other useful information are exhibited in Table 5.

Table 5: ARFIMA coefficients - WEGE3

Coef.	Value	Std. Error	T stat.	p-value
d	0.3039	0.0141	21.5817	0.0000
AR(1)	0.3200	0.0201	15.9326	0.0000
MA(1)	0.5042	0.0098	51.2308	0.0000

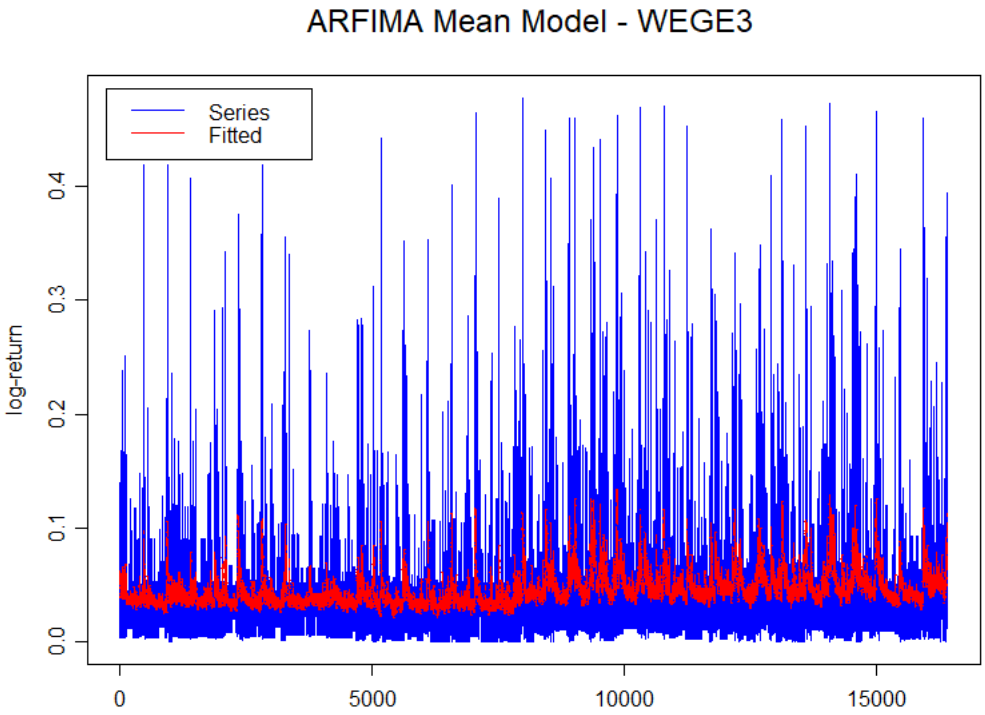

Figure 18: ARFIMA adjusted to the WEGE3 series

It is worthwhile to comment the results in Table 5. The model was estimated using maximum likelihood procedure, using as starting values the method of moments estimates for the AR and MA part. All the parameters are statistically significant at the 5% level. The estimated $d = 0.125$ parameter shows that there is a significant long memory in the series. There is also evidence of strong dependence in the short memory part of the model, particularly in the first lag, both for the AR and MA parts. The estimated model is stationary since the roots of the AR and MA polynomials lie outside the unit circle and $|d| < 0.5$. The fact that $d > 0$ means that the model is persistent, which is coherent with the stylized facts of financial time series, as expected.

Figure 18 displays the fitted ARFIMA(1,d,1) model versus the observed return series. While the ARFIMA model can describe the dynamics of the series and accounts for the long-range dependence in the mean, there is still a pattern of heteroskedasticity that the model does not account for, which is why it is necessary to consider the conditional variance model.

For the conditional variance model, one can consider the FIGARCH(1,d,1) model with innovations generated by a generalized error distribution. This model is chosen as the most parsimonious while ensuring that there are no significant autocorrelations at the significance level of 5% in the squared residuals of the ARFIMA model. The Table 6 shows the coefficients for the FIGARCH(1,d,1) model and the estimated degrees of freedom for the GED distribution. Figure 19 exhibits four plots related to the FIGARCH fitting: the residuals of the ARFIMA model, the standardized residuals of the FIGARCH model, the conditional standard deviation and the conditional variance.

Table 6: FIGARCH coefficients - WEGE3

Coef.	Value	Std. Error	T stat.	p-value
Cst x 10^4	0.4202	0.1573	2.6710	0.0076
d-Figarch	0.2440	0.0421	5.7950	0.0000
ARCH(1)	0.3786	0.0825	4.5900	0.0000
GARCH(1)	0.4440	0.0883	5.0290	0.0000
G.E.D.(DF)	1.0817	0.0194	55.8700	0.0000

Some points must be emphasized based on Table 6. The long memory parameter $d = 0.2440$ in conditional variance is bigger than the one in the conditional mean, however this does not necessarily indicate that there is a stronger dependence. As Caporin (2002) stressed, the d parameter operates in a different fashion for the FIGARCH process: as $d \to 0$, the memory of the process increases, due to the representation of the effect of a random shock in equation 1.91. It should be pointed out, from the estimated values in Table 6, that the positivity constraint for the FIGARCH (1,d,1) model is observed, because $0.200025 < 0.378579 < 0.585326$ and $0.000144052 < 0.0792865$, following the conditions in Bollerslev and Mikkelsen (1996).

The optimization procedure used was the BFGS, with implementation described by Laurent (2018) in OxMetrics-G@RCH^{TM} software.

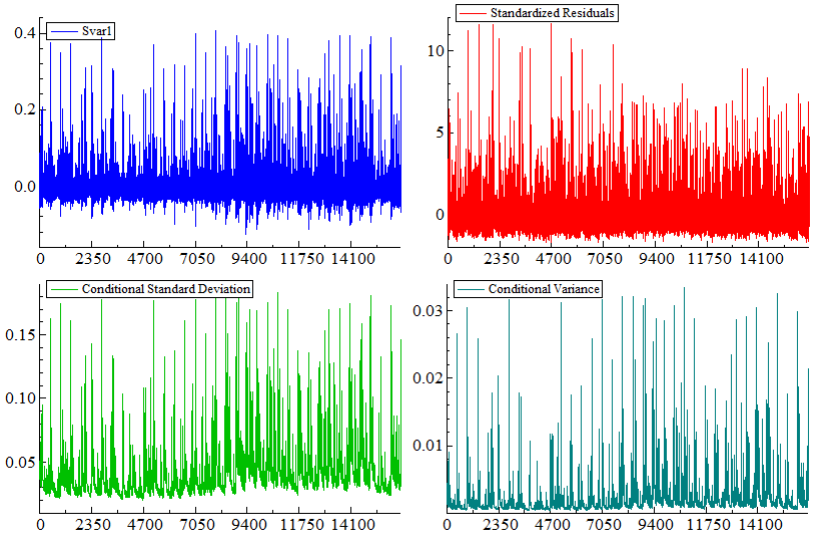

Figure 19: Adjusted FIGARCH for WEGE3 series with GED distribution

Figure 20 exhibits the return series with 2 SD superimposed, which essentially shows that most of the variation of the time series lie within 2 conditional standard deviation. This corroborates that the estimated conditional standard deviation in Figure 19 is coherent with the data, because it reflects the increase in volatility during the period of clusters and it captures the heteroskedastic behaviour of the model.

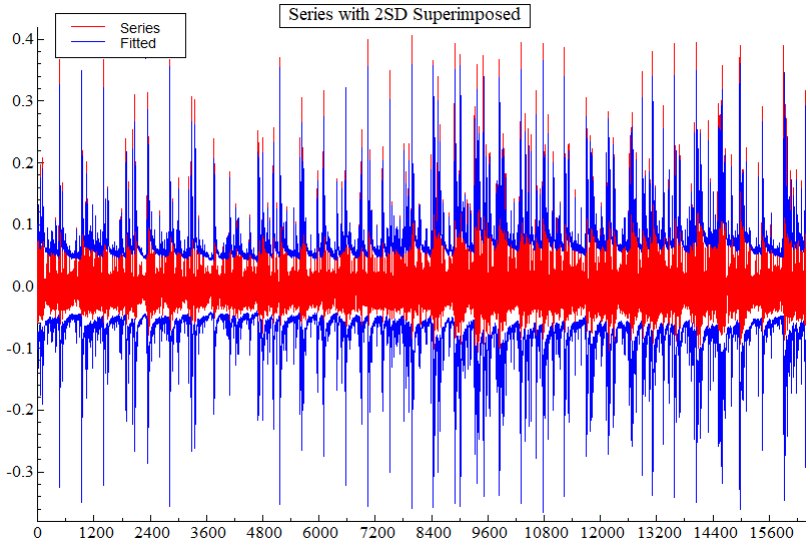

Figure 20: Series with 2 SD Superimposed - WEGE3

The Table 7 displays the Box-Pierce test for the residuals of the models. The result indicates that the chosen models are enough to describe the dynamics of the serial dependence of this time series.

Table 7: P-value of the Box-Pierce test for the residuals - WEGE3

Lag	Mean res.	Var. std. res.
5	0.4774	0.4672
10	0.7641	0.8806
15	0.1583	0.8950

The RMSE of the model after removing the price jumps with wavelets calculated by equation 4.2 is evaluated as 0.008539516. Considering the serial correlation structure exhibited in Figure 16, one can fit an ARFIMA(3,d,0)-GARCH(1,1) model using the GED distribution, which produces a RMSE of 0.1510499, as shown in table 8. Therefore, the model removing jumps with wavelets is significantly more accurate in describing the volatility of the intraday series.

Table 8: Comparison of RMSE of models - WEGE3

Model	RMSE
Removing jumps	0.0085395
Not removing jumps	0.1510499

From the results presented in this section, it is possible to state that the methodology for modelling the volatility of these financial time series turned out to successfully account for the serial correlation in the log-return series. The wavelet decomposition and shrinkage proved also interesting in terms of identification of the jumps, in such a way that the outliers are not identified arbitrarily, but by a non-parametric procedure, which generates a much more precise estimate in terms of RMSE.

4.4.2 Models for SMLL Index

In this subsection, the time series used is the homogeneously sampled SMLL Index series, sample every other minute, with length $T = 16,384$. The series starts as well in 2020-01-02 10:03:00 and ends in 2020-02-20 10:08:00, which corresponds to approximately one and a half month of trade days (36 days).

The SMLL Index is the small cap index of the Brazilian Exchange and OTC. The aim of the SMLL Index is to reflect not only the price of the stocks which integrate the theoretical portfolio used to make the index, but also to show the impact of the earnings of these companies. Small cap is a term used to refer to stocks of small companies traded in the exchange, usually with market value ranging from BRL 300 million to BRL 2, and are less liquid than the big companies. Investors might be interested in small caps as compared to the so-called “blue-chips” (bigger companies in the stock exchange) because

of the significant potential growth that is hardly going to be matched by larger companies, with the trade-off of higher risk.

Consider the Figure 21, which exhibits both the intraday prices and the log-returns of the SMLL index. The absolute log-return series corroborates with this, because there are several price jumps, and it displays the traits expected of financial time series: it is possible to notice volatility clusters and presence of jumps.

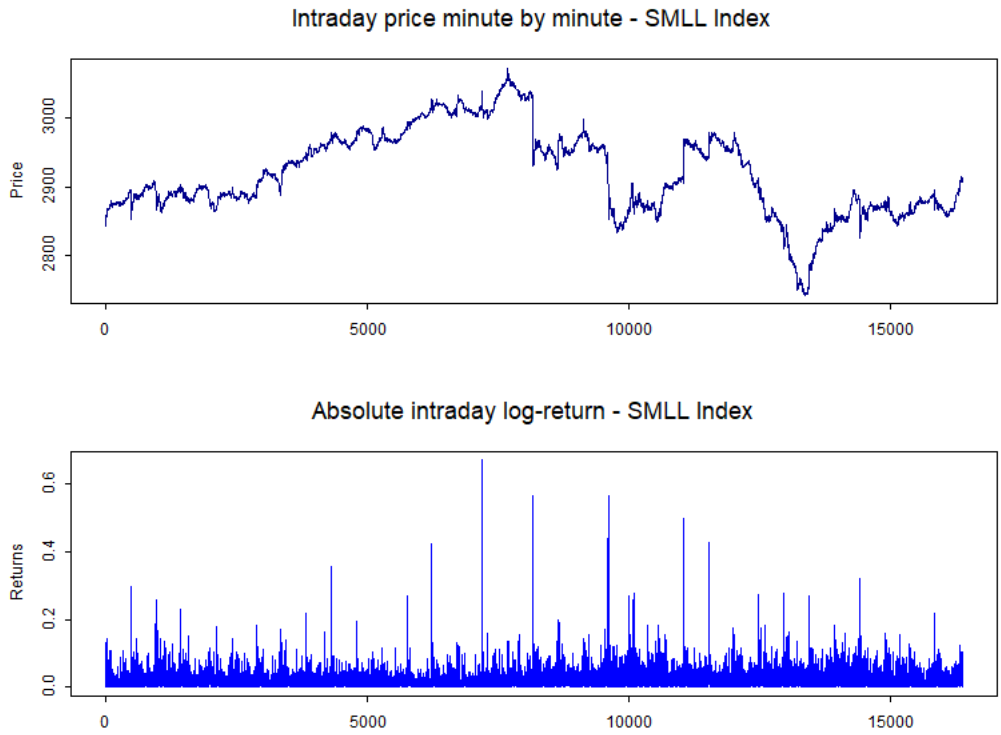

Figure 21: SMLL Index Intraday reconstructed price series and absolute log-returns

It is interesting to point out that a price jump has not necessarily anything to do with a fixed magnitude (say 2% or 3%) from a time stamp to the other, but is it rather more a particular trait that varies from a time series to the other. For instance, a price change of around 0.6% for the WEGE3 series as in Figure 14 does not seem to be anomalous, whereas for the SMLL index as in Figure 21 it corresponds to the largest value in the whole series. Since this would depend on the dynamics of the series, it is interesting to use wavelets as a non-parametric method to identify these price jumps. Generally, in terms of magnitude, as compared for instance with the WEGE3 series in the previous section, the jumps are much smaller.

The Table 9 displays the ten largest absolute 100× log-returns for SMLL Index. As one can see, 8 out of 10 correspond to overnight changes in prices, with exception being two price changes that were intraday, in the end of day, most likely to close auctions.

Table 9: SMLL absolute log-returns - top 10

Ticker	Date	Time	Price	Abs.logret
SMLL	20200122	175900	3037.7400	0.6708
SMLL	20200130	100700	2879.8200	0.5668
SMLL	20200127	100900	2933.5300	0.5666
SMLL	20200127	100700	2951.2500	0.5188
SMLL	20200127	100200	3005.5000	0.5187
SMLL	20200204	100700	2949.9600	0.4968
SMLL	20200130	100200	2918.2200	0.4373
SMLL	20200205	100800	2978.6700	0.4276
SMLL	20200120	175900	3026.3500	0.4219
SMLL	20200127	100500	2977.2700	0.3999

Consider then the Figure 22 which displays the result of the discrete wavelet transform of the SMLL Index log-returns. The figure exhibits on the left side, the discrete wavelet transform of the SMLL Index log-returns using the Haar wavelet and, on the right, the wavelet coefficients after the threshold procedure is applied. The result of the procedure is that the shrinkage procedure cleans out the "noisy" part of the return series, leaving then just the identified jump series.

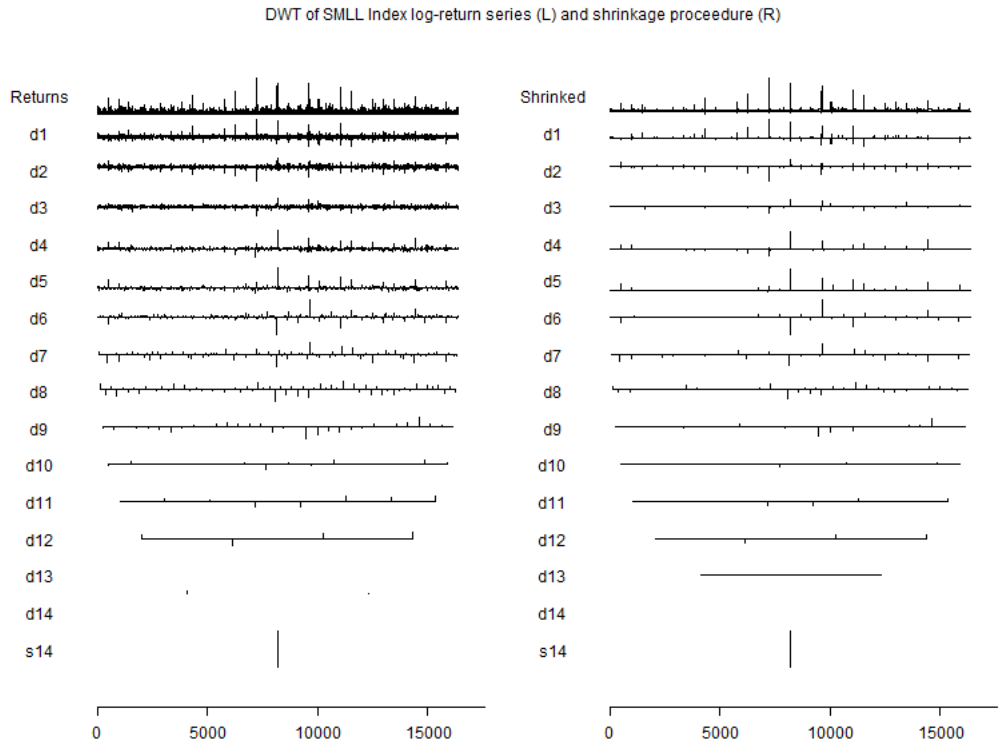

Figure 22: DWT and Shrinkage procedure - SMLL Index log-returns

The jumps series in Figure 22 was obtained using the wavelet thresholding procedure, with the MAD shrinkage function, as stated in the methodology section. In order to attain the jump series, it is applied the inverse discrete wavelet transform to the data. Figure 23 exhibits the identified pattern of jumps in red plotted together with the intraday log-returns of the SMLL Index.

Figure 23 also displays in the bottom a plot of the series after removing those identified jumps. The series still holds the patterns as expected due to the stylized facts of financial time series, meaning that there is volatility clusters and there seems to be conditional heteroskedasticty, with periods of higher variance and periods of smaller variance.

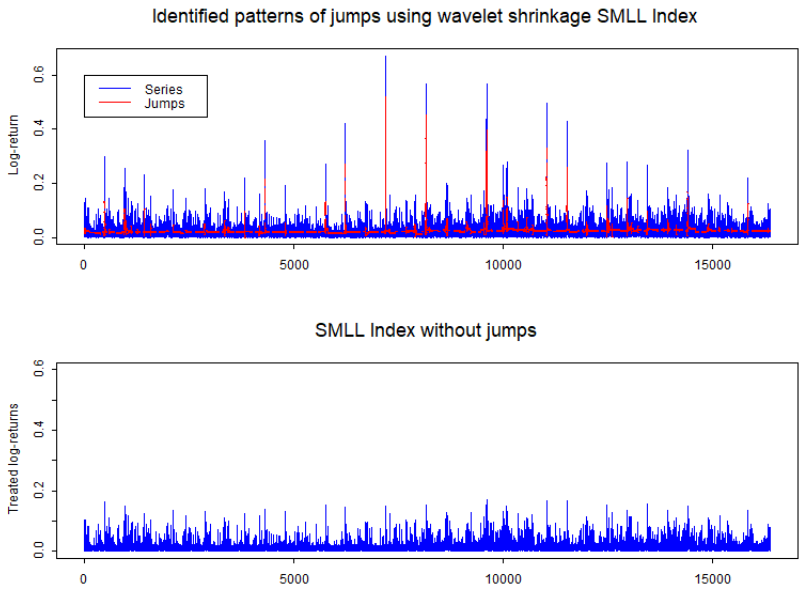

Figure 23: SMLL Index Identified jump series using MAD shrinkage function

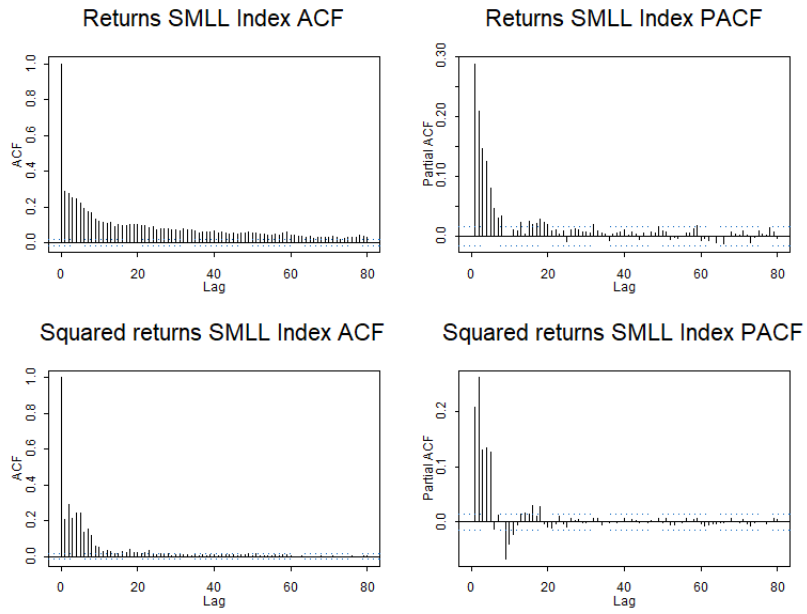

Figure 24: ACF and PACF of SMLL Index log-returns without removing the jumps

Figures 24 and 25 exhibit, respectively, the ACF and PACF of the SMLL Index log-return series without removing the referred to anomalous price changes and after removing the price jumps with the wavelet procedure. As expected, it is possible to see that the additive outliers overshadow the true serial correlation behaviour in the time series, making it appear that there is no long memory in the conditional variance, because most of the structure of the serial correlation lie within the Bartlett limits, making it statistically non-significant at the 5% level.

Additionally, as seen in Figure 24, the presence of additive outliers generates spurious serial correlations, like the negative partial autocorrelation calculated near lag 10 in the conditional variance, which vanishes after removing the jumps with the wavelet procedure, which is coherent to the literature.

If evaluating the ACF and PACF of the series without removing the jumps, one could suggest a GARCH(1,1) or GARCH(2,2) model in order to sufficiently account for the serial dependence structure showed in Figure 24. However, after removing the price jumps, the outshadowed serial correlation structure reappears, as it can be seen in Figure 25, which clearly indicates a pattern of serial autocorrelation, where there are statistically significant autocorrelations still in lag 40, considering a 5% significance level: this suggests that it is necessary to consider long memory models to describe this behaviour.

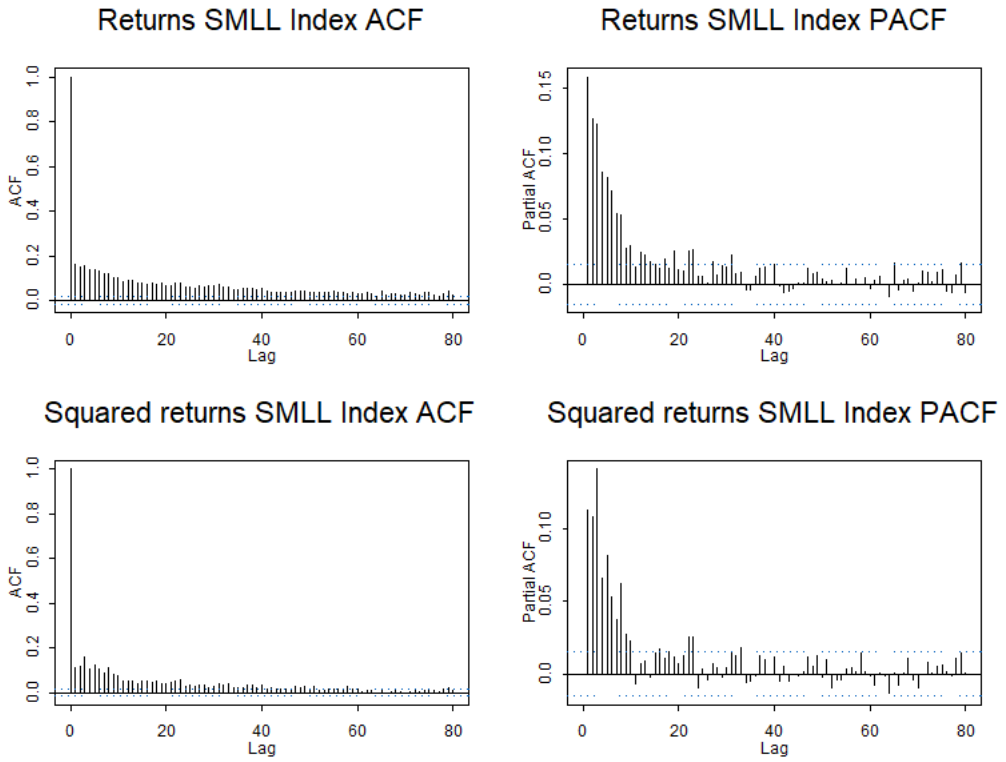

Figure 25: ACF and PACF of SMLL Index treated log-return series

As it was done in the previous section, two models will be considered for comparison of the aforementioned methodology with the traditional approach: a model only removing the jumps using DWT and a model without removing the jumps, compared in terms of RMSE.

A model which suits the data well in this case was the ARFIMA(0,d,2). This was chosen as the most parsimonious one to which there are no remaining statistically significant autocorrelations at the 5% level. The estimated coefficients can be seen in Table 10. Analyzing that, one can see that the ARFIMA model is stationarity and invertible, because $|d|< 0.5$ and the roots of the moving average polynomial lie outside the unit circle. It is also possible to see that there is a strong dependence in terms of long memory, because $d = 0.3137$ indicates a stronger dependence as compared for instance with the one in Table 18, because is closer to 0.5, corroborating the noticed behaviour in Figure 25.

Table 10: ARFIMA coefficients - SMLL Index

Coef.	Value	Std. Error	T stat.	p-value
d	0.3137	0.0110	28.5564	0.0000
MA(1)	0.2302	0.0132	17.3850	0.0000
MA(2)	0.0572	0.0103	5.5475	0.0000

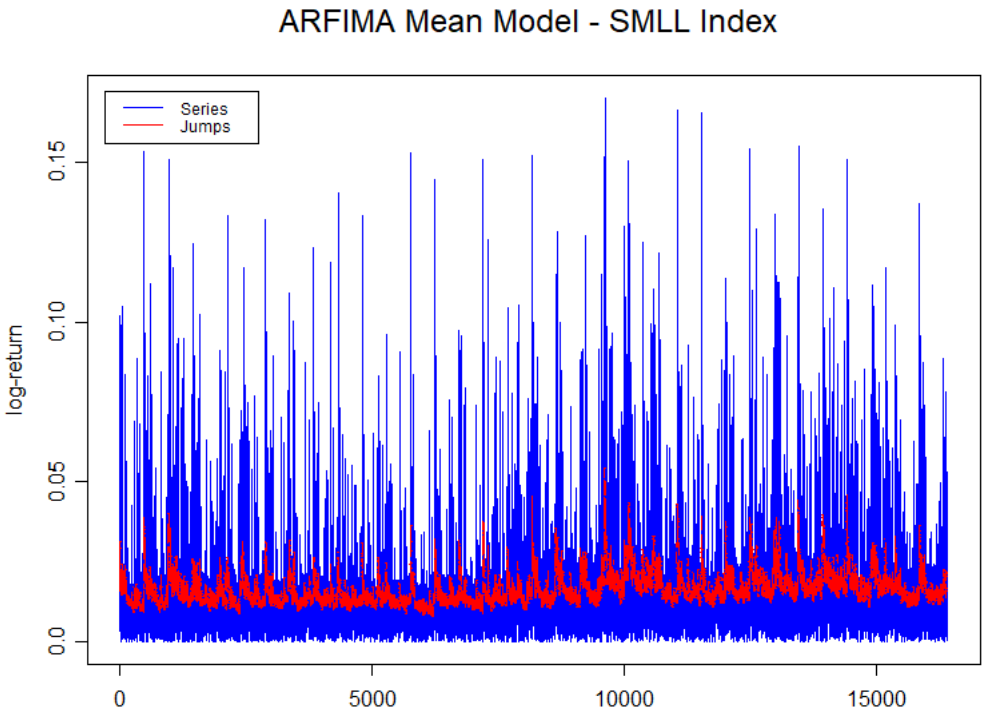

Figure 26: ARFIMA adjusted to the SMLL Index return series

It is also possible to see that the short run dynamics of the series needed to be accounted for, because of the MA(2) order in the model, with statistically significant coefficients at the 5% level. The short-run persistence is mild, evaluating the magnitude of the coefficients for θ_1 and θ_2 in the moving-average model. Figure 26 displays the fitted values of the model versus the log-return series. From that, one can see that, while the ARFIMA model successfully describes the dynamics and effects of shocks in the conditional mean, there seems still to exist an heteroskedastic behaviour which needs to be accounted for. In that case, one can consider long memory GARCH models, to account both for the long-run and short-run dynamics of the time series in the conditional variance.

In order to do so, the FIGARCH(1,d,2) was considered to model the dynamics of the time series. This model was chosen as the most parsimonious, whilst ensuring that all the serial correlation in the squared series was modelled and considered null at the 5%

significance level.

Table 11: FIGARCH coefficients - SMLL Index

	Value	Std. Error	T stat.	p-value
Cst x 10^4	0.0023	0.0001	0.0071	0.9943
ARCH(1)	0.0561	0.0072	7.7895	0.0000
BETA(1)	0.3551	0.0267	13.3179	0.0000
BETA(2)	0.5022	0.0075	66.9818	0.0000
d-Figarch	0.6011	0.0671	8.9517	0.0000
Asymmetry	1.4247	0.0922	15.4582	0.0000
Tail	3.3680	1.3159	2.5594	0.0105

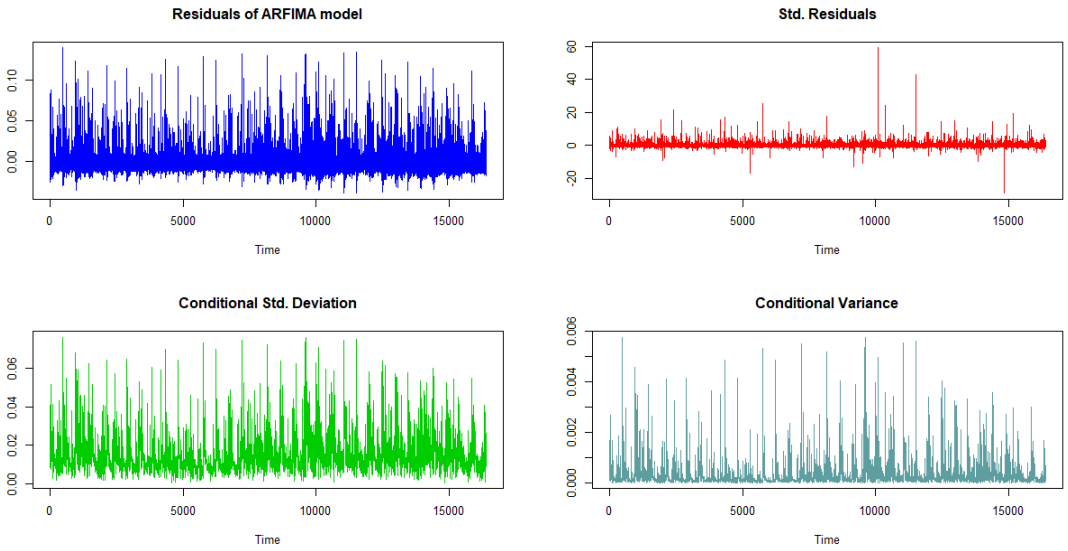

Figure 27: Adjusted FIEGARCH for SMLL Index series with Skewed Student distribution

The model was estimated using the "rugarch" package for R, developed and maintained by Ghalanos (2020). The numerical optimization procedure was the SOLNP, as implemented by Ghalanos and Theussl (2015). The model assumed a skewed Student t distribution for the innovation term to account for the skewness in the standardized residuals of the model.

It is important to point out that, even though the model in Table 11 is not second-order stationary, because $d > 0.5$, the model is stationary in the wide sense, because $0 < d < 1$. This situation of a model not being second-order stationary but strictly stationary also happens with the IGARCH models, as stated by Francq and Zakoian (2019).

As for the positivity condition for the conditional variance, the question remains open for the FIGARCH(p,d,q), when $q \geq 2$. It is also important to note that, even thought the constant is statistically non-significant, it is necessary to keep it in order to ensure the positivity condition for the FIGARCH model, which requires at least that $\alpha > 0$.

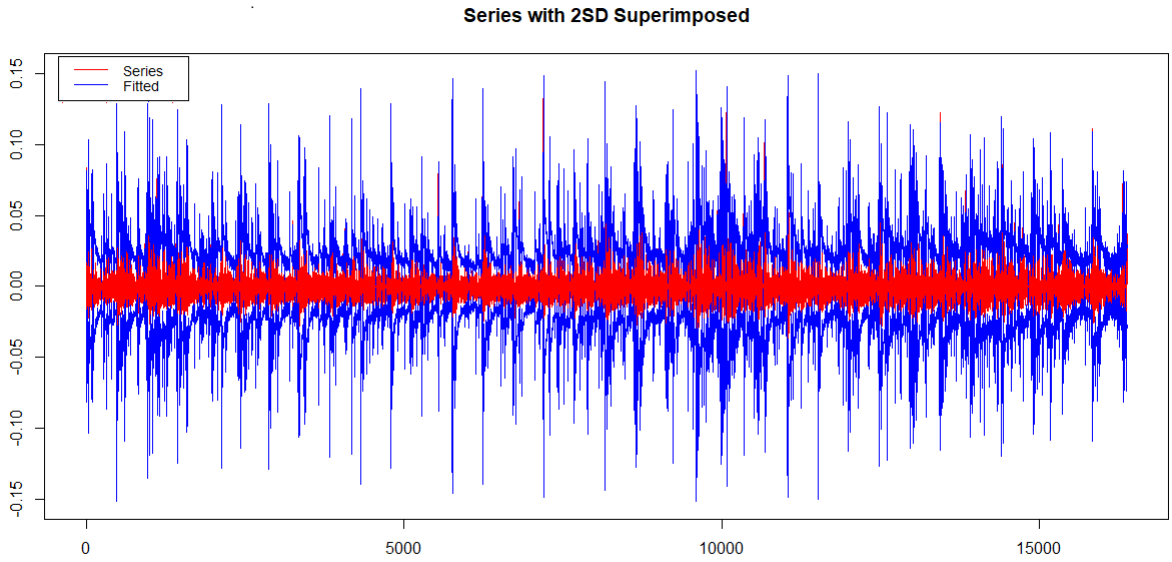

Figure 28: Series with 2 SD Superimposed - SMLL Index

The Figure 27 displays the fitted model's conditional standard deviation, conditional variance and standardized residuals, whereas Figure 28 exhibits the series with 2 conditional SD superimposed. Although there exists jumps in the standardized residuals, the Box-Pierce test indicates that there is no autocorrelations yet to be modelled at the 5% level, as one can see by the Table 12, both for the conditional mean and conditional variance.

From Figure 28 one can see that most of the series' variability lie within the interval of two times the conditional standard deviation. Figure 28 also indicates that the estimated conditional standard deviation was also able to capture the dynamics of the heteroskedasticity in the time series.

From the result in Table 12, it is possible to notice that the model accounts well for the dependence structure in the conditional mean and conditional variance, because at the significance level of 5%, the residuals of the model behave statistically as a white noise.

The RMSE of the model using the wavelet methodology to identify and remove the jumps, calculated by equation 4.2, is evaluated as 0.0009907396. Considering the serial correlation structure exhibited in Figure 24, one can fit an ARFIMA(2,d,2)-EGARCH(1,1)

Table 12: P-value of the Box-Pierce test for the residuals - SMLL Index

Lag	Mean model res.	Var. model std. res.
5	0.4423	0.9991
10	0.5652	0.9928
15	0.7382	0.9999

model using the skewed Student t distribution for the innovations, which produces a RMSE of 0.007944408. Therefore, the model removing jumps with wavelets has smaller RMSE than the traditional approach. This is also exhibited in table 13.

Table 13: Comparison of RMSE of models - SMLL Index

Model	RMSE
Removing jumps	0.00099075
Not removing jumps	0.00794441

In general, from the fitted models and its results displayed in this section, one can affirm that the methodology for modelling the volatility of this log-return series was adequate to account for the serial correlation in the log-return series. The DWT method was also interesting in terms of identification of the jumps.

4.4.3 Models for ETH/USD

Consider in this subsection the ETH/USD time series traded in Gemini Exchange, hereinafter referred only as ETH series, displayed in Figure 29. The time series is sampled every other minute, it starts in 2020-03-01 at 0:00 and ends in 2020-03-12 at 09:03, with $T = 16,384$ observations. An interesting fact is that there is no such thing as "trading hours" in the cryptocurrency exchanges, meaning that the stocks and assets are traded 24 hours a day.

Etherum (ETH) is a cryptocurrency of the Etherum platform and applications, traded in cryptocurrency exchanges around the world. Etherum was founded by Vitalik Buterin in 2015 and, while in the beginning one could only send ETH from an Ethereum account to another, as for now one can use ETH for payments, trade ETH with other tokens and even earn interest on your ETH.

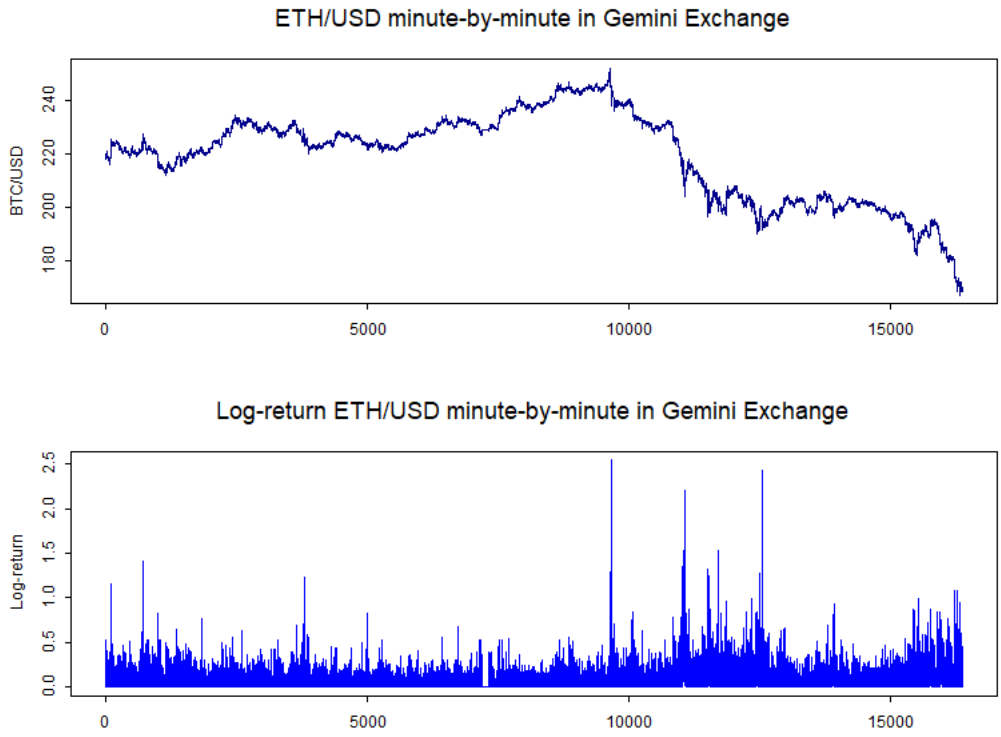

Figure 29: ETH/USD Intraday series and absolute log-returns

In Figure 29 the ETH/USD, around observation 7,200, there is little or no price change, which reflects to something similar to a gap in the log-return series around timestamp 7200. While this might seem like a gap, in fact there was little price change or no price change during the time span of around two hours, from 2020-03-06 00:03 to 2020-03-06 at 01:54 in the Gemini Exchange. This, however, corresponds to no integrity problem of the data, but is a result rather more of the fact that there are no trade hours in the cryptocurrency exchanges, which causes these "anomalies".

Consider the ETH/USD series (first plot) in Figure 29. An economic reason for the downfall in the exchange rate series after March/2020 (around observation 10.000) might be the result of uncertainty generated in financial markets due to the COVID-19 outbreak in the beginning of 2020 and the its spread around the world. Since cryptocurrencies are traded worldwide by different market players and is unregulated, the effect of such events might be noticed much faster than in exchange rates of actual currencies.

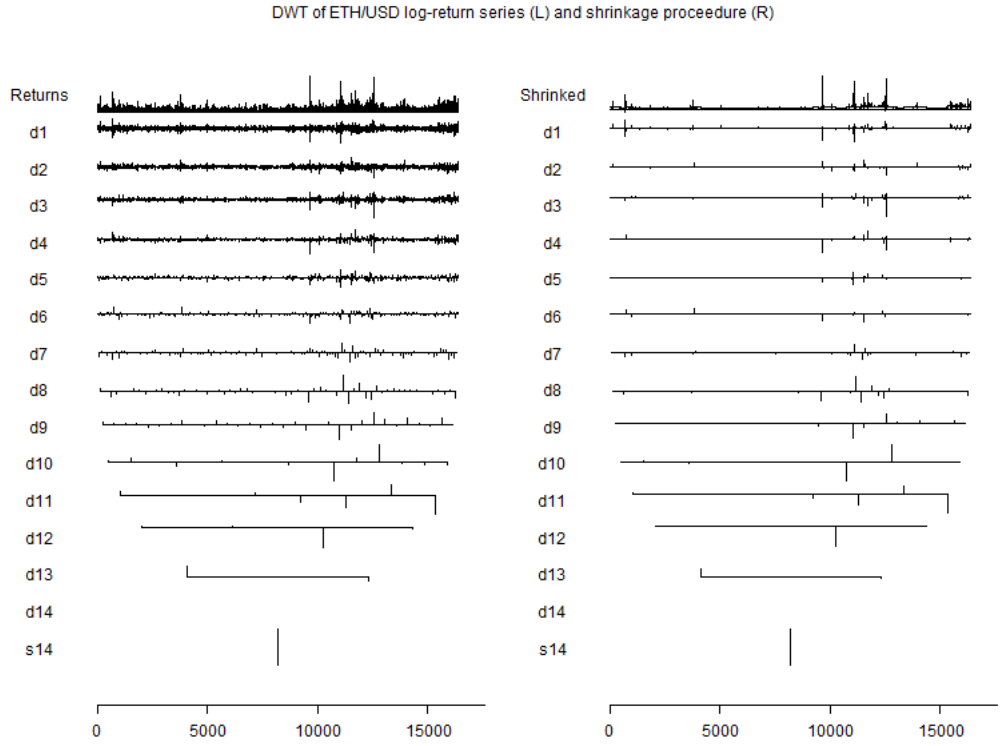

Figure 30: DWT and Shrinkage procedure - ETH/USD log-returns

A previous analysis of the time series will be to perform the DWT, shrinkage and inverse DWT procedures, which is exhibited in the Figure 30: on the left one can see the DWT of the ETH/USD log-return series using the Haar wavelets in the left plot whereas on the right, the result of the shrinkage procedure in the DWT of the series, using the MAD function.

It is also interesting to notice that, since ETH/USD is a cryptocurrency and therefore unregulated by any central bank, the series is expected to be much more volatile than series of exchange rates or even stocks traded in the Brazilian Exchange and OTC. Also, because there is no such thing as a "trade day" in cryptocurrencies, which are traded 24 hours a day, there can be no overnight jumps, as seen in the log-return series of stocks: as a consequence, less price jumps are expected, which is coherent to the identified jumps using the wavelet shrinkage procedure.

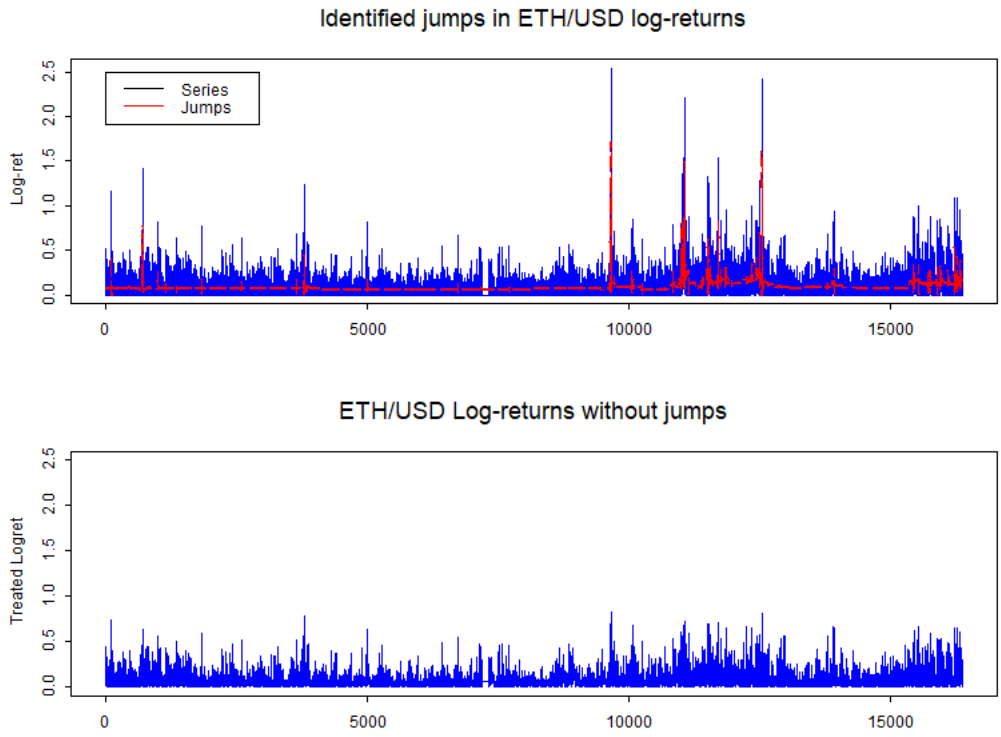

Figure 31: ETH/USD Identified jump series using MAD shrinkage function

Consider then the identified pattern of jumps for the time series as in Figure 31. The jumps are are indicated in red, and the log-return series after removing the price jumps, hereinafter referred only as the returns, are in blue. It is important to notice that, while some of the jumps which are identified by visual inspection were also identified with the wavelet series, as volatility increases in the end of the series, it is possible to see that some of the price changes were considered jumps. This is why the shrinkage and DWT provides an interesting non-parametric method to deal with abrupt changes in the time series.

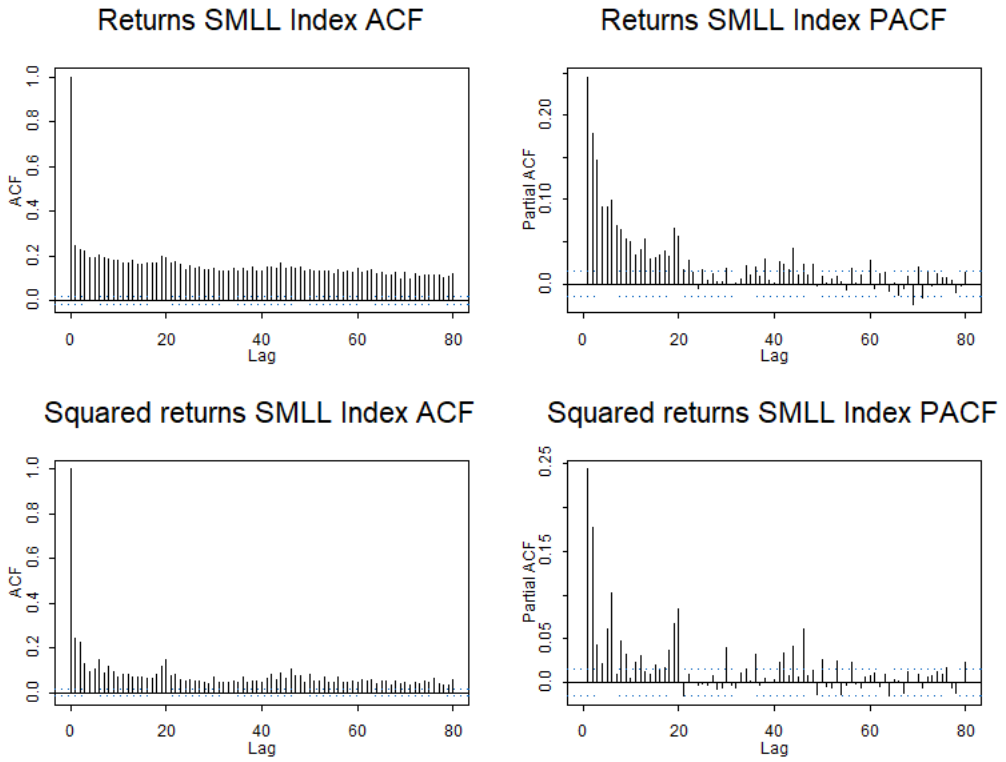

Figure 32: ACF and PACF of ETH/USD log-return series without removing the jumps

Figure 32 and 33 exhibit, respectively, the sample autocorrelations and partial autocorrelations of the ETH/USD log-return series without removing the referred to anomalous price changes and after removing the price jumps with the wavelet procedure. In this case, the presence of these outliers is not enough to completely wash out the serial dependence structure in the conditional mean and conditional variance, as described in the literature referenced in the methodology section. However, the magnitude of the dependence is affected, with a much stronger serial dependence in the conditional variance series in series after removing the jumps. As one can see in Figure 33, the decay is much slower in the PACF of the squared returns, which will affect and outshadow the behaviour of the fractional difference parameter.

Evaluating the ACF and PACF of the series without removing the jumps, one could suggest an ARFIMA-FIGARCH(1,1) model in order to sufficiently account for the serial dependence structure showed in Figure 32. However, after removing the price jumps, the outshadowed serial correlation structure reappears, as it can be seen in Figure 33, it is expected for the d parameter in the FIGARCH model to be significantly greater than the parameter to the fitted model after removing the jumps. This would happen because, as discussed by Caproin (2002), as $d \to 0$, the memory of the process increases.

Following the procedure in the previous sections, two models will be considered as

means of comparison: a model only removing the jumps using DWT and one following traditional approach as in the previous sections. In the end of this section, the RMSE of these models will be compared in order to evaluate the results.

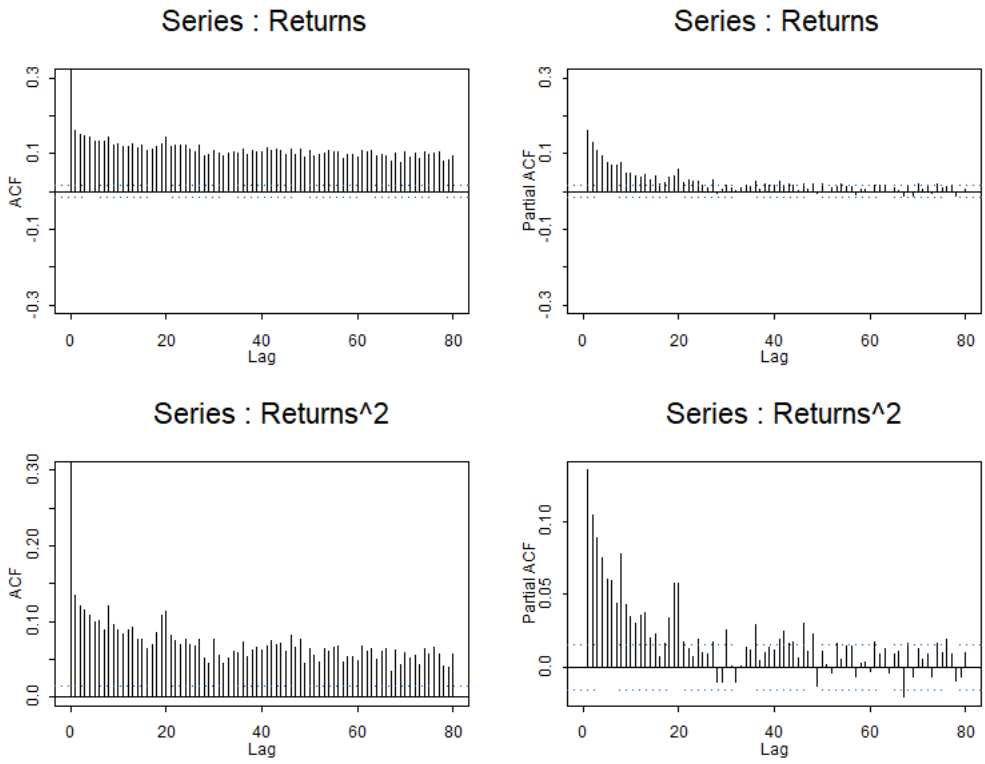

Figure 33: ACF and PACF of ETH/USD treated log-return series

In order to account for the described traits of the return series exhibited in Figure 33, one can consider the ARFIMA(3,d,1) model for the conditional mean. The orders of the model were chosen by ensuring that it is the most parsimonious whilst ensuring that the conditional mean had no autocorrelations left to be modelled, meaning that the autocorrrelations were statistically null at the 5% confidence level. Table 14 shows the coefficients estimated for the model, the standard errors for the coefficients and the T-statistics, from which one can test whether these coefficients are statistically equal to zero.

From the results in Table 14, one can see that the model has a strong dependence both in the long-run dynamics and in the short-run dynamics, due to the large orders of the AR(1) and MA(1) coefficients, and due to $d = 0.1249$ being statistically significant at the 5% level. This explains the slow decay of the ACF of the return series. It is important to mention that the model holds the stationarity and invertibility conditions, because $|d| < 0.5$ and the roots of the autorregressive and moving average polynomials lie outside the unit circle.

Table 14: ARFIMA coefficients - ETH/USD

Coef.	Value	Std. Error	T stat.	p-value
d	0.1249	0.0132	9.4656	0.0000
AR(1)	0.9235	0.0152	60.6018	0.0000
AR(2)	0.0510	0.0123	4.1315	0.0000
AR(3)	0.0235	0.0105	2.2365	0.0253
MA(1)	0.9876	0.0013	770.3636	0.0000

Figure 34 exhibits the ARFIMA model in red plotted versus the actual return series in blue. While the ARFIMA model captures the dynamics of the conditional mean series, even accounting for higher periods of volatility, there is still heteroskedastic behaviour which needs to be accounted for in the model.

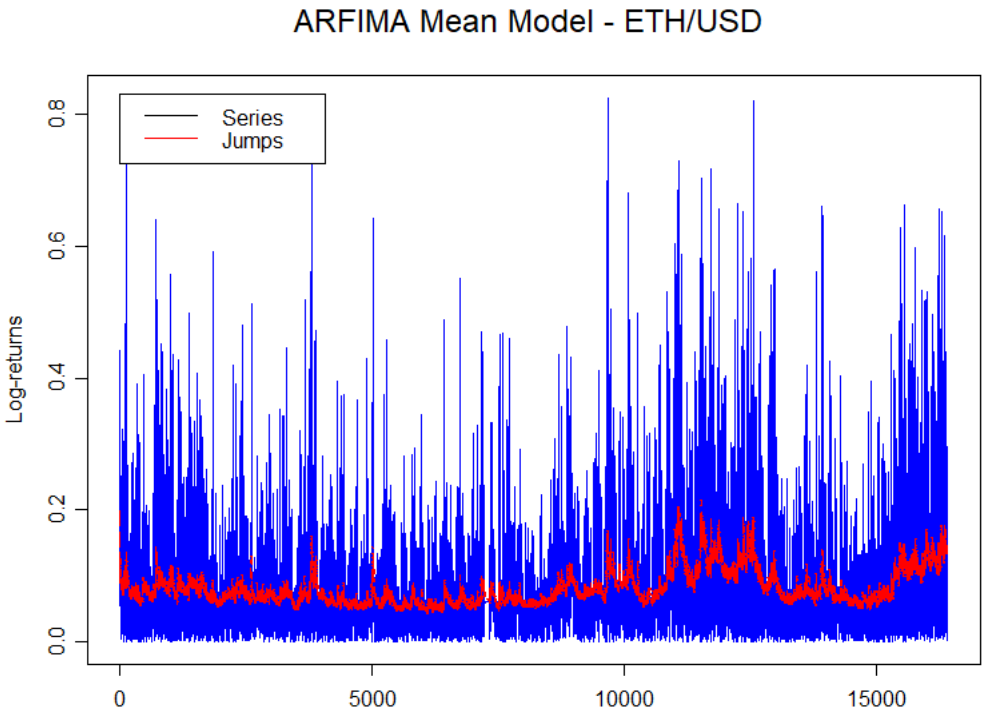

Figure 34: ARFIMA adjusted to the ETH/USD series

To account for the conditional heteroskedasticity and the long memory in the conditional variance, one can consider the FIGARCH(1,d,1). The order of the FIGARCH model was chosen due to the fact that it was that it was the most parsimonious choice which generated standardized residuals which are statistically null at the 5% significance level. The model was estimated using the OxMetrics-G@RCH^{TM} software developed by Laurent (2018), using the BFGS algorithm for numerical optimization, with its imple-

mentation in OxMetrics-G@RCH^{TM}.

Table 15: FIGARCH coefficients - ETH/USD

Coef.	Value	Std. Error	T stat.	p-value
Cst x 10^4	1.5421	0.3186	4.8410	0.0000
d-Figarch	0.2427	0.0138	17.5600	0.0000
ARCH(1)	0.3691	0.0343	10.7500	0.0000
GARCH(1)	0.5642	0.0338	16.6700	0.0000
Asymmetry	1.5188	0.0967	15.6700	0.0000
Tail	3.4415	0.1376	25.0200	0.0000

Table 15 displays results for the estimated model. The innovations term of the model was chosen to be a skewed student distributions, and the estimated parameters were $\xi = 0.4131$ and $\nu = 3.4415$, in order to account for the skeweness and excess kurtosis. Analyzing the results from the estimated values in Table 15, it is possible to state that the positivity constraint for the FIGARCH (1,d,1) model is observed, because $0.321439 < 0.36911 < 0.585751$ and $-0.00231014 < 0.0268956$. The model is also stationary, because it follows the conditions in Bollerslev and Mikkelsen (1996) and $d < 0.5$. It is also worth to note that the estimated values are statistically significant at the 5% level.

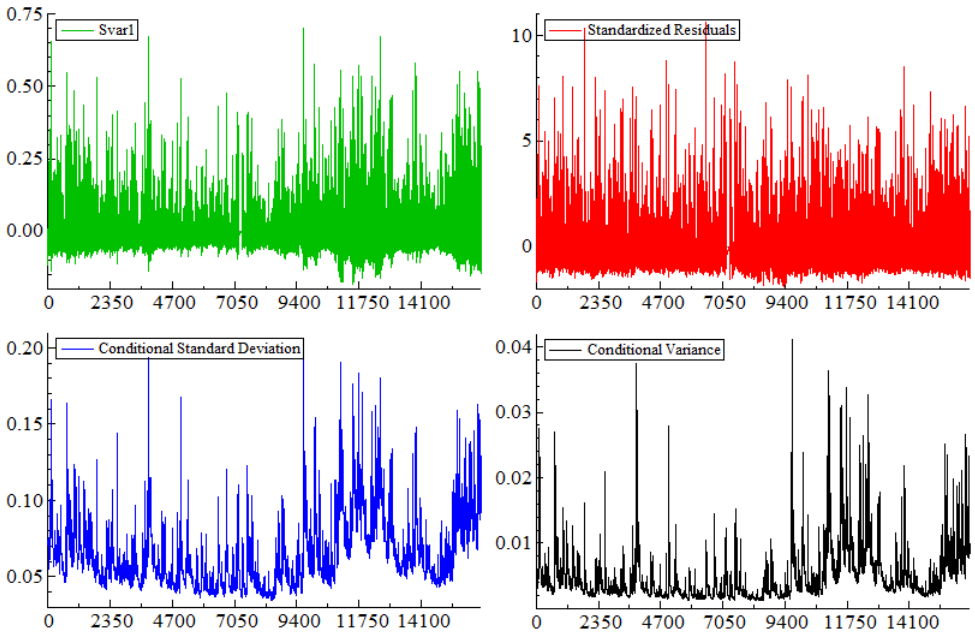

Figure 35: Adjusted FIGARCH for ETH/USD series with Skewed Student distribution

Figure 35 exhibits the conditional standard deviation and conditional variance estimated considering the FIGARCH(1,d,1) model in Table 35. It is possible to see, by

comparing the conditional standard deviation to the time series, that the volatility is captured successfully, because in periods of higher volatility clusters, the conditional standard deviation follows the pattern of increased volatility.

Figure 36 displays the series with 2 conditional standard deviation superimposed. It is possible to see that most of the observed series lie within the interval of two conditional SD, and that the periods of increased volatility are well described by the estimated conditional variance. This indicates that the model coherently describes the dynamics of this time series.

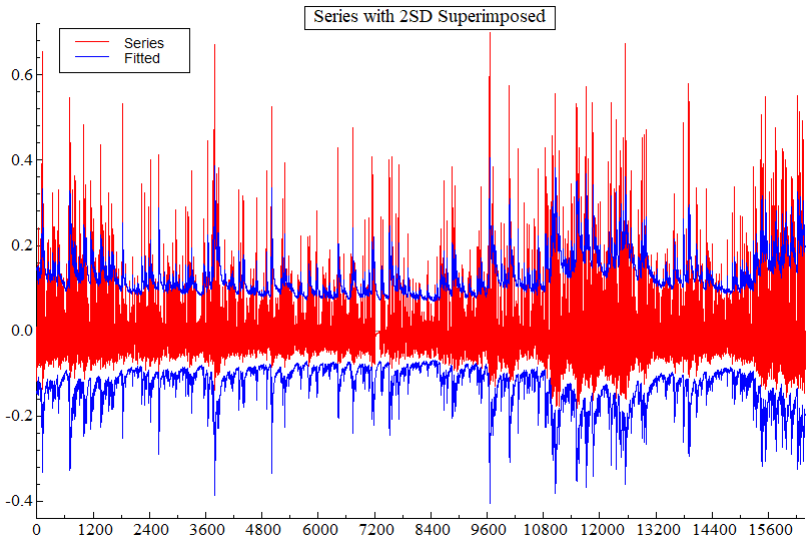

Figure 36: Series with 2 SD Superimposed - ETH/USD

Another measure to evaluate the fit of the model is the p-value of the Box-Pierce test performed in the residuals of the model. From that, one concludes that the model accounts coherently for the dynamics of the time series at the 5% significance level for both the conditional mean and conditional variance.

Table 16: P-value of the Box-Pierce test for the residuals - ETH/USD

Lag	Mean model res.	Var. model std. res.
5	0.9339	0.0863
10	0.5985	0.0823
15	0.7600	0.1182

The RMSE of the model using the wavelet methodology to identify and remove the jumps, calculated by equation 4.2, is evaluated as 0.02024787. Considering the serial correlation structure exhibited in Figure 32, one can fit an ARFIMA(3,d,3)-IGARCH(1,1) model using the skewed Student t distribution for the innovations or an ARFIMA(3,d,3)-FIGARCH(1,d,1). Respectively, their RSME calculated by equation 4.1 is given by 0.0793478

and 0.0789962, as shown in table 17. Therefore, the model removing jumps with wavelets has a significantly smaller RMSE than the one generated by the traditional approach.

Table 17: Comparison of RMSE of models - ETH/USD

Model	RMSE
Removing jumps	0.0202479
Not removing jumps (1)	0.0793478
Not removing jumps (2)	0.0789962

By assessing the overall results of the fitted models and other results displayed in this section, it is possible to state that methodology for modelling the volatility of this log-return series was adequate to account for both the serial dependence in the the jump series and the log-returns after removing the jumps.

4.4.4 Models for VALE3

For this subsection, the stock considered is the VALE3 stocks traded in the Brazilian Exchange and OTC, homogeneously sampled minute-by-minute, starting starts in 2020-01-02 10:03:00 and ending in 2020-02-20 10:08:00, which corresponds to approximately one and a half month of trade days (36 days) and $T = 16,384$ points.

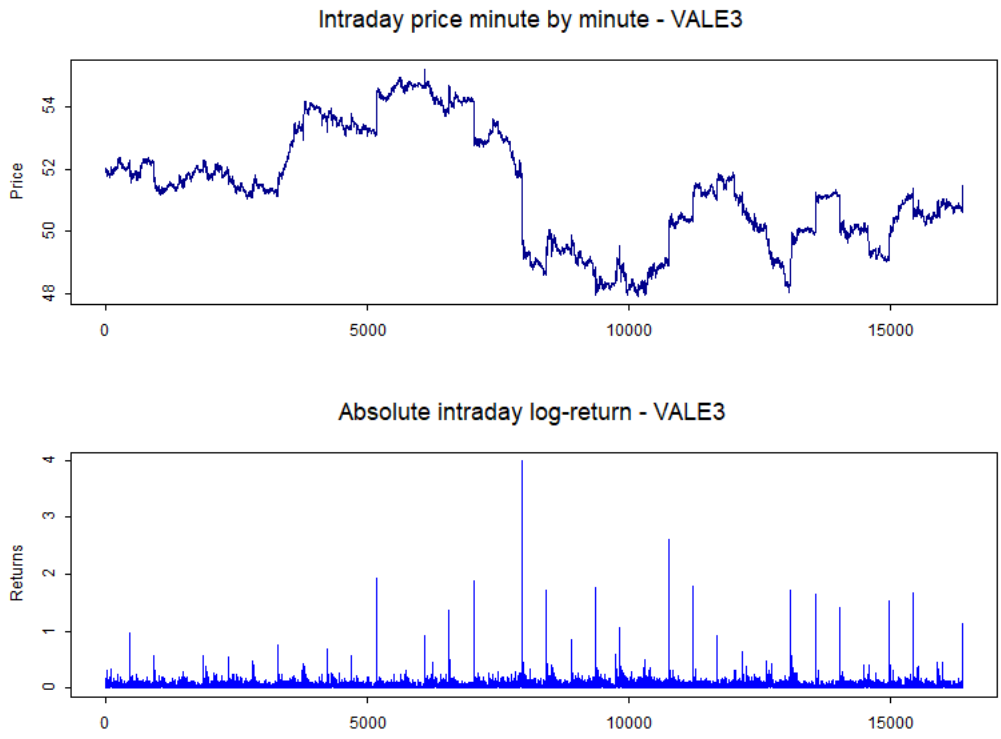

Figure 37: VALE3 Intraday price series and absolute log-returns

The symbol VALE3 corresponds to the ordinary shares of Vale S.A traded in the Brazilian Exchange and OTC (B3) since 2007. Vale S.A. is a corporation headquartered in Rio de Janeiro, Brazil, engaged mainly in the metals and mining business, as well as logistics and solutions for ports and container terminals. The corporation is the largest producer of iron ore and nickel in the world and has operations covering more than 30 countries. It was established in 1942, named "Companhia Vale do Rio Doce" (shortly referred to as CVRD) prior to 2007, after which the company began a rebranding strategy and changed its name to "Vale".

The Figure 37 exhibits two plots: the upper one is the intraday prices of VALE3 in the aforementioned timespan and the second one is the intraday log-returns. The absolute log-return series also display the traits expected in financial time series, such as the presence of volatility clusters, as in the previous cases. It is possible to see that the largest returns are around 2.5% from one time stamp to the other, up to 4%. Usually these anomalous returns correspond to overnight price changes, as it happens in this dissertation with other stocks traded in the Brazilian Exchange and OTC. Table 18 displays the largest values of r_t for VALE3. As one can see, all of those correspond to overnight changes in prices and 15 out of 20 correspond roughly to a percent change of around 1% or 2% in the prices from one time stamp to the upcoming.

Table 18: VALE3 absolute log-returns - top 20

Ticker	Date	Time	Price	Abs.logret
VALE3	20200127	100800	49.7544	3.9816
VALE3	20200204	100700	50.2934	2.6172
VALE3	20200117	101300	54.1236	1.9209
VALE3	20200123	100900	53.1708	1.8826
VALE3	20200205	101100	51.2076	1.7824
VALE3	20200130	101100	47.9837	1.7693
VALE3	20200128	100800	49.4561	1.7272
VALE3	20200211	101000	49.1385	1.7186
VALE3	20200218	100800	50.5436	1.6804
VALE3	20200212	100900	50.9285	1.6577
VALE3	20200217	100300	49.8410	1.5371
VALE3	20200213	100900	50.3511	1.4233
VALE3	20200122	100900	54.6625	1.3649
VALE3	20200220	100600	51.4001	1.1297
VALE3	20200131	100400	49.0423	1.0541
VALE3	20200103	100900	51.7754	0.9803
VALE3	20200121	101000	54.6914	0.9283
VALE3	20200206	100800	51.4675	0.9205
VALE3	20200129	100800	49.6967	0.8557
VALE3	20200113	100500	51.7465	0.7654

In Figure 39, the absolute log-returns with the identified pattern of jumps and the series after removing the jumps from the original time series are shown. This jump series was obtained using the wavelet shrinkage procedure described in the methodology section, using the universal rule. The shrinkage function used was the L_2 norm, in the LP parameterization proposed by Bruce and Gao (1996).

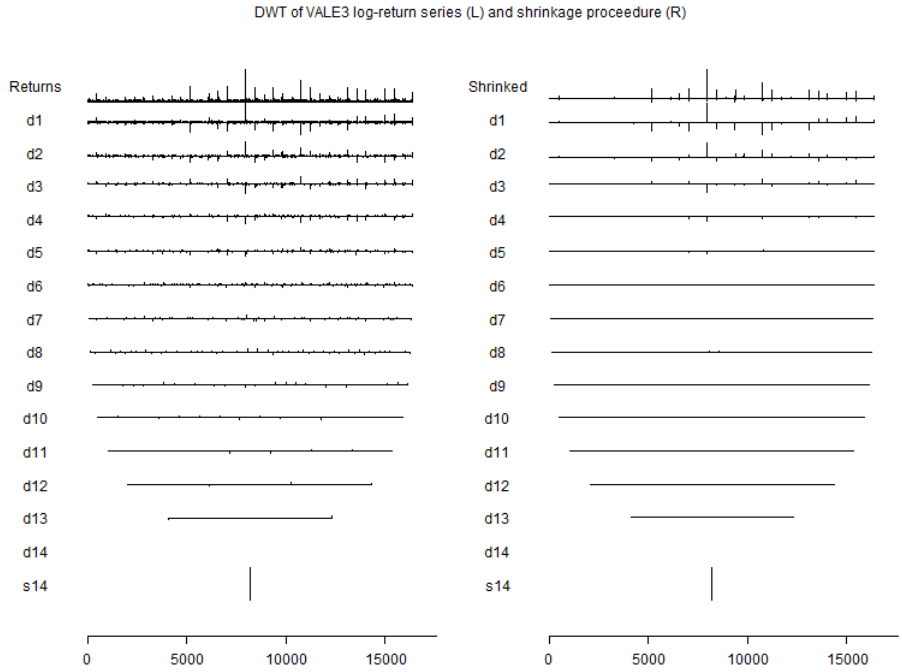

Figure 38: DWT and Shrinkage procedure - VALE3 log-returns

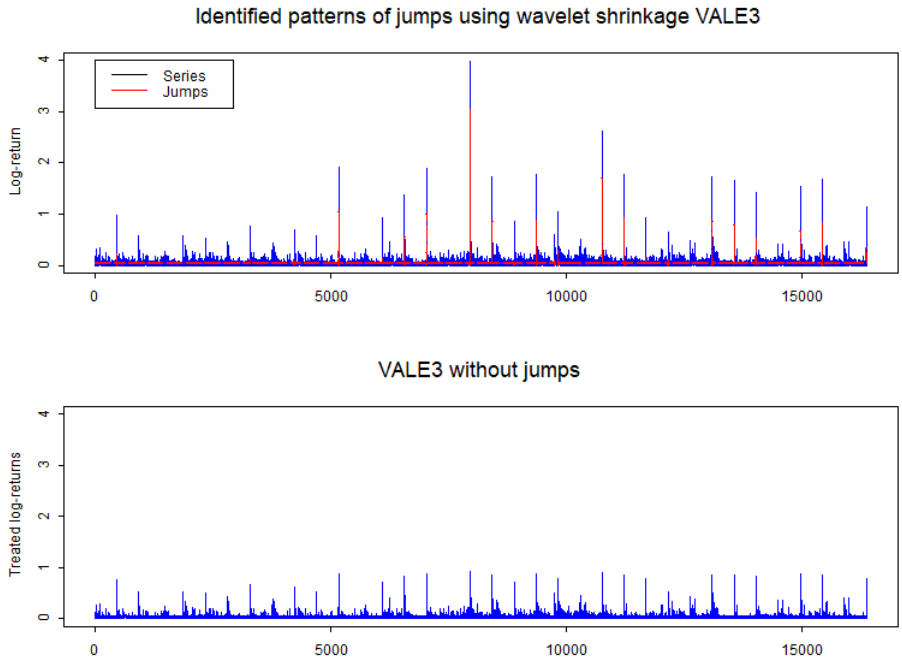

Figure 39: VALE3 Identified jump series using L2 norm shrinkage function

The plots in Figure 38 exhibit the results of both the DWT using the Haar wavelet and the shrinkage function using the L_2 norm. On the left side, one can see the discrete wavelet transform of the VALE3 log-returns and on the right side the wavelet coefficients after the shrinkage procedure is applied. The jump series is then generated after applying the inverse DWT to the shrinked data.

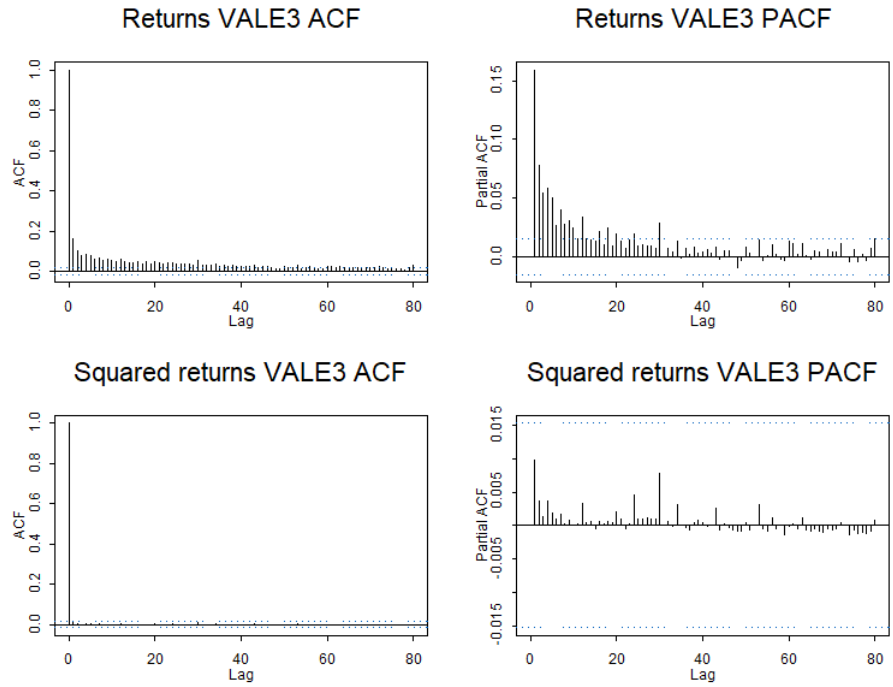

Figure 40: ACF and PACF of VALE3 treated log-return series without removing jumps

Figures 40 and 41 exhibit the ACF and PACF of the VALE3 log-return series without removing the referred anomalous price changes and after removing the price jumps with the wavelet procedure, respectively. In this case, the presence of these outliers was enough to completely dominate the serial dependence structure in the conditional variance, as Mendes (2000) and Carnerno, Peña and Ruiz (2004) describe.

Indeed, evaluating the autocorrelation and partial autocorrelation of the squared return series, as in Figure 40, there seems to exist no serial dependence, because all the calculated partial autocorrelations and autocorrelations lie within the Bartlett limits, thus indicating that, at 5% significance level, they are statistically null. However, after removing the jumps with the wavelet procedure, as one can see in Figure 41, the decay is indeed much slower in the ACF and PACF of the squared returns, which are statistically significant at the 5% level up to lag 10.

If evaluating the ACF and PACF of the series without removing the jumps, one could suggest the usage only of an ARFIMA mode in order to sufficiently account for the serial dependence structure showed in Figure 40. However, after removing the price jumps, the outshadowed serial correlation structure reappears, as it can be seen in Figure 41, it is expected to fit an ARFIMA-GARCH(1,1) model to the time series.

In order to compare the proposed methodology with the traditional approach, one can consider two models: a model only removing the jumps using DWT and a model following the traditional approach, considering no serial dependence in the conditional

variance.

Consider then the log-return series after removing the identified jumps, referred to as the return series. Figure 41 shows the ACF and PACF for the log-return series after removing the jumps, which is hereinafter only referred to as the returns, and for the squared returns. In that case, it is possible to see that the ACF and PACF present the traits of a long-range dependence in the return series: in fact, the ACF has a slow, seemingly hyperbolic decay, and there are non-null partial autocorrelations at the significance level of 5% in lags close to 30. This indicates that long memory models for the conditional mean can be considered in order to deal with this particular correlation structure.

It is interesting to mention that, even though there seems to exist serial correlation in the conditional variance, the dependence does not seem as strong as in the conditional mean. In fact, it is expected that the conditional variance model should have a less strong dependence than the conditional mean model.

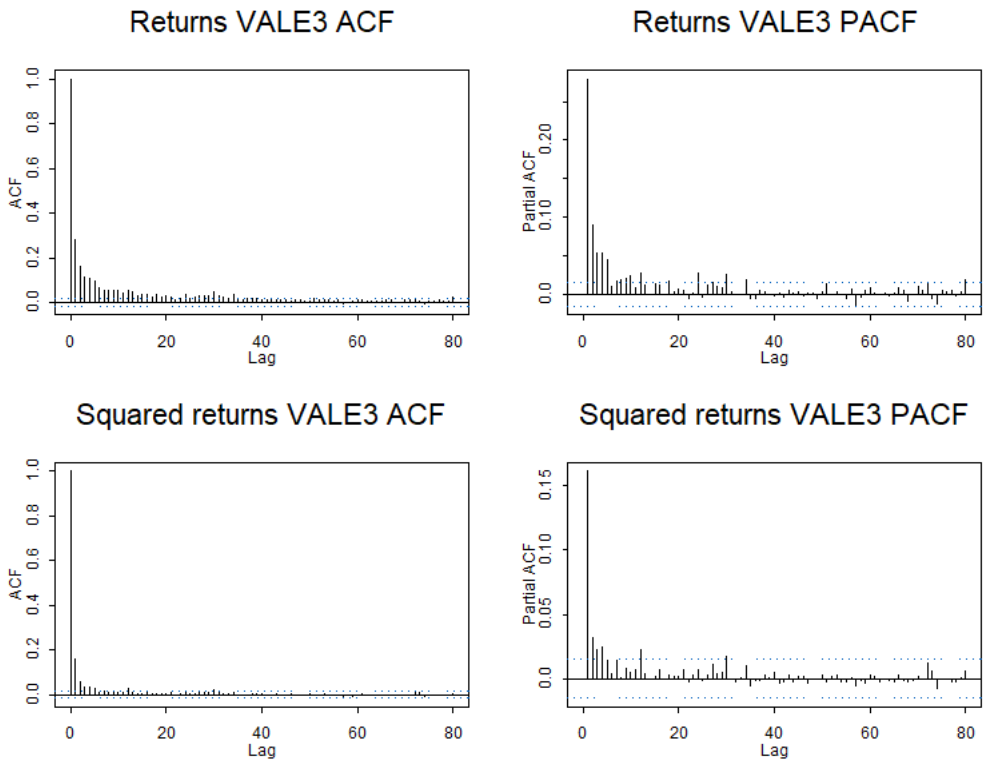

Figure 41: ACF and PACF of VALE3 treated log-return series

To assess if it is necessary to consider a long memory conditional variance model for this particular time series, one can perform the long memory tests in the squared time series, for instance the GPH test and the rescaled range test, which are implemented in the S+FinMetrics modules for S-PLUS. The result for the Modified R/S Test is a test statistics of 1.3888, which indicates that there is no evidence to reject the null-hypothesis

of no long-term dependence at the significance level of 5%. The result for the GPH Test is an estimated value of $d = 0.0348$, with a test statistic of 0.5714. This indicates that one cannot reject the null hypothesis of $d = 0$ at the 5% level. As a consequence, for the conditional variance model, there is no need to consider the long-memory model, only for the conditional mean.

In order to account for the serial dependence in the conditional mean, a model which can be considered is the ARFIMA(1,d,0). This model is chosen as the most parsimonious while ensuring that there are no significant autocorrelations at the significance level of 5%. The parameters for the ARFIMA model and other useful information are exhibited in Table 19.

Table 19: ARFIMA coefficients - VALE3

Coef.	Value	Std. Error	T stat.	p-value
d	0.1888	0.0315	6.0008	0.0000
AR(1)	-0.3369	0.0190	-17.7328	0.0000
AR(2)	-0.3112	0.0113	-27.5470	0.0000
MA(1)	-0.3875	0.0333	-11.6343	0.0000
MA(2)	-0.3305	0.0179	-18.4988	0.0000

From the results in Table 19, it is possible to see that the model is stationary and invertible, because $|d| < 0.1888$ and the roots of the autorregressive and moving average polynomials lie outside the unit circle. In fact, while the magnitude of the long-run dependence is strong (measured by the d parameter), the ARMA parameter indicates a moderate to strong short-run dependence in the conditional mean for the treated log-returns of these time series.

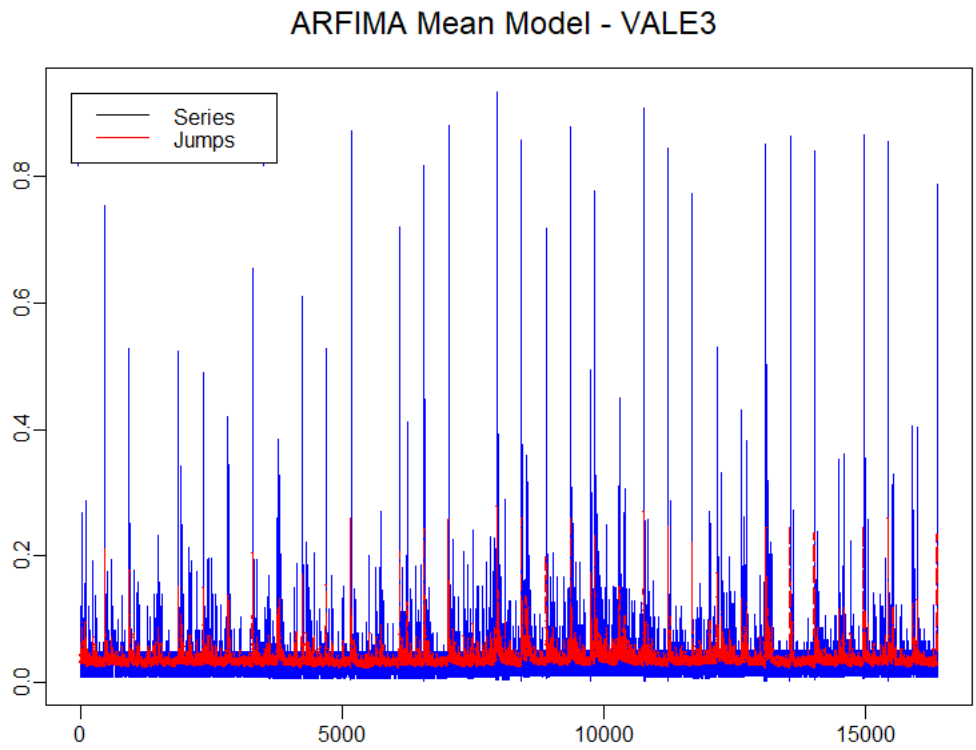

Figure 42: ARFIMA adjusted to the VALE3 series

Figure 42 displays the fitted model for the conditional mean in red, plotted versus the residuals time series. From that, it is possible to see that the model successfully describes the correlation structure of the mean model. However, there is still conditional heteroskedasticity in the series which needs to be dealt with.

To account for the conditional heteroskedasticity in the series, one can consider the EGARCH(1,1) model. This was chosen not only as the most parsimonious model, but also as means to account eventually for asymmetry effects which might be observed in this time series. The parameters for the model were estimated using the R package rugarch developed by Ghalanos (2020).

Table 20: EGARCH coefficients - VALE3

Coef.	Value	Std. Error	T stat.	p-value
Cst	-0.851	0.004	-192.847	0
ARCH(1)	0.204	0.005	37.593	0
GARCH(1)	0.876	0.0003	2923	0
GAMMA	0.106	0.006	19.166	0
Asymmetry	3.143	0.049	64.682	0
Tail	4.076	0.158	25.838	0

The numeric optimization procedure used was the GOSOLNP, as proposed by Hu, Shonkwiler and Spruill (1994) and implemented in R by Theussl and Ghalanos (2015). The coefficients estimated for the model can be seen in Table 20. In the EGARCH model estimated, the γ represents the parameter to measure the asymmetrical leverage effect.

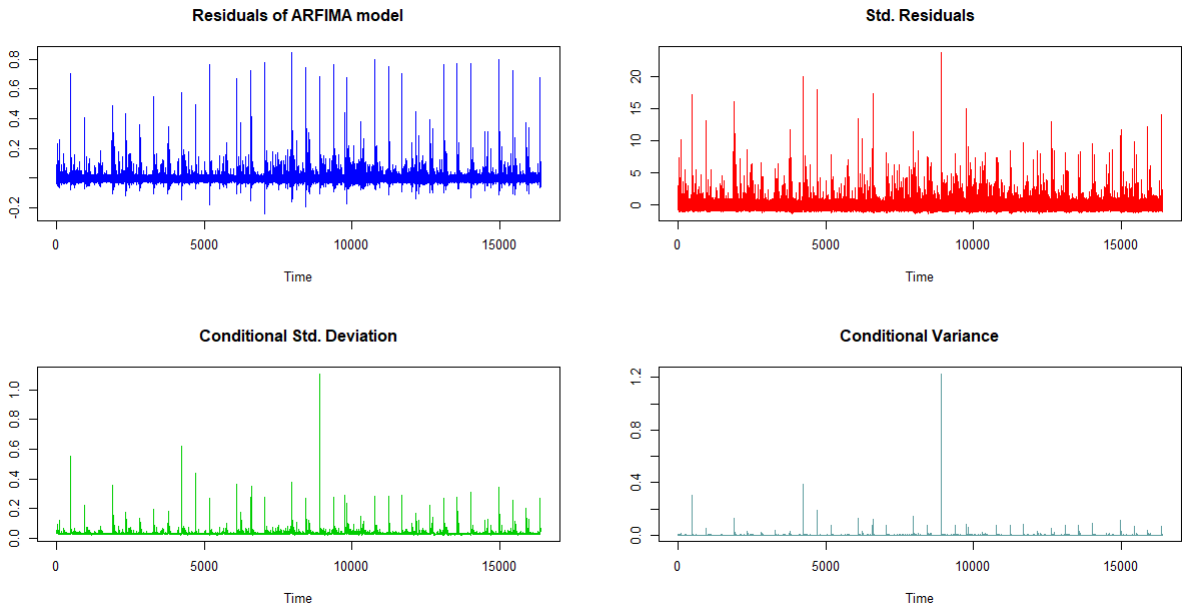

Figure 43: Adjusted EGARCH for VALE3 series with Skewed Student distribution

From the results in the table, one can see that all the coefficients are statistically significant at the 5% significance level. It is also possible to see that the coefficients for the EGARCH model (namely GAMMA) are statistically significant, which means that the model incorporates the asymmetry leverage effect. It is also worth mentioning that the EGARCH model always fulfills the positivity constraint for the conditional variance and can have negative coefficients.

Figure 43 exhibits the estimated conditional variance, conditional standard deviation and standardized residuals for the EGARCH(1,1) model with coefficients as in Table 20. The conditional standard deviation, as expected in the EGARCH model, incorporates a leverage effect. When comparing the conditional standard deviation for the EGARCH model with the observed time series, it is possible to see that most of the observed points of the series lie within the interval of two times the conditional standard deviation, as in Figure 44, which indicates that the model can successfully describe the volatility of the series.

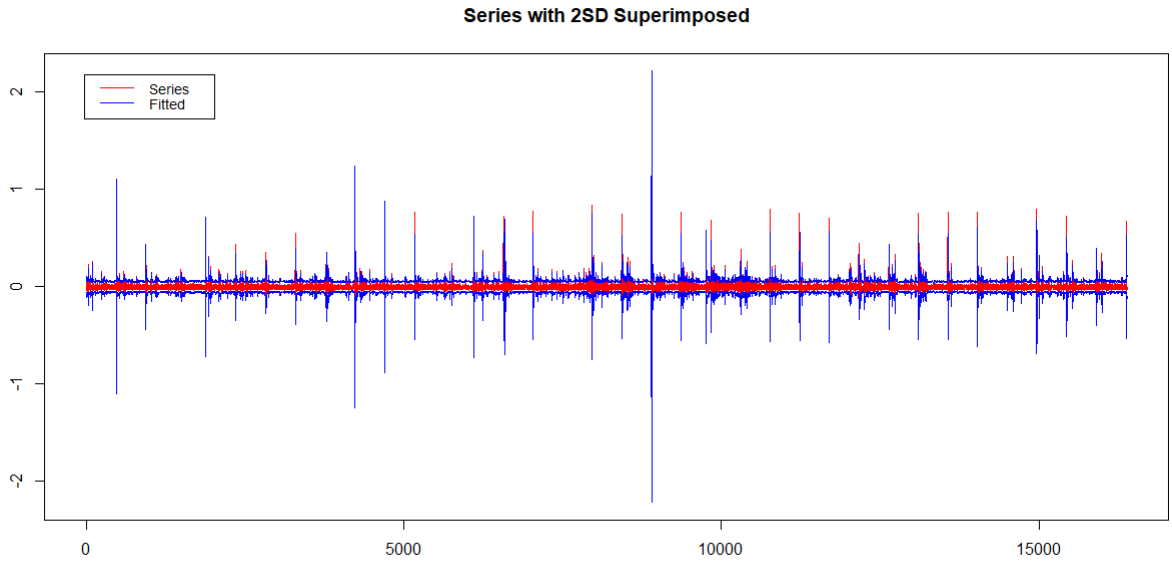

Figure 44: Series with 2 SD Superimposed - VALE3

Table 21: P-value of the Box-Pierce test for the residuals - VALE3

Lag	Mean model res.	Var. model std. res.
5	0.2156	0.64445
10	0.2737	0.9849
15	0.2695	0.9956

In order to assess whether the proposed model accounted coherently for the heteroskedasticity and the serial correlation observed in the series, one can perform the Box-Pierce test in the standardized residuals of the model, as displayed in Table 21. From that, one can state that there are no remaining serial correlations yet to be modelled which are statistically non-null at the 5% significance level for both the conditional mean and conditional variance.

Table 22: Comparison of RMSE of models - VALE3

Model	RMSE
Removing jumps	0.0235208
Not removing jumps	0.1501252

Consider the results exhibited in 22. The RMSE of the model using the wavelet methodology to identify and remove the jumps, calculated by equation 4.2, is evaluated as 0.02352077. Considering the serial correlation structure exhibited in Figure 24, one can fit an ARFIMA(1,d,1) model. In that case, the RMSE is calculated as 0.1501252.

A few comments are worth regarding the RMSE of the models and the choice of the model for the traditional approach. In order to ensure that there were no ARCH effects yet to be modelled, it was performed an LM Test for ARCH effects, as described by Zivot and Wang (2005). The resulting test statistics was 3.5761 and the p-value was close to 1, which is a strong evidence for no ARCH effects. Moreover, a Ljung-Box test on the squared residuals indicates no serial correlation which is statistically significant at 5% level. This is an example of the effects of jumps or additive outliers in $\hat{\rho}(j)$.

Evaluating the results of each of the models and other results discussed in this section, it is possible to state that, in general, the methodology for the log-return series was adequate to account for the serial dependence in the log-return series after removing the jumps. The usage of wavelets for the identification of the jumps resulted to be an interesting application.

4.4.5 Models for BOVA11

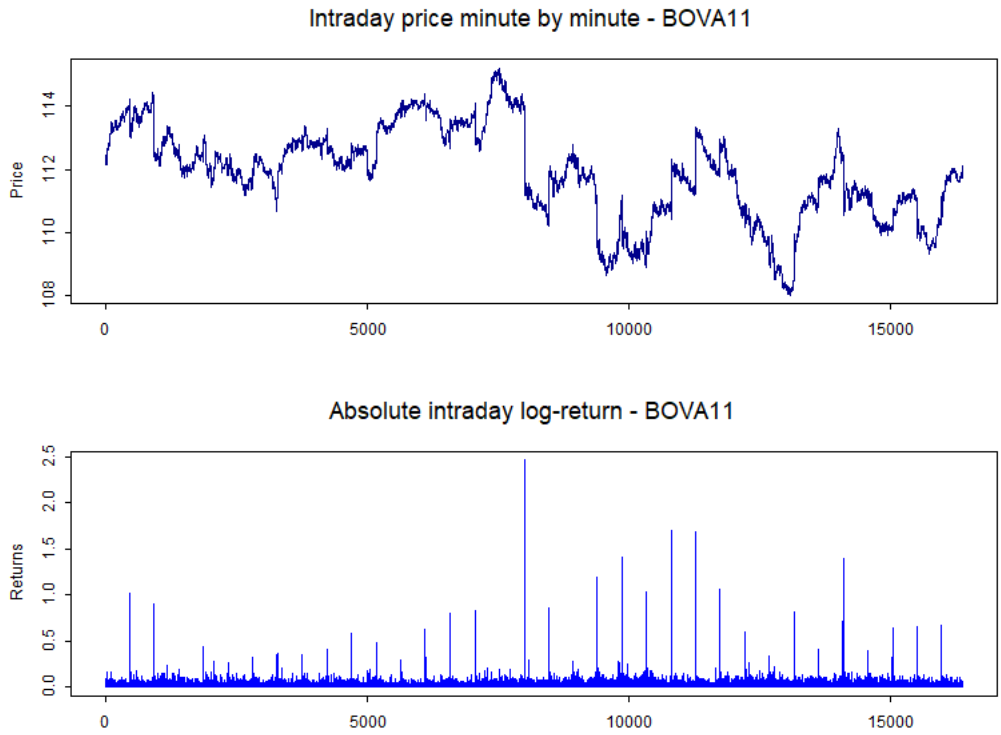

Figure 45: BOVA11 Intraday price series and absolute log-returns

For this subsection, consider the BOVA11 series, which is homogeneously sampled minute-by-minute. The series starts in 2020-01-02 10:03:00 and ends in 2020-02-20 10:08:00, which corresponds to approximately one and a half month of trade days (36 days) and $T = 16,384$ points.

The symbol BOVA11 correspond to the exchange traded fund (ETF) traded in the Brazilian Exchange and OTC (B3), which aims to replicate the behaviour of the Ibovespa. The ETF is a stock portfolio with more than 60 stocks and is managed by BlackRock Asset Management. Among the ETFs traded in B3, BOVA11 is the most traded in terms of trade volume.

Table 23: BOVA11 absolute log-returns - top 10

Ticker	Date	Time	Price	Abs.logret
BOVA11	20200127	101400	111.2300	2.4598
BOVA11	20200204	100700	112.3100	1.6972
BOVA11	20200205	100500	113.1000	1.6852
BOVA11	20200131	100500	109.6300	1.4129
BOVA11	20200213	100500	110.8900	1.3970
BOVA11	20200130	100500	109.5600	1.1976
BOVA11	20200206	100600	112.9600	1.0591
BOVA11	20200203	100700	110.0300	1.0323
BOVA11	20200103	101000	113.0900	1.0118
BOVA11	20200106	100400	112.7800	0.9004

The plot in the Figure 45 exhibits the intraday price minute-by-minute and the intraday log-returns. There are significant price jumps, which reflect in the absolute log-return series. It also displays the volatility clustering pattern which is expected for series with conditional heteroskedasticity.

As for the price jumps, consider the Table 23, which exhibits the 10 highest absolute log-returns in the period analyzed. From the column "Time" in the table, it is possible to state that all of these price jumps happened during opening auctions, meaning that these were overnight jumps, which is a common aspect for intraday log-return series of stocks traded in the Brazilian Exchange and OTC. It is also interesting to state that 9 out of 10 among the greatest absolute log-returns are roughly above 1% from one time stamp to the other. It is expected, from visual inspection and from Table 23 that the jumps should be values around 1% or bigger.

In order to identify the price jumps for the BOVA11 series, we shall consider the wavelet shrinkage procedure described in the methodology. Figure 46 displays two different plots of results related to the wavelet identification procedure. The first one, on the left, exhibits the discrete wavelet transform of the BOVA11 log-returns using the Haar wavelet, whereas the second one on the right, the wavelet coefficients after the threshold procedure is applied, with MAD threshold function.

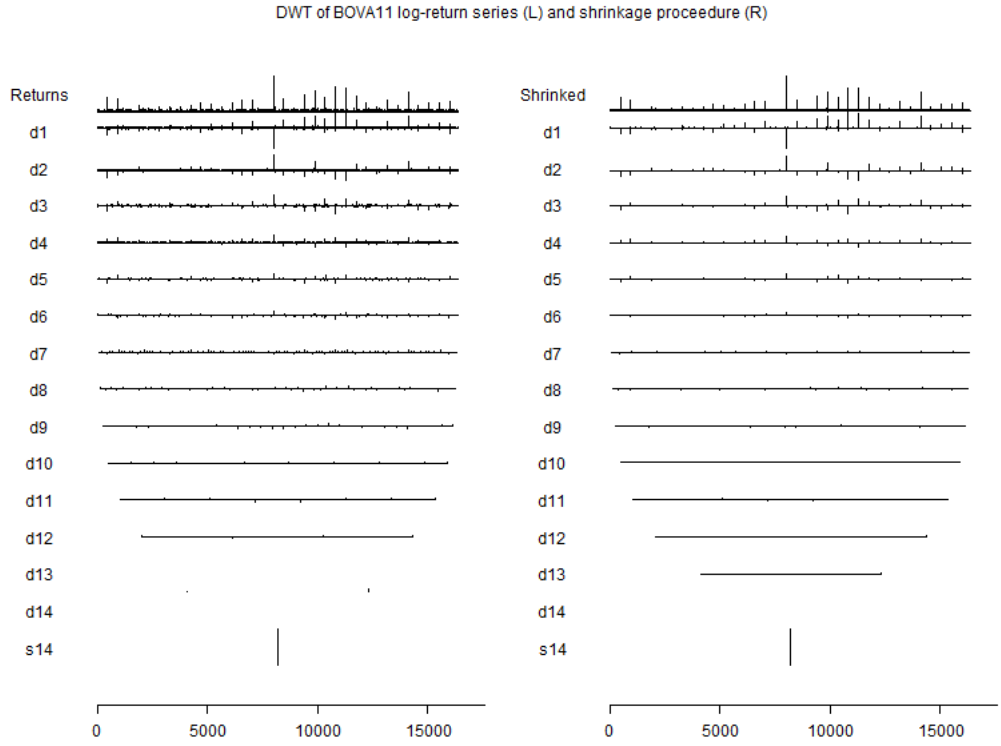

Figure 46: DWT and Shrinkage procedure for BOVA11 log-returns

It is interesting to note that the series seemed, by visual inspection, to be very volatile, and the results do indicate the presence of several price jumps in the time series. In order to obtain the jump series, it is applied the inverse discrete wavelet transform (IDWT) to the DWT series after the thresholding procedure.

In Figure 46, the identified pattern of jumps in correspond to the wavelet reconstruction of the shrinked DWT procedure of the log-return series as in Figure 47 in red. Analyzing the result of the procedure, it is noteworthy that not only the values close to 1%, which by visual inspection and evaluating the information in Table 23 seemed anomalous, were identified, but also other values close to 0.5% values were identified as jumps too. The resulting series of log-returns after removing the jumps plotted in the bottom of Figure 47 seem, by visual inspection, to keep the properties expected for financial time series, with heteroskedastic behaviour and volatility clusters.

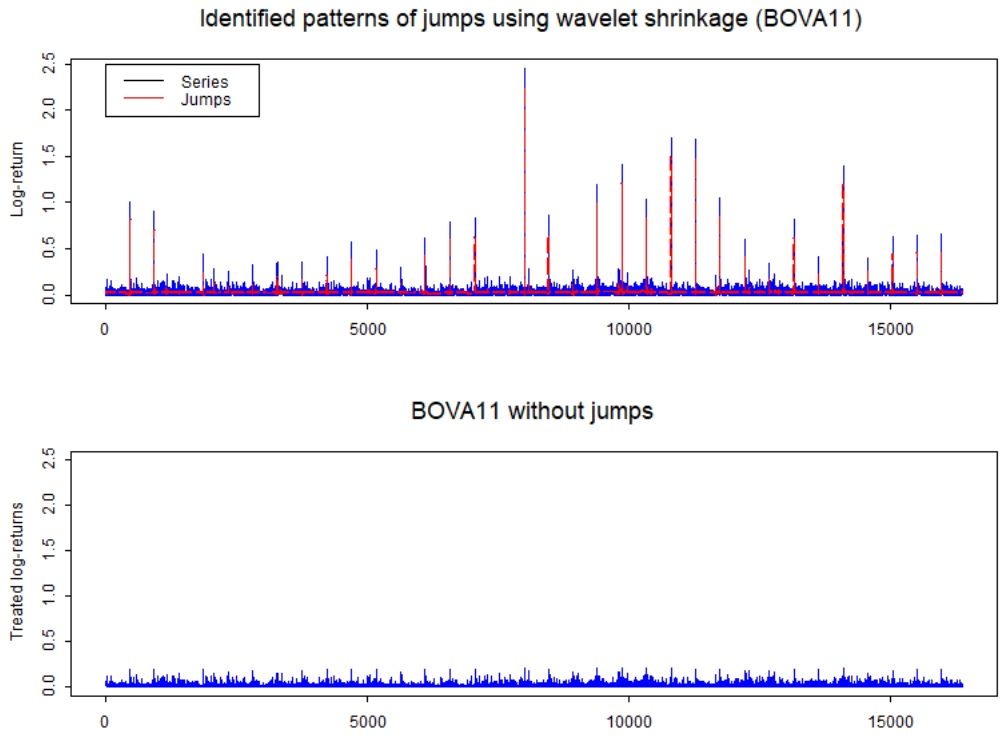

Figure 47: BOVA11 Identified jump series using MAD shrinkage function

Consider now Figure 48 and Figure 49: they exhibit the ACF and PACF of the BOVA11 log-return series without removing the jumps and the ACF and PACF after removing the jumps with the wavelet procedure, respectively. In this series, the presence of jumps which behave as additive outliers has completely outshadowed the autocorrelation structure in the conditional variance, as described in literature. By visual inspection of the ACF and PACF in Figure 48, there seems to exist no serial correlation, because all the calculated partial autocorrelations and autocorrelations lie within the Bartlett limits, thus indicating that, at 5% significance level, they are statistically null. Interestingly however, after removing the jumps with the wavelet procedure, the decay is indeed much slower in the ACF and PACF of the squared returns, which are statistically significant at the 5% level up to lag 30, even indicating the presence of long memory for the conditional variance. This can be seen in Figure 49.

This particular case is indeed interesting, because it shows how additive outliers can completely dominate the behaviour of a series: as a result, the identification procedure could mislead one into identifying a sub-optimal model to describe the dynamics of the series. In fact, analyzing the ACF and PACF of the series without removing the jumps, one could suggest the usage only of an ARFIMA mode in order to sufficiently account for the serial dependence structure showed in Figure 48. However, after removing the price jumps, the outshadowed serial correlation structure reappears, as it can be seen in Figure

49, it is expected to fit an ARFIMA-FIGARCH model to the time series.

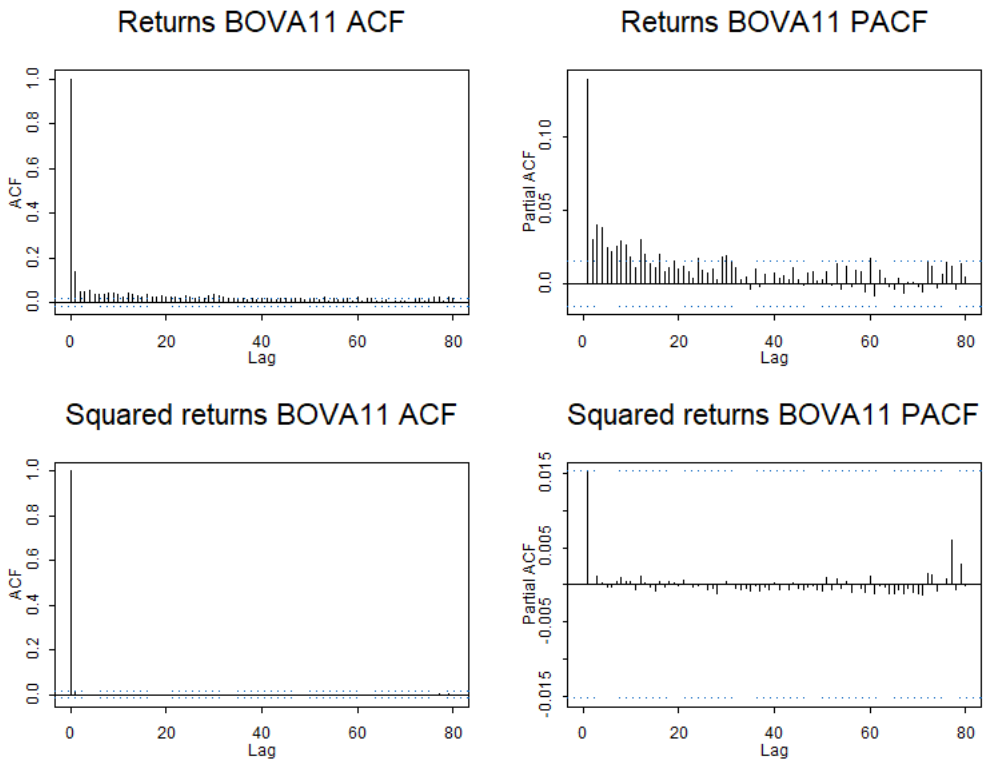

Figure 48: ACF and PACF of BOVA11 log-return series without removing the jumps

The slow decay of the ACF and PACF shown in Figure 49 indicates the presence of long-memory. In fact, there are still non-null autocorrelations at 5% level around lag 30 in the ACF and around lag 20 in the PACF for both the conditional mean and conditional variance. To account for the long memory in the conditional mean, one can consider the ARFIMA(1,d,1) model, chosen as the most parsimonious one, whilst ensuring all serial correlation were statistically null at the 5% significance level.

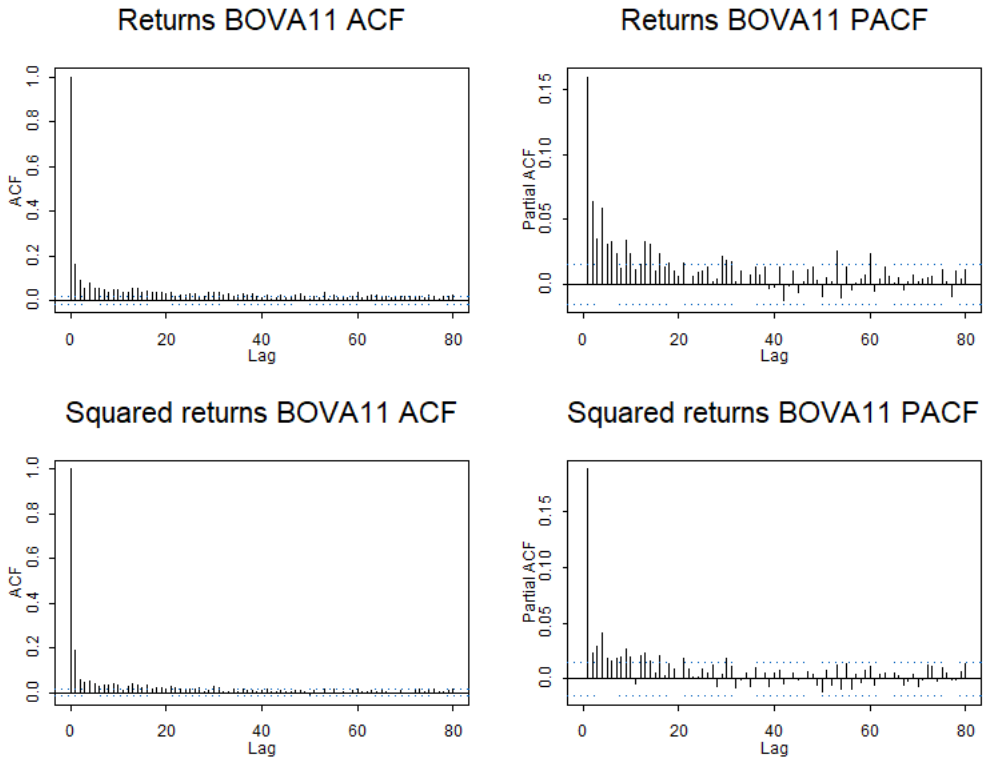

Figure 49: ACF and PACF of BOVA11 treated log-return series

The estimated coefficients for the ARFIMA(1,d,1) are displayed in the Table 24, as well as the standard errors for the coefficients and the T-statistics, from which one can test whether these coefficients are statistically equal to zero.

Table 24: ARFIMA coefficients - BOVA11

Coef.	Value	Std. Error	T stat.	p-value
d	0.2271	0.0216	10.5359	0.0000
AR(1)	0.6214	0.0212	29.2928	0.0000
MA(1)	0.7130	0.0060	118.1928	0.0000

It is important to state that the coefficients estimated in table 24 are statistically significant at the 5% level. It is also worth to note that the model is stationary and invertible because $|d| = |0.227| < 0.5$ and the roots of the autorregressive and moving average polynomial lie outside the unit circle. The model described by Table 24 suggests not only a mild dependence in the long-run dynamics, but there is also strong dependence in the short-run dynamics, evaluating the ARMA coefficients.

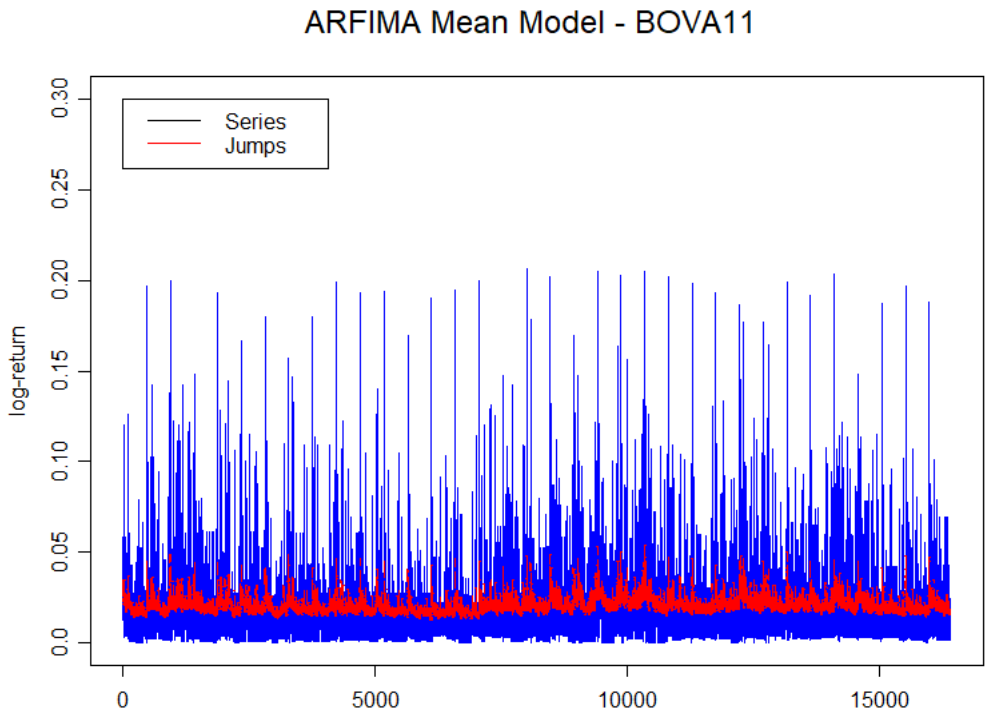

Figure 50: ARFIMA adjusted to the BOVA11 series

The Figure 50 exhibits the fitted model in red versus the observed BOVA11 series. From that, one can see that the model cannot account for the dynamics of the conditional variance, even though the conditional mean can be described by the ARFIMA model. Indeed, there seems to exist conditional heteroskedasticity, corroborating what was seen in Figure 49. This is again an indication that it is necessary to consider conditional variance models to account for the heteroskedasticity.

Table 25: FIGARCH coefficients - BOVA11

Coef.	Value	Std. Error	T stat.	p-value
Cst x 10^4	0.1412	0.00001	1.6803	0.0929
ARCH(1)	0.1209	0.0029	42.0982	0.0000
GARCH(1)	0.1658	0.0010	165.0187	0.0000
GARCH(2)	0.4876	0.0022	223.8580	0.0000
d-Figarch	0.2548	0.0009	274.8219	0.0000
Asymmetry	1.5057	0.0469	32.1047	0.0000
Tail	3.5132	0.1908	18.4164	0.0000

Consider then the FIGARCH(1,d,2) in order to account for the conditional heteroskedasticity and the long memory in the conditional variance. The order of the FIGARCH model was chosen due to the fact that it was that it was the most parsimonious

choice which generated standardized residuals which are statistically null at the 5% significance level. The model was estimated using the rugarch package for R, developed and maintained by software developed by Ghalanos (2020). The numeric optimization method used for that was the GOSOLNP method, as proposed by Hu, Shonkwiler and Spruill (1994) and implemented in R by Theussl and Ghalanos (2015).

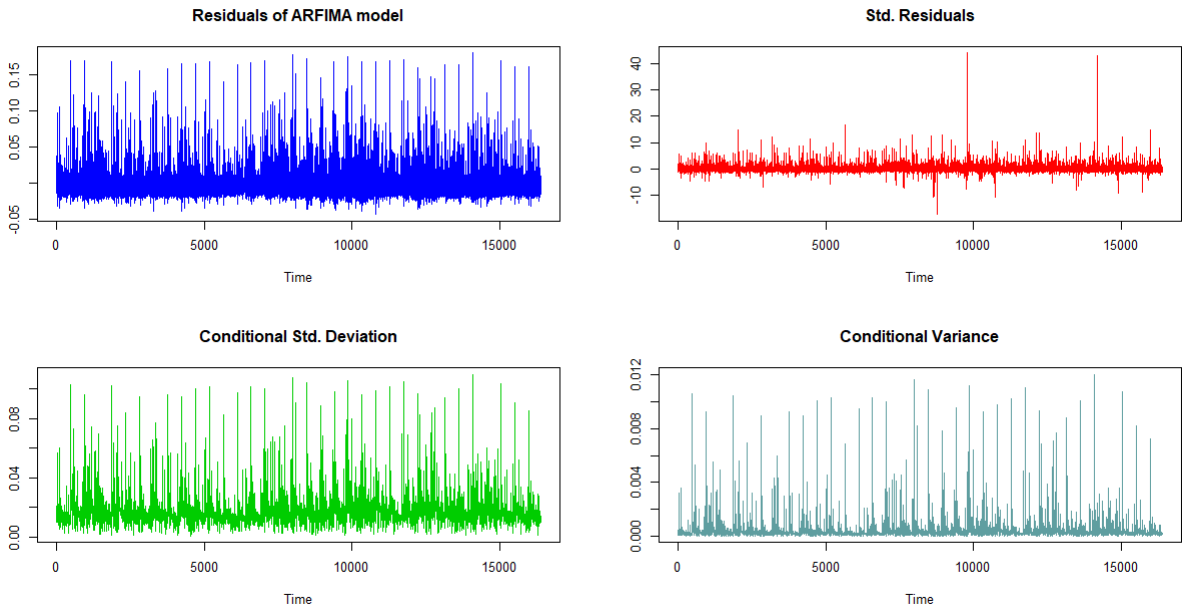

Figure 51: Adjusted FIGARCH for BOVA11 series with Skewed Student distribution

The Table 25 exhibits the estimated coefficients for the FIGARCH(1,d,2) model. It is worth mentioning that almost all coefficients are statistically significant at the 5% level, with the exception being the constant term, which is statistically significant at 10% level. Because having the constant term positive is a condition for the positivity for the FIGARCH(1,d,1), it is chosen to keep it, even though it is only significant at 10% significance level.

It is also worth mentioning that the positivity constraints for the FIGARCH(1,d,2) remains as an open question: therefore, it is not possible to verify whether this model fitted meets or not certain positivity constraints, as done previously for each FIGARCH(1,d,1) model. It is worth mentioning that $d < 0.5$, which is a condition to ensure that the FIGARCH model is second order stationary. The ARCH and GARCH coefficients do not sum up to 1, and there is indication of meaningful dependence of both long-run and short-run dependence.

Figure 52 exhibits the return series with the 2 times the estimated conditional standard deviation superimposed, which essentially shows that most of the variation of the time series lie within 2 conditional standard deviation. This corroborates that the

estimated conditional standard deviation in Figure 51 is coherent with the data, because it reflects the increase in volatility during the period of clusters and it captures the heteroskedastic behaviour of the model.

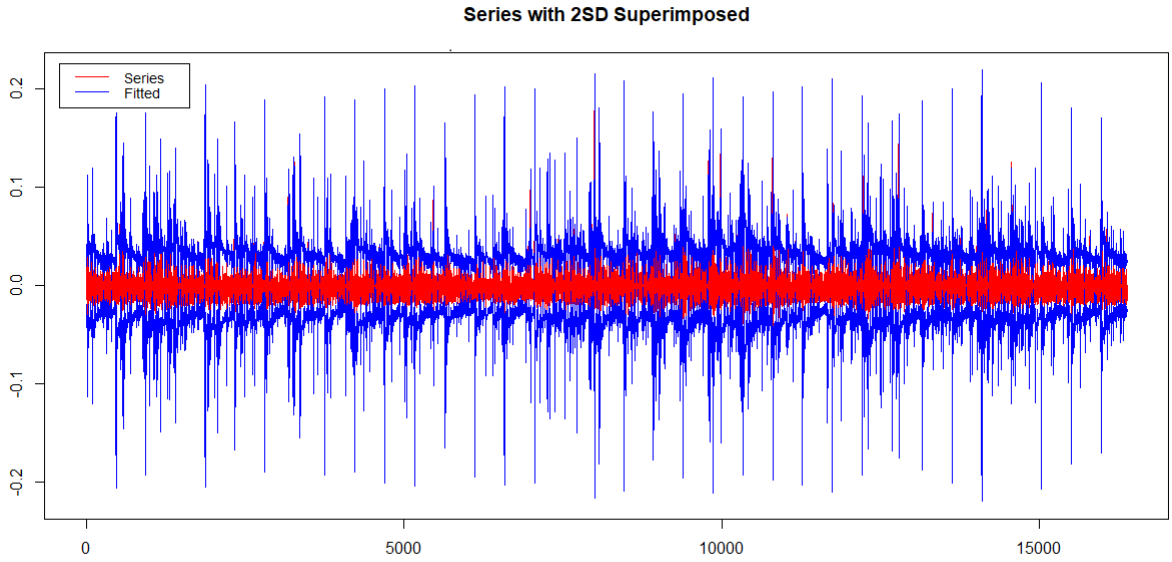

Figure 52: Series with 2 SD Superimposed - BOVA11

To corroborate the results stated in terms of goodness of fit for the chosen models, one can consider the Box-Pierce of the residuals of the fitted models, displayed in Table 26. From that, it is possible to state that at the 5% significance level the residuals behave statistically like a white noise.

This, together with all the other information provided when analyzing the fit of the model, make possible to state that the chosen models successfully account for the serial correlation structure observed in the return series.

Table 26: P-value of the Box-Pierce test for the residuals - BOVA11

Lag	Mean model res.	Var. model std. res.
5	0.988	0.993
10	0.472	0.955
15	0.286	0.998
20	0.182	1.000

The RMSE of the model using the wavelet methodology to identify and remove the jumps, calculated by equation 4.2 equals 0.00165136. For the serial correlation structure exhibited in Figure 24, one can fit an ARFIMA(1,d,1) model, which produces a RMSE of 0.09888833, which is also displayed in 27.

Table 27: Comparison of RMSE of models - BOVA11

Model	RMSE
Removing jumps	0.0016514
Not removing jumps	0.0988883

A few comments are worth regarding the RMSE of the models and the choice of the model for the traditional approach. As for the case of BOVA11, in order to ensure that there were no ARCH effects yet to be modelled, it was performed an LM Test for ARCH effects, as described by Zivot and Wang (2005) and implemented in S+FinMetrics. The resulting test statistics was 10.3839 and the p-value was close to 1, which is a strong evidence for no ARCH effects. Moreover, a Ljung-Box test on the squared residuals indicates no serial correlation which is statistically significant at 5% level. This is yet another example of the effects of jumps or additive outliers in $\hat{\rho}(j)$, and why it is necessary to remove the presence of jumps.

At last, assessing the results of each of the models and analyzing the information in the provided tables and complemented by the figures in this section, it is possible to state that the methodology for modelling both the return series and the jump series adequately accounted for the serial dependence displayed in each case. Nevertheless, the usage of wavelets for the identification of the jumps turned out as an interesting tool for the procedure, providing significant results.

4.4.6 Models for FLRY3

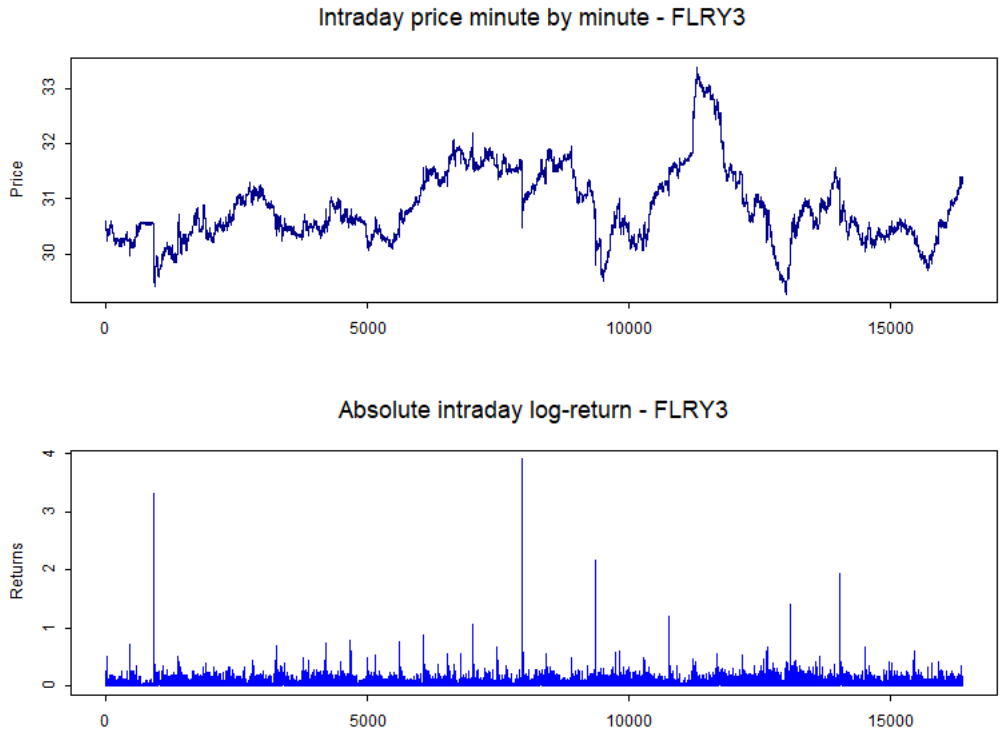

Figure 53: FLRY3 Intraday price series and absolute log-returns

In this subsection, the time series used is the intraday price of FLRY3, once again homogeneously sampled minute-by-minute. As in the case of WEGE3, the series starts in 2020-01-02 10:03:00 and ends in 2020-02-20 10:08:00, which corresponds to approximately one and a half month of trade days (36 days) with $T = 16,384$ observations.

The symbol FLRY3 correspond to the ordinary shares of Fleury S.A traded in the Brazilian Exchange and OTC (B3) since 2007. The Fleury Group is a medicine and healthcare organizations in Brazil, with nearly 100 years of existence. The group is specialized in medical services in the field of diagnostics, treatments, clinical analysis, health management and medical assistance. The three main business lines of Fleury Group are patient service units, reference labs and diagnostics operations in hospitals.

It is possible to notice in Figure 53 that the absolute log-return series displays the traits expected traits of volatility clusters and presence of jumps, even though they are not as frequent as in the other series used in this application chapter.

The Table 28 displays the 10 largest absolute $100\times$ log-returns for FLRY3. Unlikely in the case of WEGE3, for the FLRY3 log-return series, as one can see, these returns are either at the end of the trade day or at the beginning of the trade day. As compared to other series used in the applications chapter, from visual inspection, there seems to

Table 28: FLRY3 absolute log-returns - top 10

Ticker	Date	Time	Price	Abs.logret
FLRY3	20200127	100300	30.4746	3.9072
FLRY3	20200103	174900	29.4855	3.3315
FLRY3	20200130	100900	29.8576	2.1736
FLRY3	20200213	100400	30.5333	1.9375
FLRY3	20200210	175400	30.2885	1.4000
FLRY3	20200204	101000	31.4244	1.1912
FLRY3	20200123	100500	31.8455	1.0705
FLRY3	20200122	175400	32.1883	0.8862
FLRY3	20200120	175400	31.5615	0.8726
FLRY3	20200203	175400	31.0523	0.8166

be less price jumps in the FLRY3 log-returns in Figure 53 with a magnitude of 2% or more. However, other values which are "smaller" might correspond to an anomalous price changes: this is the reason why the usage of wavelets is important to identify which values correspond to a a jump and which do not, and the non-parametric method is useful in that case, because each time series will have its particularities in order to understand what characterizes a jump and what does not.

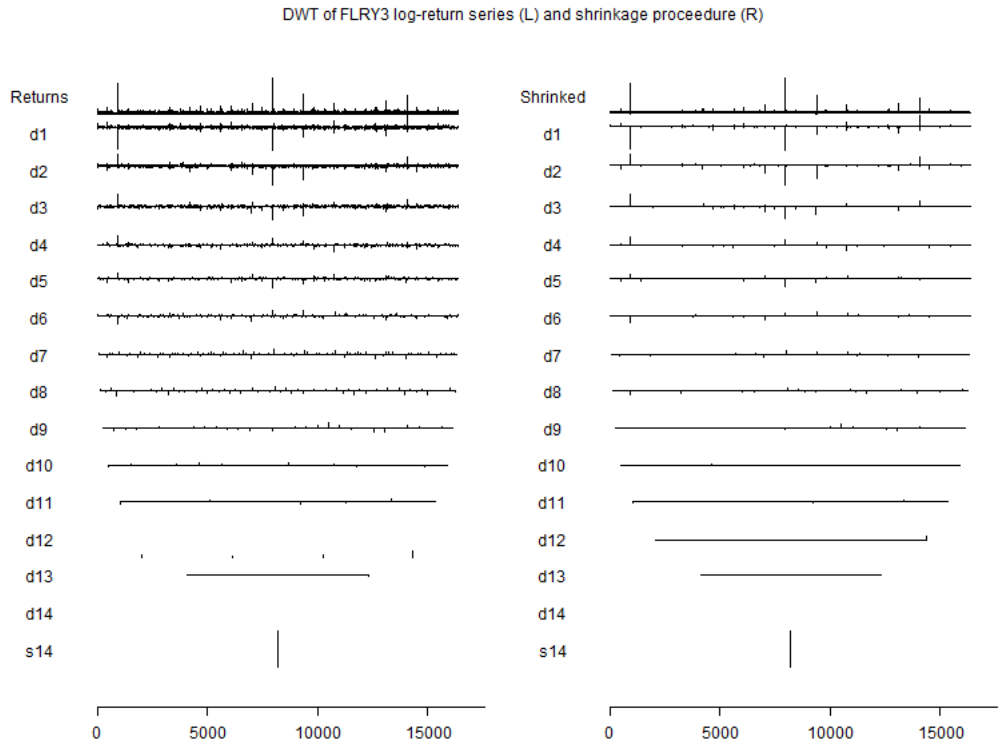

Figure 54: DWT and Shrinakge procedure - FLRY3 log-returns

Figure 54 exhibits the discrete wavelet transform in the left plot and the result of the shrinkage procedure in the right plot. In order to estimate these jumps, the Haar wavelet was used and the MAD function. The identified jump series is in that case much smoother and there are less jumps: some values closer to 1% in the case of FLRY3 log-returns are considered as jumps

Figure 55 displays the identified pattern of jumps versus the original plot and the log-return series discounting the jumps from the original log-return series. Considering the plot in first figure, one can see that there are not so many jumps in the case of FLRY3 log-returns series. It is also possible to see that the treated log-return series displays the expected patterns of volatility clusters and heteroskedasticity, which are stylized facts of financial time series.

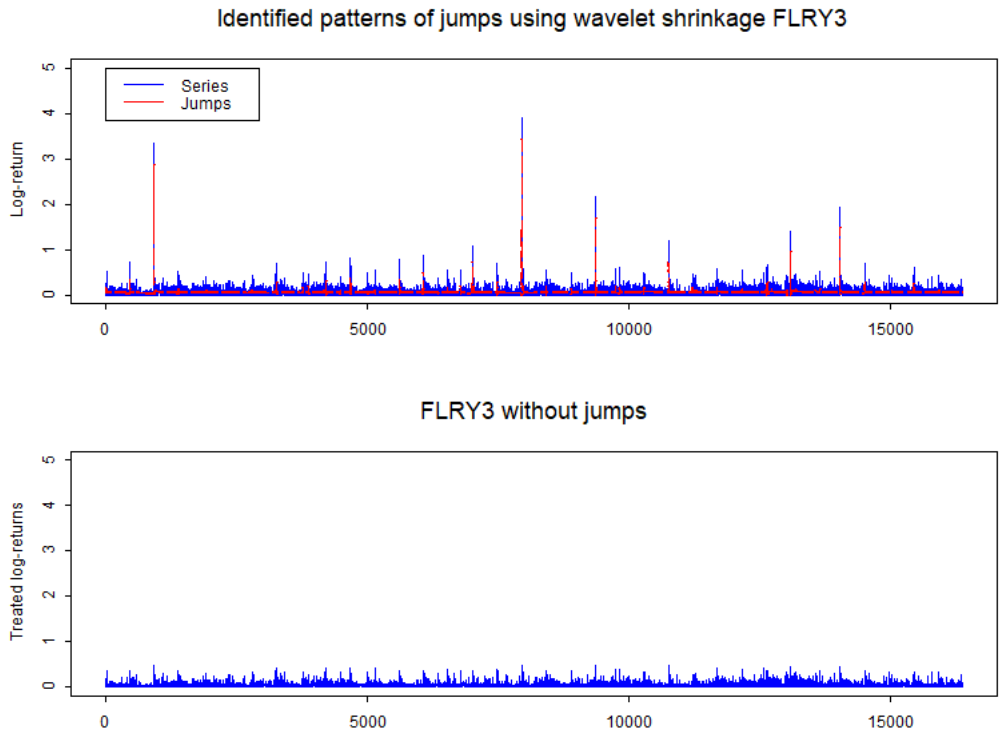

Figure 55: FLRY3 Identified jump series using MAD shrinkage function

Figure 56 and 57 exhibit the ACF and PACF of the FLRY3 log-return series before removing the identified jumps and changes and after removing the price jumps with the wavelets, respectively. In that case, it is possible to see that the presence of outliers in the series dominated the behaviour of the conditional volatility, overshadowing the true serial correlation behaviour in the log-returns time series, making it appear that there is no long range dependence in the squared returns. However, after removing the price jumps, as seen in Figure 57, the decay is much slower in the ACF and PACF of the squared

returns than before removing the jumps, exhibiting traits which indicate long memory models both in the conditional mean and conditional variance.

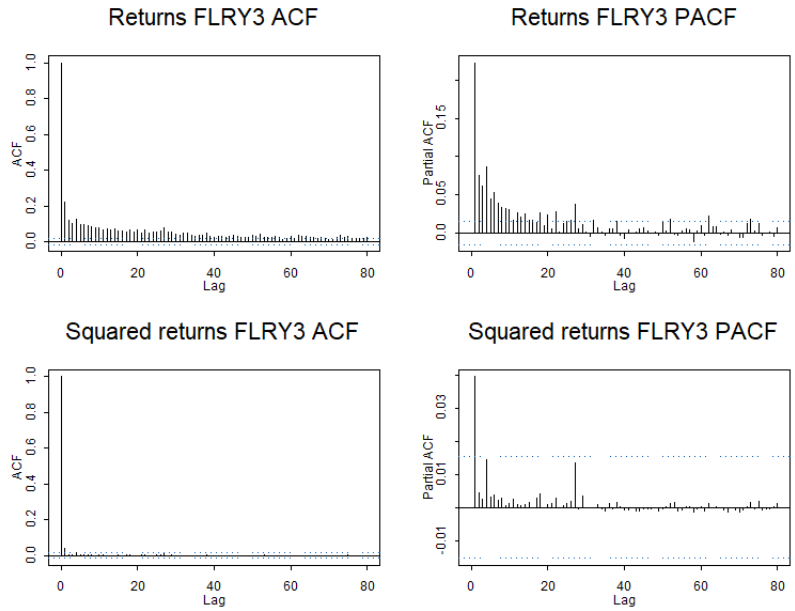

Figure 56: ACF and PACF of FLRY3 treated log-return series without removing the jumps

If evaluating the ACF and PACF of the series without removing the jumps, one could suggest the usage only of an ARFIMA-ARCH or purely and ARFIMA model in order to sufficiently account for the serial dependence structure showed in Figure 56.

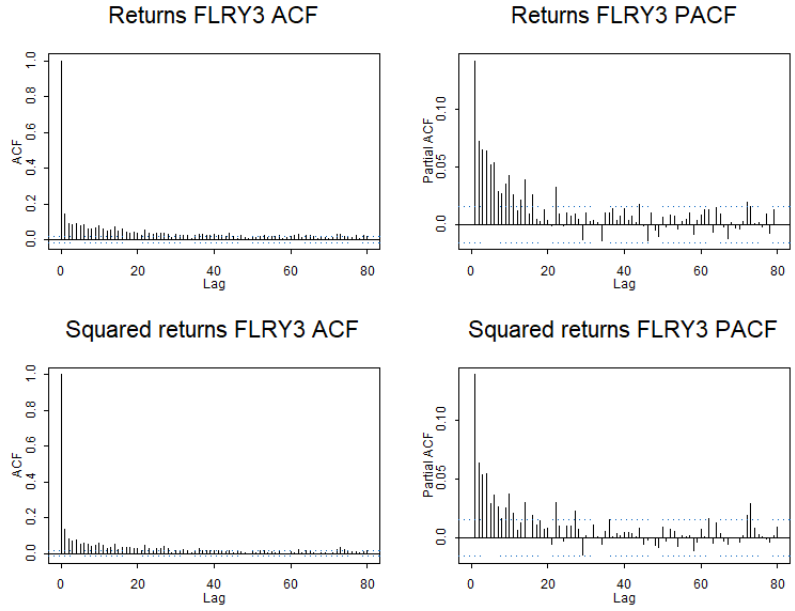

Figure 57: ACF and PACF of FLRY3 treated log-return series

A possible model to account for the serial dependence strcutre shown in Figure 57 is then is to fit an ARFIMA model for the conditional mean and then fit a FIGACRH model to the conditional variance.

The Table 29 shows the estimated coefficients for the ARFIMA model. The chosen model in that case was an ARFIMA(1,d,1) with a $d = 0.2584$ parameter for long-memory. The model was chosen because it was the most parsimonious one which accounted for all the serial correlation of the conditional mean, at the significance level of 5%.

Table 29: ARFIMA coefficients - FLRY3

Coef.	Value	Std. Error	T stat.	p-value
d	0.2583	0.0160	16.1713	0.0000
AR(1)	0.4264	0.0198	21.5753	0.0000
MA(1)	0.5780	0.0080	72.3469	0.0000

From the results pointed out in Table 29, one can see that the process is stationary and invertible, because $|d| = |0.2583| < 0.5$ and the roots of the autoregressive and moving average polynomials lie outside the unit circle. This indicates a persistent process in the long-run dynamics, however it was necessary to consider the AR and MA part for the model in order to account for the short-run dynamics of the process.

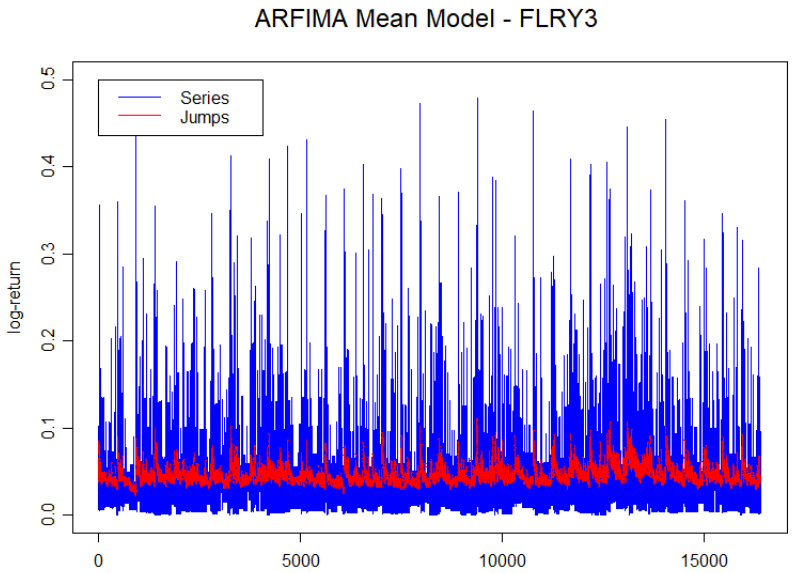

Figure 58: ARFIMA adjusted to the FLRY3 series

Figure 58 displays the fitted values of the ARFIMA mean model versus the FLRY3 log-return series. The figure shows that, while the ARFIMA describes part of the dynamics

for the conditional mean, the original series is heteroskedastic and still needs to take the conditional variance model into account.

In order to do so, for the conditional variance, one can consider the FIGARCH(1,d,1) model fitted for the data, with the innovations term following a generalized error distribution. This model is chosen as the most parsimonious whilst ensuring that there are no significant autocorrelations at the significance level of 5% in the squared residuals of the ARFIMA model. The model was estimated using the OxMetrics-G@RCH^{TM}. The numeric optimization procedure used for the MLE was the BFGS, with implementation in Ox by Laurent (2018).

The Table 30 shows the coefficients for the FIGARCH(1,d,1) model and the estimated parameters for the skewed student distribution. Figure 59 exhibits four plots related to the FIGARCH fitting: the residuals of the ARFIMA model, the standardized residuals of the FIGARCH model, the conditional standard deviation and the conditional variance.

Table 30: FIGARCH coefficients - FLRY3

Coef.	Value	Std. Error	T stat.	p-value
Cst x 10^4	4.3884	0.6343	6.9190	0.0000
d-Figarch	0.1526	0.0122	12.5000	0.0000
ARCH(1)	-0.9850	0.0065	-152.2000	0.0000
GARCH(1)	-0.9836	0.0068	-144.6000	0.0000
G.E.D.(DF)	1.1495	0.0187	61.3500	0.0000

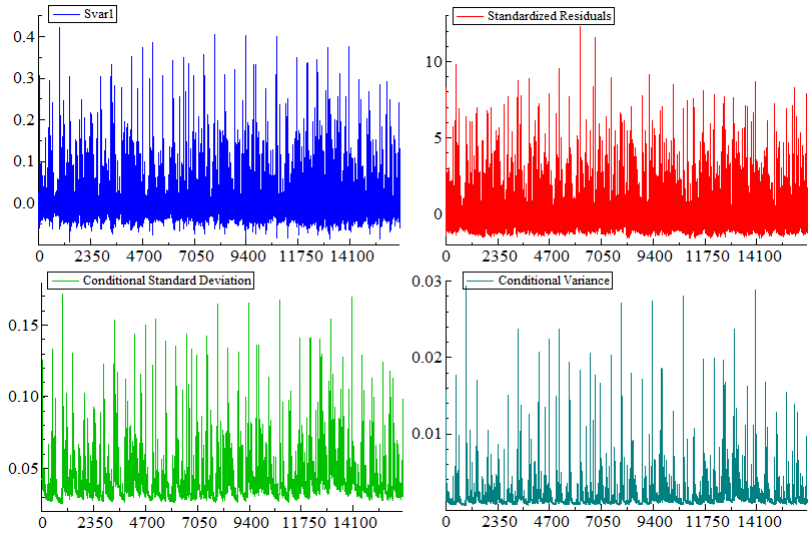

Figure 59: Adjusted FIGARCH for FLRY3 series with GED

Based on the Table 30, some results are worth to point out. First, once again,

from the estimated values, the positivity constraint for the FIGARCH (1,d,1) model is observed, using the referred conditions in Bollerslev and Mikkelsen (1996), because -1.13617<-0.985048<0.615808 and -0.214944<-0.148645 holds.

Although the model has a d parameter for the conditional variance which is smaller than the one in the conditional mean, it is important to state that, as Caporin (2002) mentions, the effect is inverse in the long memory: as $d \to 0$ in the FIGARCH model, the memory increases. This indicates that there is a strong long-run dependence in the conditional variance as well.

The dynamics of the process is described well by the model. One can see that most of the variability of the series lie within the interval of 2 conditional standard deviation, as in Figure 60. It is also possible to see that the the conditional standard deviation as in Figure 60 and 59 captures the volatility of the series, following the behaviour of the original series.

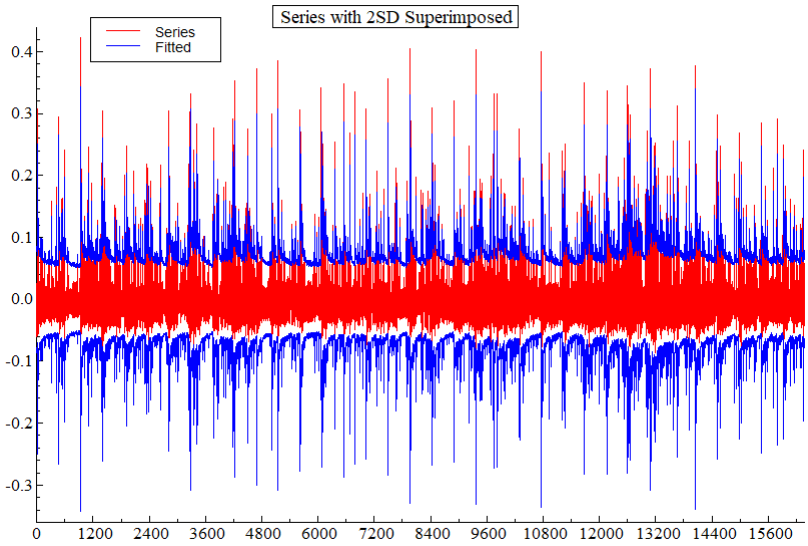

Figure 60: Series with 2 SD Superimposed - FLRY3

Table 31: P-value of the Box-Pierce test for the residuals - FLRY3

Lag	Mean model res.	Var. model std. res.
5	0.9889	0.9219
10	0.1890	0.9069
15	0.2056	0.9760

The Table 31 displays the results of the Box-Pierce test for the series, considering the conditional mean and conditional variance model. From that, one can state that the standardized residuals behave statistically as a white noise, at the significance level of

5%. This means that these ARFIMA-FIGARCH models were enough to account for the dynamics of the log-returns' serial correlation.

Table 32: Comparison of RMSE of models - FLRY3

Model	RMSE
Removing jumps	0.0075734
Not removing jumps	0.1521347

Consider the table 32. The RMSE of the model using the wavelet methodology to identify and remove the jumps, calculated by equation 4.2, is evaluated as 0.00757341. One can also consider an ARFIMA(1,d,1)-ARCH(1) model by visual inspection and evaluating the serial correlation structure exhibited in Figure 56. In that case, the RMSE of the fitted model is 0.1521347.

Overall, after fitting all the conditional mean and conditional variance models in this subsection, the methodology for modelling the volatility of the financial time series seemed satisfactory. In fact, the chosen ARFIMA-FIGARCH models chosen seemed to account for the serial dependence structure in the log-returns series. Once more the wavelet method provided means identify the jumps series with a method that reduces the subjectiveness and arbitrariness in the identification procedure.

It is also worth mentioning that, even though the tests displayed in each subsection were the Box-Pierce tests for serial dependence, it was also performed the Ljung-Box (LJUNG; BOX, 1978) test in each case and the results were similar, thus indicating that the residuals behaved statistically as a white noise at the 5% significance level for each series.

5 Conclusions and Further Considerations

Some models were presented in this dissertation in order to account for some of the stylized facts of high frequency financial time series. In order to describe the conditional heteroskedasticity, long memory in both conditional mean and conditional variance, and fat-tailedness (or skeweness when necessary), ARFIMA-FIGARCH models were considered, with innovations either generated by a Generalized Error Distribution or a Skewed Student t distribution. In order to deal with the intraday or overnight jumps in the log-return series, a non-parametric procedure based on wavelet shrinkage was described, providing means to identify and remove the jumps using wavelets.

In terms of the jump series models, for most cases of practical applications, the median of absolute deviation from the median function (shortly referred to as MAD) was used in the shrinkage procedure, mainly because it displayed the best performance in an heuristic approach. An application was also considered using the L_2 norm.

As for the models for the log-return series after removing these price jumps, in most cases ARFIMA(1,d,1) or those of lower orders were sufficient to model the serial correlation in the conditional mean, and either the FIGARCH(1,d,1) or FIGARCH(1,d,2) models were used to explain the dynamics of the long memory in the conditional variance.

The results shown in each subsection indicate that the proposed methodology generates a more precise estimate to the volatility in terms of RMSE. This is coherent to the literature by Chan (1995), Mendes (2000) and Carnerno, Peña and Ruiz (2004), that outliers or jumps in the return series would generate an estimate of the sample autocorrelations and sample partial autocorrelations that could outshadow the true behaviour of the ACF and PACF of the data generating process, to the point that it could completely dominate the behaviour of the sample ACF and PACF.

A consequence of the bias in $\hat{\rho}(j)$ is that the models identified and estimated according to this procedure will result in a poorer fit to the data, because it will not capture long-range dependence of the volatility. This is very clear when comparing the RMSE of the model only removing the jumps with wavelets and the RMSE without removing jumps.

The Table 33 above summarizes the results shown in the previous chapters. The table is ordered according to the sections discussed in the practical application section. From that, it is possible to see that the methodologies proposed generate a more precise estimate of the volatility in terms of the RMSE than the traditional approach without removing the jumps.

Table 33: Comparison of RMSE of the fitted models removing and not removing jumps

Symbol	RMSE of the different methods		
	Wavelet	Traditional	Percent Difference
BOVA11	0.00165	0.09889	-98.33%
FLRY3	0.00757	0.15213	-95.02%
WEGE3	0.00854	0.15105	-94.35%
SMLL	0.00099	0.00794	-87.53%
VALE3	0.02352	0.15012	-84.33%
ETH/USD	0.02025	0.07810	-74.07%

In the recent years financial datasets are more and more accessible and easily available, whereas also computational capacity and speed is increasing, which allows researchers and practitioners to think of more sophisticated models and alternatives to describe the traits of stylized facts of high frequency financial time series and deal with them.

There are a few topics which might arise from this dissertation for further and more detailed research. For instance, an interesting question would be if there could be a way to identify the best method for determining the value of σ in equation 2.37 in a non-heuristic approach, specifically for this procedure of jump identification for financial time series.

Another topic would be a comparison between methods for modelling the price jumps, in order to describe the dynamics of these anomalous values in a time series, such as INGARCH models (LIBOSCHIK; FOKIANOS; FRIED, 2017) or those based on Hawkes processes (HAWKES, 1971) or Lévy processes (ITO, 1984). Since these price jumps seem not to happen in regular intervals of time, not only the magnitude of the jumps needs to be understood, but also the waiting time between two jumps can be understood as a random variable, and thus could to be modelled. For a major review of Levy processes and applications in stochastic calculus, see Applebaum (2004).

A third topic of interest is to compare the DWT and shrinkage procedure with other non-parametric alternatives, such as splines or the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), and those with the parametric approaches for this type of procedure, as proposed by Torres et al (2011).

Finally, another topic for further research would be to compare the resulting models proposed in the applications chapter with other procedures to identify jumps, such as the Cramér-Von Mises test, the Kolmogorov-Smirnov Statistics and a wavelet procedure to identify jumps in high frequency time series, as discussed by Duran (2019), in order to compare the precision of the models. Other alternatives could be considered using for instance the method proposed by Wang (1995) to identify jumps and sharp cusps in time

series using wavelets.

Bibliography

ANDERSEN, T.; BOLLERSLEV, T. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, v. 4, n. 2-3, p. 115–158, 1997.

ANDERSEN, T. G. et al. *Handbook of Financial Time Series*. USA: Springer, 2009.

APPLEBAUM, D. *Levy processes and stochastic calculus*. 1. ed. Cambridge: Cambridge University Press, 2004. (Cambridge studies in advanced mathematics 93).

B3 Brasil, Bolsa, Balcão. *B3's Trading Procedure Manual*. Brazil, 2019. Disponível em: <http://www.b3.com.br/data/files/73/F2/F5/33/8BFE961023208E96AC094EA8-/B3%20S%20Trading%20Procedures%20Manual%20-%20Version%2004082019.pdf>.

B3 Brasil, Bolsa, Balcão. *Trading Tunel Calulation Methodology*. Brazil, 2019. Disponível em: <http://www.b3.com.br/en_us/solutions/platforms/puma-trading-system/for-members-and-traders/rules-and-trading-parameters/trading-rules/>.

BAILLE, R.; BOLLERSLEV, T.; MIKKELSEN, H. Fractionally integrated generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, v. 74, p. 3–30, 1996.

BAUWENS, L.; HAFNER, C.; LAURENT, S. *The Handbook of Volatility Models*. USA: John Wiley and Sons, 2012.

BELISLE, C. J. P. *Convergence theorems for a class of simulated annealing algorithms on Rd*. USA: J. Applied Probability, 29, 885-895, 1992.

BERAN, J. *Statistics for Long-Memory Processes*. USA: Chapman and Hall, 1994.

BERAN, J. *Mathematical Foundations of Time Series Analysis*. Germany: Springer International Publishing, 2017.

BERAN, J. et al. *Long Memory Process: Probabilistic Properties and Statistical Methods*. Germany: Springer-Verlag, 2013.

BLACK, F. *Studies in stock price volatility changes*. USA: Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association, Washington, 1976.

BOGGESS, A.; NARCOWICH, F. J. *A First Course in Wavelets with Fourier Analysis: Second Edition*. USA: Wiley and Sons, 2009.

BOLLERSLEV, T. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, v. 31, n. 3, p. 307–327, 1986.

BOLLERSLEV, T.; MIKKELSEN, H. Modelling and pricing long memory in stock market volatility. *Journal of Econometrics*, v. 73, p. 151–184, 1996.

BORRI, N. Conditional tail-risk in cryptocurrency markets. *SSRN Electronic Journal*, p. 1–37, 01 2018.

BOUDT, K. et al. *highfrequency: Tools for Highfrequency Data Analysis*. Austria, 2020. R package version 0.6.5.

BOUGEROL, P.; PICARD, N. Strict stationarity of generalized autoregressive processes. *The Annals of Probability*, v. 20, n. 2, p. 1714–1730, 1992.

BOX, G. E. P.; PIERCE, D. A. Distribution of autocorrelations in autorregressive moving average models. *Journal of American Statistical Association*, n. 65, p. 1509–1526, 1970.

BOX, G. P.; JENKINS, G. *Time Series Analysis: Forecasting and Control*. USA: Holden-Day, 1970.

BOX, G. P.; JENKINS, G.; REINSEL, G. C. *Time Series Analysis: Forecasting and Control, 4^{th} Edition*. USA: Wiley and Sons, 2008.

BRILLINGER, D. R. *Time Series: Data Analysis and Theory*. USA: SIAM, 2001.

BROCKWELL, P. J.; DAVIS, R. A. *Time Series: theory and methods 2nd edition*. USA: Springer, 2009.

BRUCE, A.; GAO, H. *Applied Wavelet Analysis with S-PLUS*. USA: Springer, 1996.

CALVORI, F. et al. *TAQMNGR: Manage Tick-by-Tick Transaction Data*. Austria, 2018. R package version 2018.5-1.

CAPORALE, G. M.; GIL-ALANA, L.; PLASTUN, A. Persistence in the cryptocurrency market. *Research in International Business and Finance*, 2018.

CAPORIN, M. *FIGARCH models: stationarity, estimation methods and the identification problem*. Italy, 2002.

CARNERO, M. A.; PEÑA, D.; RUIZ, E. Spurious and hidden volatility. *Universidad Carlos III, Departamento de Estadística y Econometría, Statistics and Econometrics Working Papers*, p. 1–48, 05 2004.

CHAN, W.-S. Understanding the effect of time series outliers on sample autocorrelations. *Test*, v. 14, p. 179–186, 1995.

CHANG, I. *Outliers in Time Series*. Thesis (PhD) — University of Wisconsin, 1982.

CHANG, I.; TIAO, G.; CHEN, C. Estimation of time series parameters in the presence of outliers. *Technometrics*, v. 30, p. 193–204, 05 1988.

CHEN, Y. T.; LAI, W. N.; SUN, E. Jump detection and noise separation by a singular wavelet method for predictive analytics of high-frequency data. *Computational Economics*, 01 2019.

COCCO, L.; CONCAS, G.; MARCHESI, M. Using an artificial financial market for studying a cryptocurrency market. *Journal of Economic Interaction and Coordination*, v. 12, p. 345–365, 06 2014.

CONRAD, C.; HAAG, B. Inequality constraints in the fractionally integrate garch model. *Journal of Financial Econometrics*, v. 4, p. 413–449, 06 2006.

CONT, R. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, v. 1, p. 223–236, 03 2002.

DACOROGNA, M. M. et al. *An Introduction to High-Frequency Finance*. USA: Academic Press, 2001.

DAUBECHIES, I. *Ten Lectures of Wavelets*. USA: Springer-Verlag, 1992.

DEGIANNAKIS, S.; FLOROS, C. *Modelling and Forecasting High Frequency Financial Data*. UK: Palgrave Macmillan, 2015.

DONOHO, D. L.; JOHNSTONE, I. *Wavelet Shrinkage: asymptopia?* USA: Journal of the Royal Statistical Society Series B. 57:301-369, 1995.

DOORNIK, J. A. *Object-Oriented Matrix Programming Using Ox*. UK, 2007. 3rd. Ed.

DOORNIK, J. A.; OOMS, M. *Outlier Detection in GARCH Models*. Netherlands, 2005. (TI 2005-092/4 Working paper).

DOWNIE, T. Accurate signal estimation near discontinuities. *International Journal of Wavelets, Multiresolution and Information Processing*, v. 02, p. 433–453, 11 2011.

DURAN, W. G. R. *Identifying Jumps Variations in High Frequency Time Series*. Thesis (PhD) — Mathematics and Statistics Institute, University of São Paulo, 2019.

ENDERS, W. *Applied Econometric Time Series*. USA: Wiley and Sons, 2010.

ENGLE, R.; RUSSELL, J. R. Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica*, v. 66, n. 5, p. 1127–1162, 1998.

ENGLE, R. F. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, v. 50, p. 987–1008, 1982.

ENGLE, R. F.; BOLLERSLEV, T. Modelling the persistance of conditional variance (with comments and a reply by the authors). *Econometric Reviews*, v. 5, p. 1–87, 1986.

FERNANDEZ, C.; STEEL, M. F. J. *On Bayesian Modelling of Fat Tails and Skewness*. USA: Journal of the American Statistical Association, 93, 359-371, 1998.

FIGLEWSKI, S.; WANG, X. *Studies in stock price volatility changes*. USA: SSRN Electronic Journal, 1976.

FRANCQ, C.; ZAKOIAN, J. M. *GARCH Models: Structure, Statistical Inference and Financial Applications*. USA: Wiley and Sons, 2019.

FRAZIER, M. *An Introduction to Wavelets Through Linear Algebra*. USA: Springer, 1999.

FULLER, W. A. *Introduction to Statistical Time Series*. USA: John Wiley and Sons, 1996.

GARMAN, M. B.; KLASS., M. J. *On the Estimation of Security Price Volatilities from Historical Data*. USA, 1977.

GHALANOS, A. *rugarch: Univariate GARCH models.* USA, 2020. R package version 1.4-2.

GHALANOS, A.; THEUSSL, S. *Rsolnp: General Non-linear Optimization Using Augmented Lagrange Multiplier Method*. USA, 2015. R package version 1.16.

GIUDICI, G.; MILNE, A.; VINOGRADOV, D. Cryptocurrencies: market analysis and perspectives. *Journal of Industrial and Business Economics*, 09 2019.

GOFFE, W. L.; FERRIER, G. D.; ROGERS, J. *Global Optimization of Statistical Functions with Simulated Annealing*. USA: Journal of Econometrics, 60, 65-100, 1994.

GRANGER, C. W.; JOYEUX, R. *An Introduction to Long Memory Time Series Models and Fractional Differencing*. USA: Journal of Time Series Analysis 1, 15-29, 1980.

GUTTMAN, I.; TIAO, G. Effect of correlation on the estimation of a mean in the presence of spurious observations. *La Revye Canadienne de Statistique*, v. 6, p. 229 – 247, 12 1978.

HAAR, A. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*, Springer, v. 71, n. 1, p. 38–53, 1911.

HAFERKORN, M.; DIAZ, J. Seasonality and interconnectivity within cryptocurrencies - an analysis on the basis of bitcoin, litecoin and namecoin. In: . [S.l.: s.n.], 2015. v. 217, p. 1–15.

HAMILTON, J. D. *Time Series Analysis*. USA: Princeton University Press, 1994.

HANOUSEKA, J.; KOĈENDA, E.; NOVOTNÝ, J. *The Identification of Price Jumps*. Czech Republic, 2011. (Working Paper Series 434 (ISSN 1211-3298) ISBN 978-80-7343-235-5).

HANSEN, B. E. *Autoregressive Conditional Density Estimation*. USA: International Economic Review, 33, 705-730, 1994.

HASLETT, J.; RAFTERY, A. E. Space-time modelling with long-memory dependence: Assessing ireland's wind power resource (with discussion). *Applied Statistics*, v. 38, p. 1–50, 1989.

HAUTSCH, N. *Econometric Finance of High-Frequency Data*. Boston: Springer, 2012.

HAWKES, A. G. Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 58, n. 1, p. 83–90, 1971.

HOSKING, J. R. M. *Fractional Differencing*. United Kingdom: Biometrika, 68, 165-176, 1981.

HU, A.; PARLOUR, C.; RAJAN, U. Cryptocurrencies: Stylized facts on a new investible instrument. *SSRN Electronic Journal*, p. 1049–1068, 01 2018.

HU, X. et al. *Random Restarts in Global Optimization*. 1994.

HUANG, W. et al. Price volatility forecast for agricultural commodity futures the role of high frequency data. *Journal for Economic Forecasting*, n. 4, p. 83–103, 2012.

HURST, H. E. *Long-term storage capacity of reservoirs*. USA: Transactions of the American Society of Civil Engineers, 16, 770-799, 1951.

HYNDMAN, R.; KHANDAKAR, Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, v. 26, n. 3, p. 1–50, 2008.

ITO, K. *Lectures on stochastic processes*. New York: Bombay : Tata Institute of Fundamental Research, Distributed by Springer-Verlag, 1984.

JORION, P. *Value at risk: the new benchmark for managing financial risk*. 3rd. ed. USA: McGraw Hill Professional, 2011.

KAIZOJI, T. Statistical properties of absolute log-returns and a stochastic model of stock markets with heterogeneous agents. *arXiv.org, Quantitative Finance Papers*, v. 550, 01 2006.

KERCHEVAL, A. N.; LIU, Y. Risk forecasting with garch, skewed t distributions, and multiple timescales. In: *Handbook of Modeling High-Frequency Data in Finance*. USA: Wiley, 2011. p. 163 – 218.

LAMBERT, P.; LAURENT, S. *Modelling Skewness Dynamics in Series of Financial Data*. France: Discussion Paper, Institut de Statistique, Louvain-La-Neuve, 2000.

LAMBERT, P.; LAURENT, S. *Modelling Financial Time Series using GARCH-Type Models and Skewed Student Density*. France: Mimeo, Université de Liège, 2001.

LANNOY, G.; DECKER, A. D.; VERLEYSEN, M. A supervised wavelet transform algorithm for r spike detection in noisy ecgs. In: . CCIS: Communications in Computer and Information Science, 2008. v. 25, p. 256–264.

LAURENT, S. *Estimating and Forecasting ARCH models using G@RCH TM 8: Ox Metrics*. United Kingdom: Timberlake Consultants, 2018.

LIBOSCHIK, T.; FOKIANOS, K.; FRIED, R. *tscount: an R Package for Analysis of Count time Series Following Generalized Linear Models*. USA: Journal of Statistical Software 82(5), 1-51, 2017.

LJUNG, G.; BOX, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, n. 65, p. 1279–1313, 1978.

LOPES, S. R. C.; MENDES, B. V. M. Bandwidth selection in classical and robust estimation of long memory. *International Journal of Statistics and Systems*, v. 1, n. 2, p. 167–190, 2006.

LOPES, S. R. C.; PRASS, T. S. Theoretical results on fractionally integrated exponential generalized autoregressive conditional heteroskediastic process. *Physica A: Statistical Mechanics and its Applications*, v. 401, p. 278–307, 2013.

MALLAT, S. *A Wavelet Tour of Signal Processing: the sparse way*. USA: Elsevier, 2009.

MALMSTEN, H.; TERÄSVIRTA, T. *Stylized Facts of Financial Time Series and Three Popular Models of Volatility*. Sweden, 2004.

MANDELBROT, B. B.; WALLIS, J. R. *Noah, Joseph and Operational Hydrology*. Advanced Earth and Space Science: Water Resources Research, 4, pp. 909-918, 1968.

MENDES, B. V. d. M. Assessing the bias of maximum likelihood estimates of contaminated garch models. *Journal of Statistical Computation and Simulation - J STAT COMPUT SIM*, v. 67, p. 359–376, 11 2000.

MEYER, Y. *Ondalettes*. France: Hermann, 1990.

MEYER, Y. *Wavelets: algorithms and applications*. USA: SIAM, 1993.

MORETTIN, P. A. *Ondas e Ondaletas*. São Paulo: Editora da Universidade de São Paulo, 2014.

MORETTIN, P. A. *Econometria financeira: um curso em séries temporais financeiras*. São Paulo: Blucher, 2016.

MORETTIN, P. A.; TOLOI, C. M. C. *Análise de séries temporais*. São Paulo: Blucher, 2006.

MORLET, J. *Sampling Theory and Wave Propagation*. USA: 51^{th} Annual Meeting Soc. Explor. Geophysics, 1981.

MOULINES, E.; ROUEFF, F.; TAQQU, M. S. *A wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series*. USA: The Annals of Statistics, 36 (4), 1925-1956, 2008.

NADARAJAH, S. *A generalized normal distribution*. USA: Journal of Applied Statistics 32:7, 685,694, 2005.

NASON, G. P. *Wavelet Methods in Statistics with R (Use R!)*. United Kingdom: Springer, 2008.

NELSON, D. B. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, n. 2, p. 347–370, 1991.

NELSON, D. B.; CAO, C. Q. Conditional heteroskedasticity in asset returns. *Econometrica*, v. 10, p. 229–235, 1992.

NENADIC, Z.; BURDICK, J. W. Spike detection using the continuous wavelet transform. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, v. 54, p. 74–87, 01 2005.

NOCEDAL, J.; WRIGHT, S. *Numerical Optimization*. USA: Springer, 2006.

OHARA, M. *Market Microstructure Theory*. USA: Wiley and Sons, 1998.

PALMA, W. *Long-Memory Time Series: Theory and Methods*. USA: Wiley Sons Interscience, 2007.

PARKINSON, M. The extreme value method for estimating the variance of the rate of return. *The Journal of Business*, v. 53, n. 1, p. 61–65, 1980.

PETUKHINA, A. A.; REULE, R.; HÄRDLE, W. K. *Rise of the Machines? Intraday High-Frequency Trading Patterns of Cryptocurrencies*. Germany, 2020.

PHILLIP, A.; CHAN, J.; PEIRIS, S. A new look at cryptocurrencies. *Economics Letters*, v. 163, p. 6–9, 11 2017.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: <http://www.R-project.org/>.

ROWAN, T. *Functional Stability Analysis of Numerical Algorithms*. Thesis (PhD) — Department of Computer Sciences, University of Texas at Austin, 1990.

RUSSELL, J.; ENGLE, R. Analysis of high-frequency data. In: *The Handbook of Financial Econometrics: Tools and Techniques*. USA: Elsevier Science, 2010. v. 1, p. 383–426.

SHAO, X.-D.; LIAN, Y.-J.; YIN, L.-Q. Forecasting value-at-risk using high frequency data: the realized range model. *Global Finance Journal*, v. 20, p. 128–136, 12 2009.

SHUMWAY, R. H.; STOFFER, D. S. *Time Series Analysis and its Applications with R Examples: Fourth Edition*. USA: Springer, 2016.

SIBBERTSEN, P. *Long-memory in volatilities of German stock returns*. Germany, 2001.

SOWELL, F. *Maximum Likelihood estimation of stationary univariate fractionally integrated time series models*. USA: Journal of Econometrics 53, 165-188, 1992.

SPITHOVEN, A. Theory and reality of cryptocurrency governance. *Journal of Economic Issues*, v. 53, p. 385–393, 04 2019.

STRAUMANN, D.; MIKOSCH, T. *Quasi-maximum-likelihood estimation in heteroscedastic time series: a stochastic recurrence equation approach*. USA: Annals of Statistics 34: 2449-2495, 2006.

TORRES, M. et al. A complete ensemble empirical mode decomposition with adaptive noise. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 4144–4147, 2011.

TSAY, R. S. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, v. 7, p. 1–20, 1988.

TSAY, R. S. *Analysis of Financial Time Series: Second Edition*. USA: Wiley and Sons, 2005.

VASUDEVA, R.; KUMARI, J. V. *On general error distributions*. India: ProbStat Forum, 06, 89-95, 2013.

VIDAKOVIC, B. *Statistical Modeling by Wavelets*. USA: Wiley and Sons, 1999.

VIENS, F. G.; MARIANI, M. C.; FLORESCU, I. *Handbook of Modeling High-Frequency Data in Finance*. 1. ed. USA: Wiley, 2011. (Wiley Handbooks in Financial Engineering and Econometrics).

WANG, Y. Jump and sharp cusp detection by wavelets. *Biometrika*, [Oxford University Press, Biometrika Trust], v. 82, n. 2, p. 385–397, 1995.

WASSERFALLEN, W.; ZIMMERMANN, H. The behavior of intra-daily exchange rates. *Journal of Banking Finance*, v. 9, n. 1, p. 55–72, 1985.

WORTHINGTON, A.; HIGGS, H. Systematic features of high-frequency volatility in australian electricity markets: Intraday patterns, information arrival and calendar effects. *The Energy Journal*, v. 26, p. 23–42, 10 2005.

XEKALAKI, E.; DEGIANNAKIS, S. *ARCH Models for Financial Applications*. USA: John Wiley and Sons, 2010.

XUE, Y.; GENÇAY, R.; FAGAN, S. Jump detection with wavelets for high-frequency financial time series. *Quantitative Finance*, v. 14, p. 1427–1444, 08 2014.

YAN, B.; ZIVOT, E. *Analysis of High-Frequency Financial Data with S-Plus*. USA: Working Papers UWEC-2005-03, University of Washington, Department of Economics., 2003.

YE, Y. *Interior Algorithms for Linear, Quadratic, and Linearly Constrained Non-Linear Programming*. Thesis (PhD) — Department of ESS, Stanford University, 1987.

ZAFFARONI, P. *Stationarity and Memory of ARCH models*. United Kingdom, 2000. (Discussion Paper No. EM/00/383 JEL No.: C22).

ZHANG, Y. et al. Stylised facts for high frequency cryptocurrency data. *Physica A: Statistical Mechanics and its Applications*, v. 513, 09 2018.

ZIVOT, E.; WANG, J. *Modelling Financial Time Series With S-PLUS: 2nd edition*. USA: Springer, 2005.

ZUMBACH, G.; MULLER, U. *Operators on Inhomogeneous Time Series*. Switzerland, 2000. (Document ID: GOZ.1998-11-01).

—